Advanced Probability Theory: Notes

YANG, Ze

March 3, 2016

Contents

Chapter 1

Measure Space, Prob Space

1.1 Algebraic Structures on Prob Space

1.1.1 Sigma Field

Default setting: Let S be a set.

Def. **Algebra**: A family of $A \subseteq S$, Σ_0 is an algebra if

- $S \in \Sigma_0$.
- $A^c \in \Sigma_0$. (close to **Complement**)
- n finite, $\bigcup_{i=1}^n A_i \in \Sigma_0$. (close to **Finite Union**)

Rm. 1,2,3 implies

- $\emptyset \in \Sigma_0$.
- $A \cap B, A \cup B, (A \setminus B), (A \triangle B) \in \Sigma_0$.
- $\bigcap_{i=1}^n A_i \in \Sigma_0$.

Def. **Sigma-Field**: A family of $A \subseteq S$, Σ is a sigma-field if 1,2 (algebra) and

- $\bigcup_{j=1}^\infty A_j \in \Sigma$. (close to **Countable Union**)

Def. **Generated Sigma-Field**: $C \subseteq S$, $\sigma(C)$ is generated sigma-field from C if

- $\sigma(C)$ is a sigma field.
- $C \subseteq \sigma(C)$.
- If $C \subseteq \Sigma' \neq \sigma(C)$, Σ' is another sigma field, then $\sigma(C) \subseteq \Sigma'$.

i.e. $\sigma(C)$ is the smallest sigma field that is a supset of C. Also written as,

$$\sigma(C) = \bigcap_{\Sigma\text{: sigma field}} \{\Sigma : C \subseteq \Sigma\} \tag{1.1}$$

Prop. Several Propositions.

- Intersection of sigma field is still sigma field. (**No for Union**.)

- To obtain a sigma field from union of sigma fields, define:

$$\bigvee_{\alpha\in I}\Sigma_\alpha := \sigma(\bigcup_{\alpha\in I}\Sigma_\alpha) \tag{1.2}$$

- $\sigma(\sigma(C)) = \sigma(C)$.
- $A \subseteq B \Rightarrow \sigma(A) \subseteq \sigma(B)$.

Def. **Borel Sigma Field**: S is topological space (where open sets can be defined).

$$\mathscr{B}(S) := \sigma(\{O \subseteq S : O \text{ is open}\}) \tag{1.3}$$

Rm. **Borel Sigma Field on Real Line** By construction of open sets, $\forall$ open set $O \subseteq \mathbb{R}$, O can be written as: $O = \bigcup_{k=1}^{n}(a_k, b_k)$. Therefore, Borel sigma field on real line is actually:

$$\mathscr{B}(\mathbb{R}) = \sigma(\{(a,b) : a,b \in \mathbb{R}, a < b\}) \tag{1.4}$$

Def. **Measurable Space**: Set S equipped with sigma field Σ, i.e pair (S,Σ) is a measurable space. $A \in \Sigma$ is Σ-measurable subset of S.

1.1.2 Pi System and Dynkin's D System

Def. **Pi System**: A family of $A \subseteq S$, $\mathcal{I}$ is a π-system if

- $I_1, I_2 \in \mathcal{I} \Rightarrow I_1 \cap I_2 \in \mathcal{I}$. (closed to **Finite Intersection**)

Rm. π systems are easier then sigma field. For example, $\mathbb{R}$ generated π (one notion) is family of all intervals of form $(-\infty, x]$.

$$\pi(\mathbb{R}) = \{(-\infty, x] : x \in \mathbb{R}\} \tag{1.5}$$

Def. **D System**: A family of $A \subseteq S$, $\mathcal{D}$ is Dykin's d-system if

- $S \in \mathcal{D}$.
- $A, B \in \mathcal{D}$, $A \subseteq B \Rightarrow B \setminus A \in \mathcal{D}$.
- $A_n \in \mathcal{D}, n \geq 1, A_n \nearrow A \Rightarrow \lim_{n\to\infty} A_n = A \in \mathcal{D}$. (closed to **limit from below**)

Prop. Σ is a σ-algebra $\iff$ Σ is both π-system and d-system.

Proof. $\Rightarrow$ is obvious.
$\Leftarrow$: Check against 3 defining properties. (1) by 1-d. (2) by 2-d, pick $B = S \in \Sigma$, $A^c = B \setminus A \in \Sigma$. (3) Consider

$$U_n := \bigcup_{n\geq 1} B_n = (\bigcap_{n\geq 1} B_n^c)^c \in \Sigma \tag{1.6}$$

This is ensured by 1-pi and 2-d. And $U_n \nearrow \bigcup_{n\geq 1} B_n =: U$; by 3-d, $U \in \Sigma$. ■

1.1.3 Dynkin's Lemma

Thm. (**Dynkin**) If $\mathcal{I}$ is a π-system on S, $\mathcal{D}$ is a d-system on S; $\mathcal{I} \subseteq \mathcal{D}$. **Then** $\sigma(\mathcal{I}) \subseteq \mathcal{D}$.

Proof. $d(\mathcal{I})$:=d system generated by $\mathcal{I}$.[1]. Define

$$\mathcal{D}_1 := \{B \in d(\mathcal{I}) : A \cap B \in d(\mathcal{I}), \forall A \in \mathcal{I}\} \tag{1.7}$$

By definition $\mathcal{D}_1 \subseteq d(\mathcal{I})$. Clearly $\mathcal{I} \subseteq \mathcal{D}_1$, so if $\mathcal{D}_1$ is d-system, we will have $d(\mathcal{I}) = \mathcal{D}_1$. Consider any $A \in \mathcal{I}$:

- $S \cap A = A. \Rightarrow S \in \mathcal{D}_1$.
- $(B_1 \setminus B_2) \cap A = (B_1 \cap A) \setminus (B_2) =: D$. Both sides of setminus $\in d(\mathcal{I})$. Since $d(\mathcal{I})$ is d-system, $D \in d(\mathcal{I})$.
- $B_n \nearrow U := \bigcup_{n \geq 1} B_n$. $A \cap B_n \nearrow A \cap U \in d(\mathcal{I})$. So $U \in \mathcal{D}_1$. *Check.*

Define

$$\mathcal{D}_2 := \{C \in d(\mathcal{I}) : B \cap C \in d(\mathcal{I}), \forall B \in d(\mathcal{I})\} \tag{1.8}$$

$\mathcal{I} \subseteq \mathcal{D}_2$. Similarly, we check that $\mathcal{D}_2$ is indeed a d-system. So $\mathcal{D}_2 = d(\mathcal{I})$. Now we check $\mathcal{D}_2$ is a pi-system. Consider any $B \in d(\mathcal{I})$:

- $(C_1 \cap C_2) \cap B = C_1 \cap (C_2 \cap B) =: C_1 \cap B'$. By definition of $\mathcal{D}_2$, $B' \in d(\mathcal{I})$; $C_1 \cap B' \in d(\mathcal{I})$. *Check.*

Now that $d(\mathcal{I}) =: \Sigma$ is both pi and d, it is a sigma field.
Since $\mathcal{I} \in \Sigma$, $\sigma(\mathcal{I}) \subseteq \Sigma$.
For any other d-system $\mathcal{D}' \supseteq \mathcal{I}$. Therefore

$$\mathcal{I} \subseteq \sigma(\mathcal{I}) \subseteq \Sigma := d(\mathcal{I}) \subseteq \mathcal{D}' \tag{1.9}$$

For any d-system $\mathcal{D}' \supseteq \mathcal{I}$. ∎

Rm. We claim without proof that $\sigma(E) \supseteq d(E)$ for any set E. Generated sigma field is always more complex then generated d. Dynkin's suggests that, if $E = \mathcal{I}$ is pi system, then

$$d(\mathcal{I}) = \sigma(\mathcal{I}) \quad \mathcal{I} \text{ - pi system.} \tag{1.10}$$

1.2 Measure

1.3 Events

1.3.1 Events as Sets

Def. **Events**: In prob space $(\Omega, \mathcal{F}, \mathbb{P})$, set $E \in \mathcal{F}$ is an event.

- If $\mathbb{P}(E) = 1$, say E happens **Almost Surely**. If $\mathbb{P}(E) = 0$, say E happens **Almost Nowhere**.

[1] $d(\mathcal{I})$ is d system, $d(\mathcal{I})$ subset $\mathcal{I}$ no other d system subset $d(\mathcal{I})$ supset $\mathcal{I}$

1.3.2 IO and EV

Def. **$\mathbf{E}_n$ Infinitely Often**: Sequence of events $\{E_n\} \in \mathcal{F}$, define:

$$\{E_n \ i.o.\} = \limsup_{n\to\infty} E_n := \bigcap_{n\geq 1} \bigcup_{m\geq n} E_m \tag{1.11}$$

Clearly $U_n = \bigcup_{m\geq n} E_m \searrow \{E_n \ i.o.\}$. Because it is a union of less and less sets. Therefore by continuity from above:

$$\{E_n \ i.o.\} = \lim_{n\to\infty} U_n \tag{1.12}$$

Def. **$\mathbf{E}_n$ Eventually Always**: Sequence of events $\{E_n\} \in \mathcal{F}$, define:

$$\{E_n \ e.v.\} = \liminf_{n\to\infty} E_n := \bigcup_{n\geq 1} \bigcap_{m\geq n} E_m \tag{1.13}$$

Clearly $A_n = \bigcap_{m\geq n} E_m \nearrow \{E_n \ i.o.\}$. Because it is an intersection of less and less sets. Therefore by continuity from below:

$$\{E_n \ e.v.\} = \lim_{n\to\infty} A_n \tag{1.14}$$

Prop. **Properties about limit events**

Basic:

1. $\liminf_{n\to\infty} E_n \subseteq \limsup_{n\to\infty} E_n$
2. $(\limsup_{n\to\infty} E_n)^c = \liminf_{n\to\infty} E_n^c$
3. $\lim_{n\to\infty} E_n = E \iff \limsup_{n\to\infty} E_n = \liminf_{n\to\infty} E_n = E$

Cap/Cup:

4. $(\limsup_{n\to\infty} A_n) \cup (\limsup_{n\to\infty} B_n) = \limsup_{n\to\infty}(A_n \cup B_n)$
5. $(\limsup_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} B_n) \supseteq \limsup_{n\to\infty}(A_n \cap B_n)$
6. $(\liminf_{n\to\infty} A_n) \cap (\liminf_{n\to\infty} B_n) = \liminf_{n\to\infty}(A_n \cap B_n)$
7. $(\liminf_{n\to\infty} A_n) \cup (\liminf_{n\to\infty} B_n) \subseteq \liminf_{n\to\infty}(A_n \cup B_n)$

Setminus:

8. $(\limsup_{n\to\infty} E_n) \setminus (\liminf_{n\to\infty} E_n) = \limsup_{n\to\infty}(E_n \setminus E_{n+1})$

With Measure:

9. $\mathbb{P}(\liminf_{n\to\infty} E_n) \leq \liminf_{n\to\infty} \mathbb{P}(E_n) \leq \limsup_{n\to\infty} \mathbb{P}(E_n) \leq \mathbb{P}(\limsup_{n\to\infty} E_n)$[2]
10. If $\lim_{n\to\infty} E_n = E$, then $\mathbb{P}(\lim_{n\to\infty} E_n) = \mathbb{P}(E)$.

Proofs. for some of above.

[2] First $\leq$ is Fatou's lemma, third is reverse-Fatou's lemma.

8.

9. i.e.

$$\begin{aligned}\mathbb{P}(\{E_n \ e.v\}) &\leq \liminf_{n\to\infty} \mathbb{P}(E_n) \\ &\leq \limsup_{n\to\infty} \mathbb{P}(E_n) \leq \mathbb{P}(\{E_n \ i.o\})\end{aligned} \tag{1.15}$$

$\mathbb{P}(\{E_n \ i.o\}) = \mathbb{P}(\lim_{n\to\infty} U_n)$
Cont from above (need finiteness of $\mathbb{P}$!): $\mathbb{P}(\lim_{n\to\infty} U_n) = \lim_{n\to\infty} \mathbb{P}(U_n)$.
Clearly $\mathbb{P}(U_n) \geq \sup_{n\geq m} \mathbb{P}(E_n)$.[3] Take limit both side:
$\mathbb{P}(\{E_n \ i.o\}) = \lim_{n\to\infty} \mathbb{P}(U_n) \geq \lim_{n\to\infty} \sup_{n\geq m} \mathbb{P}(E_n) =: \limsup_{n\to\infty} \mathbb{P}(E_n)$. ■

1.3.3 Fatou's Lemma

Lemma (**Reverse FATOU** - Need Finiteness of $\mathbb{P}$)

$$\limsup_{n\to\infty} \mathbb{P}(E_n) \leq \mathbb{P}(\{E_n \ i.o\}) \tag{1.16}$$

Lemma (**FATOU** - Apply for General Measure)

$$\mathbb{P}(\{E_n \ e.v\}) \leq \liminf_{n\to\infty} \mathbb{P}(E_n) \tag{1.17}$$

1.3.4 Borel-Cantelli 1st Lemma

Thm. (**BC 1**) In $(\Omega, \mathcal{F}, \mathbb{P})$, $\{E_n\} \subseteq \mathcal{F}$:

$$\sum_{n\geq 1} \mathbb{P}(E_n) < \infty \Rightarrow \mathbb{P}(\{E_n \ i.o\}) = 0 \tag{1.18}$$

Proof. Since $U_n \searrow \{E_n \ i.o\} \Rightarrow U_n \subseteq U_{n-1} \subseteq ... \subseteq U_1$.

$$\begin{aligned}\mathbb{P}(\{E_n \ i.o\}) &= \lim_{n\to\infty} \mathbb{P}(U_n) \\ &\leq \mathbb{P}(U_1) \\ &\leq \sum_{m\geq 1} \mathbb{P}(E_m) = 0 \quad \blacksquare\end{aligned} \tag{1.19}$$

[3] LHS is union, RHS is picking maximum from E_n.

Chapter 2

Mapping, RV

2.1 Measurable Function

2.2 Random Variable

2.3 Law, Distribution Function

2.4 Convergence of RV

Chapter 3

Independence

3.1 Independence: Sets

3.1.1 Indep Events

Def. **Mutually Independent Events**: Events in $\{E_n\}$ sequence are mutually indep. $\iff$ whatever $k \geq 1$, index-subsequence $\{n_1, n_2, ...n_k\}$:

$$\mathbb{P}(E_{n_1} \cap E_{n_2} \cap ... \cap E_{n_k}) = \prod_{j=1}^{k} \mathbb{P}(E_{n_j}) \tag{3.1}$$

Rm.
- $A \perp B \iff A^c \perp B \iff A^c \perp B^c$
- If $\mathbb{P}(A) = 1$ or $0 \Rightarrow A \perp \ \forall B \in \mathcal{F}$.

Def. **Pairwise Indep**: $\{E_n\}$ sequence are pairwise indep if $\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\,\mathbb{P}(E_j)$, $\forall i \neq j$.

3.1.2 Indep Sigma Field

Def. **Independent Sigma Field**: Sequence (Not necessarily finite) of sub sigma-fields $\mathcal{G}_1, \mathcal{G}_2...$ of $\mathcal{F}$ are indep. if, for any k, any subsequence of k distinct members: $\{\mathcal{G}_{n_1}, \mathcal{G}_{n_2}, ..., \mathcal{G}_{n_k}\}$ ($\{n_k\}$ distinct), any choice of set $G_i \in \mathcal{G}_i$:

$$\mathbb{P}(G_{n_1} \cap G_{n_2} \cap ... \cap G_{n_k}) = \prod_{j=1}^{k} \mathbb{P}(G_{n_j}) \tag{3.2}$$

Def. **Indep of Events - Redefine**: Events $\{E_n\}$ are indep if sigma field $\{\mathcal{E}_n\}$ are indep, where

$$\mathcal{E}_i = \{\emptyset,\ E_i,\ \Omega \setminus E_i,\ \Omega\} \tag{3.3}$$

3.1.3 Pi System Lemma

Thm. (*Study indep via generator pi systems*) $\mathcal{G}, \mathcal{H}$ are sub-sigma field of $\mathcal{F}$. $\mathcal{I}, \mathcal{J}$ are pi systems, where $\sigma(\mathcal{I}) = \mathcal{G}$, $\sigma(\mathcal{J}) = \mathcal{H}$. *Then*

$$\mathcal{G} \perp \mathcal{H} \iff \mathcal{I} \perp \mathcal{J}$$

i.e. $\forall\ I \in \mathcal{I},\ J \in \mathcal{J}$:

$$\mathbb{P}\left(I \cap J\right) = \mathbb{P}\left(I\right)\mathbb{P}\left(J\right) \tag{3.4}$$

Proof.

3.1.4 Borel-Cantelli 2nd Lemma

Thm. (**BC 2**) $\{E_n\}$ is a seq of **INDEPENDENT** events, *then*

$$\sum_{n\geq 1} \mathbb{P}\left(E_n\right) = \infty \Rightarrow \mathbb{P}\left(\{E_n\ i.o\}\right) = 1 \tag{3.5}$$

Proof. Do the complement, i.e. $\mathbb{P}\left(\{E_n^c\ e.v\}\right) = 0$.

$$\{E_n^c\ e.v\} = \liminf_{n\to\infty} E_n^c = \bigcup_{n\geq 1}\bigcap_{m\geq n} E_m^c = \bigcup_{n\geq 1} A_n \tag{3.6}$$

$$\begin{aligned} \mathbb{P}\left(A_n\right) = \mathbb{P}\left(\bigcap_{m\geq n} E_m^c\right) &= \prod_{n\geq m}\left(1 - \mathbb{P}\left(E_n\right)\right) \\ &\leq \exp\left[-\sum_{m\geq n}\mathbb{P}\left(E_n\right)\right] = 0 \end{aligned} \tag{3.7}$$

So $\mathbb{P}\left(\{E_n^c\ e.v\}\right) \leq \sum_{n\geq 1}\mathbb{P}\left(A_n\right) = 0$. ∎

3.1.5 Tail Sigma Field, Kolmogorov 0/1

Def. **Tail Sigma Field** associated with a sequence of events:

$$\mathcal{T} := \bigcap_{n\geq 1}\sigma(\{E_m\} : m \geq n) = \bigcap_{n\geq 1}\sigma(E_n, E_{n+1}, E_{n+2}, ...) \tag{3.8}$$

Thm. (**Kolmogorov 0/1**) If $\{E_n\}$ is **Indep** sequence, $\mathcal{T}$ is tail associated with $\{E_n\}$. *Then*, $\mathbb{P}\left(A\right) = 0$ or $1\ \forall A \in \mathcal{T}$.

3.2 Independence: RV

3.2.1 With Expectations

Note: This section is introduced after chapter 4.

Lemma. X, Y are indep RV, $X \in \mathcal{L}^1$, then $\forall B \subseteq \mathscr{B}(\mathbb{R})$,

$$\mathbb{E}\left[X; Y \in B\right] = \mathbb{E}\left[X\right]\cdot\mathbb{P}\left(Y \in B\right) \qquad \# \tag{3.9}$$

Proof. If $X = \mathbb{1}_A$ indicator, # is clearly true.
By linearity, # holds for $X \in SF^+$.
By (MON), # holds for $X \in m\mathcal{F}^+$.
Since $X \in \mathcal{L}^1$, so do $X^\pm$. All integrals involved in # are finite, linearlity $\Rightarrow$ # holds for any $X \in m\mathcal{F}$. ∎

Thm. (*Indep: product in expectation is expectation of product.*) If X, Y indep, $X, Y \in \mathcal{L}^1$; *then* $XY \in \mathcal{L}$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof. Assume $Y = \mathbb{1}_A$. By lemma #, for all $X \in m\mathcal{F}$:

$$\mathbb{E}[XY] = \mathbb{E}[X; A] = \mathbb{E}[X]\mathbb{P}(A) = \mathbb{E}[X]\mathbb{E}[Y] \tag{3.10}$$

Implies thm holds for Y indicator. By linearity, holds for simples.
(MON) $\Rightarrow$ holds for non-negative.
Since $X, Y \in \mathcal{L}^1$, holds for $X^\pm, Y^\pm$. All integrals involved in equation are finite.
linearity $\Rightarrow$ holds for all $Y \in m\mathcal{F}$. ■

Cor. (*Composition with Borel function*) X, Y indep (does not require integrability in X, Y themselves), and f, g are Borel functions, $f(X) \in \mathcal{L}^1$, $g(Y) \in \mathcal{L}^1$; *then*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)] \tag{3.11}$$

Proof. Apply thm above. Note that $f(X), g(Y)$ are indep RVs. Since $f(X) \in m\sigma(X)$, $g(Y) \in m\sigma(Y)$, $X \perp Y$. ■

Cor. (*Covariance*): If X, Y are serially uncorrelated, *then* $\mathrm{Cov}[X, Y] = 0$. Moreover, if process $\{X_n\} \in \mathcal{L}^2$, *then* define $S_n := \sum_1^n X_j$ as partial sum, we have $\mathrm{Var}[S_n] = \sum_1^n \mathrm{Var}[X_j]$

Chapter 4

Integration, Expectation

4.1 Integration

4.1.1 Integrability, L1 Space

Default setting: in general (abstract) measure space (S, Σ, μ).

Def. **Integrable**: $f \in m\Sigma$ is μ-integrable, denote $f \in \mathcal{L}^1(S, \Sigma, \mu)$ if both $\mu(f^+)$ and $\mu(f^-)$ are finite.
$\iff \mu(|f|) < \infty$

Prop. **Properties of $\mathcal{L}^1$ Functions**: if $f \in \mathcal{L}^1$

- $\mu(\{f = \pm\infty\}) = 0$.
- $|\mu(f)| \le \mu(|f|)$
- (*linearity*) if $f, g \in \mathcal{L}^1$, $a, b \in \mathbb{R}$ then $af + bg \in \mathcal{L}^1$.
- (*monotonicity*) if $f \le g$ *a.e*, then $\mu(f) \le \mu(g)$.

Proof. for some

(*linearity*) First show $f + g \in \mathcal{L}^1$. In that $|f + g| \le |f| + |g|$ everywhere $\Rightarrow$ $\mu(|f + g|) \le \mu(|f|) + \mu(|g|)$.
Then prove linearity. Let $h := f + g$, $h^+ - h^- = f^+ - f^- + g^+ - g^-$. Shift to obtain plus sign: $h^+ + f^- + g^- = h^- + f^+ + g^+$. $\Rightarrow \mu(h^+ + f^- + g^-) = \mu(h^- + f^+ + g^+)$. Apply linearity for $m\Sigma^+$ functions both sides. Also since $h^\pm, f^\pm, g^\pm \in \mathcal{L}^1$, we can shift things back. ∎

4.2 Convergence Theorems

Default setting: In general measure space (S, Σ, μ). $f_n : S \mapsto \bar{\mathbb{R}}$ (extended real line), $f : S \mapsto \bar{\mathbb{R}}$; $f_n, f \in m\Sigma$.

4.2.1 Monotone Convergence Thm

Thm. (**MON**) If $f_n \nearrow f$, and $\mu(f_1^-) < \infty$; *then* $\mu(f_n) \nearrow \mu(f)$.

Rm. (**MON**) still applies if $f_n \nearrow f$ *a.s.* This MON is also a more general version, which only requires one support from $\mu(f_1^-)$.

Cor. (**Nonnegative - MON**): If $f_n \in m\Sigma^+$, $f_n \nearrow f$, then $\mu(f_n) \nearrow \mu(f)$.

Cor. (**Reverse - MON**): If $f_n \searrow f$, and $\mu(f_1^-) \leq \infty$; *then* $\mu(f_n) \searrow \mu(f)$.

Proof. Define $g_n := f_1^+ - f_n$, then $g_n \geq 0$ in that $f_1^+ \geq f_n^+ \geq f_n$. Clearly $g_n \nearrow g := f_1^+ - f$. Apply (**MON**) to $\{g_n\}$:

4.2.2 Fatou's Lemma

Thm. (**FATOU**): If exists $g : S \mapsto \bar{\mathbb{R}}$, $g \in m\Sigma$, $\mu(g^-) < \infty$. And that $f_n \geq g$ uniformly $\forall n \geq 1$. *Then,*

$$\mu(\liminf_{n\to\infty} f_n) \leq \liminf_{n\to\infty}(\mu(f_n)) \tag{4.1}$$

Proof. Define $g_n := \inf_{m\geq n} f_m$, clearly

$$g_n \nearrow \sup_{n\geq 1}\inf_{m\geq n} f_m =: \liminf_{n\to\infty} f_n \tag{4.2}$$

$g_n = \inf_{m\geq n} f_m \geq g$[1] for $\forall n$. So $g_n^- \leq g^-$.

Thus $\{g_n\}$ are supported by $\mu(g_1^-) \leq \mu(g^-) < \infty$. Apply **(MON)**: $\mu(g_n) \nearrow \mu(\liminf_{n\to\infty} f_n)$

On the other hand, $g_n + g^- = g_n^+ + (g^- - g_n^-) \in m\Sigma^+$ non-negative. By definition of g_n: $0 \leq g_n + g^- \leq f_m + g^- \in m\Sigma^+$ $\forall m \geq n$. Use monotonicity/linearity for plus non-negatives:

$$\begin{aligned} \mu(g_n + g^-) &\leq \mu(f_m + g^-) \\ \mu(g_n) + \mu(g^-) &\leq \mu(f_m) + \mu(g^-) \\ \mu(g_n) &\leq \inf_{m\geq n}\mu(f_m) \end{aligned} \tag{4.3}$$

Holds for all $n \geq 1$. Let $n \to \infty$ both sides in increasingly:

$$\mu(\sup_{n\geq 1}\inf_{m\geq n} f_m) \nwarrow \mu(\inf_{m\geq n} f_m) \leq \inf_{m\geq n}\mu(f_m) \nearrow \sup_{n\geq 1}\inf_{m\geq n}\mu(f_n) \tag{4.4}$$

where nwarrow follows (**MON**), nearrow is just taking limit directly. Anyway, we have:

$$\mu(\sup_{n\geq 1}\inf_{m\geq n} f_m) \leq \sup_{n\geq 1}\inf_{m\geq n}\mu(f_n) \quad \blacksquare \tag{4.5}$$

Rm. $\leq$ in (**FATOU**) can be strict $<$. Consider: $f_n = \mathbb{1}_{[n,n+1]}$ is a moving hat to plus inf. Clearly $\liminf_{n\to\infty} f_n = 0$, because for any x, after $N > x$, $f_n(x) \equiv 0$. But $\mu(f_n) \equiv 1$. $0 = \mu(\liminf_{n\to\infty} f_n) < \liminf_{n\to\infty}\mu(f_n) = \mu(f_n) = 1$.

Thm. (**Reverse - FATOU**) If exists $g : S \mapsto \bar{\mathbb{R}}$, $g \in m\Sigma$, $\mu(g^+) < \infty$. And that $f_n \leq g$ uniformly $\forall n \geq 1$. *Then,*

$$\mu(\limsup_{n\to\infty} f_n) \geq \limsup_{n\to\infty}(\mu(f_n)) \tag{4.6}$$

[1] Since every f_m in infimum $\geq g$ uniformly.

4.2.3 Dominated Convergence Thm

Thm. (**DOM**) $f_n \xrightarrow{a.s.} f$. For some $g \in \mathcal{L}^1$, $|f_n| \leq g$ uniformly. *Then* $f_n \xrightarrow{\mathcal{L}^1} f$.
In particular $f \in \mathcal{L}^1$, $\mu(f_n) \to \mu(f)$.

Proof. Clearly $f_n \in \mathcal{L}^1$ for all n.
$|f| = \lim_{n\to\infty} |f_n| \leq g$. So $f \in \mathcal{L}^1$.
By pointwise $(a.s.)$ convergence, $\limsup_{n\to\infty} |f_n - f| = 0$.
Moreover $|f_n - f| \leq 2|g|$ uniformly. $\mu(g^+) < \infty$. Apply (**Reverse - FATOU**):

$$0 = \mu(\limsup_{n\to\infty} |f_n - f|) \geq \limsup_{n\to\infty}(\mu(|f_n - f|)) \geq 0 \tag{4.7}$$

So,

$$0 = \limsup_{n\to\infty}(\mu(|f_n - f|)) \geq \liminf_{n\to\infty} \mu(|f_n - f|) \tag{4.8}$$

i.e. $\limsup_{n\to\infty}(\mu(|f_n - f|)) = \liminf_{n\to\infty} \mu(|f_n - f|) = 0$. Therefore $f_n \xrightarrow{\mathcal{L}^1} f$. ■

Rm. $\xrightarrow{a.s.}$ **Does NOT imply** $\xrightarrow{\mathcal{L}^1}$, so the dominant condition cannot be dropped. $\xrightarrow{\mathcal{L}^1}$ **Does NOT imply** $\xrightarrow{a.s.}$ either.

4.2.4 Scheffe's Lemma

Thm. (**SCHEFFE**) $f, f_n \in \mathcal{L}^1$, $f_n \xrightarrow{a.s.} f$. *Then*

$$\mu(f_n) \to \mu(f) \iff f_n \xrightarrow{\mathcal{L}^1} f \tag{4.9}$$

Proof. $\Leftarrow$ is clear. Prove $\Rightarrow$.
Define $g_n := |f_n| + |f| - |f_n - f| \geq 0$ uniformly. Just for checking, $\mu(0^+) < \infty$.
Apply (**FATOU**) to g_n:

$$\mu(\liminf_{n\to\infty} |f_n| + |f| - |f_n - f|) \leq \liminf_{n\to\infty} \mu(|f_n| + |f| - |f_n - f|) \tag{4.10}$$

By $a.s$ convergence $\mu(\limsup_{n\to\infty} |f_n - f|) = 0$.

$$\begin{aligned} LHS &= \mu(2|f| - \limsup_{n\to\infty} |f_n - f|) \\ &= 2\mu(|f|) - \mu(\limsup_{n\to\infty} |f_n - f|) \end{aligned} \tag{4.11}$$

Note that inf is switched to sup when taking minus out.

$$RHS = 2\mu(|f|) - \limsup_{n\to\infty} \mu(|f_n - f|) \tag{4.12}$$

$f \in \mathcal{L}^1$ so it can be cancelled out.

$$0 = \mu(\limsup_{n\to\infty} |f_n - f|) \geq \limsup_{n\to\infty} \mu(|f_n - f|) \tag{4.13}$$

$\mu(|f_n - f|) \to 0$. ■

4.3 Radon-Nikodyn Thm

Def. $f\mu$ **measure**: $f \in (m\Sigma)^+$ **Non-negative!**, $f\mu$ is a new measure on measurable space (S, Σ) defined for $A \in \Sigma$ as

$$f\mu(A) := \int_A f d\mu = \mu(f\mathbb{1}_A) \tag{4.14}$$

Easy to check that this definition is indeed a measure (contable additive).

Prop. For $h \in (m\Sigma)^+$ (**Non-negative**): $(f\mu)(h) = \mu(fh)$ (#).

Proof. Let $h = \mathbb{1}_A$, $A \in \Sigma$. Then

$$(f\mu)(h) := \int_\Omega f\mathbb{1}_A d\mu = \mu(f\mathbb{1}_A) = \mu(fh) \tag{4.15}$$

holds for indicators. By linearity, (#) holds for $h \in SF^+$.
By (**MON**), (#) holds for $h \in (m\Sigma)^+$. ∎

Cor. For $h \in m\Sigma$ (**General** fucntion now!), *then*,

$$h \in \mathcal{L}^1(S, \Sigma, f\mu) \iff f \cdot h \in \mathcal{L}^1(S, \Sigma, \mu) \tag{4.16}$$

In particular, if this($h \in \mathcal{L}^1$) is the case, then $(f\mu)(h) = \mu(fh)$ (#).

Proof. $h \in \mathcal{L}^1(S, \Sigma, \mu) \iff f\mu(h^+) = \mu(fh^+) < \infty$ and $f\mu(h^-) = \mu(fh^-) < \infty$.
Since $f \in m\Sigma^+$, above $\iff \mu(fh^+) = \mu((fh)^+) < \infty$, $\mu(fh^-) = \mu((fh)^-) < \infty$.
$\iff \mu(fh) < \infty \iff f \cdot h \in \mathcal{L}^1(S, \Sigma, \mu)$. The equality is clearly true. ∎

Thm. (**Radon-Nikodyn**) If μ, λ are measures on (S, Σ), both are σ-finite. Moreover, if λ is absolutely continous wrt μ, [2] *Then*,
Exists $f \in (m\Sigma)^+$, such that $\lambda = f\mu$. Define *Radon-Nikodyn derivative* of λ wrt μ as this f. Denote

$$f =: \frac{d\lambda}{d\mu}$$

4.4 Expectation

4.4.1 Notation

Def. **Expectation**: $(\Omega, \mathcal{F}, \mathbb{P})$, $X : \Omega \mapsto \bar{\mathbb{R}}$.

$$\mathbb{E}[X] := \int_\Omega X d\mathbb{P} \tag{4.17}$$

Def. **Integrability**: $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ if $\mathbb{E}[X] < \infty$.

[2] i.e.(**ab.cont**) $\forall A \in \Sigma$, $\mu(A) = 0 \Rightarrow \lambda(A) = 0$.

4.4.2 Convergence Theorems

Default setting: in prob space $(\Omega, \mathcal{F}, \mathbb{P})$. Sequence of RVs $X_n : \Omega \mapsto \bar{\mathbb{R}}$, $X : \Omega \mapsto \bar{\mathbb{R}}$ and $X_n, X \in m\mathcal{F}$. (Note: **NOT** imposing $X_n, X \in \mathcal{L}^1$ here in general.)

Thm. (**MON**): $X_n \nearrow X$ ($\xrightarrow{a.s.}$), $\mathbb{E}[X_1^-] < \infty$; *then* $\mathbb{E}[X_n] \nearrow \mathbb{E}[X]$.

Thm. (**FATOU**): $\mathbb{E}[X^-] < \infty$, $X_n \geq X$ for all $n \geq 1$ for some X; *then* $\mathbb{E}[\liminf\limits_{n\to\infty} X_n] \leq \liminf\limits_{n\to\infty} \mathbb{E}[X_n]$. (*liminf inside < liminf outside*)

Thm. (**Revserse. FATOU**) $\mathbb{E}[X^+] < \infty$, $X_n \leq X$ for all $n \geq 1$ for some X; *then* $\mathbb{E}[\limsup\limits_{n\to\infty} X_n] \geq \limsup\limits_{n\to\infty} \mathbb{E}[X_n]$. (*limsup inside < limsup outside*)

Thm. (**DOM**) $X_n \xrightarrow{a.s.} X$, $|X_n| \leq Y$ for some $Y \in \mathcal{L}^1$; *then* $X_n \xrightarrow{\mathcal{L}^1} X$, i.e. $\mathbb{E}[|X_n - X|] \to 0$.

Thm. (**SCHEFFE**) $X_n, X \in \mathcal{L}^1$, $X_n \xrightarrow{a.s.} X$; *then* $\mathbb{E}[X_n] \to \mathbb{E}[X] \iff X_n \xrightarrow{\mathcal{L}^1} X$. ($\Leftarrow$ is trivial)

Rm. (Strengthened version of convergence thms in Prob space) $X_n \xrightarrow{a.s.} X$ in **MON, DOM, SCHEFFE** can be replaced with $X_n \xrightarrow{i.p.} X$, same result can be obtained nevertheless.

4.4.3 Lp Space

Def. $\mathcal{L}^p$ **Integrable, p-th Moment,** $\mathcal{L}^p$ **Norm**: $1 \leq p < \infty$

- Define $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ if $|X|^p \in \mathcal{L}^1$, i.e. $\mathbb{E}[|X|^p] < \infty$.
- For $X \in \mathcal{L}^p$, define $\mathbb{E}[X^p]$ as p-th moment of X.
- Define $\mathcal{L}^p$ norm of X as:

$$\|X\|_p := (\mathbb{E}[|X|^p])^{\frac{1}{p}} \tag{4.18}$$

Prop. **Properties of** $\mathcal{L}^p$

- $\mathcal{L}^p$ is a vector space in $\mathbb{R}$.
- $\|X\|_p$ satisfies defining properties of norm:
  $\|X\|_p \geq 0$.
  $\|X\|_p = 0 \Rightarrow X = 0$ *a.s.*
  $\|cX\|_p = |c|\|X\|_p$, constant c.
  $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$ (*triangle ineq.*) Equal sign achieved at: $Y = cX$, constant $c \geq 0$.
- (**Minkowski ineq.**) Another name for the triangular built-in property of vector space (as $\mathcal{L}_p$ space).
- (**Cauchy-Schwartz ineq.**) If $X, Y \in \mathcal{L}^2$, then $XY \in \mathcal{L}^1$. And $\mathbb{E}[|XY|] \leq \|X\|_2\|Y\|_2$. Equal sign achieved at $Y = cX$.
- (**Holder's ineq.**) For $1 < p, q < \infty$, and $1/p + 1/q = 1$, $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$; *then* $XY \in \mathcal{L}^1$, and $\mathbb{E}[|XY|] \leq \|X\|_p\|Y\|_q$. This a generalized version of Cauchy-Schwartz.

- **Monotonicity** of $\|\cdot\|_p$. If $1 \leq p < q < \infty$, $X \in \mathcal{L}^q$; *then* $X \in \mathcal{L}^p$. Moreover $\|X\|_p \leq \|X\|_q$. Equal sign is achieved at $X = c$ constant $a.s.$
- $\mathcal{L}^p$ is *Banach Space* i.e. $\mathcal{L}^p$ is complete under metric $d(X,Y) = \|X - Y\|_p$. In particular, $\mathcal{L}^2$ is *Hilbert Space*: $\forall X, Y \in \mathcal{L}^2$, inner product:

$$\langle X, Y\rangle_2 = \int_\Omega XY d\mathbb{P} \tag{4.19}$$

Def **Variance**: Define the second moment of quantity $X - \mathbb{E}[X]$ (centered X): $\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X] \geq 0$ by monotonicity: $(\mathbb{E}[X^2])^{\frac{1}{2}} \geq \mathbb{E}[X]$.

Rm. Monotonicity of $\|\cdot\|_p$ can be proved by Holder's ineq. taking $Y = 1$, we do need the prob space where $\mathbb{P}(\Omega) = 1$.

4.4.4 Markov's Ineq.

Non-negative valued mapping $g : \mathbb{R} \mapsto [0, +\infty]$ is *non-decreasing* Borel function ($g \in m\mathscr{B}$). *Then* for all constant $c \in \mathbb{R}$:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(X); X \geq c] = \int_{\{X \geq c\}} g(X) d\mathbb{P} \geq g(c)\mathbb{P}(X \geq c) \tag{4.20}$$

Rearrange this ineq, we estimate the upper bound of probability $\mathbb{P}(X \geq c)$ by $\mathbb{E}$ and some pre-determined function evaluated at this constant c, i.e.

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[g(X); X \geq c]}{g(c)} \leq \frac{\mathbb{E}[g(X)]}{g(c)} \tag{4.21}$$

gives **Markov's Ineq.**

Rm. This upperbound is meaningful only if at least $g(X) \in \mathcal{L}^1$. Also, the second $\leq$ uses the fact that g is non-negative valued.

EX.1 $g(X) = X$ identity. $X \in \mathcal{L}^1$, $X \geq 0$ (non-negative) *then*:

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c} \tag{4.22}$$

EX.2 Take $g(X) = |X|^p \cdot \mathbb{1}_{(0,+\infty)}$. If $X \in \mathcal{L}^p$, *then*:

$$\mathbb{P}(X \geq c) = \mathbb{P}(|X|^p \geq c^p) \leq \frac{\mathbb{E}[|X|^p]}{c^p} \tag{4.23}$$

EX.3 Take $g(X) = e^{a|X|} \cdot \mathbb{1}_{(0,+\infty)}$ for some a. If $e^{a|X|} \in \mathcal{L}^1$, *then*:

$$\mathbb{P}(X \geq c) = \mathbb{P}(e^{a|X|} \geq e^{ac}) \leq \frac{\mathbb{E}[e^{a|X|}]}{e^{ac}} \tag{4.24}$$

Prop. $X_n \xrightarrow{\mathcal{L}^p} X \Rightarrow X_n \xrightarrow{i.p} X$.

Proof. Use Markov's ineq. $\forall \epsilon > 0$

$$\mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{P}(|X_n - X|^p \geq \epsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p} \tag{4.25}$$

$X_n \xrightarrow{\mathcal{L}^p} X \Rightarrow \mathbb{E}[|X_n - X|^p] \xrightarrow{n\to\infty} 0 = RHS$. ∎

Rm.
- $X_n \xrightarrow{\mathcal{L}^p} X$ **Does not imply** $X_n \xrightarrow{a.s.} X$.
- $X_n \xrightarrow{a.s.} X$ **Does not imply** $X_n \xrightarrow{\mathcal{L}^p} X$ either. DOM, SHEFFE supports this arrow because they imposes extra condtions.
- $X_n \xrightarrow{i.p.} X$ **Does not imply** $X_n \xrightarrow{\mathcal{L}^p} X$.
- However $X_n \xrightarrow{i.p.} X$ plus some extra conditions can lead to $X_n \xrightarrow{\mathcal{L}^p} X$. Conditions can be: **DOM**, **SCHEFFE** (note that i.p. and a.s. are equivalent hypothesis for these two in $(\Omega, \mathcal{F}, \mathbb{P})$), or **Unifrom Integrable**.

4.4.5 Uniform Integrablility

Prop. (*Motivation for Unif.Integrability*)

$$X \in \mathcal{L}^1 \iff \lim_{M\to\infty} \mathbb{E}[|X|; |X| > M] = 0$$

Proof. $\Leftarrow$: Let $C := \sup_{M>1} \mathbb{E}[|X|; |X| \geq M]$. By hypothesis, this is bounded, i.e. $C < \infty$. And

$$\begin{aligned} \mathbb{E}[|X|] &= \mathbb{E}[|X|; |X| > M] + \mathbb{E}[X; |X| \leq M] \\ &\leq M + C < \infty \end{aligned} \tag{4.26}$$

$\Rightarrow$: Consider $X_M := |X| \cdot \mathbb{1}_{\{|X| \leq M\}} \nearrow |X|$. Clearly $X_1 \in \mathcal{L}^1$. By *(MON)*: $\mathbb{E}[|X|; |X| \leq M] = \mathbb{E}[|X_M|] \nearrow \mathbb{E}[|X|] < \infty$.

$$\begin{aligned} \lim_{M\to\infty} \mathbb{E}[|X|; |X| > M] &= \lim_{M\to\infty} \mathbb{E}[|X|] - \mathbb{E}[|X|; |X| \leq M] \\ &= 0 \quad ∎ \end{aligned} \tag{4.27}$$

Def. **Unifrom Integrable**: Sequence of RV $\{X_n\}$ is *U.I.* if

$$\lim_{M\to\infty} \sup_{n\geq 1} \mathbb{E}[|X_n|; |X_n| > M] = 0 \tag{4.28}$$

Rm. U.I. says that for all $\epsilon > 0$, exists M large, such that $\mathbb{E}[|X_n|; |X_n| > M] < \epsilon$ **uniformly** for all $n \geq 1$.

Prop. (*Strength of U.I. hypothesis*)

- X_n U.I. $\Rightarrow \{|X_n|\}$ is unifromly bounded in $\mathcal{L}^1$.
- $\{|X_n|\}$ is unifromly bounded in $\mathcal{L}^p$ for all $p \geq 1 \Rightarrow X_n$ U.I.

Rm. Say $\{|X_n|\}$ is *unifromly bounded* in $\mathcal{L}^p$ if: $\forall n \geq 1$, $\exists M < \infty$ is *irrelevant* to n; such that $\mathbb{E}[|X_n|^p] < M$. *OR* just:

$$\sup_n \mathbb{E}[|X_n|^p] < M \tag{4.29}$$

Proof. (1): By hypothesis, $\exists M$ large, $\sup_n \mathbb{E}[|X_n|; |X_n| > M] < \epsilon$.

$$\begin{aligned}\mathbb{E}[|X_n|] &= \mathbb{E}[|X_n|; |X_n| \le M] + \mathbb{E}[|X_n|; |X_n| > M] \\ &\le M + \sup_n \mathbb{E}[|X_n|; |X_n| > M]. \quad \blacksquare\end{aligned} \tag{4.30}$$

(2): By hypothesis, $\sup_n \mathbb{E}\left[|X_n|^p\right] < C < \infty$.

$$\begin{aligned}\mathbb{E}[|X_n|; |X_n| > M] &\le \mathbb{E}\left[\frac{|X_n|^{p-1}}{M^{p-1}} \cdot |X_n|; |X_n| > M\right] \\ &= \frac{1}{M^{p-1}} \mathbb{E}\left[|X_n|; |X_n|^p > M\right] \\ &\le \frac{\sup_n \mathbb{E}\left[|X_n|^p\right]}{M^{p-1}} \le \frac{C}{M^{p-1}} \xrightarrow{M \to \infty} 0. \quad \blacksquare\end{aligned} \tag{4.31}$$

Thm. (*the **Exact Gap** between i.p. and $\mathcal{L}^1$ convergence*)

$$X_n \xrightarrow{\mathcal{L}^1} X \iff X_n \xrightarrow{i.p} X \text{ and } \{X_n\} \text{ is U.I.} \tag{4.32}$$

Proof.

4.4.6 Jensen's Ineq.

Def. **Convex Mapping**: $\phi : \mathbb{R} \mapsto \mathbb{R}$, if $x, y \in \mathbb{R}$, $p, q \in (0, 1)$, $p + q = 1$. ϕ is a convex function *if*:

$$\phi(px + qy) \le p\phi(x) + q\phi(y) \tag{4.33}$$

Prop. **Support Line**: If *phi* is convex, then $\forall x \in \mathbb{R}$, $\exists$ a line l crosses $(x, \phi(x))$; l stays entirely below the graph of ϕ.

$X \in \mathcal{L}^1$, $\phi : \mathbb{R} \mapsto \mathbb{R}$ is a convex mapping, $\phi \in \mathcal{L}^1$; *then*:

$$\phi(\mathbb{E}[X]) \le \mathbb{E}[\phi(X)] \tag{4.34}$$

gives **Jensen's Ineq.**, average inside $\le$ average outside.

Proof. Using support line. Exists l passes $(\mathbb{E}[X], \phi(\mathbb{E}[X]))$, say $y = ax + b$, that supports ϕ; i.e. $\forall w \in \Omega$:

$$aX(w) + b \le \phi(X(w)) \tag{4.35}$$

Take expectation both sides, and notice $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b = \phi(\mathbb{E}[X])$:

$$\phi(\mathbb{E}[X]) = \mathbb{E}[aX + b] \le \mathbb{E}[\phi(X)]. \quad \blacksquare \tag{4.36}$$

Cor. **Popular Convex**: $|X|$, X^2, e^{aX}, $X^+ := \max\{X, 0\}$, $X^- := \max\{-X, 0\}$ are convex, satisfy jensen.

4.4.7 On Prob Density Function

Recall law of X, $\mathcal{L}_X(B) := \mathbb{P}(X \in B)$, $B \in \mathscr{B}(\mathbb{R})$ is a new prob measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. For default setting in this subsection, consider $f : \mathbb{R} \mapsto \mathbb{R}$ is a *Borel mapping.* And RV $X : \Omega \mapsto \mathbb{R}$.

Also recall distribution function: $F_X(x) := \mathcal{L}_X(-\infty, x]$.

Prop. (*Transference of Integrability/Integral against* $\mathbb{P}$ *and* $\mathcal{L}_X$ *measure*):

$$f(X) = f \circ X(w) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \iff f(x) \in \mathcal{L}^1(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathcal{L}_X)$$

In particular, if $f \circ X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$:

$$\int_\Omega f(X(w))d\mathbb{P} = \int_\mathbb{R} f(x)d\mathcal{L}_X \quad (\#) \tag{4.37}$$

Rm. This thing is in fact transferring the relationship on *sets* level, i.e. $\mathcal{L}_X(B) = \mathbb{P}(X \in B)$ to *integral* level.

Proof. Let $f = \mathbb{1}_B$. Define preimage $X^{-1}(B) := \{w \in \Omega : X(w) \in B\} \subseteq \Omega$.

$$LHS = \int_\Omega \mathbb{1}_B(X(w))d\mathbb{P} = \int_\Omega \mathbb{1}_{X^{-1}(B)}d\mathbb{P} = \mathbb{P}(X^{-1}(B)) \tag{4.38}$$

$$RHS = \int_\mathbb{R} \mathbb{1}_B(x)d\mathcal{L}_X = \mathcal{L}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) \tag{4.39}$$

(#) holds for indicators.
By linearity $\Rightarrow$ (#) holds for $f \in SF^+$.
By (MON) $\Rightarrow$ (#) holds for $f \in [m\mathscr{B}(\mathbb{R})]^+$.
For general $f = f^+ - f^- \in \mathcal{L}^1$, $\int f^\pm < \infty$. By linearity, (#) holds for general f. ■

Cor. ($\mathbb{E}[X]$ *and* $\mathrm{Var}[X]$) For $X \in \mathcal{L}^1$, $X \in \mathcal{L}^2$ respectively:

$$\mathbb{E}[X] = \int_\mathbb{R} x d\mathcal{L}_X \tag{4.40}$$

$$\mathrm{Var}[X] = \int_\mathbb{R} (x - E[X])^2 d\mathcal{L}_X \tag{4.41}$$

Cor. (*Law as a lower level object*) If X, Y has identical Law $\mathcal{L}$, *then* $\forall f \in m\mathscr{B}$:

- $f(X) \in \mathcal{L}^1 \Rightarrow f(Y) \in \mathcal{L}^1$
- If $f(X) \in \mathcal{L}^1$:

$$\mathbb{E}[X] = \mathbb{E}[Y] = \int_\mathbb{R} f(x)d\mathcal{L} \tag{4.42}$$

Notation. (*Lebesgue-Stieltjes version of* $\mathbb{E}[X]$ *using dist function*)

$$\mathbb{E}[X] = \int_\mathbb{R} f(x)d\mathcal{L}_X = \int_\mathbb{R} f(x)dF_X \tag{4.43}$$

Def. **Probability Density Function**: RV X has p.d.f f_X, if

- $f_X : \mathbb{R} \mapsto [0, +\infty]$ is measurable.
- The *Radon-Nikodyn derivative* of measure $\mathcal{L}_X$ with respect to lebesgue measure μ exists. I.e. $\mathcal{L}_X$ is *absolutely continous* wrt μ.

Use dx as abbr of $d\mu_{leb}$, for f_X, if exists, we have:

$$f_X = \frac{d\mathcal{L}_X}{dx} \tag{4.44}$$

Prop. (*Transference of Integrability/Integral against* $\mathbb{P}$ *and lebesgue measure via p.d.f*): If f_X exists, and $h : \mathbb{R} \mapsto \mathbb{R}$ is Borel function, we have:

$$h(X) = h \circ X(w) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \iff hf_X \in \mathcal{L}^1(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu_{leb})$$

In particular, if $h \circ X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$:

$$\int_\Omega h(X(w))d\mathbb{P} = \int_\mathbb{R} h(x)f_X(x)dx \tag{4.45}$$

Rm. **Existance of p.d.f**

- f_X exists $\Rightarrow$ F_X is continous *everywhere*.

  Assume otherwise with discontinuity $\{x_0\}$, $F(x_0^+) - F(x_0^-) > 0$. Then $\mathcal{L}_X(\{x_0\}) > 0$, not absolutely cont wrt μ_{leb}.
- f_X exists $\Rightarrow$ F_X is differentiable *a.e.*

  $F_X(y) = \int_{-\infty}^y f_X(x)dx$
  $F_X'(y) = f_X(y) \quad a.e.$
- F_X is differentiable *a.e.* **Does not imply** f_X exists.

  Counter Example: $\mathcal{L}_X = \delta_0$ is Dirac Delta function. $\mathcal{L}_X$ is not absolutely cont wrt μ_{leb}.
- F_X is differentiable *everywhere* $\Rightarrow$ f_X exists.

  $F_X'(y) = f_X(y) \quad a.e.$
- F_X is continuous *everywhere* **Does not imply** f_X exists.

  Counter Example: $\mathcal{L}_X$ is Cantor function (fractal structured), $\mathcal{L}_X$ is not absolutely cont wrt μ_{leb}.

Chapter 5

Law of Large Numbers

5.1 Terminology

Given process $\{X_n\}$, define partial sum $S_n := \sum_1^n X_j$. The **Strong/Weak Law of Large Number** is said to hold for $\{X_n\}$ in following two cases,

- In *Classical Setting*, say WLLN holds if

$$\frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{i.p} 0 \tag{5.1}$$

  Say SLLN holds if

$$\frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{a.s.} 0 \tag{5.2}$$

- In *General Setting* consider $\{a_n\} \in \mathbb{R}$, $\{b_n\} > 0$, $b_n \nearrow \infty$. WLLN for $\{X_n\}$ normalized by a_n, b_n, if

$$\frac{S_n - a_n}{b_n} \xrightarrow{i.p} 0 \tag{5.3}$$

  SLLN if

$$\frac{S_n - a_n}{b_n} \xrightarrow{a.s} 0 \tag{5.4}$$

We study the conditions under which WLLN and SLLN can hold. There are two types of them:

- Estimates/Controls on **Moments** (i.e. **Integrability**)
- Estimates/Controls on **Distributions**.

5.2 Chebyshev (WLLN1)

Thm. (**Chebyshev**) $\{X_n\}$ is a seq of RVs, satisfying

- (*Dist*) $\{X_n\}$ are uncorrelated, i.e. $\mathrm{Cov}[X_i, X_j] = 0 \ \forall i \neq j$.
- (*Moments*) $\{X_n\}$ is bounded by $\mathcal{L}^2$, i.e. $\sup_n \mathbb{E}[X_n^2] < \infty$.

Then WLLN holds for $\{X_n\}$.

Proof. WLOG assume $\mathbb{E}[X_n] = 0$.
Otherwise we can always take $Z_n = X_n - \mathbb{E}[X_n]$ be centerred X_n, which has zero means. Under this, $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\,\mathbb{E}[X_j] = 0$.

$$\begin{aligned}\mathbb{E}\left[S_n^2\right] &= \sum_{j=1}^{n} \mathbb{E}\left[X_j^2\right] + \sum_{1\le i\ne j\le n} \mathbb{E}[X_i X_j] \\ &\le n \cdot \sup_n \mathbb{E}\left[X_n^2\right]\end{aligned} \tag{5.5}$$

For all $\epsilon > 0$, using Markov's ineq,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \ge \epsilon\right) \le \frac{\mathbb{E}[S_n^2]}{n^2\epsilon^2} \le \frac{n \sup_n \mathbb{E}[X_n^2]}{n^2\epsilon^2} \xrightarrow{n\to\infty} 0 \tag{5.6}$$

So WLLN ($\xrightarrow{i.p}$) holds. ■

Rm. **Chebyshev's ineq**: If $X \in \mathcal{L}^2$, then

$$\mathbb{P}(|X - \mathbb{E}[X]| > c) \le \frac{\operatorname{Var}[X]}{c^2} \tag{5.7}$$

Says exactly same thing as Markov's.

5.3 Rajchmah (SLLN1)

Thm. (**Rajchmah**) Same hypothesis,

- (*Dist*) $\{X_n\}$ are uncorrelated, i.e. $\operatorname{Cov}[X_i, X_j] = 0\ \forall i \ne j$.
- (*Moments*) $\{X_n\}$ is bounded by $\mathcal{L}^2$, i.e. $\sup_n \mathbb{E}[X_n^2] < \infty$.

In fact we have SLLN.

Proof. WLOG assume $\mathbb{E}[X_n] = 0$. By proof of SLLN, we already have:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \ge \epsilon\right) \le \frac{M}{n\epsilon^2} \tag{5.8}$$

Where $M = \sup_n \mathbb{E}[X_n^2]$. But the whole thing (order $\frac{1}{n}$) is not summable. Consider subsequence $\{X_{n^2}\} \subseteq \{X_n\}$,

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| \ge \epsilon\right) \le \frac{\mathbb{E}\left[S_{n^2}^2\right]}{n^4\epsilon^2} \le \frac{n^2 M}{n^4\epsilon^2} = \frac{M}{n^2\epsilon^2} \tag{5.9}$$

is summable, i.e. for all $\epsilon > 0$,

$$\sum_{n\ge1} \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| \ge \epsilon\right) < \infty \tag{5.10}$$

By (BC1): $\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \epsilon \;\; i.o.\right) = 0$, which is $\iff$ $\left|\frac{S_{n^2}}{n^2}\right| \xrightarrow{a.s.} 0$. Holds for subsequence n^2.
Then define

$$D_n := \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}| \tag{5.11}$$

which somehow captures the worst deviation from n^2 subsequence.

$$\begin{aligned}\mathbb{E}\left[D_n^2\right] &\leq \sum_{k=n^2+1}^{(n+1)^2-1} \mathbb{E}\left[(S_k - S_{n^2})^2\right] \\ &= \sum_{k=n^2+1}^{(n+1)^2-1} \sum_{l=n^2+1}^{k} \mathbb{E}\left[X_l^2\right] \\ &\leq 2n \cdot 2n \cdot M = \Theta(n^2)\end{aligned} \tag{5.12}$$

Where $M = \sup_n \mathbb{E}\left[X_n^2\right]$, Second equal sign follows that $\{X_n\}$ are uncorrelated. Final leq is just counting the terms. We now has the idea to estimate D_n by its order n^2. For all $\epsilon > 0$, markov:

$$\mathbb{P}\left(\frac{|D_n|}{n^2} > \epsilon\right) \leq \frac{\mathbb{E}\left[D_n^2\right]}{n^4\epsilon^2} \leq \frac{4n^2M}{n^4\epsilon} = \frac{4M}{n^2\epsilon^2} \tag{5.13}$$

Which is summable ($\frac{1}{n^2}$).

$$\sum_{n\geq 1} \mathbb{P}\left(\frac{|D_n|}{n^2} > \epsilon\right) < \infty \tag{5.14}$$

BC1: $\mathbb{P}\left(\frac{|D_n|}{n^2} > \epsilon \;\; i.o.\right) = 0 \Rightarrow \frac{|D_n|}{n^2} \xrightarrow{a.s.} 0$.
For every $w \in \Omega$ such that both $\frac{|D_n|}{n^2} \to 0$ and $\left|\frac{S_{n^2}}{n^2}\right| \to 0$ occurs[1], for every $k \geq 1$, $\exists!\ n(k)$ such that $n^2(k) \leq k < (n(k)+1)^2$, and

$$\begin{aligned}\frac{|S_k|}{k} &\leq \frac{|S_k - S_{n^2(k)}| + |S_{n^2(k)}|}{n^2(k)} \\ &\leq \frac{|D_n|}{n^2(k)} + \frac{|S_{n^2(k)}|}{n^2(k)} \xrightarrow{k\to\infty} 0\end{aligned} \tag{5.15}$$

Which holds for $a.e.$. So $\frac{|S_k|}{k} \xrightarrow{a.s.} 0$. ■

Rm. (**Cantelli**)

- (*Dist*) $\{X_n\}$ are indep.
- (*Moments*) $\{X_n\}$ is bounded by $\mathcal{L}^4$.

Supports SLLN, much weaker then the one above.

[1]Since these two are a.s. convergence, w is in fact also a.s.

5.4 Khintchine (WLLN2) and Kolmogorov-Feller (WLLN3)

5.4.1 Equivalence of Seqs

Def. **Equivalence**: Two sequence $\{X_n\}, \{Y_n\}$ are called equivalent, if

$$\sum_{n\geq 1} \mathbb{P}(X_n \neq Y_n) < \infty \tag{5.16}$$

Prop. If X_n, Y_n are equivalent, then $\sum_{n\geq 1}(X_n - Y_n)$ converges almost everywhere. And $\forall b_n > 0$, $b_n \nearrow \infty$,

$$\frac{1}{b_n}\sum_{k=1}^{n}(X_k - Y_k) \xrightarrow{a.s.} 0 \tag{5.17}$$

Proof. BC1: $\mathbb{P}(X_n \neq Y_n \ \ i.o.) = 0 \Rightarrow \mathbb{P}(X_n = Y_n \ \ e.v.) = 1$. Which means X_n and Y_n are eventually the same.
For almost every $w \in \Omega$, $\exists N(w) > 0$, such that $\forall n > N(w)$, $X_n(w) = Y_n(w)$. So clearly

$$\sum_{n\geq 1}(X_n(w) - Y_n(w)) = \sum_{k=1}^{N(w)}(X_k(w) - Y_k(w)) < \infty \tag{5.18}$$

i.e. $\sum_{n\geq 1}(X_n(w) - Y_n(w))$ a.s. converges.
Moreover, for a.e. w,

$$\frac{1}{b_n}\sum_{n\geq 1}(X_n(w) - Y_n(w)) = \frac{1}{b_n}\sum_{k=1}^{N(w)}(X_k(w) - Y_k(w)) \xrightarrow{b_n\to\infty} 0 \tag{5.19}$$

i.e. $\frac{1}{b_n}\sum_{n\geq 1}(X_n - Y_n) \xrightarrow{a.s.} 0$. ■

5.4.2 Big O and Small o Notations

Def. **Big O and Small o Notations**: Assume $a_n \subseteq \mathbb{R}$, $\{b_n\} \subseteq \mathbb{R}^+$. $b_n \nearrow +\infty$.

- We write $a_n = O(b_n) \iff \exists c > 0$, $\exists N$ large, such that $\forall n > N$:

$$\frac{1}{c}b_n \leq a_n \leq cb_n \tag{5.20}$$

- We write $a_n = o(b_n) \iff$

$$\lim_{n\to\infty}\frac{a_n}{b_n} = 0 \tag{5.21}$$

Lemma (*Sum of converge-to-zero seq is o(n)*): If $\lim\limits_{n\to\infty} a_n = 0$, then $\sum_{j=1}^{n} a_i = o(n)$, i.e.

$$\lim_{n\to\infty}\frac{\sum_{j=1}^{n} a_j}{n} = 0 \tag{5.22}$$

Proof. Since a_n converges to zero $\Rightarrow \forall\epsilon$, $\exists N_1$, $\forall n > N_1$, $|a_n| < \frac{\epsilon}{2}$.

$$\left|\frac{1}{n}\sum_{j=1}^{n} a_j\right| \leq \frac{1}{n}\sum_{j=1}^{N_1}|a_j| + \frac{1}{n}\sum_{j=N_1+1}^{n}|a_j| \tag{5.23}$$

Clearly the second term $< \frac{\epsilon}{2}$ and the first term converges to zero, i.e. $\exists N_2 > 0$, $\forall n > N_2$: $\frac{1}{n}\sum_{j=1}^{N_1}|a_j| < \frac{\epsilon}{2}$. Take $N = \max\{N_1, N_2\}$, the whole thing $< \epsilon$. ■

5.4.3 Khintchine's WLLN

From now we are using the general sense of LLNs, specified in the first section.

Thm. (**Khintchine**) $\{X_n\}$ be a seq of RVs satisfying

- (*Dist*) $\{X_n\}$ are **pairwise indep.**, and **identically distributed**.
- (*Moments*) $m := \mathbb{E}[X_n] < \infty$. ($\mathcal{L}^1$).

Then,

$$\frac{S_n - nm}{n} \xrightarrow{i.p} 0; \quad \text{i.e.} \quad \frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{i.p} 0; \quad \text{i.e.} \quad \frac{S_n}{n} \xrightarrow{i.p} m = \mathbb{E}[X_n] \tag{5.24}$$

Proof. Consider truncated sequence $\{Y_n\}$,

$$Y_n = \begin{cases} X_n & \text{if } |X_n| \le n, \\ 0 & \text{otherwise.} \end{cases}$$

We hope that X_n and Y_n are *equivalent*. By definiton,

$$\sum_{n\ge1} \mathbb{P}(X_n \ne Y_n) = \sum_{n\ge1} \mathbb{P}(|X_n| > n) = \sum_{n\ge1} \mathbb{P}(|X_1| > n) \tag{5.25}$$

$$\begin{aligned} \infty > \mathbb{E}[|X_1|] &= \int_0^\infty \mathbb{P}(|X_1| > t)\,dt \\ &= \sum_{n\ge1} \int_{n-1}^n \mathbb{P}(|X_1| > t)\,dt \ge \sum_{n\ge1} \mathbb{P}(|X_1| > n) \end{aligned} \tag{5.26}$$

Therefore they are indeed equivalent. And clearly $\{Y_n\}$ are also pairwise indep. Define $T_n := \sum_{j=1}^n Y_j$

$$\mathrm{Var}[T_n] = \sum_{j=1}^n \mathrm{Var}[Y_j] \le \sum_{j=1}^n \mathbb{E}\left[Y_j^2\right] = \sum_{j=1}^n \mathbb{E}\left[X_j^2; |X_j| \le j\right] \tag{5.27}$$

Note that we can easily obtain a *crude* estimate of $\mathrm{Var}[T_n]$. $RHS \le \sum_1^n j \cdot \mathbb{E}[|X_j|] \le \mathbb{E}[|X_1|] \sum_1^n j = O(n^2)$, which is not sufficient. To show any WLLN, we **always** somehow need $\mathrm{Var}[S_n] = o(n^2)$.
Finer estimate is made by following. Consider l_n, such that, $0 < l_n \nearrow \infty$, $l_n < n$, and $l_n = o(n)$. For example, $l_n = \lfloor \sqrt{n} \rfloor$. Then,

$$\begin{aligned} \mathrm{Var}[T_n] &\le \left(\sum_{j=1}^{l_n} + \sum_{j=l_n+1}^{n} \right) \mathbb{E}\left[X_j^2; |X_j| \le j\right] \\ &\le \mathbb{E}[|X_1|] \sum_{j=1}^{l_n} j + \sum_{j=l_n+1}^{n} \left(\mathbb{E}\left[X_j^2; |X_j| \le l_n\right] + \mathbb{E}\left[X_j^2; l_n < |X_j| \le j\right] \right) \\ &\le \mathbb{E}[|X_1|]\, O(l_n^2) + \sum_{j=l_n+1}^{n} l_n \cdot \mathbb{E}[|X_1|] + \sum_{j=l_n+1} j \cdot \mathbb{E}[|X_1|; |X_1| > l_n] \\ &= \mathbb{E}[|X_1|] \cdot O(l_n^2) + \mathbb{E}[|X_1|] \cdot O(n l_n) + \mathbb{E}[|X_1|; |X_1| > l_n] \cdot O(n^2) \\ &= o(n^2) \end{aligned} \tag{5.28}$$

For the last equal sign to $o(n^2)$, notice that the first two terms are clearly $o(n^2)$, and since $X_1 \in \mathcal{L}^1 \Rightarrow \mathbb{E}[|X_1|; |X_1| > l_n] \xrightarrow{l_n \to \infty} 0$, so the third term is also $o(n^2)$ (actually zero at infinity). We have

$$\lim_{n\to\infty} \frac{\operatorname{Var}[T_n]}{n^2} = 0 \tag{5.29}$$

Apply (**Chebyshev**), for all $\epsilon > 0$,

$$\mathbb{P}\left(\frac{|T_n - \mathbb{E}[T_n]|}{n} > \epsilon\right) \le \frac{\operatorname{Var}[T_n]}{n^2\epsilon^2} \xrightarrow{n\to\infty} 0 \tag{5.30}$$

Implies

$$\frac{T_n - \mathbb{E}[T_n]}{n} \xrightarrow{i.p} 0 \tag{5.31}$$

Now we are going from T_n to S_n.

$$\begin{aligned}\frac{|S_n - \mathbb{E}[S_n]|}{n} &\le \frac{|S_n - T_n|}{n} + \frac{|T_n - \mathbb{E}[T_n]|}{n} + \frac{|\mathbb{E}[T_n] - \mathbb{E}[S_n]|}{n} \\ &= Q_1 + Q_2 + Q_3\end{aligned} \tag{5.32}$$

- For Q_1, since $\{X_n\}, \{Y_n\}$ are equivalent, use the property of equivalent sequence, $Q_1 \xrightarrow{a.s.} 0$.
- For Q_2, we already know $Q_2 \xrightarrow{i.p} 0$.
- For Q_3,

$$Q_3 \le \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[|X_j|; |X_j| > j] = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[|X_1|; |X_1| > j] \tag{5.33}$$

  By lemma in last subsection, sum of converge-to-zero seq is $o(n)$. Since $\lim_{n\to\infty} \mathbb{E}[|X_1|; |X_1| > n] = 0$, the sum is $o(n)$, implies $Q_3 \to 0$ (pointwise).

Pick the weakest convergence of three sub-quantities, $\frac{|S_n - nm|}{n} \xrightarrow{i.p} 0$. ■

Rm. To show any WLLN, we **always** somehow need $\operatorname{Var}[S_n] = o(n^2)$.

5.4.4 Kolmogorov-Feller's WLLN

Thm. (**Kolmogorov-Feller**) $\{X_n\}$ are **pairwise indep.**, some seq of numbers $\{b_n\}$, $0 < b_n \nearrow \infty$. $\{X_n\}$ satisfies,

$$\lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}(|X_j| > b_n) = 0 \tag{5.34}$$

$$\lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{E}\left[\frac{|X_j|^2}{b_n^2}; |X_j| \le b_n\right] = 0 \tag{5.35}$$

Then, if a_n is defined by $a_n := \sum_{j=1}^{n} \mathbb{E}[X_j; |X_j| \le b_n]$, we have

$$\frac{S_n - a_n}{b_n} \xrightarrow{i.p} 0 \tag{5.36}$$

5.5 Kolmogorov (SLLN2)

5.5.1 Kronecker's Lemma

Lemma (**Kronecker**) Two sequences of numbers, $\{x_n\} \subseteq \mathbb{R}$, $\{a_n\} \subseteq \mathbb{R}^+$, $a_n \nearrow \infty$, then

$$\sum_{n\geq 1} \frac{x_n}{a_n} \text{ Converges to finite value } \Rightarrow \frac{1}{a_n}\sum_{j=1}^{n} x_j \xrightarrow{n\to\infty} 0$$

Note the reverse direction ($\Leftarrow$) is **Not** true.

Proof. For $1 \leq n < \infty$, define

$$b_n := \sum_{j=1}^{n} \frac{x_j}{a_j} \tag{5.37}$$

By hypothesis, $\lim_{n\to\infty} b_n = b < \infty$. Let $b_0 = a_0 = 0$, clearly by definition $x_n = a_n(b_n - b_{n-1})$, so

$$\begin{aligned}
\frac{1}{a_n}\sum_{j=1}^{n} x_j &= \frac{1}{a_n}\sum_{j=1}^{n} a_j(b_j - b_{j-1}) \\
&= (b_n - b_{n-1}) + \frac{1}{a_n}\sum_{j=1}^{n-1} a_j b_j - \frac{1}{a_n}\sum_{j=1}^{n-1} a_j b_{j-1} \\
&= (b_n - b_{n-1}) + \frac{1}{a_n}\sum_{j=1}^{n-1} a_j b_j - \frac{1}{a_n}\sum_{j=0}^{n-2} a_{j+1} b_j \\
&= (b_n - b_{n-1}\frac{a_n}{a_n}) + \frac{1}{a_n}\sum_{j=1}^{n-1}(a_j - a_{j+1})b_j + \frac{a_{n-1}b_{n-1}}{a_n} \\
&= b_n - \frac{1}{a_n}(a_n - a_{n-1})b_{n-1} + \frac{1}{a_n}\sum_{j=1}^{n-1}(a_j - a_{j+1})b_j \\
&= b_n - \frac{1}{a_n}\sum_{j=1}^{n-1}(a_{j+1} - a_j)b_j \quad (\#)
\end{aligned} \tag{5.38}$$

This bunch of thing,

$$\frac{1}{a_n}\sum_{j=1}^{n} a_j(b_j - b_{j-1}) = b_n - \frac{1}{a_n}\sum_{j=1}^{n-1}(a_{j+1} - a_j)b_j \tag{5.39}$$

is actually **Abel Summation Formula** (discrete version of integration by parts). Note the telescoping sum $\frac{1}{a_n}\sum_0^{n-1}(a_{j+1} - a_j) = 1$, so $b_n = \frac{1}{a_n}\sum_0^{n-1} b_n(a_{j+1} - a_j)$, therefore

$$(\#) = \frac{1}{a_n}\sum_{j=1}^{n-1}(b_n - b_j)(a_{j+1} - a_j) \tag{5.40}$$

Since $\lim_{n\to\infty} b_n = b < \infty$, $\{b_n\}$ is Cauchy-sequence. I.e. $\forall\epsilon > 0$, $\exists N$, such that $\forall n, m > N$, $|b_n - b_m| < \epsilon$. Split (#) into two parts,

$$\begin{aligned}
(\#) &= \frac{1}{a_n}\sum_{j=1}^{n-1}(b_n - b_j)(a_{j+1} - a_j) \\
&= \frac{1}{a_n}\left(\sum_{j=1}^{N-1} + \sum_{j=N}^{n}\right)(b_n - b_j)(a_{j+1} - a_j) \\
&\leq 2\epsilon
\end{aligned} \tag{5.41}$$

Where the first part is taken care by $\frac{1}{a_n} \to 0$, second is due to $|b_n - b_j| \to 0$. ■

5.5.2 Kolmogorov's Ineq

Lemma. (**Kolmogorov's Ineq**) Seq $\{X_n\}$. Define S_n as partial sum, $\{X_n\}$ satisfy

- $\{X_n\}$ is **Mutually Indep.** (Pairwise Not sufficient!)
- $\mathbb{E}[X_n] = 0$, $\mathbb{E}[X^2] < \infty$ for all n.

Then, $\forall\epsilon > 0$,

$$\mathbb{P}\left(\max_{1\leq j\leq n}|S_j| > \epsilon\right) \leq \frac{\mathbb{E}[S_n^2]}{\epsilon^2} \tag{5.42}$$

Proof. Define $A := \{\max_{1\leq j\leq n}|S_j| > \epsilon\}$ the event inside LHS.
Define $A_j := \{|S_i| \leq \epsilon, \text{for } i = 1, 2, ..., j-1\} \cap \{|S_j| > \epsilon\}$ the larger one pops up at exactly index j.
$A_i \cap A_j = \emptyset$, for $i \neq j$ ($\{A_j\}$ are disjoint), clearly.
We have

$$A = \bigcup_{j=1}^{n} A_j \tag{5.43}$$

$$\begin{aligned}
\mathbb{E}\left[S_n^2\right] \geq \mathbb{E}\left[S_n^2; A\right] &= \sum_{j=1}^{n}\mathbb{E}\left[S_n^2; A_j\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[(S_j + (S_n - S_j))^2; A_j\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[S_j^2; A_j\right] + 2\sum_{j=1}^{n}\mathbb{E}\left[S_j(S_n - S_j); A_j\right] + \sum_{j=1}^{n}\mathbb{E}\left[(S_n - S_j)^2; A_j\right] \\
&= \sum_{j=1}^{n}\mathbb{E}\left[S_j^2; A_j\right] + \sum_{j=1}^{n}\mathbb{E}\left[(S_n - S_j)^2\right] \\
&\geq \sum_{j=1}^{n}\mathbb{E}\left[S_j^2; A_j\right] > \epsilon^2\sum_{j=1}^{n}\mathbb{P}(A_j) = \epsilon^2\mathbb{P}(A)
\end{aligned} \tag{5.44}$$

Two things in this derivation,

- The cross term $S_j(S_n - S_j)$ is removed because $S_n - S_j = \sum_{k=j+1}^{n} X_k$, indepent wrt X_l for any $l \leq j$, thus indep wrt S_j. $\mathbb{E}[S_j(S_n - S_j)] = \mathbb{E}[S_j]\,\mathbb{E}[S_n - S_j]$, and clearly $\mathbb{E}[S_n - S_j] = 0$.
- The final estimate of expectation by probability. Since A_j contains constraint $|S_j| > \epsilon$, so $\mathbb{E}\left[S_j^2; A_j\right] > \mathbb{E}[\epsilon^2; A_j] = \epsilon^2 \mathbb{P}(A_j)$.

Therefore $\mathbb{P}(A) \leq \frac{\mathbb{E}[S_n^2]}{\epsilon^2}$. ■

5.5.3 Kolmogorov's SLLN

Thm. (**Kolmogorov-Prelude**) $\{Y_n\}$ satisfy

- $\{Y_n\}$ **Mutually indep.**
- $\sum_{n\geq 1} \mathrm{Var}[Y_n] < \infty$, (automatically have $Y_n \in \mathcal{L}^2$).

Then, $\sum_{n\geq 1}(Y_n - \mathbb{E}[Y_n])$ converges almost surely.

Proof. Denote partial sum S_n. Fix some $N > 0$, consider $\{Y_{N+n} : n \geq 1\}$. Denote tail of summation $T_m := \sum_{j=1}^{m} Y_{N+j} = S_{N+m} - S_N$.
Clearly $\{T_m\}$ mutually indep, apply Kolmogorov's ineq to sequence $T_m - \mathbb{E}[T_m]$,

$$\mathbb{P}\left(\max_{1\leq j\leq m} |T_j - \mathbb{E}[T_j]| > \epsilon\right) \leq \frac{\mathrm{Var}[T_m]}{\epsilon^2} = \frac{1}{\epsilon^2}\sum_{j=N+1}^{N+m} \mathrm{Var}[Y_j] \tag{5.45}$$

We are allowed to take $m \to \infty$(?)

$$\begin{aligned}
\mathbb{P}\left(\sup_{j\geq 1} |T_j - \mathbb{E}[T_j]| > \epsilon\right) &= \mathbb{P}\left(\bigcup_{m\geq 1}\{\max_{1\leq j\leq m} |T_j - \mathbb{E}[T_j]| > \epsilon\}\right) \\
&= \lim_{m\to\infty} \mathbb{P}\left(\max_{1\leq j\leq m} |T_j - \mathbb{E}[T_j]| > \epsilon\right) \qquad (5.46)\\
&\leq \frac{1}{\epsilon^2}\sum_{j=N+1}^{\infty} \mathrm{Var}[Y_j] \xrightarrow{N\to\infty} 0
\end{aligned}$$

Convergence to 0 when $N \to \infty$ follows the hypothesis that $\sum_{n\geq 1} \mathrm{Var}[Y_n] < \infty$.
Now we have an important intermediate result,

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{j\geq 1} |T_j - \mathbb{E}[T_j]| > \epsilon\right) = 0 \tag{5.47}$$

i.e.

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{j\geq 1} |(S_{N+j} - S_N) - \mathbb{E}[S_{N+j} - S_N]| > \epsilon\right) = 0 \tag{5.48}$$

This line says that for all ϵ, we can somehow control the **maximum oscillation of tail sequence**. We will see that this statement *always* implies convergence a.s.
Consider

$$\begin{aligned}
&\{S_n - \mathbb{E}[S_n] \text{ does not converge in } \mathbb{R}\} \subseteq \{S_n - \mathbb{E}[S_n] \text{ is not Cauchy}\} \\
&\qquad = \bigcup_{k\geq 1}\bigcap_{N\geq 1}\{\sup_{j\geq N} |(S_j - \mathbb{E}[S_j]) - (S_N - \mathbb{E}[S_N])| > \frac{1}{k}\}
\end{aligned} \tag{5.49}$$

We hope that this has zero probability. Fix k

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{N\geq 1}\{\sup_{j\geq N}|(S_j-\mathbb{E}[S_j])-(S_N-\mathbb{E}[S_N])|>\frac{1}{k}\}\right)\\
&=\mathbb{P}\left(\bigcap_{N\geq 1}\{\sup_{j\geq N}|T_j-\mathbb{E}[T_j]|>\frac{1}{k}\}\right)\\
&\leq \lim_{N\to\infty}\mathbb{P}\left(\sup_{j\geq 1}|T_j-\mathbb{E}[T_j]|>\epsilon\right)=0
\end{aligned} \tag{5.50}$$

Therefore

$$\mathbb{P}\left(\{S_n-\mathbb{E}[S_n]\ \text{ converges in } \mathbb{R}\}\right)=1 \tag{5.51}$$

i.e. $\sum_{n\geq 1}(Y_n-\mathbb{E}[Y_n])$ converges almost surely. ■

Thm. (**Kolmogorov**) $\{X_n\}$ satisfies

- **Mutually indep**.
- $\sum_{n\geq 1}\mathrm{Var}[X_n]/n^2<\infty$

then,

$$\sum_{n\geq 1}\frac{X_n-\mathbb{E}[X_n]}{n}\ \text{Converges almost surely.}$$

Apply Kronecker's Lemma, we have SLLN:

$$\frac{1}{n}\sum_{n\geq 1}(X_n-\mathbb{E}[X_n])\xrightarrow{a.s.}0\ \text{ i.e. }\ \frac{S_n-\mathbb{E}[S_n]}{n}\xrightarrow{a.s.}0$$

Proof. Let $Y_n:=\frac{X_n}{n}$. Check Y_n satisfying hypothesis of prelude thm, then we have desired result by prelude thm. ■

5.6 Kolmogorov' (SLLN3)

Thm. (**Kolmogorov'**) $\{X_n\}$ is i.i.d. sequence. Following two statement holds,

$$\mathbb{E}[|X_1|]<\infty\Rightarrow\frac{S_n-\mathbb{E}[S_n]}{n}\xrightarrow{a.s.}0\ \text{ i.e. }\ \frac{S_n}{n}\xrightarrow{a.s.}\mathbb{E}[X_1]\quad(\#1) \tag{5.52}$$

$$\mathbb{E}[|X_1|]=\infty\Rightarrow\limsup_{n\to\infty}\frac{|S_n|}{n}=\infty\quad\text{a.s.}\quad(\#2) \tag{5.53}$$

Proof. First do (#2). Assume $\mathbb{E}[|X_1|]=\infty$. Fix any $A>0$, $\mathbb{E}\left[|\frac{X_1}{A}|\right]=\infty$. Then

$$\sum_{n\geq 1}\mathbb{P}\left(\left|\frac{X_1}{A}\right|>n\right)=\infty \tag{5.54}$$

(Because, in more general case,

$$\begin{aligned}
\infty=\mathbb{E}[|X|]&=\int_0^\infty\mathbb{P}(|X|>t)\,dt\\
&=\sum_{n\geq 1}\int_{n-1}^{n}\mathbb{P}(|X|>t)\,dt\leq\sum_{n\geq 1}\mathbb{P}(|X|>n-1)
\end{aligned} \tag{5.55}$$

In fact, $\mathbb{E}[|X|] < \infty \iff \sum_1^\infty \mathbb{P}(|X| > n) < \infty)$
Apply (**BC2**), $\mathbb{P}(|X_n| > nA \;\; i.o) = 1$, i.e.

$$\mathbb{P}\left(\frac{|S_n - S_{n-1}|}{n} > A \;\; i.o\right) = 1 \tag{5.56}$$

Consider,

$$\begin{aligned}\{|S_n - S_{n-1}| > nA\} &\subseteq \left\{|S_n| > \frac{n}{2}A\right\} \cup \left\{|S_{n-1}| > \frac{n}{2}A\right\} \\ &\subseteq \left\{\frac{|S_n|}{n} > \frac{A}{2}\right\} \cup \left\{\frac{|S_{n-1}|}{n-1} > \frac{A}{2}\right\}\end{aligned} \tag{5.57}$$

Two parts at RHS says same thing, so actually we have

$$\mathbb{P}\left(\frac{|S_n|}{n} > \frac{A}{2} \;\; i.o\right) = 1 \tag{5.58}$$

This is true for $\forall A > 0$. So take intersection over A, the statement still holds.

$$\begin{aligned}\bigcap_{m\geq 1}\left\{\frac{|S_n|}{n} > m \;\; i.o\right\} &\subseteq \bigcap_{m\geq 1}\left\{\limsup_{n\to\infty}\frac{|S_n|}{n} > m\right\} \\ &= \left\{\limsup_{n\to\infty}\frac{|S_n|}{n} = \infty\right\} \quad \blacksquare\end{aligned} \tag{5.59}$$

Now show (#1), assume $\mathbb{E}[|X_1|] < \infty$, truncate X_n,

$$Y_n := \begin{cases} X_n & \text{if } |X_n| \leq n, \\ 0 & \text{otherwise.}\end{cases}$$

By same argument in WLLN(Khintchine), we can come to $\sum \mathbb{P}(Y_n \neq X_n) < \infty$, i.e. X_n, Y_n are equivalent. Clearly $\{Y_n\}$ is also indep.
We want to refer to (SLLN2), i.e. $\sum \frac{1}{n^2}\mathrm{Var}[Y_n] < \infty$, consider this quantity

$$\begin{aligned}\sum_{n\geq 1}\frac{\mathrm{Var}[Y_n]}{n^2} &\leq \sum_{n\geq 1}\frac{\mathbb{E}[Y_n^2]}{n^2} = \sum_{n\geq 1}\frac{\mathbb{E}[X_n^2; |X_n| < n]}{n^2} \\ &= \sum_{n\geq 1}\frac{\mathbb{E}[X_1^2; |X_1| < n]}{n^2} \\ &= \sum_{n\geq 1}\left(\frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}\left[X_1^2; j-1 \leq |X_1| \leq j\right]\right) \\ &= \sum_{j=1}^{n}\left(\mathbb{E}\left[X_1^2; j-1 \leq |X_1| \leq j\right]\sum_{n\geq j}\frac{1}{n^2}\right) \\ &= \sum_{j=1}^{n}\mathbb{E}\left[X_1^2; j-1 \leq |X_1| \leq j\right]\cdot O\left(\frac{1}{j}\right) \\ &\leq C\sum_{j=1}^{n}\frac{1}{j}\cdot j\cdot\mathbb{E}[|X_1|; j-1 \leq |X_1| \leq j] = C\mathbb{E}[|X_1|] < \infty\end{aligned} \tag{5.60}$$

In which we swich the order of summation at the forth equal sign, noticing that $\sum_{n\geq j} 1/n^2 = O(1/j)$, and apply definition of O notation at the end, $0 < C < \infty$ is constant.
Apply (SLLN2) for $\{Y_n\}$, we have

$$\frac{\sum_{j=1}^{n} |Y_j - \mathbb{E}[Y_j]|}{n} \xrightarrow{a.s.} 0 \tag{5.61}$$

Split target quantity in similar fashion as WLLN2:

$$\begin{aligned}\frac{|S_n - \mathbb{E}[S_n]|}{n} &\leq \frac{|\sum_1^n X_j - Y_j|}{n} + \frac{|\sum_1^n Y_j - \mathbb{E}[Y_j]|}{n} + \frac{|\sum_1^n \mathbb{E}[Y_j] - \mathbb{E}[X_j]|}{n} \\ &= Q_1 + Q_2 + Q_3\end{aligned} \tag{5.62}$$

We proved $Q_2 \xrightarrow{a.s.} 0$.
By property of equivalent seqs $Q_1 \xrightarrow{a.s.} 0$.

$$Q_3 = \frac{\sum_1^n \mathbb{E}[X_j : |X_j| > j]}{n} = \frac{1}{n}\sum_1^n \mathbb{E}[X_1; |X_1| > j] \tag{5.63}$$

By lemma, $a_n \to 0 \Rightarrow \sum a_n = o(n)$. We have $Q_3 \to 0$ pointwise. Therefore $Q_1 + Q_2 + Q_3 \xrightarrow{a.s.} 0$ as desired. ■

5.7 (SLLN4)

Thm. (**SLLN4**) Let $\{X_n : n \geq 1\}$ be sequence of $\mathcal{L}^1$, indep RVs; S_n be partial sum. Let $\phi : \mathbb{R} \to \mathbb{R}$ be positive and continuous even function such that $\frac{\phi(x)}{|x|}$ is non-decreasing in x and $\frac{\phi(x)}{x^2}$ is non-increasing in x. Assume for some sequence $\{b_n : n \geq 1\}$ of positive real numbers with $b_n \nearrow \infty$,

$$\sum_{n\geq 1} \frac{\mathbb{E}[\phi(X_n)]}{\phi(b_n)} < \infty \tag{5.64}$$

Show that $\sum_{n\geq 1} \frac{X_n - \mathbb{E}[X_n]}{b_n}$ converges a.s., hence

$$\frac{S_n - \mathbb{E}[S_n]}{b_n} \xrightarrow{a.s.} 0 \tag{5.65}$$

Proof. See problem 8-4-3.

5.8 Levy's Equivalence Thm

Thm. (**Levy**) $\{X_n\}$ indep. S_n is partial sum, then

$$S_n \xrightarrow{i.p} S \iff S_n \xrightarrow{a.s.} S \tag{5.66}$$

In fact (won't prove)

$$S_n \xrightarrow{dist} S \iff S_n \xrightarrow{a.s.} S \tag{5.67}$$

Rm. Intuition is that, in general, it is so *hard* for sum of independent RV to converge that as long as it converges, it converges *in all sense.*

Proof. Only for the in.prob part. $\Rightarrow$:
By i.p; $\forall \epsilon > 0$, $\exists N$, for all $m, n > N$,

$$\mathbb{P}\left(|S_m - S_n| > \epsilon\right) \leq \mathbb{P}\left(|S_n - S| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(|S_m - S| > \frac{\epsilon}{2}\right) \leq \epsilon \tag{5.68}$$

$$\begin{aligned}\epsilon &\geq \mathbb{P}\left(|S_m - S_n| > \epsilon\right) \\ &\geq \mathbb{P}\left(|S_m - S_n| > \epsilon \ \ \& \ \max_{n+1 \leq k \leq m} |S_k - S_n| > 2\epsilon\right) \\ &= \sum_{k=n+1}^{m} \mathbb{P}\left(|S_m - S_n| > \epsilon \ \ \& \ \ |S_j - S_n| \leq 2\epsilon, \forall j = n+1, ..., k-1 \ \ \& \ \ |S_k - S_n| > 2\epsilon\right) \\ &\geq \sum_{k=n+1}^{m} \mathbb{P}\left(|S_m - S_k| \leq \epsilon \ \ \& \ \ |S_j - S_n| \leq 2\epsilon, \forall j = n+1, ..., k-1 \ \ \& \ \ |S_k - S_n| > 2\epsilon\right)\end{aligned} \tag{5.69}$$

Notice that

$$\begin{aligned}&\{|S_m - S_k| \leq \epsilon\} \in \sigma(X_{k+1}, ..., X_m) \\ &\{|S_j - S_n| \leq 2\epsilon, \forall j = n+1, ..., k-1 \ \ \& \ \ |S_k - S_n| > 2\epsilon\} \in \sigma(X_{n+1}, ..., X_k)\end{aligned} \tag{5.70}$$

Are independent, so

$$\begin{aligned}\epsilon &\geq \sum_{k=n+1}^{m} \mathbb{P}\left(|S_m - S_k| \leq \epsilon \ \ \& \ \ |S_j - S_n| \leq 2\epsilon, \forall j = n+1, ..., k-1 \ \ \& \ \ |S_k - S_n| > 2\epsilon\right) \\ &= \sum_{k=n+1}^{m} \mathbb{P}\left(|S_m - S_k| \leq \epsilon\right) \cdot \mathbb{P}\left(|S_j - S_n| \leq 2\epsilon, \forall j = n+1, ..., k-1 \ \ \& \ \ |S_k - S_n| > 2\epsilon\right) \\ &\geq (1-\epsilon) \sum_{k=n+1}^{m} \mathbb{P}\left(|S_j - S_n| \leq 2\epsilon, \forall j = n+1, ..., k-1 \ \ \& \ \ |S_k - S_n| > 2\epsilon\right) \\ &= (1-\epsilon) \cdot \mathbb{P}\left(\max_{n+1 \leq k \leq m} |S_k - S_n| > 2\epsilon\right)\end{aligned} \tag{5.71}$$

we have, $\forall \epsilon > 0$, for all $m, n > N$,

$$\mathbb{P}\left(\max_{n+1 \leq k \leq m} |S_k - S_n| > 2\epsilon\right) \leq \frac{\epsilon}{1-\epsilon} \tag{5.72}$$

Let $m \to \infty$,

$$\mathbb{P}\left(\sup_{k \geq n+1} |S_k - S_n| > 2\epsilon\right) \leq \frac{\epsilon}{1-\epsilon} \tag{5.73}$$

Therefore

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k \geq n+1} |S_k - S_n| > 2\epsilon\right) = 0 \tag{5.74}$$

Which implies

$$\mathbb{P}\left(\lim_{n\to\infty} S_n \text{ exists in } \mathbb{R}\ (<\infty)\right) = 1 \ \text{ i.e. } \ S_n \xrightarrow{a.s.} S \quad \blacksquare \tag{5.75}$$

Chapter 6

Product Space

6.1 Basic Structure

Def **Product Space**: Let $(S_1, \Sigma_1), (S_2, \Sigma_2)$ be two measurable spaces. Define

$$S := S_1 \times S_2$$

$$\Sigma := \sigma(\{B_1 \times B_2; B_i \in \Sigma_i \text{ (rectangles), i=1,2}\})$$

And coordinate maps $\rho_i : S \to S_i$, $\rho_i(s) = s_i$ for $\forall s \in S$. (S, Σ) is called product space by $(S_1, \Sigma_1) \times (S_2, \Sigma_2)$.

Rm.

- In fact $\Sigma = \sigma(\rho_1, \rho_2)$, i.e. preimage of $\rho_i \in \Sigma$, which is clearly the case, for example, pick any $B_1 \in \Sigma_1$, $\rho_1^{-1}(B_1) = B_1 \times S_2 \in \Sigma$.
- The generator set in $\sigma(\cdot)$, collection of rectangles, is a π system.

Lemma (*Measurability on prod space implies that at each, fix another coordinate.*) $(S, \Sigma) = (S_1, \Sigma_1) \times (S_2, \Sigma_2)$. Consider $m\Sigma \ni f : S \to \mathbb{R}$, then

- Fix $\bar{s_1} \in S_1$ *then* $m\Sigma_2 \ni f(\bar{s_1}, \cdot) : S_2 \to \mathbb{R}, s_2 \mapsto f(\bar{s_1}, s_2)$.
- Fix $\bar{s_2} \in S_2$ *then* $m\Sigma_1 \ni f(\cdot, \bar{s_2}) : S_1 \to \mathbb{R}, s_1 \mapsto f(s_1, \bar{s_2})$.

Proof. We use **Monotone Class Thm**. Let $\mathcal{H}$ be the class of real-valued functions, such that results in lemma holds. It suffices to show $m\Sigma \subseteq \mathcal{H}$, i.e. $\forall f \in m\Sigma$, $f \in \mathcal{H}$, lemma holds.
One can easily show $\mathcal{H}$ is a vector space[1], and $1 \in \mathcal{H}$.
Consider $\{f_n\} \subseteq \mathcal{H}$, $f_n \nearrow f$, $f_n > 0$. Then, for all $s \in S$, $f(s) = \lim_{n\to\infty} f_n(s)$,[2] $f \in m\Sigma$. Hence $\mathcal{H}$ is monotone class.
π system $\mathcal{I} = \{B_1 \times B_2, B_i \in \Sigma_i, i = 1, 2\}$, $\sigma(\mathcal{I}) = \Sigma$, for all $A \in \mathcal{I}$,

$$\mathbb{1}_A(s) = \mathbb{1}_{B_1 \times B_2}((s_1, s_2)) = \mathbb{1}_{B_1}(s_1) \cdot \mathbb{1}_{B_2}(s_2) \tag{6.1}$$

Clearly, $\mathbb{1}_A$ is Σ_i measurable fixing the other coordinate, i.e. $\mathbb{1}_A \in \mathcal{H}$. By monotone class thm, $m(\sigma(\mathcal{I})) \in \mathcal{H}$. ■

[1] Since linearity preserves measurability.
[2] Since limiting preserves measurability.

6.2 Product Measure, Fubini's Thm

Motivation: We want to define measure on product space (S, Σ).

Def. **Slice Integral**: Assume μ_i is finite measure on (S_i, Σ_i). For pointwise mapping $f : S \to \mathbb{R}$ for either $f \in b\Sigma$ or $(m\Sigma)^+$, for all $s_1 \in S_1$ and $s_2 \in S_2$, define two slice integrals of f:

$$I_1^f(s_1) := \int_{S_2} f(s_1, s_2)\mu_2(ds_2) \tag{6.2}$$

$$I_2^f(s_2) := \int_{S_1} f(s_1, s_2)\mu_1(ds_1) \tag{6.3}$$

Lemma (Integrate slice against another coordinate) Assume $f \in b\Sigma$, then $I_i^f \in b\Sigma_i$, $i = 1, 2$ And

$$\int_{S_1} I_1^f(s_1)\mu_1(ds_1) = \int_{S_2} I_2^f(s_2)\mu_2(ds_2) \quad (\dagger) \tag{6.4}$$

i.e.

$$\int_{S_1}\int_{S_2} f(s_1, s_2)d\mu_1 d\mu_2 = \int_{S_2}\int_{S_1} f(s_1, s_2)d\mu_2 d\mu_1 \tag{6.5}$$

Proof. Let $\mathcal{H}$ be class of bounded functions s.t. lemma holds. Verify that $\mathcal{H}$ is a monotone class (1,2 omitted here, for 3, $f_n \nearrow f$, $\dagger$ holds on f by (**DOM**))
Choose same π system $\mathcal{I}$ $(B_1 \times B_2)$, indicator $\mathbb{1}_A$:

$$I_1^{\mathbb{1}_A}(s_1) = \int_{S_2} \mathbb{1}_A(s_1, s_2)d\mu_2 = \int_{S_2} \mathbb{1}_{B_1}(s_1) \cdot \mathbb{1}_{B_2}(s_2)d\mu_2 = \mathbb{1}_{B_2}(s_2)\mu_2(B_2) \tag{6.6}$$

Similarly

$$I_2^{\mathbb{1}_A}(s_2) = \mathbb{1}_{B_1}(s_1)\mu_1(B_1) \tag{6.7}$$

$(\dagger)$ integrate out remaining coordinate, both are $\mu_1(B_1)\mu_2(B_1)$. Therefore $\mathbb{1}_A \in \mathcal{H}$. By monotone class thm, $\sigma(\mathcal{I}) = b\Sigma \subseteq \mathcal{H}$. ∎

Cor. (**Tonelli**) $f \in (m\Sigma)^+$, then $\dagger$ holds for $I_i^f \in (m\Sigma)^+$.

Proof. For each $k > 0$, define $f_k := f \wedge k := f \cdot \mathbb{1}_{\{f \le k\}}$. Clearly $f_k \in b\Sigma$, moreover $f_k \nearrow f$. Apply lemma for f_k, we have $I_i^{f_k} \in b\Sigma_i$.
Since $f = \lim_{k\to\infty} f_k$, by (**MON**) $\Rightarrow I_i^f = \lim_{k\to\infty} I_i^{f_k}$, $i = 1, 2$. So $I_i^f \in (m\Sigma)^+$. ∎

Thm. (**Fubini**) Measure space (S_i, Σ_i, μ_i), $i = 1, 2$, μ_i are finite measure. Define (S, Σ) same as section 1, define $\mu : S \to \mathbb{R}$, s.t. for all $A \in \Sigma$,

$$\mu(A) := \int_{S_1} I_1^{\mathbb{1}_A} d\mu_1 = \int_{S_2} I_2^{\mathbb{1}_A} d\mu_2 \tag{6.8}$$

Denote $(S, \Sigma, \mu) = (S_1, \Sigma_1, \mu_1) \times (S_2, \Sigma_2, \mu_2)$, denote $\mu = \mu_1 \times \mu_2$,
Then

- μ is a measure on (S, Σ) (contable additive).
- μ is the unique measure on (S, Σ), such that $\mu(B_1 \times B_2) = \mu_1(B_1) \cdot \mu_2(B_2)$.

· If $f \in (m\Sigma)^+$, then

$$\int_S f d\mu = \int_{S_1} I_1^f d\mu_1 = \int_{S_2} I_2^f d\mu_2 \quad (\#) \tag{6.9}$$

· If $f \in \mathcal{L}^1(S, \Sigma, \mu)$, then $I_i^f \in \mathcal{L}^1(S_i, \Sigma_i, \mu_i)$, and $(\#)$ holds

Proof. • Part-1, μ is measure.
Pick $A, B \in \Sigma$, disjoint $\Rightarrow \mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B$. By definition

$$\mu(A \cup B) := \int_{S_1} (I_1^{\mathbb{1}_A} + I_1^{\mathbb{1}_B}) d\mu_1 =: \mu(A) + \mu(B) \tag{6.10}$$

So we have finite additivity. Now consider $\mu(U), U := \bigcup_{n\geq 1} E_n, U_n := \bigcup_{j=1}^n E_j, E_n \in \Sigma$. We have $\mathbb{1}_{U_n} \nearrow \mathbb{1}_U$. By (**MON**): $I_1^{\mathbb{1}_{U_n}} \nearrow I_1^{\mathbb{1}_U}$.
By (**MON**) again: $\int I_1^{\mathbb{1}_{U_n}} \to \int I_1^{\mathbb{1}_U}$.
Therefore

$$\mu(U) := \int_{S_1} I_1^{\mathbb{1}_U} = \lim_{n\to\infty} \int_{S_1} I_1^{\mathbb{1}_{U_n}} = \int_{S_1} \lim_{n\to\infty} I_1^{\mathbb{1}_{U_n}} = \int_{S_1} \sum_{n\geq 1} I_1^{\mathbb{1}_{E_n}} = \sum_{n\geq 1} \mu(E_n) \quad \blacksquare \tag{6.11}$$

• Part-2, μ is unique.
If μ' is another measure satisfies hypothesis $(\mu(B_1 \times B_2) = \mu_1(B_1) \cdot \mu_2(B_2))$. Clearly $\mu = \mu'$ on $\mathcal{I}$, rectangles. $\mathcal{I}$ is π system. By π-system thm, $\mu = \mu'$ on Σ. $\blacksquare$

• Part-3, $(\#)$ eq for $f \in (m\Sigma)^+$.
The second equal sign is clear, (**Tonelli**), show the first one.
For $f = \mathbb{1}_A$,

$$\int_S f d\mu = \int_S \mathbb{1}_A d\mu = \mu(A) := \int_{S_1} I_1^{\mathbb{1}_A} d\mu_1 \tag{6.12}$$

Holds just by definition of μ.
For $f \in SF^+$, by linearity, # holds.
For $f \in (m\Sigma)^+$ by (**MON**), # holds. $\blacksquare$

• Part-4, $(\#)$ eq for $f \in \mathcal{L}^1$.
$f = f^+ - f^-$, $f \in \mathcal{L}^1 \Rightarrow f^\pm < \infty$ $a.s.$ So # holds for $f^\pm$. All relevant integrals are finite, we can rearrange terms by linearity. So # holds for f. $\blacksquare$

Rm. Remarks on (**Fubini**)

1. The condition in statement says μ_i are finite. We actually have Fubini for μ_i that are σ-finite.
2. Since we can extend product of two to product of finitely many, Fubini holds for $n < \infty$ product space, i.e. $\prod_{k=1}^n (S_k, \Sigma_k, \mu_k)$.
3. Lemma in section 1 says measurability on product space implies that at each factor space. But other direction is not true. i.e. $f(\bar{s_1}, \cdot) \in m\Sigma_1, f(\cdot, \bar{s_2}) \in m\Sigma_2$ **Does Not Imply** $f \in m\Sigma$.

4. Fubini says integrability on product space implies that at each factor space. But other direction is not true. i.e. $I_i^f \in \mathcal{L}^1(S_i, \Sigma_i, \mu_i)$ **Does Not Imply** $f \in \mathcal{L}^1(S, \Sigma, \mu)$.

Two examples of 3 and 4:

6.3 Joint Distribution, Joint Law

Def. **Joint Distribution**: Prob space $(\Omega, \mathcal{F}, \mathbb{P})$, real valued RV X, Y, define joint distribution function as

$$F_{(X,Y)}(x, y) := \mathbb{P}\left(X \leq x, Y \leq y\right) \tag{6.13}$$

Def. **Joint Law**: Define $\mathcal{L}_{(X,Y)}$ as joint law of (X, Y). $\mathcal{L}_{(X,Y)}$ is then a prob measure on product image space $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$, s.t. for all $A \in \mathscr{B}(\mathbb{R}^2)$,

$$\mathcal{L}_{(X,Y)}(A) := \mathbb{P}\left((X, Y) \in A\right) \tag{6.14}$$

Def. **Joint PDF**: If $\mathcal{L}_{(X,Y)}$ is absolutely continuous with respect to lebesgue measure on $\mathbb{R}^2$ (denote as $dxdy$), then the joint pdf of (X, Y) exists, denote $f_{(X,Y)}$, $f_{(X,Y)} \in m\mathscr{B}(\mathbb{R}^2)$, and is defined as Radon-Nikodym derivative of joint law wrt lebesgue measure on product image space,

$$f_{(X,Y)} := \frac{d\mathcal{L}_{(X,Y)}}{dxdy} \tag{6.15}$$

Prop. If $f_{(X,Y)}$ is joint pdf, then by (**Fubini**), then

$$f_X(x) := \int_{\mathbb{R}} f_{(X,Y)}(x, y)dy \quad \text{is pdf of } X.$$

$$f_Y(y) := \int_{\mathbb{R}} f_{(X,Y)}(x, y)dx \quad \text{is pdf of } Y.$$

6.3.1 Joint * of Indep RVs

Prop. X, Y are RV with respective cdf and law $\mathcal{L}_X, \mathcal{L}_Y$; F_X, F_Y. *Then TFAE*[3]:

- X, Y are independent.
- $\mathcal{L}_{(X,Y)} = \mathcal{L}_X \times \mathcal{L}_Y$.
- $F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$ for all $(x, y) \in \mathbb{R}^2$.
- (If respective pdf f_X, f_Y exists) $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ for a.e. $(x, y) \in \mathbb{R}^2$.

Statement four is special, in that respective pdf may not exist. And there is allowance for a.e. form every (x, y), because integration eliminates aberrant null sets.

[3] Jargon: The followings are equivalent ($\iff$).

Proof. Proof is straightforward, noticing all four statements $\iff$ $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y)$. ■

Prop. X, Y indep, $X + Y$ is a new RV. Then Law of $X + Y$ is given by

$$\mathcal{L}_{X+Y}(c) = \int_{\mathbb{R}} \mathcal{L}_Y([-\infty, c - x])\mathcal{L}_X(dx) = \int_{\mathbb{R}} \mathcal{L}_X([-\infty, c - y])\mathcal{L}_Y(dy)$$

Proof.

$$\begin{aligned}
\mathcal{L}_{X+Y}(c) = \mathbb{P}(X + Y \leq c) &= \iint_{\{(x,y):x+y\leq c\}} d\mathcal{L}_{(X,Y)} \\
&= \iint_{\{(x,y):x+y\leq c\}} d(\mathcal{L}_X \times \mathcal{L}_Y) \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbb{1}_{(-\infty,c-x]}(y)\mathcal{L}_Y(dy)\right] \mathcal{L}_X(dx) \\
&= \int_{\mathbb{R}} \mathcal{L}_Y([-\infty, c - x])\mathcal{L}_X(dx) \quad ■
\end{aligned} \tag{6.16}$$

6.3.2 Convolutions

Def. **Convolution of Function**: for $f \in \mathcal{L}^1$, g is bounded, define

$$(f * g)(x) := \int_{\mathbb{R}} f(x - y)g(y)dy$$

Def. **Convolution of Measure**: Given two finite measures μ, ν on (S, Σ), $\mu * \nu = \nu * \mu$ is a measure, for all $A \in (S, \Sigma)$, given by

$$(\mu * \nu)(A) := \int_S \mu(A - s)\nu(ds) = \int_S \nu(A - s)\mu(ds) =: (\nu * \mu)(A)$$

Where $A - s$ is s translation of A, i.e. $A - s = \{t \in S, t + s \in A\}$.

Rm. By prop in last section, we actually have: (when X, Y indep)

$$\mathcal{L}_{X+Y} = \mathcal{L}_X * \mathcal{L}_Y$$

If f_X, f_Y exists, we have

$$f_{X+Y} = f_X * f_Y$$

6.4 Product of Countably Many Spaces

6.4.1 Product Measure

We are now considering product of countably many spaces. i.e. $\prod_{n\geq1}(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$. Define

$$\Omega := \prod_{n\geq1} \Omega_n$$

representative element $w = (w_1, w_2, ...), w_i \in \Omega_i$.
Consider **Cylinder Sets** E, defined by

$$E := \prod_{n \geq 1} F_n = \prod_{k=1}^{N} F_{n_k} \times \prod_{j \notin \{n_k\}_1^N} \Omega_j \tag{6.17}$$

Where $F_{n_k} \subseteq \Omega_{n_k}$, other $F_j = \Omega_j$ for $j \notin \{n_k\}$. This is saying that *All but finitely many factors of E are Ωs.*
Define

$$\Sigma_0 := \left\{ \bigcup_{k \geq 1}^{K} E^{[k]} : E^{[k]} \text{ are disjoint cylinder sets} \right\} \tag{6.18}$$

It can be shown (omitted) that Σ_0 is an algebra. Let $\mathcal{F} := \sigma(\Sigma_0)$.
Define set function $\mathbb{P} : \Sigma_0 \to [0, 1]$, such that for all $A = \bigcup_{k \geq 1}^{K} E^{[k]} \in \Sigma_0$,

$$\mathbb{P}(A) := \sum_{k=1}^{K} \left[\prod_{j \geq 1} \mathbb{P}_j (F_j) \right] \tag{6.19}$$

Where $\mathbb{P}_j$ is measure on factor space. Then one can prove (omitted) that $\mathbb{P}$ is well-defined, $\mathbb{P}$ is a measure (countable additive).
Thus by Caratheodory extension thm, $\mathbb{P}$ can be uniquely extended to $\mathcal{F} = \sigma(\Sigma_0)$.
So we define $(\Omega, \mathcal{F}, \mathbb{P}) := \prod_{n \geq 1} (\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$.

6.4.2 Kolmogorov Extension Thm

Thm. (**Prelude**) Let $\{\mu_n : n \geq 1\}$ be a countable sequence of prob measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Then there exists $(\Omega, \mathcal{F}, \mathbb{P})$ and a seq of **indep** RVs $\{X_n\}$, such that $\mathcal{L}_{X_n} = \mu_n$ for all $n \geq 1$.

Proof. Previous result, for all $n \geq 1$, exists $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, RV $Y_n : \Omega_n \to \mathbb{R}$, s.t $Y_n = \mu_n$. Construct such spaces and Y_n for all n, and together, define

$$(\Omega, \mathcal{F}, \mathbb{P}) := \prod_{n \geq 1} (\Omega_n, \mathcal{F}_n, \mathbb{P}_n) \tag{6.20}$$

Define in product space $X_n : \Omega \to \mathbb{R}$, $w \in \Omega \mapsto Y_n(w_n) \in \mathbb{R}$, i.e. $X_n(w) = Y(w_n)$. Now for all $B \in \mathscr{B}(\mathbb{R})$, by definition of product measure in countably product space,

$$\begin{aligned} \mathbb{P}(X_n \in B) &= \prod_{j=1}^{n-1} \mathbb{P}_j (\Omega_j) \cdot \mathbb{P}_n (Y_n \in B) \cdot \prod_{j=n+1}^{\infty} \mathbb{P}_j (\Omega_j) \\ &= \mathbb{P}_n (Y_n \in B) = \mu_n(B) \end{aligned} \tag{6.21}$$

Then show $\{X_n\}$ indep, i.e. $\forall L \geq 1$, $n_1, n_2, ..., n_L$ disjoint,

$$\begin{aligned}\mathbb{P}\left(X_{n_1} \in B_1, X_{n_2} \in B_2, ..., X_{n_L} \in B_L\right) &= \mathbb{P}\left(\prod_{j \notin {n_k}_1^L} \Omega_j \times \prod_{l=1}^{L} \{Y_{n_l} \in B_l\}\right) \\ &= \prod_{l=1}^{L} \mathbb{P}\left(Y_{n_l} \in B_l\right) \\ &= \prod_{l=1}^{L} \mathbb{P}\left(X_{n_l} \in B_l\right) \quad (indep) \ \blacksquare\end{aligned} \tag{6.22}$$

Thm. (**Kolmogorov's Extension**) For $n \geq 1$, $\mu^{(n)}$ is prob measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. For $1 \leq m \leq n$, define $\pi_{m,n}$ as extension mapping, $\forall B \in \mathscr{B}(\mathbb{R}^m)$, $\pi_{m,n}(B) := \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n, (x_1, x_2, ..., x_m) \in B\}$, i.e.

$$\pi_{m,n}(B) = B \times \mathbb{R}^{n-m}$$

Assume $\mu^{(n)}$ satisfies consistency condition:
$\forall n \geq 1$, $\forall 1 \leq m \leq n$, $\forall B \in \mathscr{B}(\mathbb{R}^m)$,

$$\mu^{(n)}(\pi_{m,n}(B)) = \mu^{(n)}(B)$$

Then, exists prob space $(\Omega, \mathcal{F}, \mathbb{P})$, sequence of RV (*Not necessarily indep*) $\{X_n : n \geq 1\}$, such that $\mu^{(n)} = \mathcal{L}_{(X_1, X_2, ..., X_n)}$.

Rm. Thm prelude is a particular case of (**Kolmogorov**).

Chapter 7

Conditioning and Martingale

7.1 Conditional Expectation

Def. **Conditional Expectation**: Define $(\Omega, \mathcal{F}, \mathbb{P})$ be probability space and $X : \Omega \to \mathbb{R}$ RV, $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ algebra Then, $Y \in \mathcal{L}^1$ is the conditional expectation of X given $\mathcal{G}$ (actually an RV), denoted by $Y := \mathbb{E}[X|\mathcal{G}]$ if

- $Y \in m\mathcal{G}$.
- For all $A \in \mathcal{G}$,

$$\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$$

Rm. The intuition of $\mathbb{E}[X|\mathcal{G}]$ is, given the partial information contained in $\mathcal{G}$, the best prediction of X on whole space.

Rm. Conditional expectation can be defined for X, Y not necessarily in $\mathcal{L}^1$. It is ok as long as for all $A \in \mathcal{G}$, integral of X, Y on A are defined.

Rm. The defining condition can be replaced by if $\mathcal{G} = \sigma(\mathcal{I})$, where $\mathcal{I}$ is a π system, then $\forall A \in \mathcal{I}$, the integrals are equal. Because $A \in \mathcal{G} \mapsto \int_A X d\mathbb{P}$ can be viewed as a signed measure on $\mathcal{G}$, we can apply π system lemma.

Prop. (**Monotonicity**) If $X_1 \leq X_2 \;\; a.s.$ then $Y_1 := \mathbb{E}[X_1|\mathcal{G}] \leq \mathbb{E}[X_2|\mathcal{G}] =: Y_2 \;\; a.s.$

Proof. Let $A := \{Y_2 > Y_1\}$, clearly $A \in \mathcal{G}$, by definition

$$\int_A Y_1 d\mathbb{P} = \int_A X_1 d\mathbb{P} \leq \int_A X_2 d\mathbb{P} = \int_A Y_2 d\mathbb{P} \tag{7.1}$$

$$\int_A (Y_1 - Y_2) d\mathbb{P} \leq 0 \tag{7.2}$$

But $(Y_1 - Y_2) > 0$ on A, so $\mathbb{P}(A) = 0$. ■

Thm. (**Existence and Uniqueness**) Given $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} \subseteq \mathcal{F}$, $X \in \mathcal{L}^1$, then $\mathbb{E}[X|\mathcal{G}]$ exists and is unique a.s.

Proof. Prove uniqueness first, assume Y_1, Y_2 both satisfies definition. Since $X = X$, by monotonicity, $Y_1 \leq Y_2$; $Y_2 \leq Y_1$. ■

Then existence. We have two approaches.

Version 1. (*Radon-Nikodyn thm*) the idea is that we view conditional expectation as a signed measure.
Define $\mu_{\mathcal{G}}^X$ on $\mathcal{G}$, such that $\forall A \in \mathcal{G}$,

$$\mu_{\mathcal{G}}^X(A) := \int_A X d\mathbb{P} \tag{7.3}$$

One can check this is a measure. Besides, when $\mathbb{P}(A) = 0$, $\mu_{\mathcal{G}}^X(A) = 0$. Moreover, $\mu_{\mathcal{G}}^X(A)$ is absolutely continous wrt $\mathbb{P}\lceil_{\mathcal{G}}$ (probability measure restricted on $\mathcal{G}$). Apply **Radon-Nikodyn**, $\exists Y \in m\mathcal{G}$,[1] s.t.

$$Y = \frac{d\mu_{\mathcal{G}}^X}{d\mathbb{P}\lceil_{\mathcal{G}}} \quad \text{i.e. the R-N derivative} \tag{7.4}$$

So, for all $A \in \mathcal{G}$, (view Y as the density)

$$\int_A X d\mathbb{P} =: \mu_{\mathcal{G}}^X(A) = \int_A Y d\mathbb{P}\lceil_{\mathcal{G}} = \int_A Y d\mathbb{P} \tag{7.5}$$

Version 2. ($\mathcal{L}^2$ *projection*) We first assume $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) =: \mathcal{L}^2(\mathcal{F})$. Then for $\mathcal{G} \subseteq \mathcal{F}$, $\mathcal{L}^2(\mathcal{G}) = \{Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) : Y \in m\mathcal{G}\}$ is a Hilbert space, and a subspace of $\mathcal{L}^2(\mathcal{F})$. Because

- $\mathcal{L}^2(\mathcal{G})$ is complete. Given any Cauchy $\{Y_n\}$ in it, $\{Y_n\}$ admits a limit in $\mathcal{L}^2(\mathcal{G})$, itself. Because
- $\{Y_n\}$ is also a Cauchy in $\mathcal{L}^2(\mathcal{F}) \Rightarrow \exists Y_\infty \in \mathcal{F}$, s.t. $Y_n \xrightarrow{\mathcal{L}^2} Y_\infty \Rightarrow Y_n \xrightarrow{i.p} Y_\infty$.
- Exists subsequence $\{Y_{n_k}\}$, $Y_{n_k} \xrightarrow{a.s.} Y_\infty$. Since $Y_{n_k} \in m\mathcal{G}$, a.s. convergence preserves measurability, so $Y_\infty \in m\mathcal{G}$, i.e. $Y_\infty \in \mathcal{L}^2(\mathcal{G})$.

For any $X \in \mathcal{L}^2(\mathcal{F})$, consider projection of X onto $\mathcal{L}^2(\mathcal{G})$, denoted by $P_{\mathcal{G}}X$, by projection, we mean

- $P_{\mathcal{G}}X \in m\mathcal{G}$.
- $(X - P_{\mathcal{G}}X)$ is orthogonal to $P_{\mathcal{G}}X$, i.e. for all $Y \in \mathcal{L}^2(\mathcal{G})$:

$$\int_\Omega Y(X - P_{\mathcal{G}}X) d\mathbb{P} = 0 \tag{7.6}$$

For any $A \in \mathcal{G}$, take $Y = \mathbb{1}_A$, we have

$$\int_\Omega \mathbb{1}_A(X - P_{\mathcal{G}}X) d\mathbb{P} = 0 \tag{7.7}$$

The conditional expection is exactly $P_{\mathcal{G}}X$.
Now for general $X \in \mathcal{L}^1$, take $X_n^\pm \in SF^+$, such that $X_n^\pm \nearrow X^\pm$. By simple function we have $X_n^\pm \in \mathcal{L}^2$ for free. By previous arguments we define $\mathbb{E}[X_n^\pm|\mathcal{G}] := P_{\mathcal{G}}X_n^\pm$.
$P_{\mathcal{G}}X_n^\pm \nearrow Y^\pm$ for some $Y^\pm \in m\mathcal{G}$ (since limit transfers measurability). We can verify by (**MON**) that $Y^\pm$ has defining property of $\mathbb{E}[X^\pm|\mathcal{G}]$.
Finally since everything are finite, by linearity, $Y = Y^+ - Y^- =: \mathbb{E}[X|\mathcal{G}]$. ■

[1] Note: here we get correct measurability of Y for free.

Ex Examples of conditional expection.

1. $\mathcal{G} = \sigma(A) = \sigma(\mathbb{1}_A) = \{\emptyset, \Omega, A, A^c\}$, $A \in \mathcal{F}$. Then for every $X \in \mathcal{L}^1$,

$$\mathbb{E}\left[X|\mathcal{G}\right] = \mathbb{1}_A \frac{\mathbb{E}\left[X; A\right]}{\mathbb{P}\left(A\right)} + \mathbb{1}_{A^c} \frac{\mathbb{E}\left[X; A^c\right]}{\mathbb{P}\left(A^c\right)} \tag{7.8}$$

Rm. Since $\mathbb{E}\left[X|\mathcal{G}\right] \in m\sigma(A) = m\sigma(\mathbb{1}_A$, think about it, $\mathbb{E}\left[X|\mathcal{G}\right]$ must be somehow a function *composed* with $\mathbb{1}_A$.

2. (*Conditioning of events*) If $X = \mathbb{1}_B$, $B \in \mathcal{F}$, $(\mathbb{E}\left[B|A\right])$

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_B|\sigma(A)\right] &= \mathbb{1}_A \frac{\mathbb{P}\left(A \cap B\right)}{\mathbb{P}\left(A\right)} + \mathbb{1}_{A^c} \frac{\mathbb{P}\left(B \cap A^c\right)}{\mathbb{P}\left(A^c\right)} \\ &= \mathbb{1}_A \mathbb{P}\left(B|A\right) + \mathbb{1}_{A^c} \mathbb{P}\left(B|A^c\right)\end{aligned} \tag{7.9}$$

3. (*Conditioning of RVs*) $\mathcal{G} = \sigma(Y)$, X, Y have joint pdf $f_{(X,Y)}$, $\mathbb{E}\left[X|\mathcal{G}\right] =: \mathbb{E}\left[X|Y\right]$, define conditional pdf of X given Y as

$$f_{X|Y}(x|y) := \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \quad \text{if } f_Y(y) \neq 0 \text{, else } 0 \tag{7.10}$$

Assume $h : \mathbb{R} \to \mathbb{R}$ is Borel function s.t. $h(X) \in \mathcal{L}^1$, define

$$g(y) := \int_{\mathbb{R}} h(x) f_{X|Y}(x,y) dx \tag{7.11}$$

Then conditional expection of X given Y is g composed with Y. (again, c.f. remark in example 1, since $\mathbb{E}\left[X|Y\right] \in m\sigma(Y)$, it must be a function composed with Y.)

$$\mathbb{E}\left[h(X)|Y\right] = g(Y) = \int_{\mathbb{R}} h(x) f_{X|Y}(x,Y) dx \tag{7.12}$$

7.2 Properties

7.2.1 Simple properties

Prop. (**Expectation**) a special case of tower property:

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]\right] = \mathbb{E}\left[X\right] \tag{7.13}$$

Prop. If $X \in m\mathcal{G}$, then

$$\mathbb{E}\left[X|\mathcal{G}\right] = X \tag{7.14}$$

Prop. (**Linearlity**)

$$\mathbb{E}\left[aX + bY|\mathcal{G}\right] = a\mathbb{E}\left[X|\mathcal{G}\right] + b\mathbb{E}\left[Y|\mathcal{G}\right] \tag{7.15}$$

Prop. (**Monotonicity**) If $X_1 \leq X_2$, then

$$\mathbb{E}\left[X_1|\mathcal{G}\right] \leq \mathbb{E}\left[X_2|\mathcal{G}\right] \tag{7.16}$$

7.2.2 Conditional Convergence Thms

Prop. (**cMON**) If $X_n \nearrow X$, $X_n, X \in \mathcal{L}^1$; then $\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}]$.

Proof. Take $X_n - X_1 \nearrow X - X_1$, clearly $X_n - X_1 \in (m\mathcal{F})^+$. Define $Y_n := \mathbb{E}[X_n|\mathcal{G}]$, for all $A \in \mathcal{G}$,

$$\begin{aligned}\int_A (X - X_1) d\mathbb{P} &= \lim_{n\to\infty} \int_A (X_n - X_1) d\mathbb{P} \quad (\textbf{MON}) \\ &= \lim_{n\to\infty} \int_A (\mathbb{E}[X_n|\mathcal{G}] - \mathbb{E}[X_1|\mathcal{G}]) d\mathbb{P} \quad (\text{definition}) \\ &= \int_A \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] - \mathbb{E}[X_1|\mathcal{G}]\, d\mathbb{P} \quad (\textbf{MON}) \text{ again}\end{aligned} \tag{7.17}$$

Cancel out X_1, we have

$$\int_A X d\mathbb{P} = \int_A \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]\, d\mathbb{P} \tag{7.18}$$

So by definition, $\mathbb{E}[X|\mathcal{G}] := \lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$. ■

Prop. (**cFatou**) If $X_n \geq 0$, then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n|\mathcal{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \tag{7.19}$$

Prop. (**cDOM**) If $|X_n| \leq Y \in \mathcal{L}^1$, $X_n \xrightarrow{a.s.} X$, then $\mathbb{E}[X_n|\mathcal{G}] \xrightarrow{a.s.} \mathbb{E}[X|\mathcal{G}]$.

Prop. (**cJensen**) $\phi : \mathbb{R} \to \mathbb{R}$, *convex*. $\phi(x) \in \mathcal{L}^1$. Then $\mathbb{E}[\phi(X)|\mathcal{G}] \geq \phi(\mathbb{E}[X|\mathcal{G}])$.

Cor. If $X \in \mathcal{L}^p$, then $|\mathbb{E}[X|\mathcal{G}]|^p \leq \mathbb{E}[|X|^p|\mathcal{G}]$. Moreover we take p norm of both sides,

$$(\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|^p])^{\frac{1}{p}} \leq (\mathbb{E}[\mathbb{E}[|X|^p|\mathcal{G}]])^{\frac{1}{p}} = \mathbb{E}[|X|^p]^{\frac{1}{p}} \tag{7.20}$$

i.e., $X \in \mathcal{L}^p$ automatically guarantees that $\mathbb{E}[X|\mathcal{G}] \in \mathcal{L}^p$, and

$$\|\mathbb{E}[X|\mathcal{G}]\|_{\mathcal{L}^p} \leq \|X\|_{\mathcal{L}^p} \tag{7.21}$$

7.2.3 Tower Property

Prop. (**Tower property**) Suppose $\mathcal{H} \subseteq \mathcal{G}$ is a sub σ algebra, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] \tag{7.22}$$

Proof. The first equal sign is trivial, because $\mathbb{E}[X|\mathcal{H}] \in m\mathcal{H} \subseteq m\mathcal{G}$. By property 2, $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}]$ can be taken out from outer expectation.
The second one. For all $A \in \mathcal{H} \subseteq \mathcal{G}$,

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \int_A \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\, d\mathbb{P} \tag{7.23}$$

The first equal sign follows that $A \in \mathcal{G}$, second follows that $A \in \mathcal{H}$. By definition, $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$. ■

7.2.4 Taking out what is known

Prop. Suppose $Z \in m\mathcal{G}$ and $XZ \in \mathcal{L}^1$, then $\mathbb{E}[XZ|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$.

Proof. Follow the definition, it suffices to show that for all $A \in \mathcal{G}$,

$$\int_A XZd\mathbb{P} = \int_A Z\mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} \quad (\dagger) \tag{7.24}$$

Where $Z \in m\mathcal{G}$. First we assume $Z = \mathbb{1}_B$ for $B \in \mathcal{G}$, then

$$LHS = \int_{A\cap B} Xd\mathbb{P} = \int_{A\cap B} \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = RHS \tag{7.25}$$

Equal sign in the middle follows the definition, where $A \cap B \in \mathcal{G}$.
By linearity, dagger holds for all $Z \in S\mathcal{G}^+$ (simple function measurable on $\mathcal{G}$).
By (**MON**), holds for all $Z \in (m\mathcal{G})^+$ with $X^\pm$.[2]

$$|XZ| = (X^+ + X^-)(Z^+ + Z^-) < \infty \tag{7.26}$$

So $X^\pm Z^\pm \in \mathcal{L}^1$ for any combinations of plus minus, thus all integrals involved in dagger are finite, by linearity, dagger holds for general X, Z. ■

7.2.5 Independence condition

Prop. (*Drop the independent sigma algebra*) If $\mathcal{H} \subseteq \mathcal{F}$ is another sub sigma algebra; $\mathcal{H}$ is indep. of $\sigma(\mathcal{G}, \sigma(X))$, then

$$\mathbb{E}[X|\sigma(\mathcal{G},\mathcal{H})] = \mathbb{E}[X|\mathcal{G}] \quad (\triangle) \tag{7.27}$$

In particular, if $\mathcal{H}$ is indep of $\sigma(X)$,

$$\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X] \tag{7.28}$$

Proof. Define

$$\mathcal{I} := \{G \cap H : G \in \mathcal{G}, H \in \mathcal{H}\} \tag{7.29}$$

One can verify that $\mathcal{I}$ is a pi system. Moreover $\sigma(I) = \sigma(\mathcal{G},\mathcal{H})$. Examine eq triangle, we can see that LHS is the conditional expectation of X given $\sigma(\mathcal{G},\mathcal{H})$. So it suffices to estabilish: for all $A \in \sigma(\mathcal{G},\mathcal{H})$

$$\int_A Xd\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} \tag{7.30}$$

It futher suffices to show this only on pi system. For $A \in \mathcal{I}$, say $A = G \cap H$ for $G \in \mathcal{G}, H \in \mathcal{H}$. We have

$$\begin{aligned}
\int_{G\cap H} Xd\mathbb{P} = \int_\Omega \mathbb{1}_G\mathbb{1}_H Xd\mathbb{P} &= \mathbb{E}[\mathbb{1}_G\mathbb{1}_H X] \\
&= \mathbb{E}[\mathbb{1}_H]\cdot\mathbb{E}[\mathbb{1}_G X] = \mathbb{P}(H)\int_G Xd\mathbb{P} \\
&= \mathbb{P}(H)\int_G \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} \quad \text{(By definition for } G \in \mathcal{G}\text{)} \\
&= \int_{G\cap H} \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P}
\end{aligned} \tag{7.31}$$

Apply extension theorem, for $A \in \sigma(\mathcal{I})$, this also holds. ■

[2] we don't know the sign of X, so we pose constraint to $X^\pm$ such that XZ is positive.

Prop. (*Two coordinates*) Assume X, Y indep RVs, law $\mathcal{L}_X, \mathcal{L}_Y$. $h : \mathbb{R}^2 \to \mathbb{R}$ is Borel function s.t. $h(X,Y) \in \mathcal{L}^1$. Define function γ^h,

$$\gamma^h(x) := \mathbb{E}\left[h(x,Y)\right] \tag{7.32}$$

(taking expectation wrt second coordinate; integrate second coordinate out). Then, $\mathbb{E}\left[h(X,Y)|\sigma(X)\right] = \gamma^h(X)$. ($\gamma^h(X) \in m\sigma(X)$ follows this.)

Rm. This proposition is saying, for borel function $h(X,Y)$, the best predition of h given $\sigma(X)$ is just integrate Y out.

Proof. It is sufficient to show that for all $A \in \sigma(X)$ (the preimage set, for $B \in \mathscr{B}(\mathbb{R})$, $A = \{w : X(w) \in B\}$)

$$\int_A h(X,Y)d\mathbb{P} = \int_A \gamma^h(X)d\mathbb{P} \tag{7.33}$$

We start from LHS, $A := \{w \in \Omega : X(w) \in B\}$

$$\begin{aligned}
\int_{\{w:X(w)\in B\}} h(X,Y)d\mathbb{P} &= \iint_{B\times\mathbb{R}} h(x,y)d\mathcal{L}_{(X,Y)} \\
&= \iint_{B\times\mathbb{R}} h(x,y)d(\mathcal{L}_X \times \mathcal{L}_Y) \quad \text{(using indep.)} \\
&= \int_B \left(\int_{\mathbb{R}} h(x,y)d\mathcal{L}_Y\right) d\mathcal{L}_X \\
&= \int_B \gamma^h(x)d\mathcal{L}_X = \int_A \gamma^h(X)d\mathbb{P} \quad \blacksquare
\end{aligned} \tag{7.34}$$

7.3 Martingale

Def. **Stochastic Process**: A sequence of RVs from initial state X_0, $\{X_n : n \geq 0\}$ is called a stochastic process.

7.3.1 Filtration, Adaptedness

Def. **Filtration**: given $(\Omega, \mathcal{F}, \mathbb{P})$. $\{\mathcal{F}_n : n \geq 0\}$ is a filtration if

- $\mathcal{F}_n \in \mathcal{F}$ is sub sigma algebra.
- $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \geq 0$ (*nested*).

Def. **Filtered Space**: The probability space equipped with a filtration structure, i.e. $(\Omega, \mathcal{F}, \{\mathcal{F}_n : n \geq 0\}, \mathbb{P})$ is a filtered space.

Def. **Adaptedness**: A stochastic process $\{X_n : n \geq 0\}$ on filtered space is adapted if $X_n \in m\mathcal{F}_n$.
In particular, process $\{X_n\}$ is always adapted wrt the *Natural Filtration* $\{\mathcal{F}_n : n \geq 0\}$, where $\mathcal{F}_n := \sigma(X_0, X_1, ..., X_n)$.

7.3.2 Martingale, Sub/Sup Martingale

Def. **Martingale**: Given $(\Omega, \mathcal{F}, \mathbb{P})$. $\{\mathcal{F}_n : n \geq 0\}$, a adapted process $\{X_n : n \geq 0\}$ is a martingale if

- $X_n \in \mathcal{L}^1$.
- $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$, for all $n \geq 0$.

Ex. Popular examples.

- *Partial sum of a indep, 0-mean sequence of RVs forms a martingale.* Rigorously, $\{Y_n : n \geq 1\}$ is indep, $\mathbb{E}[Y_n] = 0$. $X_0 := 0$, $X_n := \sum_{j=1}^n Y_j$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$, $\mathcal{F}_n := \sigma(Y_1, Y_2, ..., Y_n)$, then $\{X_n\}$ is martingale wrt $\{\mathcal{F}_n\}$, we can check:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_n + Y_{n+1}|\mathcal{F}_n] = X_n + \mathbb{E}[Y_{n+1}] = X_n \tag{7.35}$$

- In addition to the first example, if $\mathrm{Var}[Y_n] = 1$, then $\{X_n^2 - n : n \geq 0\}$ is martingale. (X_n^2 is square of partial sum). On top of this one, if Y_n are i.i.d standard normal, then $\forall \lambda \in \mathbb{R}$, $\{e^{\lambda X_n - \frac{\lambda^2 n}{2}} : n \geq 0\}$ is martingale.
- If $X \in \mathcal{L}^1$, define $X_n := \mathbb{E}[X|\mathcal{F}_n]$, $\{X_n : n \geq 0\}$ is martingale. Check it inserting X_n, use tower property:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X|\mathcal{F}_n] =: X_n \tag{7.36}$$

Def. **Sub-Martingale**: $\{X_n : n \geq 0\}$ is a sub-martingale if $X_n \in \mathcal{L}^1$ and $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$. Similarly we define **Sup-Martingale**: $\mathbb{E}[X_{n+1}|\mathcal{F}_n]$.

Rm. Martingale is the model of *fair game*, sub-martingale says the future is better than present, the game is biased for us. Sup martingale says game is biased against us. Given $\{X_n\}$ a Sup-Martingale, than $\{-X_n\}$ is a sub-martingale.

Rm. (*Any future is same as one step forward*): For (sub) martingale, $\forall m \geq n+1$ (any future), by tower property and definition,

$$\mathbb{E}[X_m|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X_m|\mathcal{F}_{m-1}]|\mathcal{F}_n] = \mathbb{E}[X_{m-1}|\mathcal{F}_n] \tag{7.37}$$

Repeat this until $\mathcal{F}_{n+1}$, we get $\mathbb{E}[X_m|\mathcal{F}_n] = X_n$.

Thm. (**Composition with Convex Function**) Given $\{X_n : n \geq 0\}$ is adapted, let $\phi : \mathbb{R} \to \mathbb{R}$ convex, such that $\phi(X_n) \in \mathcal{L}^1$ $\forall n \geq 0$. If either

- $\{X_n : n \geq 0\}$ is a martingale.
- $\{X_n : n \geq 0\}$ is a submartingale, ϕ is non-decreasing

Then $\{\phi(X_n) : n \geq 0\}$ is a submartingale.

Proof. By (**cJensen**), $\forall n \geq 0$:

$$\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] \geq \phi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = \phi(X_n) \quad \text{For the first condition.} \tag{7.38}$$

For the second condition, $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$, since ϕ is non-decreasing, we have $\phi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) \geq \phi(X_n)$. ■

Cor. Use thm above:

- If $\{X_n : n \geq 0\}$ is martingale, then $\{|X_n|^p : n \geq 0\}$ is a submartingale for all $p \geq 1$.
- If $\{X_n : n \geq 0\}$ is submartingale, then $\{X_n^+ : n \geq 0\}$ is submartingale.
- If $\{X_n : n \geq 0\}$ is *non-negative* submartingale, then $\{X_n^p : n \geq 0\}$ is submartingale.

Proof. First one is clear. For the second one, view ϕ as $X_n^+ = \mathbb{1}_{(0,+\infty)} X_n$. non-decreasing. Same argument for third.

7.3.3 Doob's Decomposition Thm

Thm. (**Doob Decomposition**) $\{X_n : n \geq 0\}$ is a submartingale, then there exists a process $\{Y_n : n \geq 0\}$ such that

- $Y_0 = 0$, $Y_n \in \mathcal{L}^1$, $Y_{n+1} \in m\mathcal{F}_n$ forall $n \geq 0$, i.e. $\{Y_n : n \geq 0\}$ is a **previsable** process. (Y_{n+1} is known at n).
- Y_n is non-decreasing, i.e. $Y_n \leq Y_{n+1}$ a.s.
- $M_n := X_n - Y_n$ is a martingale.
- If Y_n exists, it's unique.

Proof. First show the uniqueness. Assume Y_n exists, assume not unique, i.e. exists another $\{W_n : n \geq 0\}$ also satisfies 1,2,3. Define $\Delta := Y_n - W_n$, clearly $\Delta_0 = 0$. Manipulate Δ_n:

$$\Delta_n = Y_n - W_n = (X_n - W_n) - (X_n - Y_n) \tag{7.39}$$

By linearity, and by (3), Δ_n is martingale. Hence $\Delta_n = \mathbb{E}\left[\Delta_{n+1} | \mathcal{F}_n\right]$. However since $Y_{n+1}, W_{n+1} \in m\mathcal{F}_n$, Δ_n is also previsible, $\Delta_{n+1} \in m\mathcal{F}_n$.

$$\Delta_n = \mathbb{E}\left[\Delta_{n+1} | \mathcal{F}_n\right] = \Delta_{n+1} = \ldots = \Delta_0 \equiv 0 \quad \blacksquare \tag{7.40}$$

(We can come up with a remark: if a process is a martingale and also previsible, then it is a constant.)
Now show the existance of Y_n. $Y_0 = 0$, for $n \geq 0$, define

$$Y_{n+1} := \sum_{j=0}^{n} \left(\mathbb{E}\left[X_{j+1} | \mathcal{F}_j\right] - X_j\right) \tag{7.41}$$

The increment part of submartingale. Since $\mathbb{E}\left[X_{j+1} | \mathcal{F}_j\right] \in m\mathcal{F}_j$, clearly $Y_{n+1} \in m\mathcal{F}_n$. By property of submartingale, every term in the summation is positive, so

$Y_n \le Y_{n+1}$. Now only need to prove $X_n - Y_n$ is martingale.

$$\begin{aligned}
\mathbb{E}\left[X_{n+1} - Y_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[X_{n+1} - \sum_{j=0}^{n}\left(\mathbb{E}\left[X_{j+1}|\mathcal{F}_j\right] - X_j\right)|\mathcal{F}_n\right] \\
&= \mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] - \sum_{j=0}^{n}\left(\mathbb{E}\left[X_{j+1}|\mathcal{F}_j\right] - X_j\right) \\
&= \mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] - \left(\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] - X_n\right) - \sum_{j=0}^{n-1}\left(\mathbb{E}\left[X_{j+1}|\mathcal{F}_j\right] - X_j\right) \\
&= X_n - \sum_{j=0}^{n-1}\left(\mathbb{E}\left[X_{j+1}|\mathcal{F}_j\right] - X_j\right) = X_n - Y_n \quad \blacksquare
\end{aligned} \tag{7.42}$$

7.4 Stopping Time

Def. **Stopping Time**: Given $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$. A Random variable $\tau : \Omega \to \{0, 1, 2, ..., \infty\}$ is a stopping time if $\{\tau \le n\} \in \mathcal{F}_n$. Using $\{\tau > n\}, \{\tau < n\}, \{\tau \ge n\}, \{\tau = n\}$ are equivalent, if τ is a stopping time, all these sets $\in \mathcal{F}_n$.

Def. (*Sigma algebra with stopping time subscript*): τ is a stopping time, $\mathcal{F}_\tau := \{A \in \mathcal{F}, A \cap \{\tau \le n\} \in \mathcal{F}_n \ \ \forall n \ge 0\}$. One can verify that $\mathcal{F}_\tau$ is a sigma algebra, but note that $\mathcal{F}_\tau \ne \sigma(\tau)$.

Def. (*RV with stopping time subscript*): $\{X_n : n \ge 0\}$ is adapted, for $w \in \Omega$ define

$$X_\tau(w) := X_{\tau(w)}(w) := \begin{cases} X_n(w) & \text{if } \tau(w) = n < +\infty \\ \begin{cases} \lim\limits_{n\to\infty} X_n(w) & \text{if } X_n \text{ admits limit} \\ \text{undefined} & \text{if limit does not exist} \end{cases} & \text{if } \tau(w) = +\infty \end{cases}$$

7.4.1 Simple Properties of Stopping Time

Prop.

1· If τ is stopping time, n is any fixed positive integer, then $\tau \wedge n := \min\{\tau, n\}$ is a stopping time.

2· If τ_1, τ_2 are stopping times, then $(\tau_1 \wedge \tau_2)$, $(\tau_1 + \tau_2)$, $(\tau_1 \vee \tau_2)$ are all stopping times.

3· If $\{X_n : n \ge 0\}$ is adapted, $\mathbb{P}(\tau < \infty) = 1$ then $X_\tau \in m\mathcal{F}_\tau$.

Proof. Since $\tau < \infty$ *a.s*, X_τ is defined a.s. Working out only a pi system is enough, for all $x \in \mathbb{R}$, we want to show that $\{X_\tau \le x\} \in \mathcal{F}_\tau$. By definition

$$\{X_\tau \le x\} \cap \{\tau \le n\} = \bigcup_{j=0}^{n} \{\tau = j, X_j \le x\} \in \mathcal{F}_n \quad \blacksquare \tag{7.43}$$

4· τ_1, τ_2 are stopping times, then $\mathcal{F}_{\tau_1 \wedge \tau_2} = \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$.

5· τ_1, τ_2 are stopping times, deterministically $\tau_1 \le \tau_2$, then $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$.

Proof. First show (5). $\forall A \in \mathcal{F}_{\tau_1}$, It suffices to show that $A \in \mathcal{F}_{\tau_2}$,i.e. $\forall n \in \bar{\mathbb{N}}$, $A \cap \{\tau_2 \leq n\} \in \mathcal{F}_n$. This is true, since $\{\tau_2 \leq n\} \subseteq \{\tau_1 \leq n\}$,

$$A \cap \{\tau_2 \leq n\} = A \cap \{\tau_1 \leq n\} \cap \{\tau_2 \leq n\} \in \mathcal{F}_n \tag{7.44}$$

Because $A \cap \{\tau_1 \leq n\} \in \mathcal{F}_n$. ■

For (4), $LHS \subseteq RHS$ is clear, since $LHS \subseteq \mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}$. Only need to show ($\supseteq$) for all $A \in \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$,

$$\begin{aligned} A \cap \{\tau_1 \wedge \tau_2 \leq n\} &= A \cap (\{\tau_1 \leq n\} \cup \{\tau_2 \leq n\}) \\ &= (A \cap \{\tau_1 \leq n\}) \cup (A \cap \{\tau_2 \leq n\}) \end{aligned} \tag{7.45}$$

6· $\{X_n : n \geq 0\}$ is adapted, τ is stopping time, then $\{X_{\tau\wedge n} : n \geq 0\}$ is also an adapted process.

7.4.2 Doob's Stopping Time Thm

Thm. (**Doob**) If $\{X_n : n \geq 0\}$ is a martingale/submartingale, τ is a stopping time, then $\{X_{\tau\wedge n} : n \geq 0\}$ is still a martingale/submartingale.

Proof. Clearly, $X_{n\wedge\tau} \in \mathcal{L}^1$, because $X_{n\wedge\tau} = \sum_{j=0}^{n} \mathbb{1}_{(\tau=j)} X_j + \mathbb{1}_{(\tau>n)} X_n$. Now we show $\{X_{\tau\wedge n} : n \geq 0\}$ is a martingale. Concretely, we want to show $\mathbb{E}\left[X_{(n+1)\wedge\tau} | \mathcal{F}_n\right] = X_{n\wedge\tau}$. For all $A \in \mathcal{F}_n$,

$$\begin{aligned} \int_A X_{(n+1)\wedge\tau} d\mathbb{P} &= \int_{A\cap\{\tau\leq n\}} X_\tau d\mathbb{P} + \int_{A\cap\{\tau>n\}} X_{n+1} d\mathbb{P} \\ &= \int_{A\cap\{\tau\leq n\}} X_\tau d\mathbb{P} + \int_{A\cap\{\tau>n\}} X_{n+1} d\mathbb{P} \quad \text{(since } X_n \text{ is martingale)} \\ &= \int_A X_{\tau\wedge n} d\mathbb{P} \quad \blacksquare \end{aligned} \tag{7.46}$$

7.4.3 Hunt's Thm

Thm. (**Hunt**) $\{X_n : n \geq 0\}$ is a martingale/submartingale. τ_1, τ_2 are stopping times, $\tau_1 \leq \tau_2$. If one of following conditions holds

- τ_1, τ_2 are bounded, i.e. $\exists T > 0$, $\tau_1, \tau_2 \leq T$.
- $\{X_n : n \geq 0\}$ is uniformly integrable. And τ_1, τ_2 are finite a.s.
- $\mathbb{E}[\tau_1] \leq \mathbb{E}[\tau_2] < \infty$. And exists constant $k > 0$, s.t. $|X_{n+1} - X_n| \leq k$, $\forall n \geq 0$.

Then, $\mathbb{E}[X_{\tau_2} | \mathcal{F}_{\tau_1}] = X_{\tau_1}$. ($\geq$ for submartingale case)

7.4.4 Wald's Identity

7.5 Random Walk

7.6 Martingale Convergence

7.6.1 Doob's Upcrossing Inequility

7.6.2 Martingale Convergence Thm 1 (MCT1)

7.6.3 Martingale Convergence Thm 2 (MCT2)

7.6.4 Doob's Maximal Inequility

7.6.5 Martingale Convergence Thm 3 (MCT3)

7.6.6 Converse MCT2

7.6.7 Generalized 0-1 Law

Chapter 8

Problems

Problem 1. (**Equivalent Generating pi of Borel on Real Line**) Show that

$$\begin{aligned}\mathscr{B}(\mathbb{R}) &= \sigma(\{[a,b) : a,b \in \mathbb{R}, a < b\}) \\ &= \sigma(\{[a,b] : a,b \in \mathbb{R}, a < b\}) \\ &= \sigma(\{(-\infty,x) : x \in \mathbb{Q}\}) \\ &= \sigma(\{(-\infty,x] : x \in \mathbb{Q}\})\end{aligned} \tag{8.1}$$

Proof. Clearly, RHS $\subseteq \mathscr{B}(\mathbb{R})$. It's sufficient to show $\supseteq$. The target is to rewrite original pi (a,b) to be these 4 alternative pi. But for the first one we just show both.

.

$$\begin{aligned}(a,b) &= \bigcup_{n\geq 1}[a+\frac{1}{n},b] \quad \Rightarrow \mathscr{B}(\mathbb{R}) \subseteq RHS1 \\ [a,b) &= \bigcap_{n\geq 1}[a,b+\frac{1}{n}] \quad \Rightarrow RHS1 \subseteq \mathscr{B}(\mathbb{R})\end{aligned} \tag{8.2}$$

$$(a,b) = \bigcup_{n\geq 1}[a+\frac{1}{n},b-\frac{1}{n}] \Rightarrow \mathscr{B}(\mathbb{R}) \subseteq RHS2 \tag{8.3}$$

3 and 4; For any $a \in \mathbb{R}$, $\exists\{q_n\}, n \geq 1$ be a seq of rationals s.t. $q_n \nearrow a$ (increasingly) So,

$$(-\infty,a) = \bigcup_{n\geq 1}(-\infty,q_n) \nwarrow (-\infty,q_n) \tag{8.4}$$

Therefore we also find $p_n \nearrow b$, $\{p_n\} \subseteq \mathbb{Q}$:

$$\begin{aligned}[a,b) &= (-\infty,b) \setminus (-\infty,a) \\ &= \bigcup_{n\geq 1}(-\infty,q_n) \cap \left(\bigcup_{n\geq 1}(-\infty,q_n)\right)^c\end{aligned} \tag{8.5}$$

Implies $RHS3 \subseteq RHS1 \subseteq \mathscr{B}(\mathbb{R})$. For 4:

$$(-\infty,x) = \bigcup_{n\geq 1}(-\infty,x-\frac{1}{n}) \quad \Rightarrow RHS4 \subseteq RHS3 \subseteq \mathscr{B}(\mathbb{R} \tag{8.6}$$

Problem 2. (**Singletons are not enough to generate Borel sigma**) Show $\mathscr{B}(\mathbb{R})$ is not generated by all singletons of $\mathbb{R}$. I.e show that

$$\mathscr{B}(\mathbb{R}) \neq \sigma(\{x\}, x \in \mathbb{R}) := \mathcal{S} \tag{8.7}$$

Proof. Define

$$\begin{aligned} \mathcal{A} &:= \{\emptyset\} \cup \{\bigcup_{n\geq 1}\{r_n\} : r_n \in \mathbb{R}\} \\ \mathcal{B} &:= \{B \in \mathbb{R} : B^c \in \mathcal{A}\} \end{aligned} \tag{8.8}$$

i.e. $\mathcal{A}$ is collection of countable unions of singletons. $\mathcal{B}$ is collection of complements of things in $\mathcal{A}$. We claim that $\Sigma := \mathcal{A} \cup \mathcal{B}$ is a sigma-field.

- $\emptyset \in \Sigma$.
- $\forall A \in \Sigma$, $A^c \in \Sigma$.
- Consider countablly many $A_n \in \Sigma, n \geq 1$. A_n should be either in $\mathcal{A}$ or $\mathcal{B}$. Denote $\mathcal{I} := \{i : A_i \in \mathcal{A}\}$; $\mathcal{J} := \{j : A_j \in \mathcal{B}\}$ as indices sets marking whether collection A_n belongs. Then,

$$\bigcup_{n\geq 1} A_n = \left(\bigcup_{i\in\mathcal{I}} A_i\right) \cup \left(\bigcup_{j\in\mathcal{J}} A_j\right) =: U_1 \cup U_2 =: U \tag{8.9}$$

  where $U_1 \in \mathcal{A} \subseteq \Sigma$. $U_2 = (\bigcap_{j\in\mathcal{J}} A_j^c)^c$, $\bigcap_{j\in\mathcal{J}} A_j \in \mathcal{A}$. So $U_2 \in \mathcal{B} \subseteq \Sigma$. So $U \in \mathcal{A} \cup \mathcal{B} = \Sigma$. **Check: Σ is a sigma field**.

Clearly all singletions contained in $\mathcal{A}$, therefore Σ. So $\sigma(\{x : x \in \mathbb{R}\}) \subseteq \Sigma$. But $\mathscr{B}(\mathbb{R}) \supset (0,1) \notin \Sigma$. ■

Problem 3. (**Defining properties of Measure**) $S = (0,1]$, define

$$\Sigma := \left\{\bigcup_{i=1}^{k}(a_i, b_i] : k \in \mathbb{N}, 0 \leq a_1 \leq b_1 \leq a_2 \leq ... \leq a_k \leq b_k \leq 1\right\} \tag{8.10}$$

(Shown) Σ is sigma field. Define $\mu : \Sigma \mapsto [0,\infty]$, for $A \in \Sigma$,

$$\mu(A) = \begin{cases} 1 & \text{if } A \supseteq (\frac{1}{2}, \frac{1}{2}+\epsilon] \text{ for some } \epsilon > 0, \\ 0 & \text{otherwise} \end{cases} \tag{8.11}$$

Show (1) μ is finite additive. (2) μ is not countable additive.

Proof. For $A_n \in \Sigma, n = 1,2,...,N$. A_n disjoint.
Then there is at most one A_k s.t. $A_k \supseteq (\frac{1}{2}, \frac{1}{2}+\epsilon]$ i.e. $\mu(A_k) = 1$ and $\mu(A_j) = 0$ for $j \neq k$. Clearly $\mu(\bigcup_{j=1}^{N} A_j) = \sum_{j=1}^{N} \mu(A_j)$. ■
For the second part, it suffices to show μ is not continous (from above) at emptyset. Pick $\{A_n\}$, $A_n := (\frac{1}{2}, \frac{1}{2}+\frac{1}{n}]$. Clearly $A_n \searrow \emptyset$. But $\epsilon := \frac{1}{2n}$ for any $n \geq 1$, $\mu(A_n) \equiv 1$. ■

Problem 4. (**Indep.**) $S = (0,1]$,

8.1 Prob Space

8.2 RV

8.3 Expectation

Problem 1. On (S, Σ, μ) $f_n, g_n \in \mathcal{L}^1(S, \Sigma, \mu)$. $|f_n| \leq g_n$ for all $n \geq 1$. $\forall s \in S$, $f_n \to f$, $g_n \to g$.

Show that if $\mu(g_n) \to \mu(g) < \infty$, *then* $\mu(f)$ is defined, and $\mu(f_n) \to \mu(f)$

Proof. $|f_n| \leq g_n \Rightarrow g_n + f_n \geq 0$ and $g_n - f_n \geq 0$.
Apply (**FATOU**), and by linearly of Fatou's LHS:

$$\begin{aligned} \mu(g) + \mu(f) &= \mu(\liminf_{n\to\infty}(g_n + f_n)) \leq \liminf_{n\to\infty} \mu(g_n + f_n) \\ &= \mu(g) + \liminf_{n\to\infty} \mu(f_n) \end{aligned} \tag{8.12}$$

$$\begin{aligned} -\mu(g) + \mu(f) &= -\mu(\liminf_{n\to\infty}(g_n - f_n)) \geq -\liminf_{n\to\infty} \mu(g_n - f_n) \\ &= -\mu(g) + \limsup_{n\to\infty} \mu(f_n) \end{aligned} \tag{8.13}$$

Since $g \in \mathcal{L}^1$, $\mu(g)$ can be cancelled out from both sides:

$$\liminf_{n\to\infty} \mu(f_n) \leq \limsup_{n\to\infty} \mu(f_n) \leq \mu(f) \leq \liminf_{n\to\infty} \mu(f_n) \tag{8.14}$$

Therefore $\mu(f) := \liminf_{n\to\infty} \mu(f_n)$ is defined. Moreover $\lim_{n\to\infty} \mu(f_n)$ exists, and $\mu(f) = \lim_{n\to\infty} \mu(f_n)$. ∎

Problem 2. $(\Omega, \mathcal{F}, \mathbb{P})$, $X_n, X \in \mathcal{L}^1$, $X_n \xrightarrow{i.p} X$, $\mathbb{E}[X_n] \to \mathbb{E}[X]$, show that

$$X_n \xrightarrow{\mathcal{L}^1} X$$

(Strengthened **SCHEFFE**)

Proof. Assume opposite, NOT $X_n \xrightarrow{\mathcal{L}^1} X$, i.e. $\exists \epsilon > 0$ and subsequence $\{n_l\}$, such that $\mathbb{E}[X_{n_l} - X] \geq \epsilon$. (#)
Clearly $X_{n_l} \xrightarrow{i.p} X$. By theorem, there exists a further subsequence $X_{n_{l_m}}$ such that $X_{n_{l_m}} \xrightarrow{a.s.} X$. Moreover $\mathbb{E}\left[X_{n_{l_m}}\right] \xrightarrow{m\to\infty} \mathbb{E}[X]$ and $X_{n_{l_m}} \in \mathcal{L}^1$ for any subscript, because $\{X_{n_{l_m}}\} \subseteq \{X_n\}$.
Apply original (**Scheffe**) to $X_{n_{l_m}}$, we have $X_{n_{l_m}} \xrightarrow{\mathcal{L}^1} X$, i.e. $\forall \epsilon > 0$, $\exists M$, for all $n > M$, $\mathbb{E}\left[X_{n_{l_m}} - X\right] < \epsilon$, which contradicts (#). ∎

Problem 3.

$$\{X_n\}, \{Y_n\} \text{ are uniformly integrable } \Rightarrow \{X_n + Y_n\} \text{ is uniformly integrable}$$

Proof. For $M > 0$, consider:

$$\begin{aligned}
&\sup_n \mathbb{E}\left[|X_n + Y_n|; |X_n + Y_n| > M\right] \\
\leq &\sup_n \mathbb{E}\left[|X_n + Y_n|; |X_n| > \frac{M}{2} \;\&\; |Y_n| > \frac{M}{2} \;\&\; |X_n + Y_n| > M\right] + \\
&\sup_n \mathbb{E}\left[|X_n + Y_n|; |X_n| \leq \frac{M}{2} \;\&\; |Y_n| > \frac{M}{2} \;\&\; |X_n + Y_n| > M\right] + \\
&\sup_n \mathbb{E}\left[|X_n + Y_n|; |X_n| > \frac{M}{2} \;\&\; |Y_n| \leq \frac{M}{2} \;\&\; |X_n + Y_n| > M\right]
\end{aligned} \tag{8.15}$$

In which first term $\leq \sup_n \mathbb{E}\left[|X_n|; |X_n| > \frac{M}{2}\right] + \sup_n \mathbb{E}\left[|Y_n|; |Y_n| > \frac{M}{2}\right]$,
Second term $\leq 2\sup_n \mathbb{E}\left[|Y_n|; |Y_n| > \frac{M}{2}\right]$,
Third term $\leq 2\sup_n \mathbb{E}\left[|X_n|; |X_n| > \frac{M}{2}\right]$.

$$\begin{aligned}
LHS &\leq 3\sup_n \mathbb{E}\left[|X_n|; |X_n| > \frac{M}{2}\right] + 3\sup_n \mathbb{E}\left[|Y_n|; |Y_n| > \frac{M}{2}\right] \\
&\xrightarrow{M\to\infty} 3 \times 0 + 3 \times 0 = 0 \quad \blacksquare
\end{aligned} \tag{8.16}$$

Problem 4. Non-trivial RV X ($\mathbb{P}(X > 0) > 0$). Show that if $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, then for every $\lambda \in [0, 1]$,

$$\mathbb{P}(|X| \geq \lambda\mathbb{E}\left[|X|\right]) \geq \frac{(1-\lambda)^2\mathbb{E}^2\left[|X|\right]}{\mathbb{E}\left[X^2\right]}$$

Proof. Consider

$$\begin{aligned}
\mathbb{E}\left[|X|\right] &= \mathbb{E}\left[|X| \cdot 1; |X| \geq \lambda\mathbb{E}\left[|X|\right]\right] + \mathbb{E}\left[|X|; |X| < \lambda\mathbb{E}\left[|X|\right]\right] \\
&\leq \mathbb{E}^{\frac{1}{2}}\left[X^2; |X| \geq \lambda\mathbb{E}\left[|X|\right]\right] \cdot \mathbb{E}^{\frac{1}{2}}\left[1^2; |X| \geq \lambda\mathbb{E}\left[|X|\right]\right] + \lambda\mathbb{E}\left[|X|\right] \\
&\leq \mathbb{E}^{\frac{1}{2}}\left[X^2\right] \cdot \mathbb{P}^{\frac{1}{2}}\left(|X| \geq \lambda\mathbb{E}\left[X\right]\right) + \lambda\mathbb{E}\left[|X|\right]
\end{aligned} \tag{8.17}$$

Where the first leq applys (Holders) ineq. Rearrange terms we have

$$(1-\lambda)\mathbb{E}\left[X\right] \leq \mathbb{E}^{\frac{1}{2}}\left[X^2\right]\mathbb{P}^{\frac{1}{2}}\left(|X| \geq \lambda\mathbb{E}\left[|X|\right]\right) \tag{8.18}$$

Take square both sides,

$$\mathbb{P}\left(|X| \geq \lambda\mathbb{E}\left[|X|\right]\right) \geq \frac{(1-\lambda)^2\mathbb{E}^2\left[|X|\right]}{\mathbb{E}\left[X^2\right]} \quad \blacksquare \tag{8.19}$$

Problem 5. $\{X_n\} \in \mathcal{L}^2$; suppose $\mathbb{E}\left[X_iX_j\right] = 0$ for $i \neq j$, and $\sup_n \mathbb{E}\left[X_n^2\right] < \infty$. Show that for every $\alpha > \frac{1}{2}$:

$$\frac{\sum_{j=1}^n X_j}{n^\alpha} \xrightarrow{i.p} 0$$

Proof. By (**Markov**):

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n^\alpha}-0\right|>\epsilon\right) &= \mathbb{P}\left(\left(\frac{S_n}{n^\alpha}\right)^2>\epsilon^2\right) \\
&< \epsilon^{-2}\mathbb{E}\left[\frac{S_n^2}{n^{2\alpha}}\right] \\
&= \epsilon^{-2}n^{-2\alpha}\cdot\mathbb{E}\left[\sum_{j=1}^n X_n^2+\sum_{i\neq j}X_iX_j\right] \\
&\leq \epsilon^{-2}n^{-2\alpha}\cdot n\sup_n \mathbb{E}\left[X_n^2\right] \\
&= n^{1-2\alpha}\frac{\sup_n \mathbb{E}\left[X_n^2\right]}{\epsilon^2}
\end{aligned} \tag{8.20}$$

Since $\sup_n \mathbb{E}[X_n^2]<\infty$, we conclude that for all $\epsilon>0$, if $\alpha>1/2$, eq (4.11)$\xrightarrow{n\to\infty}0$; i.e.

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{S_n}{n^\alpha}-0\right|>\epsilon\right)=0 \tag{8.21}$$

We conclude that $\frac{S_n}{n^\alpha}\xrightarrow{i.p}0$. ∎

Problem 6. $\{X_n\}$: identically distributed RV. $\mathbb{E}[X_1^2]<\infty$.
Show: (1) for all $\epsilon>0$:

$$\lim_{n\to\infty} n\cdot\mathbb{P}\left(|X_1|\geq\epsilon\sqrt{n}\right)=0$$

(2):

$$\frac{1}{\sqrt{n}}\max_{1\leq k\leq n}|X_k|\xrightarrow{i.p}0$$

(1) *Proof.* $X_1^2\in\mathcal{L}^1\Rightarrow\lim_{n\to\infty}\mathbb{E}[X_1^2;X_1^2>n]=\lim_{n\to\infty}\mathbb{E}[X_1^2;|X_1|>\sqrt{n}]=0$. To be precise, $\forall\delta>0$, $\exists N$ large, s.t. $\forall n>N$: $\mathbb{E}[X_1^2;|X_1|>\sqrt{n}]<\delta$. So, for **Any Fixed** $\epsilon>0$, $\exists N'=\frac{N}{\epsilon^2}$ s.t. $\forall n>N'$:

$$\begin{aligned}
\mathbb{E}\left[X_1^2;|X_1|>\epsilon\sqrt{n}\right] &< \mathbb{E}\left[X_1^2;|X_1|>\epsilon\sqrt{\frac{N}{\epsilon^2}}\right] \\
&= \mathbb{E}\left[X_1^2;|X_1|>\sqrt{N}\right]\leq\delta
\end{aligned} \tag{8.22}$$

i.e. for **every fixed** $\epsilon>0$, $\lim_{n\to\infty}\mathbb{E}[X_1^2;|X_1|>\epsilon\sqrt{n}]=0$.

$$\begin{aligned}
\mathbb{E}\left[X_1^2;|X_1|>\epsilon\sqrt{n}\right] &= \int_{|X_1|>\epsilon\sqrt{n}}X_1^2d\mathbb{P} \\
&> (\epsilon\sqrt{n})^2\cdot\mathbb{P}\left(|X_1|>\epsilon\sqrt{n}\right)
\end{aligned} \tag{8.23}$$

i.e.

$$n\cdot\mathbb{P}\left(|X_1|>\epsilon\sqrt{n}\right)<\epsilon^{-2}\cdot\mathbb{E}\left[X_1^2;|X_1|>\epsilon\sqrt{n}\right] \tag{8.24}$$

Let $n\to\infty$, we get $\lim_{n\to\infty}n\cdot\mathbb{P}(|X_1|>\epsilon\sqrt{n})<\epsilon^{-2}\cdot 0=0$ as desired. ∎

(2) *Proof.* For any fixed ϵ, by the fact that $\{X_n\}$ have same law: $\mathbb{P}(X_k > c) = \mathbb{P}(X_1 > c)$ for all $c \in \mathbb{R}$, all $1 \leq k \leq n$.

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{\sqrt{n}} \max_{1\leq k\leq n} |X_k| - 0\right| > \epsilon\right) &= \mathbb{P}\left(\max_{1\leq k\leq n} |X_k| > \epsilon\sqrt{n}\right) \\
&= \mathbb{P}\left(\{X_k > \epsilon\sqrt{n} \text{ for some } 1 \leq k \leq n\}\right) \\
&= \mathbb{P}\left(\bigcup_{k=1}^{n}\{X_k > \epsilon\sqrt{n}\}\right) \\
&\leq \sum_{k=1}^{n} \mathbb{P}\left(X_k > \epsilon\sqrt{n}\right) \\
&= n \cdot \mathbb{P}\left(X_1 > \epsilon\sqrt{n}\right) \\
&\xrightarrow{n\to\infty,\text{By (1)'s result}} 0 \quad \blacksquare
\end{aligned} \tag{8.25}$$

Problem 7. $\{X_n\}$ seq of indep. RVs. $\mathbb{E}[X_n] = 0$, $\mathrm{Var}[X] = 1$ uniformly. Show that for every $Y \in \mathcal{L}^2$,

$$\mathbb{E}[X_n Y] \to 0$$

Proof. By $\mathbb{E}[X] = 0, \mathrm{Var}[X] = 1 \Rightarrow \mathbb{E}[X^2] = 1$.
Define $Y_n := \sum_{k=1}^{n} \mathbb{E}[X_k Y] X_k$, $\forall n \geq 1$, consider second moment

$$\begin{aligned}
\mathbb{E}\left[Y_n^2\right] &= \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{E}^2[X_k Y] X_k^2 + \sum_{1\leq i\neq j\leq n} \mathbb{E}[X_i Y]\mathbb{E}[X_j Y] X_i X_j\right] \\
&= \sum_{k=1}^{n} \mathbb{E}^2[X_k Y] + \sum_{1\leq i\neq j\leq n} \mathbb{E}[X_i Y]\mathbb{E}[X_j Y]\mathbb{E}[X_i]\mathbb{E}[X_j] \\
&= \sum_{k=1}^{n} \mathbb{E}^2[X_k Y]
\end{aligned} \tag{8.26}$$

Which follows that $\{X_n\}$ are independent, $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$ for $i \neq j$.
Now it suffices to show that $\mathbb{E}[Y_n^2] < \infty$ when $n \to \infty$, i.e. $\sup_n \mathbb{E}[Y_n^2] < \infty$.
Consider

$$\mathbb{E}[Y Y_n] = \mathbb{E}\left[Y \sum_{k=1}^{n} \mathbb{E}[X_k Y] X_k\right] = \sum_{k=1}^{n} \mathbb{E}^2[X_k Y] = \mathbb{E}\left[Y_n^2\right] \tag{8.27}$$

And

$$\begin{aligned}
0 \leq \mathbb{E}\left[(Y - Y_n)^2\right] &= \mathbb{E}\left[Y^2\right] - 2\mathbb{E}[Y Y_n] + \mathbb{E}\left[Y_n^2\right] \\
&= \mathbb{E}\left[Y^2\right] - 2\mathbb{E}\left[Y_n^2\right] + \mathbb{E}\left[Y_n^2\right] = \mathbb{E}\left[Y^2\right] - \mathbb{E}\left[Y_n^2\right]
\end{aligned} \tag{8.28}$$

Which implies $\mathbb{E}[Y_n^2] \leq \mathbb{E}[Y^2]$, i.e. $\sup_n \mathbb{E}[Y_n^2] \leq \mathbb{E}[Y^2] < \infty$, since $Y \in \mathcal{L}^2$ by hypothesis. Therefore

$$\sum_{k=1}^{\infty} \mathbb{E}^2[X_k Y] < \infty \tag{8.29}$$

So $\mathbb{E}[X_k Y] \xrightarrow{n\to\infty} 0$. $\blacksquare$

Problem 8. Show that following formula of the standard Gaussian rv: $X \sim N(0,1)$, then

$$\mathbb{E}\left[X^n\right] = \begin{cases} 0 & n \text{ odd,} \\ (n-1)!! & n \text{ even.} \end{cases}$$

Further, for every $k \geq 0$, $\mathbb{E}\left[|X|^{2k+1}\right] = 2^k k! \sqrt{2/\pi}$.

(1) *Proof.* For standard gaussian, we have density function:

$$\phi(x) := f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{8.30}$$

Notice that $\phi' = -x\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = -x\phi$.
For $n \geq 2$, applying integration by parts,

$$\begin{aligned} \mathbb{E}\left[X^{n-1}\right] &= \int_{\mathbb{R}} x^{n-1}\phi(x)dx \\ &= \left.\frac{x^n}{n-1}\right|_{-\infty}^{+\infty} - \int_{\mathbb{R}} \frac{x^n}{n-1}\phi'(x)dx \\ &= \left.\frac{x^n \cdot e^{-\frac{x^2}{2}}}{\sqrt{2\pi}(n-1)}\right|_{-\infty}^{+\infty} + \frac{1}{n-1}\int_{\mathbb{R}} x^{n+1}\phi(x)dx \\ &= \frac{1}{n-1}\mathbb{E}\left[X^{n+1}\right] \end{aligned} \tag{8.31}$$

So $\mathbb{E}\left[X^{n+1}\right] = (n-1)\mathbb{E}\left[X^{n-1}\right]$, $n \geq 2$.
Since $\mathbb{E}[X] = 0$, $\mathbb{E}\left[X^2\right] = 1 \Rightarrow$

$$\mathbb{E}\left[X^n\right] = \begin{cases} 0 & n \text{ odd,} \\ (n-1)!! & n \text{ even.} \end{cases} \quad \blacksquare$$

(2) *Proof.* Similar as (1),

$$\begin{aligned} \mathbb{E}\left[|X|^{n-1}\right] &= 2\int_{\mathbb{R}^+} x^{n-1}\phi(x)dx \\ &= 2\left[\left.\frac{x^n}{n-1}\right|_0^{+\infty} - \int_0^{+\infty} \frac{x^n}{n-1}\phi'(x)dx\right] \\ &= \frac{1}{n-1}\cdot 2\int_0^{+\infty} x^{n+1}\phi(x)dx \\ &= \frac{1}{n-1}\mathbb{E}\left[|X|^{n+1}\right] \end{aligned} \tag{8.32}$$

Since $\mathbb{E}[|X|] = \sqrt{2/\pi}$, $\mathbb{E}\left[|X|^2\right] = \mathbb{E}\left[X^2\right] = 1 \Rightarrow$

$$\mathbb{E}\left[|X|^n\right] = \begin{cases} \sqrt{\frac{2}{\pi}}\cdot(n-1)!! & n \text{ odd,} \\ (n-1)!! & n \text{ even.} \end{cases}$$

Take $n = 2k+1$ (odd), clearly $\mathbb{E}\left[|X|^{2k+1}\right] = 2^k k! \sqrt{2/\pi}$. $\blacksquare$

Problem 9. $X \in m\mathcal{F}^+$, show that

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

Proof. Firstly note that $X \in \mathbb{R}$ can be approached from below or above, i.e. $X = X^- = X^+$

$$X(w) = \int_0^{X(w)^-} 1 \cdot dt = \int_0^{X(w)^+} 1 \cdot dt \tag{8.33}$$

$$\begin{aligned}
\mathbb{E}[X] &= \int_\Omega \left[\int_0^{X(w)^-} 1 \cdot dt\right] d\mathbb{P} \\
&= \int_\Omega \left[\int_0^\infty \mathbb{1}_{[-\infty, X(w))}(t) \cdot dt\right] d\mathbb{P} \\
&= \int_0^\infty \left[\int_\Omega \mathbb{1}_{[-\infty, X(w))}(t) \cdot d\mathbb{P}\right] dt \\
&= \int_0^\infty \left[\int_\Omega \mathbb{1}_{\{t < X(w)\}}(w) \cdot d\mathbb{P}\right] dt \\
&= \int_0^\infty \mathbb{P}(X > t)\, dt
\end{aligned} \tag{8.34}$$

The interchangeability of two integrals wrt t and $\mathbb{P}$ follows (*Tonelli*), since X is non-negative.
To prove the second equal sign with $\mathbb{P}(X \geq t)$, we just replace upper bound of integration form of $X(w)$ with $X(w)^+$. And indicator will become $\mathbb{1}_{[-\infty, X(w)]}$. ■

Problem 10. $\{X_n\}$ identically distributed. $\mathbb{E}[|X_n|] < \infty$. Show that

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\max_{1\leq j\leq n} |X_j|\right] = 0$$

Proof. By result of (9),

$$\frac{1}{n}\mathbb{E}\left[\max_{1\leq j\leq n} |X_j|\right] = \int_0^\infty \frac{1}{n} \cdot \mathbb{P}\left(\max_{1\leq j\leq n} |X_j| > t\right) dt \tag{8.35}$$

Denote $f_n := n^{-1}\mathbb{P}\left(\max_{1\leq j\leq n} |X_j| > t\right)$, clearly $f_n \to 0$. It suffices to show

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\max_{1\leq j\leq n} |X_j|\right] = \lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n = 0 \tag{8.36}$$

For all $n \geq 1$, $\forall t \geq 0$, consider

$$\begin{aligned}
f_n := \frac{1}{n}\mathbb{P}\left(\max_{1\leq j\leq n} |X_j| > t\right) &\leq \frac{1}{n}\sum_{j=1}^n \mathbb{P}(|X_j| > t) \\
&= \frac{1}{n} \cdot \sum_{j=1}^n \mathbb{P}(|X_1| > t) = \mathbb{P}(|X_1| > t)
\end{aligned} \tag{8.37}$$

Which follows that $\{X_n\}$ are identically distributed. Take supremum wrt n,

$$\sup_n \frac{1}{n}\mathbb{P}\left(\max_{1\le j\le n}|X_j|>t\right)\le \mathbb{P}\left(|X_1|>t\right)=:g \tag{8.38}$$

By result of (9), $\mathbb{E}\left[X_1\right]<\infty \Rightarrow$ LHS$\in\mathcal{L}^1$. So f_n is bounded by $g\in\mathcal{L}^1$.
Apply (**DOM**), we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\max_{1\le j\le n}|X_j|\right] &= \lim_{n\to\infty}\int_0^\infty \frac{1}{n}\cdot\mathbb{P}\left(\max_{1\le j\le n}|X_j|>t\right)dt \\ &= \int_0^\infty \lim_{n\to\infty}\frac{1}{n}\mathbb{P}\left(\max_{1\le j\le n}|X_j|>t\right)dt = 0 \quad \blacksquare\end{aligned} \tag{8.39}$$

8.4 LLN

Problem 1. (WLLN3) Let $\{X_n : n\ge 1\}$ be a sequence of pairwise indep RV on $(\Omega,\mathcal{F},\mathbb{P})$, and S_n is partial sum. Let $\{b_n : n\ge 1\}$ be seq of positive real numbers such that $b_n\nearrow\infty$, suppose

$$\lim_{n\to\infty}\sum_{j=1}^n \mathbb{P}\left(|X_j|>b_n\right)=0 \tag{8.40}$$

$$\lim_{n\to\infty}\sum_{j=1}^n \mathbb{E}\left[\frac{|X_j|^2}{b_n^2};|X_j|\le b_n\right]=0 \tag{8.41}$$

If we set

$$a_n:=\sum_{j=1}^n \mathbb{E}\left[X_j;|X_j|\le b_n\right] \tag{8.42}$$

Then

$$\frac{S_n-a_n}{b_n}\xrightarrow{i.p}0 \tag{8.43}$$

1. For every $n\ge 1$ and $1\le j\le n$, truncate X_n at b_n, i.e. define

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

   Let $T_n:=\sum_{j=1}^n Y_{n,j}$. Show $\lim\limits_{n\to\infty}\mathbb{P}\left(S_n\ne T_n\right)=0$

2. Show $\mathrm{Var}\left[T_n\right]=o(b_n^2)$ as $n\to\infty$. Further show that

$$\frac{T_n-\mathbb{E}\left[T_n\right]}{b_n}\xrightarrow{i.p}0 \tag{8.44}$$

3. Show WLLN3 based on 1,2.

Proof. (1) Since S_n is partial sum of $\{X_j\}$, and T_n is partial sum of $\{Y_{n,j}\}$. So

$\{S_n \neq T_n\} \subseteq \{Y_{n,j} = X_j, \forall 1 \leq j \leq n\}^\complement = \{Y_{n,j} \neq X_j, \exists 1 \leq j \leq n\}$.

$$
\begin{aligned}
\mathbb{P}\left(S_n \neq T_n\right) &= \mathbb{P}\left(\{Y_{n,j} \neq X_j, \exists 1 \leq j \leq n\}\right) = \mathbb{P}\left(\bigcup_{j=1}^{n} \{Y_{n,j} \neq X_j\}\right) \\
&\leq \sum_{j=1}^{n} \mathbb{P}\left(Y_{n,j} \neq X_j\right) = \sum_{j=1}^{n} \mathbb{P}\left(|X_j| > b_n\right)
\end{aligned}
\tag{8.45}
$$

Take limit on both sides, notice that RHS is given by hypothesis (1):

$$
\lim_{n\to\infty} \mathbb{P}\left(S_n \neq T_n\right) \leq \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}\left(|X_j| > b_n\right) = 0 \quad \blacksquare \tag{8.46}
$$

Proof. (2) Since $\{X_n\}$ are pairwise indep, it is clear that for any fixed n, $\{Y_{n,j}\}$ are also pairwise indep. So $\mathrm{Var}\left[T_n\right] = \sum_{j=1}^{n} \mathrm{Var}\left[Y_{n,j}\right]$.

$$
\sum_{j=1}^{n} \mathrm{Var}\left[Y_{n,j}\right] \leq \sum_{j=1}^{n} \mathbb{E}\left[Y_{n,j}^2\right] = \sum_{j=1}^{n} \mathbb{E}\left[X_j^2; |X_j| \leq b_n\right] \tag{8.47}
$$

For any fixed n, b_n is constant with respect to summation and expectation.

$$
\begin{aligned}
\mathrm{Var}\left[T_n\right] &\leq \sum_{j=1}^{n} \mathbb{E}\left[X_j^2; |X_j| \leq b_n\right] = \sum_{j=1}^{n} \mathbb{E}\left[b_n^2 \cdot \frac{X_j^2}{b_n^2}; |X_j| \leq b_n\right] \\
&= b_n^2 \sum_{j=1}^{n} \mathbb{E}\left[\frac{X_j^2}{b_n^2}; |X_j| \leq b_n\right]
\end{aligned}
\tag{8.48}
$$

i.e.

$$
\frac{\mathrm{Var}\left[T_n\right]}{b_n^2} \leq \sum_{j=1}^{n} \mathbb{E}\left[\frac{X_j^2}{b_n^2}; |X_j| \leq b_n\right] \tag{8.49}
$$

Take limit on both sides, by the second hypothesis, we get exactly the definition of $\mathrm{Var}\left[T_n\right] = o(b_n^2)$.

$$
\lim_{n\to\infty} \frac{\mathrm{Var}\left[T_n\right]}{b_n^2} \leq \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{E}\left[\frac{X_j^2}{b_n^2}; |X_j| \leq b_n\right] = 0 \tag{8.50}
$$

Apply **Markov**'s ineq, for all $\epsilon > 0$:

$$
\begin{aligned}
\mathbb{P}\left(\frac{|T_n - \mathbb{E}\left[T_n\right]|}{b_n} > \epsilon\right) &= \mathbb{P}\left(\frac{|T_n - \mathbb{E}\left[T_n\right]|^2}{b_n^2} > \epsilon^2\right) \\
&\leq \frac{\mathbb{E}\left[|T_n - \mathbb{E}\left[T_n\right]|^2\right]}{b_n^2 \cdot \epsilon^2} \\
&= \frac{\mathrm{Var}\left[T_n\right]}{b_n^2} \cdot \frac{1}{\epsilon^2} \xrightarrow{n\to\infty} 0
\end{aligned}
\tag{8.51}
$$

i.e.

$$
\frac{T_n - \mathbb{E}\left[T_n\right]}{b_n} \xrightarrow{i.p} 0 \quad \blacksquare \tag{8.52}
$$

Proof. (3) Notice that, by its definition, $a_n = \mathbb{E}[T_n]$, so

$$\begin{aligned}\frac{|S_n - a_n|}{b_n} = \frac{|S_n - \mathbb{E}[T_n]|}{b_n} &\le \frac{|S_n - T_n|}{b_n} + \frac{|T_n - \mathbb{E}[T_n]|}{b_n} \\ &:= Q_1 + Q_2\end{aligned} \tag{8.53}$$

Since $S_n \ne T_n$ on $\mathbb{P}$-null set when $n \to \infty$, $Q_1 \xrightarrow{a.s.} 0$. And we have shown that $Q_2 \xrightarrow{i.p} 0$. So the their summation $\xrightarrow{i.p} 0$. ■

Problem 2. Let $\{X_n : n \ge 1\}$ be a sequence of i.i.d. RV with common distribution

$$\mathbb{P}(X_1 = k) = \mathbb{P}(X_1 = -k) = \frac{c}{k^2 \log k}, k = 3, 4, ... \tag{8.54}$$

where c is a constant and $c = \frac{1}{2}(\sum_{k\ge 3} \frac{1}{k^2 \log k})^{-1}$. Let S_n be partial sum.

1. Show $\frac{S_n}{n} \xrightarrow{i.p} 0$.
2. Show that $\mathbb{P}\left(\frac{|S_n|}{n} > \frac{1}{2}\ i.o.\right) = 1$. Therefore, this is an example for which WLLN holds but SLLN does not hold.
3. Show
$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{n} = \infty\right) = \mathbb{P}\left(\liminf_{n\to\infty} \frac{S_n}{n} = -\infty\right) = 1 \tag{8.55}$$
i.e. the amplitude of oscillation of $\frac{S_n}{n}$ is unbounded.

Proof. (1) Check for WLLN3, let $b_n := n$, firstly

$$\begin{aligned}\sum_{j=1}^{n} \mathbb{P}(|X_j| > n) = n\mathbb{P}(|X_1| > n) &= n \sum_{k\ge n+1} \frac{2c}{k^2 \log k} \\ &\le \frac{n}{\log n} \sum_{k\ge n+1} \frac{2c}{k^2} \le \frac{n}{\log n} \int_n^\infty \frac{2c}{x^2} dx \\ &= \frac{2cn}{\log n} \cdot (-\frac{1}{x})\Big|_n^\infty = \frac{2c}{\log n} \xrightarrow{n\to\infty} 0\end{aligned} \tag{8.56}$$

Secondly

$$\begin{aligned}\sum_{j=1}^{n} \mathbb{E}\left[\frac{X_j^2}{n^2}; |X_j| \le n\right] &= n\mathbb{E}\left[\frac{X_1^2}{n^2}; |X_1| \le n\right] \\ &= n \sum_{k=3}^{n} \frac{k^2}{n^2} \cdot \frac{2c}{k^2 \log k} = \frac{2c}{n} \cdot \sum_{k=3}^{n} \frac{1}{\log k}\end{aligned} \tag{8.57}$$

Now we estimate $\sum_{k=3}^{n} \frac{1}{\log k}$, consider

$$li(n) - li(3) = \int_3^n \frac{dx}{\log x} < \sum_{k=3}^{n} \frac{1}{\log k} < \int_4^{n+1} \frac{dx}{\log x} = li(n+1) - li(4) \tag{8.58}$$

Where $li(n) := \int_0^n dx/\log(x)$. Use the estimation[1] of $li(n)$, we have

$$\sum_{k=3}^{n} \frac{1}{\log k} \sim li(n) = O\left(\frac{n}{\log n}\right) \tag{8.59}$$

Therefore,

$$\frac{2c}{n} \cdot \sum_{k=3}^{n} \frac{1}{\log k} = O\left(\frac{1}{\log n}\right) \xrightarrow{n\to\infty} 0 \tag{8.60}$$

So the condtions for WLLN3 holds. Apply WLLN3, define

$$a_n := \sum_{j=1}^{n} \mathbb{E}\left[X_j; |X_j| \le n\right] = 0 \tag{8.61}$$

$$\frac{S_n - a_n}{b_n} = \frac{S_n}{n} \xrightarrow{i.p} 0 \ \blacksquare \tag{8.62}$$

Proof. (2) It is clear that

$$\mathbb{E}\left[|X_1|\right] = \sum_{k\ge 3} k \cdot \frac{2c}{k^2 \log k} = \sum_{k\ge 3} \frac{2c}{k \log k} = \infty \tag{8.63}$$

- Fix any $A > 0$, $\mathbb{E}\left[|\frac{X_1}{A}|\right] = \infty$.
- Follow the proof of second part of (**SLLN3**) on lecture, $\Rightarrow \sum_{j\ge 1} \mathbb{P}\left(|X_1| > jA\right) = \infty$. Since $\{X_n\}$ are i.i.d, $\Rightarrow \sum_{j\ge 1} \mathbb{P}\left(|X_j| > jA\right) = \infty$
- By (**BC2**), $\mathbb{P}\left(|X_n| > nA \ \ i.o.\right) = 1$, i.e.

$$\mathbb{P}\left(\frac{|S_n - S_{n-1}|}{n} > A \ \ i.o\right) = 1 \tag{8.64}$$

Since $\{\frac{|S_n - S_{n-1}|}{n} > A\} \subseteq \{\frac{|S_n|}{n} > \frac{A}{2}\} \cup \{\frac{|S_{n-1}|}{n-1} > \frac{A}{2}\} = \{\frac{|S_n|}{n} > \frac{A}{2}\}$. Take $A = 1$, we have

$$\mathbb{P}\left(\frac{|S_n|}{n} > \frac{1}{2} \ \ i.o\right) = 1 \ \blacksquare \tag{8.65}$$

Proof. (3) By (**SLLN3**), second part, $\mathbb{E}\left[|X_1|\right] = \infty \Rightarrow$

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{|S_n|}{n} = \infty\right) = 1 \tag{8.66}$$

Define $X_n' := -X_n$, $S_n' = \sum X_n'$. Since $\{X_n\}$ is **symmetrically** distributed about 0. X_n and X_n' are essentially identically distributed, so do S_n and S_n'. Therefore,

$$\begin{aligned}\left\{\frac{|S_n|}{n} > m \ i.o\right\} &= \left\{\frac{S_n}{n} > m \ i.o\right\} \cup \left\{\frac{S_n'}{n} > m \ i.o\right\} \\ &= \left\{\frac{S_n}{n} > m \ i.o\right\} \subseteq \left\{\limsup_{n\to\infty} \frac{S_n}{n} > m\right\}\end{aligned} \tag{8.67}$$

[1] From wikipedia.

By (2), LHS has probability 1 holds for $\forall m > 1$, take intersection with respect to m, we have

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{S_n}{n}=\infty\right)=1 \tag{8.68}$$

For the infimum side, note that S_n, S'_n are identically distributed,

$$\begin{aligned}\mathbb{P}\left(\liminf_{n\to\infty}\frac{S_n}{n}=-\infty\right)&=\mathbb{P}\left(\limsup_{n\to\infty}\frac{-S_n}{n}=\infty\right)\\&=\mathbb{P}\left(\limsup_{n\to\infty}\frac{S'_n}{n}=\infty\right)=\mathbb{P}\left(\limsup_{n\to\infty}\frac{S_n}{n}=\infty\right)=1 \quad\blacksquare\end{aligned} \tag{8.69}$$

Problem 3. (SLLN4) Let $\{X_n : n \geq 1\}$ be sequence of $\mathcal{L}^1$, indep RVs; S_n be partial sum. Let $\phi : \mathbb{R} \to \mathbb{R}$ be positive and continuous even function such that $\frac{\phi(x)}{|x|}$ is non-decreasing in x and $\frac{\phi(x)}{x^2}$ is non-increasing in x. Assume for some sequence $\{b_n : n \geq 1\}$ of positive real numbers with $b_n \nearrow \infty$,

$$\sum_{n\geq 1}\frac{\mathbb{E}\left[\phi(X_n)\right]}{\phi(b_n)}<\infty \tag{8.70}$$

Show that $\sum_{n\geq 1}\frac{X_n-\mathbb{E}[X_n]}{b_n}$ converges a.s., hence

$$\frac{S_n-\mathbb{E}\left[S_n\right]}{b_n}\xrightarrow{a.s.}0 \tag{8.71}$$

Proof. We start from ϕ.

- Since $\frac{\phi(x)}{|x|}$ is non-decreasing in x, for $|X_n| \geq b_n$, we have

$$\frac{\phi(b_n)}{b_n}\leq\frac{\phi(X_n)}{|X_n|} \tag{8.72}$$

Besides since ϕ is positive, everything above are all positve, thus we can rearrange it without changing sign, i.e.

$$\frac{|X_n|}{b_n}\leq\frac{\phi(X_n)}{\phi(b_n)} \tag{8.73}$$

Take expectation on bothsides, note that we have constrained ourselves by $|X_n| \geq b_n$,

$$\frac{\mathbb{E}\left[|X_n|;|X_n|\geq b_n\right]}{b_n}\leq\frac{\mathbb{E}\left[\phi(X_n);|X_n|\geq b_n\right]}{\phi(b_n)}\leq\frac{\mathbb{E}\left[\phi(X_n)\right]}{\phi(b_n)}\quad(\triangle) \tag{8.74}$$

- Since $\frac{\phi(x)}{x^2}$ is non-increasing in x, for $|X_n| \leq b_n$, we have

$$\frac{\phi(b_n)}{b_n^2}\leq\frac{\phi(|X_n|)}{|X_n|^2}=\frac{\phi(X_n)}{|X_n|^2}\quad\text{i.e.}\quad\frac{|X_n|^2}{b_n^2}\leq\frac{\phi(X_n)}{\phi(b_n)} \tag{8.75}$$

The equal sign from $\phi(|X_n|)$ to $\phi(X_n)$ follows that ϕ is a even function.
Take expectation on bothsides, note that we have constrained ourselves by $|X_n| \leq b_n$,

$$\frac{\mathbb{E}\left[|X_n|^2;|X_n|\leq b_n\right]}{b_n^2}\leq\frac{\mathbb{E}\left[\phi(X_n);|X_n|\leq b_n\right]}{\phi(b_n)}\leq\frac{\mathbb{E}\left[\phi(X_n)\right]}{\phi(b_n)}\quad(\dagger) \tag{8.76}$$

Now trancate X_n at the level of b_n. Define

$$Y_n = \begin{cases} X_n & \text{if } |X_n| \le b_n, \\ 0 & \text{otherwise.} \end{cases}$$

And define $T_n := \sum_1^n Y_n$. By same argument as before, X_n, Y_n are equivalent. Moreover $\{Y_n\}$ are also indep.
Consider sequence $\{\frac{Y_n}{b_n}\}$ (clearly also indep.),

$$\begin{aligned} \sum_{n\ge1} \mathrm{Var}\left[\frac{Y_n}{b_n}\right] &= \sum_{n\ge1} \frac{\mathrm{Var}\left[Y_n\right]}{b_n^2} \le \sum_{n\ge1} \frac{\mathbb{E}\left[Y_n^2\right]}{b_n^2} \\ &= \sum_{n\ge1} \frac{\mathbb{E}\left[X_n^2; |X_n| \le b_n\right]}{b_n^2} \le \sum_{n\ge1} \frac{\mathbb{E}\left[\phi(X_n)\right]}{\phi(b_n)} < \infty \end{aligned} \tag{8.77}$$

The last $\le$ is due to (†). Apply (**SLLN2-Prelude**) to $\frac{Y_n}{b_n}$ then apply (**Kronecker**) $\Rightarrow$

$$\frac{1}{b_n} \sum_{n\ge1} (Y_n - \mathbb{E}\left[Y_n\right]) \xrightarrow{a.s.} 0 \quad \text{i.e.} \quad \frac{T_n - \mathbb{E}\left[T_n\right]}{b_n} \xrightarrow{a.s.} 0 \ \ (\#) \tag{8.78}$$

Finally consider

$$\begin{aligned} \frac{|S_n - \mathbb{E}\left[S_n\right]|}{b_n} &\le \frac{|S_n - T_n|}{b_n} + \frac{|T_n - \mathbb{E}\left[T_n\right]|}{b_n} + \frac{|\mathbb{E}\left[T_n\right] - \mathbb{E}\left[S_n\right]|}{b_n} \\ &= Q_1 + Q_2 + Q_3 \end{aligned} \tag{8.79}$$

Since X_n, Y_n are equivalent, $b_n \nearrow \infty \Rightarrow Q_1 \xrightarrow{a.s.} 0$.
By (#), $Q_2 \xrightarrow{a.s.} 0$.
For Q_3,

$$Q_3 = \frac{1}{b_n} \sum_{n\ge1} \mathbb{E}\left[|X_n|; |X_n| \ge b_n\right] \tag{8.80}$$

By (△),

$$\sum_{n\ge1} \frac{\mathbb{E}\left[|X_n|; |X_n| \ge b_n\right]}{b_n} \le \sum_{n\ge1} \frac{\mathbb{E}\left[\phi(X_n)\right]}{\phi(b_n)} < \infty \tag{8.81}$$

Apply again (**Kronecker**), $Q_3 \xrightarrow{a.s.} 0$. Therefore,

$$\frac{|S_n - \mathbb{E}\left[S_n\right]|}{b_n} = Q_1 + Q_2 + Q_3 \xrightarrow{a.s.} 0 \quad \blacksquare \tag{8.82}$$

Problem 4. (Inverting Laplace Transform) Let f be bounded continuous function on $[0,\infty)$, Laplace transform of f is the function L on $(0,\infty)$ by

$$L(\lambda) := \int_0^\infty e^{-\lambda x} f(x)dx \tag{8.83}$$

Let $\{X_n\}$ be indep RVs with exponential dist of rate λ, S_n be partial sum. So $\mathbb{P}(X > x) = e^{-\lambda x}$, $\mathbb{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}[X] = \frac{1}{\lambda^2}$.

1. Show
$$(-1)^{n-1}\frac{\lambda^n L^{(n-1)}(\lambda)}{(n-1)!} = \mathbb{E}[f(S_n)] \tag{8.84}$$

2. f can be recovered from L by: for $y > 0$
$$f(y) = \lim_{n\to\infty}(-1)^{n-1}\frac{(\frac{n}{y})^n L^{(n-1)}(\frac{n}{y})}{(n-1)!} \tag{8.85}$$

Proof. (1) Denote pdf of X by ϕ_X, we claim that for $\{X_n\}$ i.i.d. exponential(λ), the pdf of partial sum evaluated at any $x > 0$ is

$$\phi_{S_n}(x) = \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!} \quad (\#) \tag{8.86}$$

We prove by induction. Basic case $n = 1$, $\phi_{S_1} = \phi_X = \lambda e^{-\lambda x}$. Assume (#) holds for n, then for $n+1$:

$$\begin{aligned}
\phi_{S_{n+1}}(x) = (\phi_X * \phi_{S_n})(x) &= \int_0^\infty \phi_X(x-y)\phi_{S_n}(y)dy \\
&= \int_0^\infty \lambda e^{-\lambda(x-y)}\lambda e^{-\lambda y}\frac{(\lambda y)^{n-1}}{(n-1)!}dy \\
&= \lambda e^{-\lambda x}\int_0^\infty \lambda^n \frac{y^{n-1}}{(n-1)!}dy \\
&= \lambda e^{-\lambda x}\frac{(\lambda x)^n}{n!}
\end{aligned} \tag{8.87}$$

Now look at LHS of equation to prove. Since $\partial_\lambda^{n-1}(e^{-\lambda x}f(x))$ exists and is continuous, we are allowed to take ∂_λ^{n-1} inside integral.

$$\begin{aligned}
(-1)^{n-1}\frac{\lambda^n L^{(n-1)}(\lambda)}{(n-1)!} &= (-1)^{n-1}\frac{\lambda^n}{(n-1)!}\int_0^\infty \partial_\lambda^{n-1}(e^{-\lambda x})f(x)dx \\
&= \int_0^\infty \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}f(x)dx = \int_0^\infty f(x)\phi_{S_n}(x)dx \\
&= \mathbb{E}[f(S_n)] = RHS \quad \blacksquare
\end{aligned} \tag{8.88}$$

Proof. (2) By (**WLLN2**), since $\{X_n\}$ i.i.d. exponential, $X_n \in \mathcal{L}^1$, we have

$$\frac{S_n}{n} \xrightarrow{i.p} \mathbb{E}[X_1] = \frac{1}{\lambda} \quad \text{i.e. } S_n \xrightarrow{i.p} \frac{n}{\lambda} =: y \tag{8.89}$$

Composition with continuous function f preserves convergence in probability, so $f(S_n) \xrightarrow{i.p} f(y)$.
Since f is bounded (by some $g \in \mathcal{L}^1[0,\infty)$?), by (**DOM**): $f(S_n) \xrightarrow{\mathcal{L}^1} f(y)$, i.e. for any fixed y such that $\lambda = \frac{n}{y}$,

$$f(y) = \mathbb{E}\left[f(y)\right] = \lim_{n\to\infty} \mathbb{E}\left[f(S_n)\right] = \lim_{n\to\infty} (-1)^{n-1} \frac{(\frac{n}{y})^n L^{(n-1)}(\frac{n}{y})}{(n-1)!} \quad \blacksquare \tag{8.90}$$

Problem 5. Let $\{X_n : n \geq 1\}$ be sequence of i.i.d RV with common distribution

$$\mathbb{P}\left(X_1 = k\right) = p_k \text{ where } p_k \in (0,1), 1 \leq k \leq L, \text{ and } \sum_{k=1}^{L} p_k = 1 \tag{8.91}$$

For every $n \geq 1$ and $1 \leq k \leq L$, let S_n be partial sum and $N_k^{(n)} := \sharp\{j : 1 \leq j \leq n, X_j = k\}$. (i.e. the number of X_j among the first n terms of sequence which take value k). Show that, if

$$P(n) := \prod_{k=1}^{L} p_k^{N_k^{(n)}} \tag{8.92}$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \cdot \log(P(n)) \quad \text{exists a.s. (find it.)} \tag{8.93}$$

Proof. Define

$$Y_{k,j} = \begin{cases} 1 & \text{if } X_j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly for any fixed $1 \leq k \leq L$, $\{Y_{k,j} : j \geq 1\}$ is a sequence of i.i.d RVs due to the fact that $\{X_j\}$ are i.i.d. And $N_k^{(n)} = \sum_{j=1}^{n} Y_{k,j}$ is a partial sum of $Y_{k,j}$.
Fix k, for all $j \geq 1$,

$$\mathbb{E}\left[Y_{k,j}\right] = \mathbb{E}\left[Y_{k,1}\right] = 1 \cdot \mathbb{P}\left(X_1 = k\right) = p_k < \infty \tag{8.94}$$

So by (**SLLN3**),

$$\frac{N_k^{(n)}}{n} \xrightarrow{a.s.} \mathbb{E}\left[Y_{k,1}\right] = p_k \tag{8.95}$$

Therefore,

$$\begin{aligned} \frac{1}{n} \cdot \log(P(n)) &= \frac{1}{n} \sum_{k=1}^{L} N_k^{(n)} \log p_k = \sum_{k=1}^{L} \frac{N_k^{(n)}}{n} \log p_k \\ &\xrightarrow{a.s.} \sum_{k=1}^{L} p_k \log p_k \end{aligned} \tag{8.96}$$

i.e. $\lim\limits_{n\to\infty} \frac{1}{n} \cdot \log(P(n))$ exists almost surely. It equals to $\sum_{k=1}^{L} p_k \log p_k$ with 1 probability.

Problem 6. Let $\{X_n : n \geq 1\}$ be sequence of i.i.d RVs with $\mathbb{E}[|X_1|] < \infty$, and S_n be partial sum. Show that if $\mathbb{E}[X_1] \neq 0$,

$$\frac{\max\limits_{1\leq k\leq n} |X_k|}{|S_n|} \xrightarrow{a.s.} 0 \tag{8.97}$$

Proof. For all $\epsilon > 0$,

$$\begin{aligned}
\infty > \mathbb{E}[|X_1|] &= \int_0^\infty \mathbb{P}(|X_1| > t)\, dt \\
&= \left(\int_0^\epsilon + \int_\epsilon^{2\epsilon} + \int_{2\epsilon}^{3\epsilon} + ...\right) \mathbb{P}(|X_1| > t)\, dt \\
&= \sum_{n\geq 1} \int_{(n-1)\epsilon}^{n\epsilon} \mathbb{P}(|X_1| > t)\, dt \\
&\geq \sum_{n\geq 1} \epsilon \cdot \mathbb{P}(|X_1| > n\epsilon) \\
&= \epsilon \cdot \sum_{n\geq 1} \mathbb{P}\left(\frac{|X_1|}{n} > \epsilon\right)
\end{aligned} \tag{8.98}$$

Therefore $\sum_{n\geq 1} \mathbb{P}\left(\frac{|X_1|}{n} > \epsilon\right) < \infty$.
By (**BC1**), $\mathbb{P}\left(\frac{|X_1|}{n} > \epsilon \;\; i.o.\right) = 0$ for all $\epsilon > 0 \Rightarrow \frac{|X_1|}{n} \xrightarrow{a.s.} 0$.
Now consider

$$\frac{\max\limits_{1\leq k\leq n} |X_k|}{|S_n|} = \frac{\max\limits_{1\leq k\leq n} |X_k|}{n} \cdot \frac{n}{|S_n|} \tag{8.99}$$

For the second factor, apply (**SLLN3**, since mutually indep, $\mathbb{E}[|X_1|] < \infty$),

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}[X_1] \neq 0 \tag{8.100}$$

So,

$$\frac{n}{|S_n|} \xrightarrow{a.s.} \left|\frac{1}{\mathbb{E}[X_1]}\right| < \infty \quad (1) \tag{8.101}$$

For the first factor, we already have $\mathbb{P}\left(\frac{|X_1|}{n} > \epsilon \;\; i.o.\right) = 0$. For any ϵ,

$$\begin{aligned}
\mathbb{P}\left(\frac{\max\limits_{1\leq k\leq n} |X_k|}{n} > \epsilon \;\; i.o.\right) &= \mathbb{P}\left(\bigcup_{k=1}^n \left\{\frac{|X_k|}{n} > \epsilon \;\; i.o.\right\}\right) \\
&= \sum_{k=1}^n \mathbb{P}\left(\frac{|X_k|}{n} > \epsilon \;\; i.o.\right) \\
&= n \cdot \mathbb{P}\left(\frac{|X_1|}{n} > \epsilon \;\; i.o.\right) = 0
\end{aligned} \tag{8.102}$$

Therefore $\frac{\max\limits_{1\leq k\leq n} |X_k|}{n} \xrightarrow{a.s.} 0$ (2). By (1) and (2),

$$\frac{\max\limits_{1\leq k\leq n} |X_k|}{|S_n|} \xrightarrow{a.s.} 0 \cdot \left|\frac{1}{\mathbb{E}[X_1]}\right| = 0 \;\; \blacksquare \tag{8.103}$$

Problem 7. Let $\{X_n\}$ be i.i.d RVs, $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = 1$. S_n is partial sum. Show that for every $c \in \mathbb{R}$ and $n \geq 1$,

$$\mathbb{P}\left(\max_{1\leq j\leq n} S_j \geq c\right) \leq 2\mathbb{P}\left(S_n \geq c - \sqrt{2n}\right) \tag{8.104}$$

Proof. For any $c \in \mathbb{R}$,

$$\begin{aligned}
\frac{1}{2}RHS &= \mathbb{P}\left(S_n \geq c - \sqrt{2n}\right) \\
&\geq \mathbb{P}\left(S_n \geq c - \sqrt{2n} \text{ and } \max_{1\leq j\leq n} S_j \geq c\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(S_n \geq c - \sqrt{2n} \text{ and } S_j < c, \forall j = 1, 2, ..., k-1 \text{ and } S_k \geq c\right) \\
&\geq \sum_{k=1}^{n} \mathbb{P}\left(S_k - S_n \leq \sqrt{2(n-k)} \text{ and } S_j < c, \forall j = 1, 2, ..., k-1 \text{ and } S_k \geq c\right) \quad (\dagger)
\end{aligned} \tag{8.105}$$

The last geq sign holds, because given $\{S_k - S_n \leq \sqrt{2(n-k)}$ and $S_k \geq c\}$, we have $\sqrt{2(n-k)} \geq S_k - S_n \geq c - S_n$.
$\Rightarrow S_n \geq c - \sqrt{2(n-k)} \geq c - \sqrt{2n}$, i.e. this event implies the original one:

$$\{S_k - S_n \leq \sqrt{2(n-k)} \text{ and } S_k \geq c\} \subseteq \{S_n \geq c - \sqrt{2n} \text{ and } S_k \geq c\} \tag{8.106}$$

Since $\{S_j < c, \forall j = 1, 2, ..., k-1 \text{ and } S_k \geq c\} \in \sigma(X_1, X_2, ..., X_k)$
And $\{S_k - S_n \leq \sqrt{2(n-k)}\} \in \sigma(X_{k+1}, X_{k+2}, ..., X_n)$, these two events are independent, so,

$$\begin{aligned}
(\dagger) &= \sum_{k=1}^{n} \mathbb{P}\left(S_k - S_n \leq \sqrt{2(n-k)}\right) \cdot \mathbb{P}\left(S_j < c, \forall j = 1, 2, ..., k-1 \text{ and } S_k \geq c\right) \\
&\geq \sum_{k=1}^{n} \mathbb{P}\left(|S_k - S_n| \leq \sqrt{2(n-k)}\right) \cdot \mathbb{P}\left(S_j < c, \forall j = 1, 2, ..., k-1 \text{ and } S_k \geq c\right)
\end{aligned} \tag{8.107}$$

By (**Markov**), note that $\{X_n\}$ are indep, $\mathbb{E}\left[X_j^2\right] = 1$, $\mathbb{E}[X_j] = 0$,

$$\begin{aligned}
\mathbb{P}\left(|S_k - S_n| > \sqrt{2(n-k)}\right) &< \frac{\mathbb{E}\left[(\sum_{j=k}^{n} X_j)^2\right]}{2(n-k)} \\
&= \frac{1}{2(n-k)} \left[\sum_{j=k}^{n} \mathbb{E}\left[X_j^2\right] + \sum_{k\leq i\neq j\leq n} \mathbb{E}[X_i X_j]\right] \\
&= \frac{1}{2(n-k)} \cdot [(n-k) + 0] = \frac{1}{2}
\end{aligned} \tag{8.108}$$

Therefore,

$$\mathbb{P}\left(|S_k - S_n| \leq \sqrt{2(n-k)}\right) \geq 1 - \frac{1}{2} = \frac{1}{2} \tag{8.109}$$

$$\begin{aligned}(\dagger) &\geq \sum_{k=1}^{n} \frac{1}{2} \cdot \mathbb{P}\left(S_j < c, \forall j = 1,2,...,k-1 \text{ and } S_k \geq c\right) \\ &= \frac{1}{2}\mathbb{P}\left(\max_{1\leq j\leq n} S_j \geq c\right) = \frac{1}{2}LHS\end{aligned} \tag{8.110}$$

So we have $LHS \leq RHS$. ■

Problem 8.

1. Let X be non-negative RV on $(\Omega, \mathcal{F}, \mathbb{P})$, $p \in (1, \infty)$, show

$$\mathbb{E}\left[X^p\right] = p\int_0^\infty t^{p-1}\mathbb{P}\left(X > t\right) dt = p\int_0^\infty t^{p-1}\mathbb{P}\left(X \geq t\right) dt \tag{8.111}$$

2. Let $\{X_n : n \geq 1\}$ is sequence of square-integrable indep random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}\left[X_n\right] = 0$ for all $n \geq 1$. Set $S_n := \sum_{j=1}^{n} X_j$ for each $n \geq 1$, assume that $\sum_{n\geq 1}\mathbb{E}\left[X_n^2\right] < \infty$. It is known from theorem that $S_n \xrightarrow{a.s.} S$ for some RV S. Moreover, due to the completeness of $\mathcal{L}^2$, we know that $S \in \mathcal{L}^2$ and $S_n \to S$ also in $\mathcal{L}^2$. Show that for every $t \geq 0$,

$$\mathbb{P}\left(\sup_{n\geq 1}|S_n|^2 > t\right) \leq \frac{1}{t}\mathbb{E}\left[S^2; \sup_{n\geq 1}|S_n|^2 > t\right] \tag{8.112}$$

3. With (1) and (2), show

$$\left(\mathbb{E}\left[\sup_{n\geq 1}|S_n|^{2p} > t\right]\right)^{\frac{1}{p}} \leq \frac{p}{p-1}\left(\mathbb{E}\left[|S|^{2p}\right]\right)^{\frac{1}{p}} \tag{8.113}$$

4. Based on (3), conclude that if $S \in L^q$ for some $q \in (2, \infty)$, then $S_n \to S$ also in $\mathcal{L}^q$.

Proof. (1), $X \in (m\mathcal{F})^+$, use result of HW3-9 (**Tonelli**), $\forall M > 0$,

$$\begin{aligned}\mathbb{E}\left[X^p; X < M\right] &= \int_{\{X<M\}} \left[\int_0^{X^p(w)^-} 1 \cdot dt\right] d\mathbb{P} \\ &= \int_\Omega \mathbb{1}_{\{X(w)<M\}}(w) \left[\int_0^\infty \mathbb{1}_{[-\infty, X^p(w))}(t) \cdot dt\right] d\mathbb{P} \\ &= \int_\Omega \left[\int_0^\infty \mathbb{1}_{\{X(w)<M\}}(w) \cdot \mathbb{1}_{[-\infty, X^p(w))}(t) \cdot dt\right] d\mathbb{P} \\ &= \int_0^\infty \left[\int_\Omega \mathbb{1}_{\{t^{\frac{1}{p}}<X(w)<M\}}(w) \cdot d\mathbb{P}\right] dt \\ &= \int_0^\infty \mathbb{P}\left(M > X > t^{\frac{1}{p}}\right) dt \\ &= \int_0^{M^p} \left[\mathbb{P}\left(X > t^{\frac{1}{p}}\right) - \mathbb{P}\left(X > M\right)\right] dt\end{aligned} \tag{8.114}$$

$\mathbb{P}\left(X > t^{\frac{1}{p}}\right)$ is montonic function w.r.t t, thus integrable on finite interval $[0, M^p]$.

$\mathbb{P}(X > M)$ is constant. Define

$$f_M(t) := \mathbb{P}\left(M > X > t^{\frac{1}{p}}\right) \nearrow \mathbb{P}\left(X > t^{\frac{1}{p}}\right) =: f(t) \tag{8.115}$$

By our argument above, $\mu(f_M(t)) = \mathbb{E}[X^p; X < M] < \infty$. By (**MON**), $\mu(f_M(t)) \to \mu(f)$, i.e.

$$\begin{aligned}
\mathbb{E}[X^p] &= \lim_{M\to\infty} \int_0^\infty \mathbb{P}\left(M > X > t^{\frac{1}{p}}\right) dt \\
&= \int_0^\infty \lim_{M\to\infty} \mathbb{P}\left(M > X > t^{\frac{1}{p}}\right) dt \\
&= \int_0^\infty \mathbb{P}\left(X > t^{\frac{1}{p}}\right) dt \qquad (\text{let } z := t^{\frac{1}{p}}) \\
&= \int_0^\infty p z^{p-1} \mathbb{P}(X > z)\, dz
\end{aligned} \tag{8.116}$$

The second equal sign is the same, just replace upper bound of integral form of $X^p(w)$ with $X^p(w)^+$, and all relevant indicators will become $\mathbb{1}_{[-\infty, X^p(w)]}$. ■

Proof. (2) The structure is similar to Kolmogorov 's inequality. Define

$$A_j := \{|S_i|^2 \leq t, \forall i = 1, 2, ..., j-1 \text{ and } |S_j|^2 > t\} \tag{8.117}$$

$$A := \{\max_{1\leq j\leq n} |S_j|^2 > t\} = \bigcup_{j=1}^n A_j \tag{8.118}$$

Note that A_j's are disjoint, then consider

$$\begin{aligned}
\mathbb{E}\left[S_n^2; A\right] &= \sum_{j=1}^n \mathbb{E}\left[S_n^2; A_j\right] = \sum_{j=1}^n \mathbb{E}\left[(S_j + (S_n - S_j))^2; A_j\right] \\
&= \sum_{j=1}^n \mathbb{E}\left[S_j^2; A_j\right] + \sum_{j=1}^n \mathbb{E}\left[(S_n - S_j)^2; A_j\right] + 2\sum_{j=1}^n \mathbb{E}\left[(S_n - S_j)S_j; A_j\right] \quad (\triangle)
\end{aligned} \tag{8.119}$$

By same argument as the proof of Kolmogorov's ineq, RV $S_j \in m\sigma(X_1, X_2, ..., X_j)$; $(S_n - S_j) \in m\sigma(X_{j+1}, ..., X_n)$, thus independent. Therefore the cross term is $2\sum_{j=1}^n \mathbb{E}[(S_n - S_j); A_j]\,\mathbb{E}[S_j; A_j] = 0$, by $\mathbb{E}[X_n] = 0$, so

$$\begin{aligned}
(\triangle) &= \sum_{j=1}^n \mathbb{E}\left[S_j^2; A_j\right] + \sum_{j=1}^n \mathbb{E}\left[(S_n - S_j)^2; A_j\right] \\
&\geq \sum_{j=1}^n \mathbb{E}\left[S_j^2; A_j\right] > t \sum_{j=1}^n \mathbb{P}(A_j) = t \cdot \mathbb{P}(A)
\end{aligned} \tag{8.120}$$

i.e.

$$\mathbb{P}\left(\max_{1\leq j\leq n} |S_j|^2 > t\right) \leq \frac{1}{t}\mathbb{E}\left[S_n^2; \max_{1\leq j\leq n} |S_j|^2 > t\right] \tag{8.121}$$

By theorem, $S_n \xrightarrow{a.s.} S$. Since X_n are **non-negative**, so $S_n^2 \nearrow S^2$. Take limit on both sides and apply (**MON**) on RHS,

$$\begin{aligned}\mathbb{P}\left(\sup_{n\geq 1}|S_n|^2 > t\right) &\leq \lim_{n\to\infty}\frac{1}{t}\mathbb{E}\left[S_n^2; \max_{1\leq j\leq n}|S_j|^2 > t\right] \\ &= \frac{1}{t}\mathbb{E}\left[S^2; \sup_{n\geq 1}|S_n|^2 > t\right] \quad \blacksquare\end{aligned} \tag{8.122}$$

Proof. (3) Since $|S_n| \geq 0$, $\sup_{n\geq 1}|S_n|^{2p} = (\sup_{n\geq 1}|S_n|^2)^p$. For non-negative RV $\sup_{n\geq 1}|S_n|^2$, apply (1), then apply (2),

$$\begin{aligned}\mathbb{E}\left[(\sup_{n\geq 1}|S_n|^2)^p\right] &= p\int_0^\infty t^{p-1}\mathbb{P}\left(\sup_{n\geq 1}|S_n|^2 > t\right)dt \\ &\leq p\int_0^\infty t^{p-1}\frac{1}{t}\mathbb{E}\left[S^2; \sup_{n\geq 1}|S_n|^2 > t\right]dt \\ &= p\mathbb{E}\left[S^2\int_0^{\sup_{n\geq 1}|S_n|^2} t^{p-2}dt\right] \\ &= \frac{p}{p-1}\mathbb{E}\left[(\sup_{n\geq 1}|S_n|^2)^{p-1}S^2\right] \quad (\triangle)\end{aligned} \tag{8.123}$$

Apply (**Holders**) to $(\triangle)$, since $\frac{1}{p} + \frac{p-1}{p} = 1$,

$$\begin{aligned}\mathbb{E}\left[\sup_{n\geq 1}|S_n|^{2p}\right] \leq (\triangle) &\leq \frac{p}{p-1}\mathbb{E}\left[((\sup_{n\geq 1}|S_n|^2)^{p-1})^{\frac{p}{p-1}}\right]^{\frac{p-1}{p}}\mathbb{E}\left[S^{2p}\right]^{\frac{1}{p}} \\ &= \frac{p}{p-1}\mathbb{E}\left[\sup_{n\geq 1}|S_n|^{2p}\right]^{\frac{p-1}{p}}\mathbb{E}\left[S^{2p}\right]^{\frac{1}{p}}\end{aligned} \tag{8.124}$$

If $\mathbb{E}\left[\sup_{n\geq 1}|S_n|^{2p}\right] < \infty$, we can divide it from both sides, which yields

$$\mathbb{E}\left[\sup_{n\geq 1}|S_n|^{2p}\right]^{\frac{1}{p}} \leq \frac{p}{p-1}\mathbb{E}\left[S^{2p}\right]^{\frac{1}{p}} \quad \blacksquare \tag{8.125}$$

Proof. (4) This is a direct result from (3). Suppose $S \in \mathcal{L}^q$ for some $q \in (2, \infty)$, let $p := \frac{q}{2} \in (1, \infty)$.

$$\mathbb{E}\left[\sup_{n\geq 1}|S_n|^q\right]^{\frac{2}{q}} \leq \frac{q}{q-2}\mathbb{E}\left[S^q\right]^{\frac{2}{q}} < \infty \tag{8.126}$$

So $\sup_{n\geq 1}|S_n| \in \mathcal{L}^q$, S_n is bounded, thus in $\mathcal{L}^q$ for all $n \geq 1$. $\blacksquare$

8.5 Martingale

Problem 1. Let $\{\mu_n : n \geq 1\}$ and $\{\nu_n : n \geq 1\}$ be two sequences of probability measures on some measurable space (S, Σ). Assume that for each $n \geq 1$, μ_n is absolutely continuous with respect to ν_n and denote the Radon-Nikodym derivative

$$Y_n := \frac{d\mu_n}{d\nu_n} \tag{8.127}$$

Set $\Omega := S \times S \times ...$, let $\mathcal{F}$ be sigma algebra generated by cylinder sets, i.e

$$\mathcal{F} := \sigma\left(\left\{\prod_{n\geq 1} F_n : F_n \subseteq S, F_n = S \text{ for all but finitely many n}\right\}\right) \tag{8.128}$$

Let $\mathbb{P}$ be the prob measure on $(\Omega, \mathcal{F})$ given by $\mathbb{P} = \bigotimes_{n\geq 1} \mu_n$, $\mathbb{Q}$ is product measure corresponding to ν, $\mathbb{Q} = \bigotimes_{n\geq 1} \nu_n$.

1. Define $\mathbb{P}_n := \bigotimes_{j=1}^n \mu_j$, $\mathbb{Q}_n := \bigotimes_{j=1}^n \nu_j$, show $\mathbb{P}_n$ is absolutely continous wrt $\mathbb{Q}_n$, (i.e. $\mathbb{Q}_n(A) = 0 \Rightarrow \mathbb{P}_n(A) = 0$). Further show that if define $X_n(w) := \prod_1^n Y_j(w_j)$, then $X_n = \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}$ is R-N derivative of $\mathbb{P}_n$ wrt $\mathbb{Q}_n$.
2. Let $X_0 = 1$. Show $\{X_n : n \geq 0\}$ is a martingale wrt natural filtration associated with $\{X_n : n \geq 0\}$; and $\lim_{n\to\infty} X_n$ exists $\mathbb{Q} - a.s.$
3. Show $\mathbb{P}(X > 0)$ is either 0 or 1.
4. Show either $\mathbb{P}, \mathbb{Q}$ are continuous wrt to each other, or they are entirely singular wrt each other.

Proof. (1) We first show that $\mu \ll \nu$, $\mu' \ll \nu' \Rightarrow \mu \times \mu' \ll \nu \times \nu'$. For any two pairs of measures.

For $A \in S \times S'$, given $(\nu \times \nu')(A) = 0$, we want to show $(\mu \times \mu')(A) = 0$. For $w \in S, w' \in S'$, define

$$\begin{aligned} I^{\mathbb{1}_A}(\bar{w}) &:= \int_{S'} \mathbb{1}_A(\bar{w}, w')\mu'(dx) = \mu'(\{w' \in S' : (\bar{w}, w') \in A\}) := \mu'(A'(\bar{w})) \\ J^{\mathbb{1}_A}(\bar{w}) &:= \int_{S'} \mathbb{1}_A(\bar{w}, w')\nu'(dx) = \nu'(\{w' \in S' : (\bar{w}, w') \in A\}) := \nu'(A'(\bar{w})) \end{aligned} \tag{8.129}$$

Since all μ, ν's are probability measures (finite), then by (**Fubini**),

$$\begin{aligned} (\nu \times \nu')(A) &:= \int_S J^{\mathbb{1}_A}(w)\nu(dw) = 0 \quad (\triangle) \\ (\mu \times \mu')(A) &:= \int_S I^{\mathbb{1}_A}(w)\mu(dw) \end{aligned} \tag{8.130}$$

By $(\triangle)$, $J^{\mathbb{1}_A}(w) = 0$ $a.s.$ w, i.e. $\nu'(A'(\bar{w})) = 0$ $a.s.$ $\bar{w}$. Define $O_\nu := \{w \in S : \nu'(A'(w)) = 0\}$, then $\nu(O_\nu) = 1$.

By $\mu' \ll \nu'$, $O_\nu \subseteq O_\mu := \{w \in S : \mu'(A'(w)) = 0\}$, hence $\mu(O_\mu) = 1$;

$\mu(S \setminus O_\mu) = 0$; in another word $I^{\mathbb{1}_A}(w) = 0$ $a.s.$ w. Therefore

$$\begin{aligned}(\mu \times \mu')(A) &:= \int_S I^{\mathbb{1}_A}(w)\mu(dw) \\ &= \left(\int_{O_\mu} + \int_{S \setminus O_\mu}\right) I^{\mathbb{1}_A}(w)\mu(dw) \\ &\leq 0 + 1 \cdot \mu(S \setminus O_\mu) = 0\end{aligned} \tag{8.131}$$

Now take $\mu, \mu' = \mu_1, \mu_2 \Rightarrow \mu_1 \times \mu_2 \ll \nu_1 \times \nu_2$.
Then take $\mu := \mu_1 \times \mu_2, \mu' = \mu_3 \Rightarrow \mu_1 \times \mu_2 \times \mu_3 \ll \nu_1 \times \nu_2 \times \nu_3$. Do this recursively, finally we conclude that for finite n,

$$\mathbb{P}_n := \bigotimes_{j=1}^{n} \mu_j \ll \bigotimes_{j=1}^{n} \nu_j =: \mathbb{Q}_n \quad \blacksquare \tag{8.132}$$

By definition, for $A_n \in S$, $\mu_n(A_n) = \mathbb{E}^{\nu_n}[Y_n \mathbb{1}_{A_n}]$. Define $X_n(w) := \prod_{j=1}^n Y_j(w_j)$ for $w \in S^n =: \Omega$, consider measurable $A := A_1 \times ... \times A_n \subseteq \Omega$,

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}_n}[X_n(w)\mathbb{1}_A(w)] &= \mathbb{E}^{\mathbb{Q}_n}\left[\prod_{j=1}^n Y_j(w_j) \prod_{j=1}^n \mathbb{1}_{A_j}(w_j)\right] \\ &= \int \cdots \int_{S^n} \left(\prod_{j=1}^n Y_j(w_j)\mathbb{1}_{A_j}(w_j)\right) d\mathbb{Q}_n \\ &= \int \cdots \int_{S^{n-1}} \left(\prod_{j=1}^{n-1} Y_j(w_j)\mathbb{1}_{A_j}(w_j) \left(\int_{A_n} Y_n(w_n) d\nu_n\right) d\mathbb{Q}_{n-1}\right) \\ &= \int \cdots \int_{S^{n-1}} \left(\prod_{j=1}^{n-1} Y_j(w_j)\mathbb{1}_{A_j}(w_j) \left(\int_{S} \mathbb{1}_{A_n}(w_n) d\mu_n\right) d\mathbb{Q}_{n-1}\right) \\ &= ... = \int \cdots \int_{S^n} \left(\prod_{j=1}^n \mathbb{1}_{A_j}(w_j)\right) d\bigotimes_{j\geq 1}^{n} \mu_j \\ &= \int \cdots \int_{S^n} \mathbb{1}_A(w_1, ..., w_n)\; d\bigotimes_{j\geq 1}^{n} \mu_j \\ &= \int_\Omega \mathbb{1}_A(w) d\mathbb{P}_n = \mathbb{P}_n(A)\end{aligned} \tag{8.133}$$

So by definition of R-N derivative, at every $w \in \Omega$, $\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(w) := X_n(w)$. $\blacksquare$

(2) In filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n := \sigma(X_0, X_1, ..., X_n) : n \geq 0\}, \mathbb{Q})$, clearly Y_{n+1} is independent wrt $\mathcal{F}_n$ for all $n \geq 0$ and $X_n \in m\mathcal{F}_n$. Now consider

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}^{\mathbb{Q}}\left[Y_{n+1}\prod_{k=1}^n Y_k \middle| \mathcal{F}_n\right] \\ &= X_n \mathbb{E}^{\mathbb{Q}}[Y_{n+1} \cdot 1] = X_n \mathbb{E}^{\mathbb{Q}}[Y_{n+1} \cdot \mathbb{1}_S] \\ &= X_n \mathbb{P}(w_{n+1} \in S) \\ &= X_n \mu_{n+1}(S) = X_n\end{aligned} \tag{8.134}$$

Since μ_n is a probability measure, hence positive. For any $w \in S$, $0 \leq \mu_n(\{w\}) = \mathbb{E}^{\nu_n}[Y_n; \{w\}] = Y_n(w)$. So $Y_n \geq 0$ everywhere for all $n \geq 1$. So $X_n = \prod_{k=1}^n Y_k \geq 0$ everywhere too.
There are two cases.

- First, if $\exists Y_m = 0$ $a.s$ for some m, then clearly $X_n = 0$ $\mathbb{Q}$-as for all $n \geq m$. We can just define $X_n \xrightarrow{a.s.} X := 0$.
- Second, if the first case does not happen, then $\{X_n : n \geq 0\}$ is $\mathbb{Q}$-martingale. Hence for any $n \geq 1$, $\mathbb{E}^{\mathbb{Q}}[X_n] = \mathbb{E}^{\mathbb{Q}}[X_0] = 1 < \infty$, so $\{X_n\}$ is uniformly integrable. By (**MCT2**) $\Rightarrow \exists X \in \mathcal{L}^1$, such that $X_n \xrightarrow{a.s.} X$, $X_n \xrightarrow{\mathcal{L}^1} X$.

In both cases, X exists $\mathbb{Q}$-a.s. ■.
(3) In $(\Omega, \mathcal{F}, \{\mathcal{F}_n := \sigma(X_0, X_1, ..., X_n) : n \geq 0\}, \mathbb{P})$, by definition,

$$X := \lim_{n\to\infty} X_n = \prod_{n\geq 1} Y_n \tag{8.135}$$

Note that $Y_n = \frac{X_n}{X_{n-1}}$, for all $n \geq 1$ if everything is positive. Hence for any $m \geq 1$, X can be regarded as

$$X = X_m \prod_{n\geq m+1} Y_n = \begin{cases} X_m \prod_{n\geq m+1} \frac{X_n}{X_{n-1}} & \text{If } X_n > 0\ \forall n \geq m \\ 0 & \text{If } X_n = 0\ \exists n \geq m \end{cases} \in m\sigma(X_m, X_{m+1}, ...) \tag{8.136}$$

So $\forall m \geq 1$:

$$\{X > 0\} \in \sigma(X_m, X_{m+1}, ...) \tag{8.137}$$

i.e.

$$\{X > 0\} \in \bigcap_{m\geq 1} \sigma(X_m, X_{m+1}, ...) =: \mathcal{T}_{X_n} \tag{8.138}$$

$\{X > 0\}$ is an event that is a member in the tail sigma algebra associated with $\{X_n\}$. By (**Kolmogorov 0-1 Law**), $\mathbb{P}(X > 0) = 0$ or 1. ■

(4) **Part-1**. Define

$$\mathcal{I} = \left\{ F, F \in \bigcup_{n\geq 1} \mathcal{F}_n \right\} \tag{8.139}$$

Then $\mathcal{I}$ is a pi system. Because $\forall F_i \in \mathcal{F}_i, F_j \in \mathcal{F}_j$, we have $(F_i \cap F_j) \in \mathcal{F}_{i\vee j} \subseteq \mathcal{I}$. Moreover, $\mathcal{F}$ is generated by $\mathcal{I}$, which is clear since $\sigma(\mathcal{I}) = \bigvee_{n\geq 1} \mathcal{F}_n =: \mathcal{F}$.
Now assume $\mathbb{Q}(X > 0) = 1$, i.e. case two in (2), where $\{X_n : n \geq 0\}$ is a $\mathbb{Q}$-martingale.
For any $n \geq 1$, any $A \in \mathcal{F}_n$,

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{E}^{\mathbb{Q}}[X_n; A] = \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[X_{n+1}|\mathcal{F}_n]; A\right] \\ &= \mathbb{E}^{\mathbb{Q}}[X_{n+1}; A] = \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[X_{n+2}|\mathcal{F}_{n+1}]; A\right] = ... \\ &= \mathbb{E}^{\mathbb{Q}}[X; A] \end{aligned} \tag{8.140}$$

I.e. two measures $X\mathbb{Q} = \mathbb{P}$ on all $A \in \mathcal{F}_n$. This is true for all $n \geq 1$. Hence $X\mathbb{Q} = \mathbb{P}$ for $A \in \mathcal{I}$. By extension theorem, finally we know $X\mathbb{Q} = \mathbb{P}$ on

$F \in \mathcal{F} = \sigma(\mathcal{I})$.
So, by definition, $\frac{d\mathbb{P}}{d\mathbb{Q}} := X$, so $\mathbb{P} \ll \mathbb{Q}$.

Part-2. Since $\nu_n \ll \mu_n$ is also assumed to be true, we have $d\mu_n/d\nu_n = 1/Y_n$. We can also show $\mathbb{Q}_n \ll \mathbb{P}_n$ by exactly same arguement as in (1). Similarly, we construct the reverse R-N derivative

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} := \frac{1}{X_n} \tag{8.141}$$

as we done in (1). Now consider

$$\begin{aligned} \mathbb{E}^{\mathbb{P}}\left[\frac{1}{X_{n+1}}\middle|\mathcal{F}_n\right] &= \mathbb{E}^{\mathbb{P}}\left[\frac{1}{Y_{n+1}}\prod_{k=1}^{n}\frac{1}{Y_k}\middle|\mathcal{F}_n\right] \\ &= \frac{1}{X_n}\mathbb{E}^{\mathbb{P}}\left[\frac{1}{Y_{n+1}}\cdot 1\right] = \frac{1}{X_n}\mathbb{E}^{\mathbb{P}}\left[\frac{1}{Y_{n+1}}\cdot \mathbb{1}_S\right] \\ &= \frac{1}{X_n}\mathbb{Q}\left(w_{n+1} \in S\right) \\ &= \frac{1}{X_n}\nu_{n+1}(S) = \frac{1}{X_n} \end{aligned} \tag{8.142}$$

We have already shown that $X_n, Y_n \geq 0$ everywhere for all $n \geq 1$. By same argument in (2) there are also two cases in reverse direction. Define $Z_n := 1/X_n$:

- First, if $\exists Y_m = 0$ ν_m-as for some m, then $1/Y_m = \infty$ μ_m-as. Clearly $Z_n = \infty$ $\mathbb{P}$-as for all $n \geq m$. We can define $Z_n \xrightarrow{a.s.} Z := \infty$.
- Second, if the first case does not happen, then $\{Z_n : n \geq 0\}$ is $\mathbb{P}$-martingale. Hence for any $n \geq 1$, $\mathbb{E}^{\mathbb{P}}[Z_n] = \mathbb{E}^{\mathbb{P}}[Z_0] = 1 < \infty$. Clearly $\{Z_n\}$ is uniformly integrable. By (**MCT2**) $\Rightarrow \exists Z \in \mathcal{L}^1$, such that $Z_n \xrightarrow{a.s.} X$, $Z_n \xrightarrow{\mathcal{L}^1} Z$.

Now assume $\mathbb{Q}(X > 0) = 1$, then $Z_n = 1/X_n < \infty$ for all $n \geq 1$. $\mathbb{E}^{\mathbb{P}}[Z_n]$ exists. For all $A \in \mathcal{F}_n$, by same argument as part-1:

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[Z; A] \tag{8.143}$$

By same pi-system argument, $Z\mathbb{P} = \mathbb{Q}$ on $F \in \mathcal{F} = \sigma(\mathcal{I})$.
So, by definition, $\frac{d\mathbb{Q}}{d\mathbb{P}} := Z$, so $\mathbb{Q} \ll \mathbb{P}$.

In summary:

- If $X > 0$ $\mathbb{Q}$-as, then $\frac{1}{X} < \infty$ $\mathbb{P}$-as. $X = d\mathbb{P}/d\mathbb{Q}; \frac{1}{X} = d\mathbb{Q}/d\mathbb{Q}$, which implies $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$.
- If $X = 0$ $\mathbb{Q}$-as, $\mathbb{Q}(X > 0) = 0$. But $\mathbb{P}(X > 0) = \mathbb{E}^{\mathbb{Q}}[X] = 1$.

There are only these two cases since $X \geq 0$. Dichotomous states. ∎

Problem 2. X, Y RV, $\mathcal{G}$ is sub sigma algebra, show

1. If $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, then $\operatorname{Var}[\mathbb{E}[X|\mathcal{G}]] \leq \operatorname{Var}[X]$.
2. If X is integrable, Y is bounded, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Y] = \mathbb{E}[X\mathbb{E}[Y|\mathcal{G}]] \tag{8.144}$$

3. If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}[X^2|\mathcal{G}] = Y^2$, $\mathbb{E}[X|\mathcal{G}] = Y$, then $X = Y$ $a.s.$

Proof. (1)

$$\operatorname{Var}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}\left[\mathbb{E}^2[X|\mathcal{G}]\right] - \mathbb{E}[\mathbb{E}[X|\mathcal{G}]]^2 \quad (\triangle) \tag{8.145}$$

By (**cJensen**), x^2 is convex, so $\mathbb{E}^2[X|\mathcal{G}] \leq \mathbb{E}[X^2|\mathcal{G}]$, by monotonicity of integral, and also note that $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$:

$$(\triangle) \leq \mathbb{E}\left[\mathbb{E}\left[X^2|\mathcal{G}\right]\right] - \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \operatorname{Var}[X] \quad \blacksquare \tag{8.146}$$

(2) For $A \in \mathcal{G}$, X integrable, let $Z := \mathbb{1}_A$, then

$$\mathbb{E}[XZ] = \int_A X d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\, Z] \tag{8.147}$$

Denote equility $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\, Z]$ as $(\dagger)$.
By linearity, $(\dagger)$ holds for $Z \in S\mathcal{G}^+$.
By (**MON**), $(\dagger)$ holds for $Z \in m\mathcal{G}^+$.
Now suppose $Z \in b\mathcal{G}$, i.e. $\exists\, 0 < M < \infty$, $|Z| \leq M$. Write $Z = Z^+ - Z^-$, then both positive and negative parts should be bounded by M, i.e. $Z^\pm \in b\mathcal{G}^+$. Hence for $Z^\pm$:

$$\mathbb{E}\left[XZ^+\right] + \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\, Z^-\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\, Z^+\right] + \mathbb{E}\left[XZ^-\right] \tag{8.148}$$

$\mathbb{E}[XZ^\pm] \leq M\mathbb{E}[X] < \infty$ since $X \in \mathcal{L}^1$, all integrals involved in the formula above are finite. By linearity:

$$\mathbb{E}\left[XZ^+\right] - \mathbb{E}\left[XZ^-\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\, Z^+\right] - \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\, Z^-\right] \tag{8.149}$$

i.e. $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\, Z]$ for $X \in \mathcal{L}^1$, $Z \in b\mathcal{G}$. Now for any Y bounded, $Z := \mathbb{E}[Y|\mathcal{G}] \in m\mathcal{G}$ and is bounded. Hence $\mathbb{E}[X\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,\mathbb{E}[Y|\mathcal{G}]]$.
• Now consider $\mathbb{E}[YW] = \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]\, W]$ $(\triangle)$ for Y bounded, $|Y| < M$.
Again $(\triangle)$ holds for $W := \mathbb{1}_A$, $A \in \mathcal{G}$.
By linearity, $(\triangle)$ holds for $W \in S\mathcal{G}^+$.
By (**MON**), $(\triangle)$ holds for $W \in m\mathcal{G}^+$.
For $W \in \mathcal{L}^1$, write $W = W^+ - W^-$, for $W^\pm$:

$$\mathbb{E}\left[YW^+\right] + \mathbb{E}\left[\mathbb{E}[Y|\mathcal{G}]\, W^-\right] = \mathbb{E}\left[\mathbb{E}[Y|\mathcal{G}]\, W^+\right] + \mathbb{E}\left[YW^-\right] \tag{8.150}$$

$\mathbb{E}[YW^\pm] \leq M\mathbb{E}[W^\pm] < \infty$ all integrals involved in the formula above are finite. By linearity:

$$\mathbb{E}\left[YW^+\right] - \mathbb{E}\left[YW^-\right] = \mathbb{E}\left[\mathbb{E}[Y|\mathcal{G}]\, W^+\right] - \mathbb{E}\left[\mathbb{E}[Y|\mathcal{G}]\, W^-\right] \tag{8.151}$$

i.e. $\mathbb{E}[YW] = \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]W]$ for $W \in m\mathcal{G}$, $W \in \mathcal{L}^1$, Y bounded. Let $W := \mathbb{E}[X|\mathcal{G}]$, we have $\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]\mathbb{E}[X|\mathcal{G}]]$. We conclude that

$$\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]Y] \tag{8.152}$$

for $X \in \mathcal{L}^1$, Y bounded. ■

(3) by hypothesis

$$\begin{aligned}\mathbb{E}\left[(X-Y)^2\right] &= \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] - 2\mathbb{E}[XY] \\ &= \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] - 2\mathbb{E}[\mathbb{E}[XY|\mathcal{G}]] \\ &= \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] - 2\mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]] \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[Y^2\right] = 0\end{aligned} \tag{8.153}$$

That implies $(X-Y)^2 = 0$ *a.s.*, hence $X = Y$ *a.s.* ■

Problem 3. X, Y RVs with joint distribution being bivariate centered Gaussian $N(0, C)$, with mean $\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$, covariance matrix $C = \left(\begin{smallmatrix}a & c\\ c & b\end{smallmatrix}\right)$, where $a, b > 0$, $ab - c^2 > 0$. That is, the joint density of (X, Y) is given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{\det C}} \exp\left(-\frac{1}{2}\mathbf{x}^T C^{-1}\mathbf{x}\right) \tag{8.154}$$

1. Determine $\mathbb{E}[X|Y]$.
2. Determine $\mathbb{E}[\exp(X - a/2)|Y]$.

Proof. (1) Define $Z := X - \frac{c}{b}\cdot Y$, by linearity Z is still a centerred Gaussian, $\mathbb{E}[Z] = 0$. We have

$$\operatorname{Cov}[Z,Y] = \operatorname{Cov}[X,Y] - \frac{c}{b}\cdot \operatorname{Var}[Y] = c - \frac{c}{b}\cdot b = 0 \tag{8.155}$$

Moreover, since $\operatorname{Cov}[Z,Y] = \mathbb{E}[ZY] - \mathbb{E}[Z]\mathbb{E}[Y] = \mathbb{E}[ZY] = 0$, by result in hint, since Z, Y has joint bivariate centerred Gaussian distribution $\Rightarrow Z, Y$ independent. Hence we can write

$$\begin{aligned}\mathbb{E}[X|Y] &= \mathbb{E}\left[Z + \frac{c}{b}\cdot Y|Y\right] = \mathbb{E}[Z|Y] + \mathbb{E}\left[\frac{c}{b}\cdot Y|Y\right] \\ &= \mathbb{E}[Z] + \frac{c}{b}\cdot Y = \frac{c}{b}\cdot Y \quad ■\end{aligned} \tag{8.156}$$

(2) Still using $Z = X - \frac{c}{b}\cdot Y$, $\mathbb{E}[Z] = 0$, $\operatorname{Var}[Z] = a + \frac{c^2}{b^2}\cdot b - 2\cdot\frac{c}{b}\cdot c = a - \frac{c^2}{b}$. Therefore, $\exp(Z) \sim \ln\mathcal{N}(0, a - \frac{c^2}{b})$, by wikipedia,

$$\mathbb{E}[\exp(Z)] = \exp\left(\mu + \frac{1}{2}\sigma^2\right) = \exp\left(\frac{a}{2} + \frac{c^2}{2b}\right) \tag{8.157}$$

We can write

$$\begin{aligned}
\mathbb{E}\left[\exp\left(X-\frac{a}{2}\right)|Y\right] &= \frac{1}{\exp(\frac{a}{2})}\mathbb{E}\left[\exp\left(Z+\frac{c}{b}\cdot Y\right)|Y\right] \\
&= \frac{1}{\exp(\frac{a}{2})}\mathbb{E}\left[\exp(Z)\exp\left(\frac{c}{b}\cdot Y\right)|Y\right] \\
&= \frac{\mathbb{E}[\exp(Z)]}{\exp(\frac{a}{2})}\exp\left(\frac{c}{b}\cdot Y\right) \\
&= \exp\left(\frac{cY}{b}+\frac{c^2}{2b}\right) \quad \blacksquare
\end{aligned} \tag{8.158}$$

Problem 4. T is a stopping time such that for some $N \in \mathbb{N}$, and some $\epsilon > 0$, we have, for every n:

$$\mathbb{P}(T \leq n+N|\mathcal{F}_n) > \epsilon, \quad a.s. \tag{8.159}$$

Show by induction using $\mathbb{P}(T > kN) = \mathbb{P}(T > kN; T > (k-1)N)$ that for $k = 1, 2, 3...$

$$\mathbb{P}(T > kN) \leq (1-\epsilon)^k \tag{8.160}$$

Show that $\mathbb{E}[T] < \infty$.

Proof. Since $\mathbb{P}(T \leq n+N|\mathcal{F}_n) > \epsilon$, for all $A \in \mathcal{F}_n$, we have

$$\int_A \mathbb{1}_{\{T\leq n+N\}} d\mathbb{P} \geq \int_A \epsilon d\mathbb{P} \tag{8.161}$$

Since T is a stopping time, clearly $\{T > n\} \in \mathcal{F}_n$, so

$$\mathbb{P}(n < T \leq n+N) = \int_{\{T>n\}} \mathbb{1}_{\{T\leq n+N\}} d\mathbb{P} \geq \int_{\{T>n\}} \epsilon d\mathbb{P} = \epsilon\mathbb{P}(T > n) \tag{8.162}$$

Hence, for every n,

$$\begin{aligned}
\mathbb{P}(T > n+N) &= \mathbb{P}(n < T) - \mathbb{P}(n < T \leq n+N) \\
&\leq (1-\epsilon)\cdot\mathbb{P}(n < T)
\end{aligned} \tag{8.163}$$

Pick $n := (k-1)N$, we have

$$\mathbb{P}(T > kN) \leq (1-\epsilon)\cdot\mathbb{P}(T > (k-1)N) \tag{8.164}$$

Note that $\mathbb{P}(T > 0) = 1$, hence for the basic case ($k = 1$) we have $\mathbb{P}(T > N) \leq (1-\epsilon)\cdot 1$. Then for any $k > 1$, proceed recursively for $2, 3, .., k$, we have $\mathbb{P}(T > kN) \leq (1-\epsilon)^k$ as desired. Now we bound T by:

$$T \leq \sum_{k=0}^{\infty}(k+1)N\cdot\mathbb{1}_{\{(kN<T\leq(k+1)N\}} \leq N\sum_{k=0}^{\infty}(k+1)\cdot\mathbb{1}_{\{(kN<T\}} \tag{8.165}$$

Take expectation both sides

$$\begin{aligned}
\mathbb{E}[T] &\leq \mathbb{E}\left[N\sum_{k=0}^{\infty}(k+1)\cdot\mathbb{1}_{\{(kN<T\}}\right] \\
&= N\sum_{k=0}^{\infty}(k+1)\cdot\mathbb{P}(kN < T) = N\sum_{k=0}^{\infty}(k+1)(1-\epsilon)^k
\end{aligned} \tag{8.166}$$

clearly, for $0 < \epsilon < 1$, the summation above converges, hence $\mathbb{E}[T] < \infty$ *a.s.*. ∎

Problem 5. Let $\{X_n : n \geq 0\}$ be i.i.d RVs with common distribution $\mathbb{P}(X_n = 1) = p$, $\mathbb{P}(X_n = -1) = q = 1 - p$, $0 < p < 1$. Define $S_0 := 0$, S_n partial sum. Then say $\{S_n : n \geq 1\}$ is a $(p-q)$ *random walk* on $\mathbb{Z}$. In particular if $p = q = 0.5$, $\{S_n : n \geq 1\}$ is a *symmetric random walk*. Given two positive integers a, b, consider

$$\tau := \inf\{n \geq 1 : S_n = -a \text{ or } S_n = b\} \tag{8.167}$$

1. Show that $\mathbb{E}[\tau] < \infty$.
2. Assume $p \neq 1/2$, compute $\mathbb{P}(S_\tau = -a)$.
3. Assume $p \neq 1/2$, compute $\mathbb{E}[\tau]$.
4. Assume $p = 1/2$, $a = b$, compute $\mathbb{E}[e^{t\tau}]$ for $t \leq 0$.

Proof. (1) Since a, b finite, the walking band has finite width $a + b < \infty$. Consider any staring time position $n \geq 0$, $S_n \in (-a, b)$, we have $\{\tau \leq a + b + n\} \supseteq \bigcap_{k=n+1}^{n+a+b}\{X_k = 1\}$. That is, S_τ must hits b before $\tau = (n + a + b)$ if it takes $(a + b)$ consecutive positive steps from $(n + 1)$. Hence for all $n \geq 1$, let $\{\mathcal{F}_n : n \geq 0\}$ be natural filtration associated with $\{X_n : n \geq 0\}$,

$$\mathbb{P}(\tau \leq a + b + n|\mathcal{F}_n) \geq p^{a+b} > 0 \tag{8.168}$$

Clearly, $\{\tau \leq a + b + n\} \in \mathcal{F}_n$. Apply problem 4's conclusion, with constant $N := a + b, \epsilon := p^{a+b}$, we conclude that $\mathbb{E}[\tau] < \infty$ *a.s.*. ∎

(2) Stay in the same filtered space for the rest of the proof, i.e. $(\Omega, \mathcal{F}, \{\mathcal{F}_n : n \geq 0\}, \mathbb{P})$, where $\mathcal{F}_n := \sigma(X_0, X_1, ..., X_n)$. Clearly $S_n \in \mathcal{F}_n$. As hint suggests, consider $(\frac{q}{p})^{S_n}(\in m\mathcal{F}_n)$. Noticing that $(\frac{q}{p})^{X_{n+1}}$ is independent wrt $\mathcal{F}_n$, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{q}{p}\right)^{S_{n+1}}\middle|\mathcal{F}_n\right] &= \mathbb{E}\left[\left(\frac{q}{p}\right)^{S_n}\left(\frac{q}{p}^{X_{n+1}}\right)\middle|\mathcal{F}_n\right] \\
&= \left(\frac{q}{p}\right)^{S_n}\mathbb{E}\left[\left(\frac{q}{p}\right)^{X_{n+1}}\right] \\
&= \left(\frac{q}{p}\right)^{S_n}\mathbb{E}\left[\frac{q}{p}\cdot\mathbb{1}_{\{X_{n+1}=1\}} + \frac{p}{q}\cdot\mathbb{1}_{\{X_{n+1}=-1\}}\right] \\
&= \left(\frac{q}{p}\right)^{S_n}\cdot(p+q) = \left(\frac{q}{p}\right)^{S_n}
\end{aligned} \tag{8.169}$$

Hence, define $Z_n := \left(\frac{q}{p}\right)^{S_n}$, $\{Z_n : n \geq 0\}$ is a martingale. Consider $|Z_{n+1} - Z_n|$:

$$\begin{aligned} |Z_{n+1} - Z_n| &= \left| \left(\frac{q}{p}\right)^{S_n} \left[\left(\frac{q}{p}\right)^{X_{n+1}} - 1 \right] \right| \\ &\leq \begin{cases} \left(\frac{q}{p}\right)^b \left(\frac{q}{p} + 1\right), & \text{If } q \geq p \\ \left(\frac{p}{q}\right)^a \left(\frac{p}{q} + 1\right), & \text{If } q < p. \end{cases} \\ &\leq \max \left\{ \left(\frac{q}{p}\right)^b \left(\frac{q}{p} + 1\right), \left(\frac{p}{q}\right)^a \left(\frac{p}{q} + 1\right) \right\} < \infty \end{aligned} \tag{8.170}$$

Also by (1)'s result, $\mathbb{E}[\tau] < \infty$. Apply (**Hunt**'s, case-3): $\mathbb{E}[Z_\tau] = \mathbb{E}[Z_0] = 1$. Now since $\mathbb{P}(S_\tau = -a) + \mathbb{P}(S_\tau = b) = 1$, and

$$1 = \mathbb{E}[Z_\tau] = \left(\frac{q}{p}\right)^{-a} \cdot \mathbb{P}(S_\tau = -a) + \left(\frac{q}{p}\right)^b \cdot \mathbb{P}(S_\tau = b) \tag{8.171}$$

we get

$$\mathbb{P}(S_\tau = -a) = \frac{p^b q^a - q^a q^b}{p^a p^b - q^a q^b} \quad q \neq p \quad \blacksquare \tag{8.172}$$

(3) Consider

$$\mathbb{E}[S_{n+1} | \mathcal{F}_n] = S_n + \mathbb{E}[X_{n+1}] = S_n + p - q \tag{8.173}$$

Substract $(p-q)(n+1)$ from both sides, we get

$$\mathbb{E}[S_{n+1} - (p-q)(n+1) | \mathcal{F}_n] = S_n - (p-q)n \tag{8.174}$$

Hence define $\{M_n : n \geq 0\} := \{S_n - (p-q)n : n \geq 0\}$, M_n is a martingale. Moreover, for any fixed $N \geq 0$, $(\tau \wedge N), (\tau \wedge 0)$ are bounded stopping times. Apply (**Hunt**'s, case-1): $\mathbb{E}[M_{\tau\wedge N}] = \mathbb{E}[M_{\tau\wedge 0}] = \mathbb{E}[M_0] = 0$. Therefore for any fixed $N \geq 0$:

$$\begin{aligned} (p-q)\mathbb{E}[\tau \wedge N] &= \mathbb{E}[S_{\tau\wedge N}] \\ &= \mathbb{E}[S_\tau; \tau \leq N] + \mathbb{E}[S_n; \tau > N] \end{aligned} \tag{8.175}$$

In which the fisrt part $\mathbb{E}[S_\tau; \tau \leq N] = \mathbb{E}\left[S_\tau \cdot \mathbb{1}_{\{\tau \leq N\}}\right] \nearrow$(by **MON**) $\mathbb{E}[S_\tau]$.
The second part $\mathbb{E}[S_n; \tau > N] \leq (a \vee b)\mathbb{P}(\tau > N) \xrightarrow{N\to\infty} 0$.
So we are allowed to take $N \to \infty$ on both sides,

$$\begin{aligned} \mathbb{E}[\tau] &= \frac{1}{p-q}\mathbb{E}[S_\tau] \\ &= \frac{1}{p-q}\left(-a \cdot \frac{q^a(p^b - q^b)}{p^a p^b - q^a q^b} + b \cdot \frac{p^b(p^a - q^a)}{p^a p^b - q^a q^b}\right) \\ &= \frac{bp^b(p^a - q^a) - aq^a(p^b - q^b)}{(p-q)(p^a p^b - q^a q^b)} \quad \blacksquare \end{aligned} \tag{8.176}$$

(4) For any fixed $r \in \mathbb{R}$, consider e^{rS_n}, clearly $e^{rS_n} \in m\mathcal{F}_n$.

$$\mathbb{E}\left[e^{rS_{n+1}} \middle| \mathcal{F}_n\right] = e^{rS_n}\mathbb{E}\left[e^{rX_{n+1}}\right] = e^{rS_n}\frac{e^r + e^{-r}}{2} \tag{8.177}$$

Divide both sides by $\cosh^{n+1} r$,

$$\mathbb{E}\left[e^{rS_{n+1}} \operatorname{sech}^{n+1} r \middle| \mathcal{F}_n\right] = e^{rS_n} \operatorname{sech}^n r \tag{8.178}$$

Hence $\{e^{rS_n} \operatorname{sech}^n r : n \geq 0\}$ is a martingale. Similar as (3), for any fixed $N \geq 0$, $\tau \wedge n, \tau \wedge 0$ are bounded stopping times. Apply (**Hunt**'s, case-1), we have

$$\mathbb{E}\left[e^{rS_{\tau\wedge N}} \operatorname{sech}^{\tau\wedge N} r\right] = \mathbb{E}\left[e^{rS_0} \operatorname{sech}^0 r\right] = 1 \tag{8.179}$$

In LHS, for all $N > 0, r \in \mathbb{R}, r < \infty$, note that $\operatorname{sech} r \leq 1$.
We have followings:

- $e^{rS_{\tau\wedge N}} \operatorname{sech}^{\tau\wedge N} r \xrightarrow{a.s.} e^{rS_\tau} \operatorname{sech}^\tau r$.
- $e^{rS_{\tau\wedge N}} \operatorname{sech}^{\tau\wedge N} r \leq e^{rS_\tau} \cdot 1$. Moreover, by symmetry: $\mathbb{P}(S_\tau = \pm a) = \frac{1}{2}$, hence we can compute $\mathbb{E}\left[e^{rS_\tau}\right] = \frac{e^{ra}+e^{-ra}}{2} = \cosh ra < \infty$, i.e. $e^{rS_\tau} \in \mathcal{L}^1$.

Apply (**DOM**): $e^{rS_{\tau\wedge N}} \operatorname{sech}^{\tau\wedge N} r \xrightarrow{\mathcal{L}^1} e^{rS_\tau} \operatorname{sech}^\tau r$. Hence

$$\begin{aligned}
1 &= \mathbb{E}\left[e^{rS_\tau} \operatorname{sech}^\tau r\right] \\
&= \mathbb{E}\left[e^{rS_\tau} \operatorname{sech}^\tau r; S_\tau = a\right] + \mathbb{E}\left[e^{rS_\tau} \operatorname{sech}^\tau r; S_\tau = -a\right] \\
&= e^{ra}\mathbb{E}\left[\operatorname{sech}^\tau r; S_\tau = a\right] + e^{-ra}\mathbb{E}\left[\operatorname{sech}^\tau r; S_\tau = -a\right] \\
&= \frac{e^{ra}}{2}\mathbb{E}\left[\operatorname{sech}^\tau r\right] + \frac{e^{-ra}}{2}\mathbb{E}\left[\operatorname{sech}^\tau r\right] \quad \text{(Since distribution of } S_\tau \text{ is symmetric)} \\
&= \cosh(ra) \cdot \mathbb{E}\left[\operatorname{sech}^\tau r\right]
\end{aligned} \tag{8.180}$$

Change variable, denote $x := \operatorname{sech} r$, $r = \operatorname{arcsech} x = \log\left(\frac{1}{x} + \sqrt{\frac{1}{x^2} - 1}\right)$,

$$\operatorname{sech}(a \operatorname{arcsech} x) = \mathbb{E}\left[x^\tau\right] \tag{8.181}$$

Hence for $t \leq 0$, $\mathbb{E}\left[e^{t\tau}\right] = \operatorname{sech}(a \cdot \operatorname{arcsech}(e^t))$. ■

Problem 6. Build a sequence $\{X_n : n \geq 1\}$, $X_n \in \mathcal{L}^1$, such that

$$\mathbb{E}\left[X_{n+1}|X_n\right] = X_n \quad \text{For all } n \geq 1, \text{ but } \quad \mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \neq X_n \quad \text{for } n \geq 2. \tag{8.182}$$

Where $\mathcal{F}_n := \sigma(X_j : 1 \leq j \leq n)$

Proof. $Z_1, Z_2 \sim \mathcal{N}(0,1)$, are two independent standard gaussians, we have $\mathbb{E}\left[Z_1\right] = 0, \mathbb{E}\left[Z_1^2\right] = 1$. Now consider $a(Z_1 + Z_2)$ and $b(Z_1 - Z_2)$ for any constant numbers $a, b < \infty$; these two are both gaussians, and has joint bavariate Gaussian distribution, moreover

$$\begin{aligned}
\operatorname{Cov}\left[a(Z_1+Z_2), b(Z_1-Z_2)\right] &= \mathbb{E}\left[ab(Z_1^2 - Z_2^2)\right] - ab\mathbb{E}\left[Z_1+Z_2\right]\mathbb{E}\left[Z_1-Z_2\right] \\
&= ab(1-1) - 0 = 0
\end{aligned} \tag{8.183}$$

Hence $a(Z_1 + Z_2)$, $b(Z_1 - Z_2)$ are independent for any $a, b < \infty$.
We construct $\{X_n : n \geq 1\}$ as follows

$$X_n = \begin{cases} 2^{\frac{n+1}{2}} \cdot Z_1, & \text{if n is odd} \\ 2^{\frac{n}{2}} \cdot (Z_1 - Z_2), & \text{if n is even} \end{cases} \tag{8.184}$$

That is, $\{X_n : n \geq 1\} := \{2Z_1, 2(Z_1 - Z_2), 4Z_1, 4(Z_1 - Z_2), 8Z_1, 8(Z_1 - Z_2), ...\}$ $\mathcal{F}_2 = \sigma(2Z_1, 2(Z_1 - Z_2))$, then for any $n \geq 2$, $X_n \in m\mathcal{F}_2 \subseteq m\mathcal{F}_3 \subseteq ... \subseteq m\mathcal{F}_{n-1}$. (actually equal signs). Now check required properties of X, for $n \geq 2$:

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] = X_{n+1} \neq X_n \tag{8.185}$$

For $n \geq 1$, $n+1$ odd:

$$\begin{aligned}
\mathbb{E}\left[X_{n+1}|X_n\right] &= \mathbb{E}\left[2^{\frac{n+2}{2}} Z_1 \middle| 2^{\frac{n}{2}}(Z_1 - Z_2)\right] \\
&= \mathbb{E}\left[2^{\frac{n}{2}}(Z_1 + Z_2) + 2^{\frac{n}{2}}(Z_1 - Z_2) | 2^{\frac{n}{2}}(Z_1 - Z_2)\right] \\
&= \mathbb{E}\left[2^{\frac{n}{2}}(Z_1 + Z_2)\right] + \mathbb{E}\left[2^{\frac{n}{2}}(Z_1 - Z_2) | 2^{\frac{n}{2}}(Z_1 - Z_2)\right] \\
&= 0 + 2^{\frac{n}{2}}(Z_1 - Z_2) = X_n
\end{aligned} \tag{8.186}$$

For $n \geq 1$, $n+1$ even:

$$\begin{aligned}
\mathbb{E}\left[X_{n+1}|X_n\right] &= \mathbb{E}\left[2^{\frac{n+1}{2}}(Z_1 - Z_2) \middle| 2^{\frac{n+1}{2}} Z_1\right] \\
&= \mathbb{E}\left[2^{\frac{n}{2}} Z_1 | 2^{\frac{n}{2}} Z_1\right] - \mathbb{E}\left[2^{\frac{n}{2}} Z_2\right] \\
&= 2^{\frac{n}{2}} Z_1 - 0 = X_n \quad \blacksquare
\end{aligned} \tag{8.187}$$

Problem 7. Given filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_n : n \geq 0\}, \mathbb{P})$, let $\{Y_n : n \geq 1\}$ adapted, such that $Y_n \in \mathcal{L}^2$, $\mathbb{E}\left[Y_n|\mathcal{F}_{n-1}\right] = 0$. Further assume that $\sum_{n\geq 1} \mathbb{E}\left[Y_n^2\right]/n^2 < \infty$. Define $X_0 := 0$, $X_n := \sum_{j=1}^n Y_j/j$, S_n be partial sum of Y_n.

1. Show that $\{X_n : n \geq 0\}$ is a martingale wrt $\{\mathcal{F}_n : n \geq 0\}$.
2. Based on (1) show that SLLN holds for sequence Y_n, i.e.

$$\frac{S_n}{n} \xrightarrow{a.s.} 0 \tag{8.188}$$

Proof. (1) Since $\{\mathcal{F}_n : n \geq 0\}$ is filtration, $\{Y_n : n \geq 1\}$ is adapted, we have $Y_n \in m\mathcal{F}_n$. Moreover, for all $1 \leq j \leq n$, $\mathcal{F}_j \subseteq \mathcal{F}_n$, hence $Y_j \in m\mathcal{F}_j \subseteq m\mathcal{F}_n$, $S_n \in m\mathcal{F}_n$, $X_n \in m\mathcal{F}_n$ are also adapted.

$$\begin{aligned}
\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[\sum_{j=1}^{n+1} \frac{Y_j}{j} \middle| \mathcal{F}_n\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{n} \frac{Y_j}{j} \middle| \mathcal{F}_n\right] + \mathbb{E}\left[\frac{Y_{n+1}}{n+1} \middle| \mathcal{F}_n\right] = X_n \quad \blacksquare
\end{aligned} \tag{8.189}$$

(2) We first calculte second moment of X_n,

$$\begin{aligned}
\mathbb{E}\left[X_n^2\right] = \mathbb{E}\left[\left(\sum_{j=1}^n \frac{Y_j}{j}\right)^2\right] &= \mathbb{E}\left[\left(\sum_{j=1}^n \frac{Y_j^2}{j^2} + \sum_{i=1}^n \sum_{k=1,k\neq i}^n \frac{Y_i Y_k}{ik}\right)\right] \\
= \sum_{j=1}^n \frac{\mathbb{E}\left[Y_j^2\right]}{j^2} + \sum_{i=1}^n \sum_{k=1,k\neq i}^n \frac{\mathbb{E}\left[Y_i Y_k\right]}{ik} & \\
= \sum_{j=1}^n \frac{\mathbb{E}\left[Y_j^2\right]}{j^2} + \sum_{i=1}^n \left(\sum_{k=1,k\leq i-1}^n \frac{\mathbb{E}\left[\mathbb{E}\left[Y_i Y_k|\mathcal{F}_{i-1}\right]\right]}{ik} + \sum_{k=1,i\leq k-1}^n \frac{\mathbb{E}\left[\mathbb{E}\left[Y_i Y_k|\mathcal{F}_{k-1}\right]\right]}{ik}\right) &
\end{aligned} \tag{8.190}$$

Now look at $\mathbb{E}\left[\mathbb{E}\left[Y_iY_k|\mathcal{F}_{i-1}\right]\right]$ in the first part (where $k \le i-1$) in the second layer of the cross terms' summation. Clearly, $Y_k \in m\mathcal{F}_k \subseteq m\mathcal{F}_{i-1}$, so it can be taken out from inner conditional expectation, i.e.

$$\mathbb{E}\left[\mathbb{E}\left[Y_iY_k|\mathcal{F}_{i-1}\right]\right] = \mathbb{E}\left[Y_k\mathbb{E}\left[Y_i|\mathcal{F}_{i-1}\right]\right] = \mathbb{E}\left[Y_k \cdot 0\right] = 0 \tag{8.191}$$

Same story for the second part (where $i \le k-1$),

$$\mathbb{E}\left[\mathbb{E}\left[Y_iY_k|\mathcal{F}_{k-1}\right]\right] = \mathbb{E}\left[Y_i\mathbb{E}\left[Y_k|\mathcal{F}_{k-1}\right]\right] = \mathbb{E}\left[Y_i \cdot 0\right] = 0 \tag{8.192}$$

Hence the cross terms are actually zero. That is $\mathbb{E}\left[X_n^2\right] = \sum_{j=1}^n \frac{\mathbb{E}\left[Y_j^2\right]}{j^2} < \infty$. We conclude that $\{X_n : n \ge 0\}$ is bounded by $\mathcal{L}^2$.
By (**MCT3**), there exists $X \in \mathcal{L}^2$, such that $X_n \xrightarrow{a.s.} X$; $X_n \xrightarrow{\mathcal{L}^2} X$. Since $X \in \mathcal{L}^2$, $|X|$ must be finite, so is X. That is to say:

$$X_n := \sum_{j=1}^n \frac{Y_j}{j} \xrightarrow{a.s.} X < \infty \tag{8.193}$$

By (**Kronecker**)'s lemma,

$$\frac{1}{n}\sum_{j=1}^n Y_n = \frac{S_n}{n} \xrightarrow{a.s.} 0 \ \blacksquare \tag{8.194}$$

Problem 8. A branching process $\{Z_n : n \ge 0\}$ is constructed in following way. I.e., for a family $\{X_k^{(n)} : n, k \ge 1\}$ of i.i.d $\mathbb{Z}^+$-valued RVs, define $Z_0 := 1$, then define recursively for $n \ge 0$,

$$Z_{n+1} := \sum_{k=1}^{Z_n} X_k^{(n+1)} \tag{8.195}$$

For any one of $X_k^{(n)}$, denoted by X, $\mu := \mathbb{E}\left[X\right] < \infty$, $0 < \sigma^2 := \mathrm{Var}\left[X\right] < \infty$. Show that $M_n := Z_n/\mu^n$ is a martingale wrt filtration $\mathcal{F}_n := \sigma(Z_0, Z_1, ..., Z_n)$. Further show that

$$\mathbb{E}\left[Z_{n+1}^2|\mathcal{F}_n\right] = \mu^2 Z_n^2 + \sigma^2 Z_n \tag{8.196}$$

And deduce that $\{M_n\}$ is bounded in $\mathcal{L}^2$ iff $\mu > 1$. Show that when $\mu > 1$,

$$\mathrm{Var}\left[M_\infty\right] = \frac{\sigma^2}{\mu(\mu-1)} \tag{8.197}$$

Proof. For any $n, k \ge 1$, $X_k^{(n+1)}$ is independent to $\mathcal{F}_n = \sigma(Z_0, Z_1, ..., Z_n)$. Moreover $\{X_k^{(n+1)} : k \ge 1\}$ are i.i.d for all n. So $\mathbb{E}\left[X_k^{(n+1)}|\mathcal{F}_n\right] = \mathbb{E}\left[X_k^{(n+1)}\right] = \mu$.

Now Consider

$$\begin{aligned}
\mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[\sum_{k=1}^{Z_n} X_k^{(n+1)}\middle|\mathcal{F}_n\right] = \mathbb{E}\left[\sum_{k\geq 1} X_k^{(n+1)}\mathbb{1}_{(Z_n\geq k)}\middle|\mathcal{F}_n\right] \\
&= \sum_{k\geq 1}\mathbb{E}\left[X_k^{(n+1)}|\mathcal{F}_n\right]\cdot\mathbb{E}\left[\mathbb{1}_{(Z_n\geq k)}|\mathcal{F}_n\right] \\
&= \mu\sum_{k\geq 1}\mathbb{E}\left[\mathbb{1}_{(Z_n\geq k)}|\mathcal{F}_n\right] \quad (\text{Next: since } \mathbb{1}_{(Z_n\geq k)}\in m\mathcal{F}_n) \\
&= \mu\sum_{k\geq 1}\mathbb{1}_{(Z_n\geq k)} = \mu\sum_{k=1}^{Z_n} 1 = \mu Z_n
\end{aligned} \tag{8.198}$$

Hence, multiply both sides by $\mu^{-(n+1)}$, we get

$$\mathbb{E}\left[\frac{Z_{n+1}}{\mu^{n+1}}\middle|\mathcal{F}_n\right] = \frac{Z_n}{\mu^n} \tag{8.199}$$

i.e. $\{M_n : n\geq 0\} := \{Z_n\mu^{-n} : n\geq 0\}$ is a martingale.
Now calculate conditional second moment of Z_{n+1}. Note that $\text{Var}\left[X_k^{(n)}\right] = \sigma^2$, hence $\mathbb{E}\left[(X_k^{(n)})^2\right] = \mu^2+\sigma^2$ for any $n, k\geq 1$.

$$\begin{aligned}
\mathbb{E}\left[Z_{n+1}^2|\mathcal{F}_n\right] &= \mathbb{E}\left[\left(\sum_{k=1}^{Z_n} X_k^{(n+1)}\right)^2\middle|\mathcal{F}_n\right] = \mathbb{E}\left[\left(\sum_{k\geq 1} X_k^{(n+1)}\mathbb{1}_{(Z_n\geq k)}\right)^2\middle|\mathcal{F}_n\right] \\
&= \mathbb{E}\left[\sum_{k\geq 1}(X_k^{(n+1)})^2\mathbb{1}_{(Z_n\geq k)} + \sum_{i\geq 1}\sum_{j\geq 1, j\neq i} X_i^{(n+1)}X_j^{(n+1)}\mathbb{1}_{(Z_n\geq i\vee j)}\middle|\mathcal{F}_n\right] \\
&= (\mu^2+\sigma^2)\sum_{k\geq 1}\mathbb{1}_{(Z_n\geq k)} + \mu^2\sum_{i\geq 1}\sum_{j\geq 1, j\neq i}\mathbb{1}_{(Z_n\geq i\vee j)} \\
&= (\mu^2+\sigma^2)\sum_{k=1}^{Z_n} 1 + \mu^2\sum_{i=1}^{Z_n}\sum_{j=1, j\neq i}^{Z_n} 1 \\
&= (\mu^2+\sigma^2)Z_n + \mu^2(Z_n^2 - Z_n) \\
&= \mu^2 Z_n^2 + \sigma^2 Z_n
\end{aligned} \tag{8.200}$$

Now devide both sides by μ^{2n+2},

$$\mathbb{E}\left[M_{n+1}^2\middle|\mathcal{F}_n\right] := \mathbb{E}\left[\frac{Z_{n+1}^2}{\mu^{2n+2}}\middle|\mathcal{F}_n\right] = \frac{Z_n^2}{\mu^{2n}} + \frac{\sigma^2 Z_n}{\mu^{2n+2}} =: M_n^2 + \frac{\sigma^2 Z_n}{\mu^{2n+2}} \tag{8.201}$$

Take expectation both sides,

$$\mathbb{E}\left[M_{n+1}^2\right] = \mathbb{E}\left[M_n^2\right] + \frac{\sigma^2\mathbb{E}\left[Z_n\right]}{\mu^{2(n+1)}} \tag{8.202}$$

Expectation of Z_n is given by M_n:

$$\mathbb{E}\left[\frac{Z_n}{\mu^n}\right] = \mathbb{E}\left[\frac{Z_0}{\mu^0}\right] \quad \text{i.e. } \mathbb{E}\left[Z_n\right] = \mu^n \tag{8.203}$$

Hence

$$\begin{aligned}\mathbb{E}\left[M_n^2\right] &= \mathbb{E}\left[M_0^2\right] + \sum_{k=1}^{n} \frac{\sigma^2 \mathbb{E}\left[Z_{k-1}\right]}{\mu^{2k}} \\ &= 1 + \sum_{k=1}^{n} \frac{\sigma^2}{\mu^{k+1}} = 1 + \frac{\sigma^2}{\mu(\mu-1)} \left(1 - \frac{1}{\mu^{n+1}}\right)\end{aligned} \tag{8.204}$$

Clearly $\mathbb{E}\left[M_n^2\right]$ converges if and only if $\mu > 1$.
When $\mu \geq 1$, $\mathbb{E}\left[M_n^2\right] < 1 + \frac{\sigma^2}{\mu(\mu-1)} < \infty$, i.e. M_n is bounded by $\mathcal{L}^2$. By (**MCT3**), $\exists M \in \mathcal{L}^2$, $M_n \xrightarrow{a.s.} M$ and $M_n \xrightarrow{\mathcal{L}^2} M$, therefore

$$\mathbb{E}\left[M^2\right] = \lim_{n\to\infty} \mathbb{E}\left[M_n^2\right] = 1 + \frac{\sigma^2}{\mu(\mu-1)} \tag{8.205}$$

Note that $\mathbb{E}\left[M\right] = \mathbb{E}\left[M_0\right] = 1$, So $\mathrm{Var}\left[M\right] = \frac{\sigma^2}{\mu(\mu-1)}$. ■