Chapter 1

Exercises and questions

1.1 Exercises with Prof. Issoglio (2024/2025)

Here I gathered the exercises seen with Prof. Issoglio. Each class begins with a small recap on the topics that the exercises will be based on.

1.1.1 Exercise class 1

Revise with Kotatsu!

→ **Measurable and measure spaces**:

$$(\underbrace{E}_{\text{set}}, \underbrace{\mathcal{E}}_{\sigma\text{-algebra}}, \underbrace{\nu}_{\text{measure}})$$

An example is given by discrete spaces (E is finite and countable) or the real spaces ($E = \mathbb{R}$), ($E = \mathbb{R}^n$).

→ **Measurable functions** form $(E, \mathcal{E})$ to $(F, \mathcal{F})$. It is a function $f : E \mapsto F$ such that $\forall B \in \mathcal{F}$ we have $f^{-1}(B) \in \mathcal{E}$. This is the least possible "regularity" we can ask for to still be able to do some analysis (it is much less restrictive than continuity).
Property: if f_n are measurable, also $\liminf_n f_n$, $\limsup_n f_n$, $f_1 + f_2$, $f_1 \cdot f_2$, λf_1 are measurable.

→ **Probability spaces**: special case of a measure space where $\nu(E) = 1$. Given a real random variable $X : \Omega \mapsto \mathbb{R}$ we can consider the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}_X)$ where $\mathcal{L}_X$ is the law of X given by

$$\mathcal{L}_X := \mathbb{P}(X^{-1}(A)), \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

We have the mapping

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}_X).$$

→ **σ-algebras generated by random variables**: given $X : \Omega \mapsto \mathbb{R}$ we define

$$\sigma(X) := \{X^{-1}(A) \text{ for some } A \in \mathcal{B}(R)\}.$$

This is the *smallest* σ-algebra such that X is a random variable on $(\Omega, \sigma(X))$. By properties of p-systems and d-systems it is enough to consider a set of elements of R that generates $\mathcal{B}(\mathbb{R})$ (for example: all open sets) when checking measurability related properties. This means that

$$\begin{aligned}\sigma(X) &= \{X^{-1}(A) \text{ for some } A \in \mathcal{B}(\mathbb{R})\}\\ &= \{X^{-1}(A) \text{ for some } A \text{ open}\}\end{aligned}$$

Exercise 1

[redacted]

Let us introduce a random variable (which is a map) $X_n : \Omega \mapsto \mathbb{R}$ given by

$$X_n = \begin{cases} 1 & \text{if } \omega_n = 1 \\ 0 & \text{if } \omega_n = 0. \end{cases}$$

Now, X_n is $\mathcal{F}$-measurable (i.e. it is a random variable on $(\Omega, \mathcal{F})$) because for $\forall\, B \in \mathcal{B}(\mathbb{R})$ we have

$$X_n^{-1}(B) = \{\omega \in \Omega : X_n(\omega) \in B\}$$

and this set actually belongs to $\mathcal{F}$. Why? Because it basically is

$$\{\omega \in \Omega : X_n(\omega) = B\} = \begin{cases} \varnothing & \text{if } 0, 1 \notin B \\ \Omega & \text{if } 0, 1 \in B \\ \{\underbrace{\omega \in \Omega : X_n(\omega) = 1}_{\text{all the } \omega \text{ with } \omega_n = 1 \implies \in C}\} & \text{if } 0 \notin B, 1 \in B \\ \{\underbrace{\omega \in \Omega : X_n(\omega) = 0}_{\text{all the } \omega \text{ with } \omega_n = 0 \implies \in C}\} & \text{if } 0 \in B, 1 \notin B \end{cases}$$

So the process is measurable with respect to $\mathcal{F}$: the σ-algebra generated by C has all the ω necessary to fully explain the process. Remember that $B \subset \Omega$! Now let $S_n = \sum_{i=1}^n X_i$. Then our set A becomes

$$A = \left\{\omega : \frac{S_n}{n} \xrightarrow[n\to\infty]{} \frac{1}{2}\right\}.$$

We know that S_n is a sum of $\mathcal{F}$-measurable functions and therefore it is $\mathcal{F}$-measurable. We can write A as

$$\begin{aligned} A &= \left\{\omega \in \Omega : \frac{S_n}{n} \to \frac{1}{2}\right\} \\ &= \left\{\omega \in \Omega : \liminf_{n\to\infty} \frac{S_n}{n} = \limsup_{n\to\infty} \frac{S_n}{n} = \frac{1}{2}\right\} \\ &= \left\{\underbrace{\omega \in \Omega : \liminf_{n\to\infty} \frac{S_n}{n} = \frac{1}{2}}_{\in \mathcal{F}}\right\} \cap \left\{\underbrace{\omega \in \Omega : \limsup_{n\to\infty} \frac{S_n}{n} = \frac{1}{2}}_{\in \mathcal{F}}\right\} \end{aligned}$$

since S_n is $\mathcal{F}$-measurable, so are its lim sup and lim inf. By doing so we went from the limit, which is not a nice thing to handle[1] when it comes to measurability, to lim inf and lim sup that are more manageable objects.

[1] Unlike a gun pointed to my head by myself.

Exercise 2

[redacted]

1. If we show that $\mathcal{A}_n$ is finite, then it is enough to prove that it is an algebra, since the closure under *union, intersection and complementation* will automatically follow from this.
We easily see that $\mathcal{A}_n$ is finite because n is fixed so $\{0,1\}^n$ is finite and since the varying set in this case is $B \subset \{0,1\}^n$ then also $\mathcal{A}_n$ must be finite. For example we would get

$$\mathcal{A}_1 = \{\{\omega \in \Omega : \omega_1 \in \{0\}\}, \{\omega \in \Omega : \omega_1 \in \{1\}\}, \{\omega \in \Omega : \omega_1 \in \{0,1\}\}, \{\omega \in \Omega : \omega_1 \in \varnothing\}\}.$$

We can also check that this is an algebra:

(a) it is obvious that $\varnothing \in \mathcal{A}_n$;

(b) let $A \in \mathcal{A}_n$. This means that $\exists B \subset \{0,1\}^n$ such that

$$A = \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}$$

and so

$$\begin{aligned} A^c &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \notin B\} \\ &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B^c\} \in \mathcal{A}_n \end{aligned}$$

since $B^c \subset \{0,1\}^n$. So $\mathcal{A}_n$ is closed under intersection.

(c) Let $A_1, A_2 \in \mathcal{A}_n$. Then

$$\begin{aligned} A_1 \cup A_2 &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n \in B_1) \cup \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B_2\}\} \\ &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B_1 \cup B_2\} \in \mathcal{A}_n \end{aligned}$$

because of course $B_1 \cup B_2 \subset \{0,1\}^n$. So $\mathcal{A}_n$ is closed under union.

So we conclude that $\mathcal{A}_n$ is an algebra and since it is finite it is also a σ-algebra.

2. To prove that $\bigcup_n \mathcal{A}_n$ is *not* a σ-algebra it is sufficient to find an element $A \in \sigma\left(\bigcup_n \mathcal{A}_n\right)$ such that $A \notin \bigcup_n \mathcal{A}_n$. This is because a σ-algebra must be closed under countable unions (beside countable intersections and complementation) and if we find an element in the σ-algebra generated by the union of the sets $\mathcal{A}_n$ that is not in the union we would get that the sets are not closed under union[2]!
We notice that for each n, $\mathcal{A}_n$ does not include the set formed by the single element $\mathbf{0} = (0, 0, \dots)$ because any element $A \in \mathcal{A}_n$ contains infinitely many ωs. Thus $\{\mathbf{0}\} \notin \bigcup_n \mathcal{A}_n$. But we can write $\{\mathbf{0}\} = \bigcap_n A_n$ where

$$A_n = \{\omega \in \Omega : (\omega_1, \dots, \omega_n) = (0, \dots, 0)\}.$$

Clearly $A_n \in \mathcal{A}_n \subset \bigcup_m \mathcal{A}_m$, so $\{\mathbf{0}\} \in \sigma\left(\bigcup_n \mathcal{A}_n\right)$ since it is expressed as the countable intersection of $\bigcup_n \mathcal{A}_n$.

Exercise 3

[redacted]

[2] The only one doing an union here is me.

1
- We show that $\emptyset \in \mathcal{F}$. Notice that $\mathbb{P}(\Omega) = 1$ and so $\Omega \in \mathcal{F}$. notice that $\mathcal{F}$ is closed under complementation because if $F \in \mathcal{F}$ then either $\mathbb{P}(F) = 0$ or $\mathbb{P}(F) = 1$. Hence $\mathbb{P}(F^c) = 1 - \mathbb{P}(F) \in \{0,1\}$. Thus $\Omega^c \in \mathcal{F} \implies \emptyset \in \mathcal{F}$.
- We show that if $F \in \mathcal{F}$ then $F^c \in \mathcal{F}$... but we already did that.
- We show that if $F_n \in \mathcal{F}$ then $\bigcup_n F_n \in \mathcal{F}$. We consider two cases:
  - if $\mathbb{P}(F_n) = 0\ \forall n$ then $\mathbb{P}(\bigcup_n F_n) \leqslant \sum_n \mathbb{P}(F_n) = 0$ and therefore
  $$\bigcup_n F_n \in \mathcal{F};$$
  - if there exists at least one $\overline{n}$ such that $\mathbb{P}(F_{\overline{n}}) = 1$ then using that $F_{\overline{n}} \subseteq \bigcup_n F_n$ we get that
  $$1 = \mathbb{P}(F_{\overline{n}}) \leqslant \mathbb{P}\left(\bigcup_n F_n\right) \implies \bigcup_n F_n \in \mathcal{F}.$$

2
(a) Take any discrete finite Ω such that each element $\omega \in \Omega$ has positive probability and strictly less than 1 (for example, a uniform probability space). In this case there are no elements $A \in \mathcal{A}$ such that $\mathbb{P}(A) \in \{0,1\}$ apart form $\emptyset$ and Ω for which probabilities are respectively 0 and 1 so that $\mathcal{F} = \{\emptyset, \Omega\}$.

(b) Here we need an infinite set Ω. Let us choose $\Omega = [0,1]$ with $\mathcal{A} = \mathcal{B}([0,1])$ and $\mathbb{P} = \lambda$. In this case
$$\mathbb{P}(\{x\}) = 0 \qquad \forall x \in [0,1]$$
and therefore $\mathcal{F}$ is finite (even uncountable).

Exercise 4

[redacted]

Let us set $A_n = \{\omega \in \Omega : \mathbb{P}(\{\omega\}) > \frac{1}{n}\}$ for $n \geqslant 1$. For example:
$$\begin{aligned} A_1 &= \{\omega \in \Omega : \mathbb{P}(\{\omega\}) > 1\} = \emptyset \\ A_2 &= \left\{\omega \in \Omega : \mathbb{P}(\{\omega\}) > \frac{1}{2}\right\} \\ A_3 &= \left\{\omega \in \Omega : \mathbb{P}(\{\omega\}) > \frac{1}{3}\right\} \\ &\vdots \end{aligned}$$

Since $\mathbb{P}(\Omega) = 1$ we have $\#A_1 = 1$, $\#A_2 < 2$, $\#A_3 < 3$ so in generale
$$\#A_n \leqslant n.$$

Notice moreover that $A_1 \subset A_2 \subset A_3 \ldots$ and so
$$A = \bigcup_{n \in \mathbb{N}} A_n.$$

Thus A is the countable union of finite sets, hence it is countable.

Exercise 5

[redacted]

ⓐ Setting

$$F = \{A \cup N : A \in \mathcal{F}, N \in \mathcal{N}\}$$

we see that

$$\mathcal{F} \cup \mathcal{N} = F \subset \overline{\mathcal{F}} := \sigma(\mathcal{F} \cup \mathcal{N}). \qquad (\otimes)$$

If we prove that F is a σ-algebra then we have $\sigma(F) = F$ hence by $\otimes$ we have $F = \mathcal{F}$. We know that

- $\varnothing \in F$ because $\varnothing = \varnothing \cup \varnothing$ and $\varnothing \in \mathcal{A}, \varnothing \in \mathcal{N}$ (✓);
- if $B \in F$ then $B^c \in F$. Indeed $B = A \cup N$ for some $A \in \mathcal{F}, N \in \mathcal{N}$ so

  $$B^c = (A \cup N)^c = A^c \cap N^c$$

  by DeMorgan's law. Now we observe that since for $N \in \mathcal{N}$ there exists $M \supseteq N$ such that $M \in \mathcal{F}, \mu(M) = 0$ we can write

  $$N^c = M^c \cup (M \setminus N)$$

  .

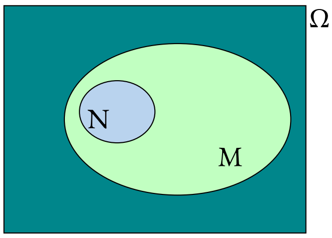


Figure 1.1: Of course, it is never too late for some sickz tickz pickz

  Then from DeMorgan Law we get that

  $$\begin{aligned} B^c &= A^c \cap (M^c \cup (M \setminus N)) \\ &= \underbrace{\underbrace{A^c}_{\in \mathcal{F}} \cap \underbrace{M^c}_{\in \mathcal{F}}}_{\in \mathcal{F}} \cup \underbrace{A^c \cap (M \setminus N)}_{\subseteq M \setminus N \subset M \text{ and } \mu(M)=0} \\ &\implies A^c \cap (M \setminus N) \in \mathcal{N} \\ &\implies B^c \in F. \end{aligned}$$

- Let $\{B_n\}_n$ be a countable family, $B_n \in F$. We prove that

  $$\bigcup_n B_n \in F.$$

  By hypothesis $\exists A_n \in \mathcal{F},\ N_n \in \mathcal{N}$ such that $B_n = A_n \cup N_n$ and therefore

  $$\begin{aligned} \bigcup_n B_n &= \bigcup_n (A_n \cup N_n) \\ &= \underbrace{\bigcup_n A_n}_{\in \mathcal{F}} \cup \underbrace{\bigcup_n N_n}_{\subseteq \bigcup_n M_n \text{ with } M_n \in \mathcal{F}, \mu(M_n)=0}. \end{aligned}$$

Moreover,

$$\mu\left(\bigcup_n M_n\right) \leqslant \sum_n \mu(M_n) = 0 \implies \bigcup_n N_n \in \mathcal{N} \implies \bigcup_n B_n \in \mathcal{F}.$$

b Recall that $\forall A, N$ we let $\overline{\mu}(A \cup N) = \mu(A)$.

b1) Let $A \in \mathcal{F}$. Then $A = A \cap \varnothing$ with $\varnothing \in \mathcal{N}$. Then

$$\overline{\mu}(A) = \overline{\mu}(A \cap \varnothing) = \mu(A).$$

b2) We have to show that

i. the DeMorgan equation is well-posed, that is

$$A \cup N = A' \cup N' \implies \overline{\mu}(A \cup N) = \overline{\mu}(A' \cup N');$$

ii. $\overline{\mu}$ is a measure on $(\Omega, \overline{\mathcal{F}})$;

iii. $\overline{\mu}$ is the only measure on $(\Omega, \overline{\mathcal{F}})$ such that the DeMorgan equation holds.

We need to prove each of these points.

i. We notice that

$$A \subset A \cup N = A' \cup N' \subset A' \cup M'$$

where $M' \supset N',\ M' \in \mathcal{F}$ and $\mu(M') = 0$. Moreover, $A \in \mathcal{F}$ and $A' \cup M' \in \mathcal{F}$. This implies

$$\mu(A) \leqslant \mu(A' \cup M') \leqslant \mu(A') + \mu(M') = \mu(A').$$

Analogously we get

$$\mu(A') \leqslant \mu(A) \implies \mu(A) = \mu(A').$$

Thus

$$\overline{\mu}(A \cup N) := \mu(A) = \mu(A') =: \overline{\mu}(A' \cup N').$$

ii. To show that $\overline{\mu}$ is a measure on $(\Omega, \overline{\mathcal{F}})$ we check:

- $\overline{\mu}(\varnothing) = \overline{\mu}(\varnothing \cup \varnothing) = \mu(\varnothing) = 0$;
- let $\{B_n\}_n \subset \overline{\mathcal{F}}$ with B_n pairwise disjoint. Then, since $B_n = \underbrace{A_n}_{\in\mathcal{F}} \cup \underbrace{N_n}_{\in\mathcal{N}}$

$$\begin{aligned}
\overline{\mu}\left(\bigcup_n B_n\right) &= \overline{\mu}\left(\bigcup_n (A_n \cup N_n)\right) \\
&= \overline{\mu}\Big(\underbrace{(\bigcup_n A_n)}_{\in\mathcal{F}} \cup \underbrace{(\bigcup_n N_n)}_{\in\mathcal{N}}\Big) \\
&= \overline{\mu}\left(\bigcup_n A_n\right) && \text{by DeMorgan eqn.} \\
&= \sum_n \mu(A_n) && \text{bc they are pair. disj.} \\
&= \sum_n \overline{\mu}(A_n \cup N_n) \\
&= \sum_n \overline{\mu}(B_n).
\end{aligned}$$

iii. To show that $\overline{\mu}$ is unique we suppose that $\exists \nu$, a measure on $(\Omega, \overline{\mathcal{F}})$ such that

$$\nu(A \cup N) = \mu(A), \qquad \forall A \in \mathcal{F}. \tag{$*$}$$

Then from DeMorgan and $*$ we get

$$\nu(A \cup N) = \overline{\mu}(A \cup N), \qquad \forall A \in \mathcal{F}, \forall N \in \mathcal{N}.$$

Since μ and ν coincide on all $F = A \cup N \in F = \overline{\mathcal{F}}$ they are the same measure.

b3) To show that the space is complete we must show that every negligible set of $\mathcal{F}$ is actually measurable ($\in \mathcal{F}$) and has measure 0. Let N be such that $N \subseteq M, M \in \overline{\mathcal{F}}$ with $\mu(M) = 0$. Since $N = \underbrace{N}_{\in\mathcal{F}} \cup \underbrace{\varnothing}_{\in\mathcal{F}}$ it means $N \in \overline{\mathcal{F}}$. Moreover

$$\overline{\mu}(N) = \overline{\mu}(N \cup \varnothing) = \mu(\varnothing) = 0.$$

Wow, this was useless.

1.1.2 Exercise class 2

Revise with Kotatsu!

→ **Independence of random variables**:

> **Definition 1.1.1**
>
> Given two random variables X_1 and X_2 we say that they are **independent** if $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

Two σ-algebras are independent if $\forall\, A_1 \in \sigma(X_1)$ and $A_2 \in \sigma(X_2)$ we have

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2).$$

An equivalent definition of independence for real-valued random variables is that

$$\mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2) = \mathbb{P}(X_1 \leqslant x_1) \cdot \mathbb{P}(X_2 \leqslant x_2) \qquad \forall\, x_1, x_2 \in \mathbb{R}.$$

To see the equivalence note that $\{(-\infty, x], x \in \mathbb{R}\}$ is a p-system for the Borel σ-algebra. Setting $C_i = \{\omega : X_1 \in (-\infty, x], x \in \mathbb{R}\}$ we have $\sigma(C_i) = \sigma(X_i)$ for $i = 1, 2$. The equivalence follows.

→ **Uniform integrability for a family of random variables**:

> **Definition 1.1.2**
>
> K is uniformly integrable if
>
> $$\lim_{b \to \infty} \sup_{X \in K} \mathbb{E}\left[|X| \mathbb{1}_{\{(|X| > b)\}}\right] = 0.$$

> **Theorem 1.1.1**
>
> K is uniformly integrable if and only if it is $L^1 -$ *bounded* and $\forall\, \varepsilon > 0\, \exists \delta > 0$ such that for $\forall$ event H we have
>
> $$\mathbb{P}(H) \leqslant \delta \implies \sup_{X \in K} \mathbb{E}\left[|X| \mathbb{1}_H\right] \leqslant \varepsilon.$$

This is known as the "$\varepsilon - \delta$ characterization" and it means that on every small set the integrals of the X are uniformly small. The "tails" of these random variables behave uniformly well, especially on small-probability sets.

→ **Transition kernels**:

> **Definition 1.1.3**
>
> Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. We say that
>
> $$K : E \times \mathcal{F} \mapsto \overline{\mathbb{R}}_+$$
>
> is a transition kernel from $(E, \mathcal{E})$ to $(F, \mathcal{F})$ if
>
> a) the map $X \mapsto K(x, B)$ (from E to $\overline{\mathbb{R}}_+$) is $\mathcal{E}$-measurable for $\forall\, B \in \mathcal{F}$;
>
> b) the map $B \mapsto K(x, B)$ (from $\mathcal{F}$ to $\overline{\mathbb{R}}_+$) is a measure on $(F, \mathcal{F})$ for every $X \in E$.

Exercise 1

Choose 2 numbers uniformly between 0 and 1, denoted by X_1 and X_2. Consider both the case with reimmission and without reimmission. Are X_1 and X_2 independent?

First we have to specify the probability space and view X_1 and X_2 as random variables, so the concept of independence makes sense.

Case with reimmission

Let

$$\Omega = [0,1] \times [0,1] \qquad \mathcal{F} = \mathcal{B}([0,1] \times [0,1]) \qquad \mathbb{P} = \lambda$$

with ($\omega \in \Omega \to \omega = (\omega_1, \omega_2)$) and

$$\begin{aligned} X_1 : \Omega &\to \mathbb{R} \\ \omega &\mapsto X_1(\omega) := \omega_1 \\ X_2 : \omega &\to \mathbb{R} \\ \omega &\mapsto X_2(\omega) := \omega. \end{aligned}$$

X_i are random variables because they are measurable. Indeed, for $A \in \mathcal{B}(\mathbb{R})$ (e.g. for $i = 1$) we have that

$$\begin{aligned} X_1^{-1}(A) &= \{\omega \in \Omega : X_1(\omega) \in A\} \\ &= \{\omega \in \Omega : \omega_1 \in A\} \\ &= \{\omega \in \Omega : \omega_1 \in A, \omega_1 \in [0,1]\} \in \mathcal{B}([0,1] \times [0,1]) \end{aligned}$$

since both A and $[0,1]$ belong in $\mathcal{B}([0,1])$. The same can be said for $i = 1$. To check the dependence (or independence) we compute the joint distribution of (X_1, X_2):

$$\mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2) = \begin{cases} 1 & \text{if } x_1, x_2 > 1 \\ x_1 x_2 & \text{if } 0 \leqslant x_1, x_2 \leqslant 1 \\ x_1 & \text{if } 0 \leqslant x_1 \leqslant 1,\ x_2 > 1 \\ x_2 & \text{if } 0 \leqslant x_2 \leqslant 1,\ x_1 > 1 \\ 0 & \text{else.} \end{cases}$$

The situation can be summarized in the following drawings:

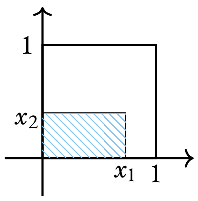


Figure 1.2: Case with $0 \leqslant x_1, x_2 \leqslant 1$

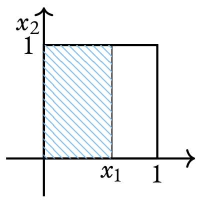


Figure 1.3: Case with $0 \leqslant x_1 \leqslant 1,\ x_2 > 1$

Calculate the marginal distribution as well:

$$\mathbb{P}(X_i \leqslant x_1) = \begin{cases} 1 & \text{if } x_i > 1 \\ x_i & \text{if} - \leqslant x_i \leqslant 1 \\ 0 & \text{else} \end{cases} \qquad \text{for } i = 1, 2.$$

So we can basically see that

$$\mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2) = \mathbb{P}(X_1 \leqslant x_1) \cdot \mathbb{P}(X_2 \leqslant x_2)$$

and therefore we have that X_1 and X_2 are independent.

Case without reimmission

In this case we must change the probability space to account for this difference. Let

$$\Omega = \{\omega \in [0,1] \times [0,1] \text{ s.t. } \omega_1 \neq \omega_2\} \qquad \mathcal{F} = \mathcal{B}([0,1] \times [0,1],\ x_1, x_2) \qquad \mathbb{P} = \lambda$$

The random variables are the same as before ($X_i(\omega),\ i = 1, 2$) and so $\mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2)$ is as above because $\mathbb{P}(x_1 = x_2) = 0$ since we are using fucking Lebesgue measure as probability measure. So also in this case X_1 and X_2 are independent. What the fuck?? And why can't I seem to be able to have centered captions anymore?

Exercise 2

[redacted]

Set

$$\lambda = \lambda_{(0,1)}.$$

ⓐ First we show that μ is a measure since it is the composition of a Borel function (q^{-1}, since q is Borel) with a measure. To show it is a probability measure we calculate

$$\begin{aligned}\mu(\overline{\mathbb{R}}) &= \lambda(q^{-1}(\overline{\mathbb{R}}))\\ &= \lambda(\{x \in [0,1] : q(x) \in \overline{\mathbb{R}}\})\\ &= \lambda(0,1) = 1.\end{aligned}$$

ⓑ Consider the following:

$$\begin{aligned}\mu((\infty, x]) &= \lambda(q^{-1}((-\infty, x]))\\ &= \lambda(\{u \in [0,1] : q(u) \in (-\infty, x]\})\\ &= \underbrace{1 - \lambda(\{u \in [0,1] : q(u) \notin (-\infty, x]\})}_{\lambda \text{ is a probability measure!}}\\ &= 1 - \lambda(\{u \in [0,1] : q(u) > x\})\\ &= \underbrace{1 - \lambda(\{u \in [0,1] : c(x) \leqslant u\})}_{\text{notice that } u \in (0,1): q(u) > x \iff u \in [0,1]: c(x) \leqslant u}\\ &= 1 - \lambda([c(x), 1))\\ &= 1 - (1 - c(x))\\ &= c(x).\end{aligned}$$

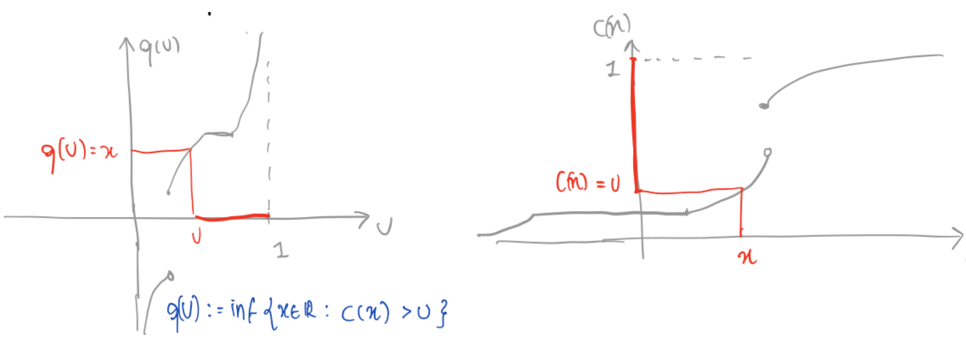


Figure 1.4: Ok but tikz are hard and long... ;) that is WHY i use them.

ⓒ We construct μ as in the previous two points. That μ is unique thanks to the theorem on uniqueness of measures (since μ is defined on the p-system $(-\infty, x]$). Vice versa, given the probability measure μ on

$\left(\overline{\mathbb{R}}, \mathcal{B}\left(\overline{\mathbb{R}}\right)\right)$ we define

$$\begin{aligned} c :& \mathbb{R} \to [0,1] \\ & x \mapsto c(x) = \mu([\infty, x]). \end{aligned}$$

We notice that c is a cumulative distribution function. Indeed:

- c is increasing (it is non-decreasing but whatever):

  $$\forall x_1 \leqslant x_2 \qquad c(x_1) = \mu[\infty, x_1] \leqslant \mu[-\infty, x_2] = c(x_2)$$

  since $[-\infty, x_1] \subseteq [-\infty, x_2]$.
- c is right continuous: let $x \in \mathbb{R}$ be fixed and let $A_n = [-\infty, x + \frac{1}{n}]$, $\forall n \geqslant 1$. Clearly $A_n \in \mathcal{B}(\mathbb{R})$ and $A_n \searrow a := [-\infty, x]$. Then by properties of measures we know that

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

  and since $\mu(A_n) = c(x + \frac{1}{n})$ and $\mu(A) = c(x)$ we conclude

  $$\lim_{n\to\infty} c\left(x + \frac{1}{n}\right) = c(x).$$

Exercise 3

[redacted]

We know that a random variable is uniformly integrable if $\lim_{b\to\infty} \sup_{X\in K} \mathbb{E}\left[|X|\mathbb{1}_{\{(|X|>b)\}}\right] = 0$. So, since we are operating over a finite set, we can swap lim and max to conclude that this is equivalent to showing

$$\max_{i=1,\dots,n} \lim_{b\to\infty} \mathbb{E}\left[|X_i|\mathbb{1}_{\{X>b\}}\right] = 0 \tag{$*$}$$

Let's fix i. We have

$$\lim_{b\to\infty} \mathbb{E}\left[|X_i|\mathbb{1}_{\{|X_i|>b\}}\right] = \mathbb{E}\underbrace{\left[\lim_{b\to\infty} |X_i|\mathbb{1}_{\{|X_i|>b\}}\right]}_{\text{this is 0 because } |X_i| \in L^1.} = 0. \tag{$**$}$$

Here we can swap expectation and limit thanks to the dominated convergence theorem, since $|X_i|$ (weakly) dominates $|X_i|\mathbb{1}_{\{|X_i|>b\}}$. Moreover, $\lim_{b\to\infty} |X_i|\mathbb{1}_{\{|X_i|>b\}}$ is 0 because being in L^1 means having "low" tails. So $**$ holds for all i and therefore $*$ holds.

Exercise 4

[redacted]

Let $K := \{X_i - X\}$. We use the $\varepsilon - \delta$ characterization for uniform integrability. So we should check that:

1. $\sup_{Y\in K} \mathbb{E}[Y] < \infty$;
2. $\forall \varepsilon \exists \delta$: if $F \in \mathcal{F}$ with $\mathbb{P}(F) \leqslant \delta$ then $\sup_{Y\in K} \mathbb{E}\left[|Y|\mathbb{1}_F \leqslant \varepsilon\right]$.

1. We have

$$\begin{aligned} \sup_{Y\in K} \mathbb{E}|Y| &= \sup_i \mathbb{E}\,|X_i - X| \\ &\leqslant \sup_i \left(\mathbb{E}\,|X_i| + \mathbb{E}\,|X|\right) \\ &\leqslant \underbrace{\mathbb{E}\,|X|}_{<\infty \text{ bc U.I.}} + \sup_i \underbrace{\mathbb{E}\,|X_i|}_{<\infty \text{ bc } \in L^1} < \infty \end{aligned}$$

The third step is simple triangle inequality...

2. Since $\{X_i\}$ is uniformly integrable, for any ε_1 there exists a δ_1 such that if $\mathbb{P}(F) < \delta_1$ then $\sup_i \mathbb{E}[|X_i|\mathbb{1}_F] \leqslant \varepsilon_1$. On the other hand, the fact that X is in L^1 means that $\{X\}$ is uniformly integrable because $\lim_{b\to\infty} \mathbb{E}\left[|X|\mathbb{1}_{\{|X|>b\}}\right]$, so for each ε_2 we can find a δ_2 such that if $\mathbb{P}(F) \leqslant \delta_2$ then $\sup_i \mathbb{E}[|X|\mathbb{1}_F] \leqslant \varepsilon_2$. Then, for any $\varepsilon > 0$, choose $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$. We need to find our δ so we say $\delta = \min\{\delta_1, \delta_2\}$. In this way, for $F \in \mathcal{F}$ such that $\mathbb{P}(F) \leqslant \delta$ (which means that is also $\leqslant \delta_1$ and $\leqslant \delta_2$) we have

$$\begin{aligned}
\sup_{Y\in K} \mathbb{E}[|Y|\mathbb{1}_F] &= \sup_i \mathbb{E}[|X_i - X|\mathbb{1}_F] \\
&\leqslant \sup_i \mathbb{E}[|X_i|\mathbb{1}_F] + \mathbb{E}[|X|\mathbb{1}_F] \\
&\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Exercise 5

[redacted]

① Let's find $\mathbb{P}_X(A)$ for each $A \in \mathcal{B}(\mathbb{R}_+)$.

$$\begin{aligned}
\mathbb{P}_X(A) = \mathbb{P}_Z(A\times\mathbb{R}) &= \int_{A\times\mathbb{R}} \mathbb{P}_Z(\mathrm{d}x,\mathrm{d}y) \\
&= \int_A\int_{\mathbb{R}} \lambda\frac{e^{-\lambda x}}{\sqrt{2\pi x}} e^{\frac{-y^2}{2x}}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_A \lambda e^{-\lambda x} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}}\,\mathrm{d}y}_{\text{density of } \mathcal{N}(0,x)}\,\mathrm{d}x \\
&= \int_A \lambda e^{-\lambda x}\cdot 1\cdot \mathrm{d}x.
\end{aligned}$$

So $\mathbb{P}_X(\mathrm{d}x) = \lambda e^{-\lambda x} \sim \mathsf{Exp}(\lambda)$.

② Since we can now write

$$\mathbb{P}_Z(\mathrm{d}x,\mathrm{d}y) = \mathbb{P}_X(\mathrm{d}x)\frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}}\,\mathrm{d}y$$

then we see that

$$K(x,\mathrm{d}y) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}}\,\mathrm{d}y$$

or, more formally, for $B \in \mathcal{B}(\mathbb{R})$ we have

$$K(x,B) = \int_B \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}}\,\mathrm{d}y.$$

Let's now check that $K(x,B)$ is a Markov transition kernel:

- it is a transition kernel because
  (a) for $B \in \mathcal{B}(\mathbb{R})$ fixed, $K(\cdot, B)$ is positive and measurable (it is even continuous);
  (b) for $x \in \mathbb{R}_+$ fixed, $K(x, \cdot)$ is a measure on $\mathcal{B}(\mathbb{R})$: in particular by inspection we see that it is a Gaussian measure with mean 0 and variance x.
- It is a Markov kernel because $K(x, \mathbb{R}) = 1 \, \forall x$ since $K(x, \cdot)$ is a probability measure $\forall x \in \mathbb{R}$.

3. Let's find $\mathbb{P}_Y(B)$ for every $B \in \mathcal{B}(\mathbb{R})$.

$$\begin{aligned}\mathbb{P}_B(B) = \mathbb{P}_Z(\mathbb{R}_+ \times B) &= \int_{\mathbb{R}_+ \times B} \lambda e^{-\lambda x} \frac{e^{\frac{y^2}{2x}}}{\sqrt{2\pi x}} \, \mathrm{d}x \, \mathrm{d}y \\ &= \int_B \left(\int_{\mathbb{R}_+} \lambda \frac{e^{-\lambda x}}{\sqrt{2\pi x}} e^{\frac{-y^2}{2x}} \, \mathrm{d}x \right) \mathrm{d}y\end{aligned}$$

Now use the hint with $\gamma = \lambda$, $\beta = \frac{y^2}{2}$ to get

$$\begin{aligned}\frac{1}{\sqrt{2\pi}} \lambda \int_{\mathbb{R}_+} e^{-\lambda x} \frac{1}{\sqrt{x}} e^{\frac{y^2}{2x}} \, \mathrm{d}x &= \frac{\lambda}{\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{\sqrt{\lambda}} \cdot e^{-2\sqrt{\lambda \cdot \frac{y^2}{2}}} \\ &= \sqrt{\frac{\lambda}{2}} e^{-\sqrt{2\lambda} \cdot |y|}.\end{aligned}$$

So our function becomes

$$\int_B \frac{1}{2} \sqrt{2\lambda} e^{-\sqrt{2\lambda |y|}} \, \mathrm{d}y.$$

Which is the probability density function of a two-sided exponential random variable. Thus $\mathbb{P}_Y(\mathrm{d}y) = \frac{1}{2}\sqrt{2\lambda} e^{-\sqrt{2\lambda}|y|} \, \mathrm{d}y$.

4. We observe that $\mathbb{P}_X \cdot \mathbb{P}_Y \neq \mathbb{P}_Z$ so X and Y are not independent.

1.1.3 Exercise class 3

Exercise 1

[redacted]

The vector is Gaussian, since it is a linear combination of Gaussian random variables.

$$\begin{aligned}W &:= (X+Y, X-Y) = (W_1, W_2) \\ Z &:= (X, Y).\end{aligned}$$

We write W in terms of Z:

$$W = \mu + \mathbf{A}Z \qquad \text{for some } \mu = (\mu_1, \mu_2) \text{ and } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Clearly $\mu = (0, 0)$ because

$$\begin{aligned}\mathbb{E}W_1 &= \mathbb{E}X + Y = \mathbb{E}X + \mathbb{E}Y = 0 \\ \mathbb{E}W_2 &= \mathbb{E}X - Y = \mathbb{E}X - \mathbb{E}Y = 0\end{aligned}$$

So now we have

$$\begin{aligned}W_1 &= X + Y = a_{11}X + a_{12}Y \\ W_2 &= X - Y = a_{21}X + a_{22}Y\end{aligned} \implies \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Let's calculate the covariance matrix Γ of W, given by

$$\Gamma = \mathbf{A}\mathbf{A}^\top = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Since $\det(\Gamma) = 4 \neq 0$ then Γ has full rank so the random variable W has density given by

$$f_W(W_1, W_2) = \frac{1}{(2\pi)^{\frac{2}{2}}} \cdot \frac{1}{\sqrt{\det(\Gamma)}} e^{-\frac{1}{2}\left(W_1 \; W_2 \right) \Gamma^{-1} \binom{W_1}{W_2}}.$$

Compute Γ^{-1}:

$$\Gamma^{-1} = \frac{1}{\det(\Gamma)} \begin{bmatrix} 2 & -0 \\ -0 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

And now plug the results inside the density formula:

$$\begin{aligned} f_W(W_1, W_2) &= \frac{1}{(2\pi)^{\frac{2}{2}}} \cdot \frac{1}{\sqrt{\det(\Gamma)}} e^{-\frac{1}{2}\left(W_1 \; W_2 \right) \Gamma^{-1} \binom{W_1}{W_2}} \\ &= \frac{1}{(2\pi)^{\frac{2}{2}}} \cdot \frac{1}{2} e^{-\frac{1}{2}\left(W_1 \; W_2 \right) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \binom{W_1}{W_2}} \\ &= \frac{1}{4\pi} e^{-\frac{1}{2}\left(\frac{1}{2}W_1^2 + \frac{1}{2}W_2^2\right)}. \end{aligned}$$

So this is the density of the random variable $W = (X + Y, X - Y)$.

Remark

Since $\Gamma_{12} = \Gamma_{21} = 0$ we have $W_1 \perp\!\!\!\perp W_2$. We can see this also because the joint distribution factorizes in the product of the marginal distributions.

$$f_W(W_1, W_2) = \underbrace{\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{1}{4}W_1^2}}_{\sim \mathcal{N}(0,2)} \cdot \underbrace{\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{1}{4}W_2^2}}_{\sim \mathcal{N}(0,2)}$$

Exercise 2

[redacted]

Take $B \in \mathcal{B}(\mathbb{R})$. We have

$$\begin{aligned} \mu_Z(B) &= \mathbb{P}(Z \in B) \\ &= \mathbb{P}(X + Y \in B) \\ &= \mathbb{E}\left[\mathbb{1}_{\{X+Y \in B\}}\right]. \end{aligned}$$

We know that $\mu_{(X,Y)} = \mu_X \cdot \mu_Y$ since $X \perp\!\!\!\perp Y$, so whe have that

$$\begin{aligned} \mu_Z(B) &= \mathbb{P}(X + Y \in B) \\ &= \mathbb{E}\left[\mathbb{1}_{\{X+Y \in B\}}\right] \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{X+Y \in B\}} \mu_{(X,Y)}(\mathrm{d}x, \mathrm{d}y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{X+Y \in B\}} \mu_X(\mathrm{d}x) \mu_Y(\mathrm{d}y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(x + y) \mu_X(\mathrm{d}x) \mu_Y(\mathrm{d}y). \end{aligned} \tag{$*$}$$

Now consider

$$\mathbb{1}_B(x+y) = \mathbb{1}_{\{B-x\}}(y) = \begin{cases} 1 & \text{if } x+y \in B \iff y \in \{B-x\} \\ 0 & \text{if } x+y \notin B \iff y \notin \{B-x\} \end{cases}.$$

Now consider $*$:

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(x+y)\mu_X(\mathrm{d}x)\mu_Y(\mathrm{d}y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{B-x\}}(y)\mu_X(\mathrm{d}x)\mu_Y(\mathrm{d}y) \\ &= \int_{\mathbb{R}} \underbrace{\left(\int_{\mathbb{R}} \mathbb{1}_{\{B-x\}}(y)\mu_Y(\mathrm{d}y) \right)}_{\int_{\{B-x\}} \mu_Y(\mathrm{d}y) = \mu_Y(B-x)} \mu_X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \mu_Y(B-x)\mu_X(\mathrm{d}x) \end{aligned}$$

Exercise 3

[redacted]

Take $B \in \mathcal{B}(\mathbb{R})$. The law of Z, μ_Z, is such that

$$\mu_Z(B) = \mathbb{P}(Z \in B)$$

and we want to show that indeed

$$\mathbb{P}(Z \in B) = \int_B f_Z(z)\,\mathrm{d}z.$$

Using what we found in exercise 2 we have that

$$\begin{aligned} \mu_Z(B) &= \int_{\mathbb{R}} \mu_Y(B-x)\mu_X(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \int_{B-x} f_Y(y)\,\mathrm{d}y f_X(x)\,\mathrm{d}x. \end{aligned}$$

If $y \in \{B-x\}$ it means that $y = t - x$ for some $t \in B$. So we can make a change of variable $y = t - x$:

$$\int_{\{B-x\}} f_Y(y)\,\mathrm{d}y = \int_B f_Y(t-x)\,\mathrm{d}t$$

which implies

$$\begin{aligned} \mu_Z(B) &= \int_{\mathbb{R}} \int_B f_Y(y)(t-x)\,\mathrm{d}t f_X(x)\,\mathrm{d}x \\ &= \int_B \left(\int_{\mathbb{R}} f_Y(t-x) f_X(x)\,\mathrm{d}x \right) \mathrm{d}t \\ &= \int_B f_Z(t)\,\mathrm{d}t. \end{aligned}$$

Exercise 4

[redacted]

Start from this set:

$$\begin{aligned} A &= \left\{ \omega : \lim_n X_n(\omega) \text{ exists} \right\} \\ &= \left\{ \omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega) \right\}. \end{aligned}$$

We know that $\mathbb{P}(A) = 1$ because we know that X_n always converges and we want to show that $A \in \tau$ because this (together with independence) would allow us to apply Kolmogorov's 0-1 law's corollary that tells us that if an event is part of a tail σ-algebra then it is a constant[3]. Let's show that $\limsup_n X_n$ is τ-measurable.

$$\limsup_n X_n = \inf_n \underbrace{\sup_{m \geqslant n} X_m}_{\text{is measurable with respect to } \bigvee_{m\geqslant n} \sigma(X_m) = \tau_n}$$

Thus $\inf_n \sup_{m \geqslant n} X_m$ is measurable with respect to

$$\bigcap_{n\geqslant 1} \bigvee_{m \geqslant n} \sigma(X_m) = \bigcap_{n \geqslant 1} \tau_n = \tau$$

so $\limsup_n X_n$ is τ-measurable. Similarly for lim inf we get that is is τ-measurable so $A \in \tau$.
We recall that X is such that

$$X(\omega) = \lim_{n \to \infty} X_n(\omega) \qquad \forall\, \omega \in A \in \tau.$$

Introduce

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

We introduced this new random variable because we are technically working with the *limit* of X_n (which is a sequence) and measurability of limits is not a beautiful thing to handle. By definition $X = Y$ $\mathbb{P}-$ a.s. (because $\mathbb{P}(A^c) = 0$) and Y is τ-measurable because $A \in \tau$. This implies that X is τ-measurable. By corollary of the Kolmogorov's 0-1 law, $X = c$ $\mathbb{P}-$ a.s. for some $c \in \mathbb{R}$.

Exercise 5

[redacted]

We guess yes because why not. Let $\varepsilon > 0$. Take the set

$$B_n := \{\omega : X_n(\omega) \geqslant n\varepsilon\}.$$

This gives us

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{E}\left[\mathbb{1}_{B_n}\right] \\ &= \int_{n\varepsilon}^{+\infty} e^{-x}\, \mathrm{d}x = e^{-n\varepsilon}. \end{aligned}$$

Thus $\sum_n \mathbb{P}(B_n) = \sum_n e^{-n\varepsilon} < \infty$. By Borel-Cantelli 1 we have

$$\mathbb{P}(\{B_n \text{ i.o.}\}) = 0.$$

Take the complement:

$$\mathbb{P}(\{B_n^c \text{ f.o.}\}) = 1.$$

For every $\omega \in \{B_n^c \text{ f.o.}\}$ there exists $m(\omega, \varepsilon)$ such that $\forall\, n > m(\omega,\varepsilon)$ then $\underbrace{X_n(\omega) < n\varepsilon}_{B_n^c}$. Set

$$M(\omega, \varepsilon) := \max\left\{X_1(\omega), X_2(\omega), \dots, X_{m(\omega,\varepsilon)}(\omega)\right\}$$

so that for $\forall\, n$ we have

$$\max\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\} \leqslant M + n\varepsilon$$

so

$$\frac{1}{n}\max\{X_1(\omega), X_2(\omega), \underbrace{\dots}_{X_m(\omega), X_{m+1}(\omega), \dots}, X_n(\omega)\} \leqslant \frac{M(\omega,\varepsilon)}{n} + \varepsilon.$$

[3] If you have already read the theory then you know how this thing caused me a mental breakdown.

Check the limit:

$$\begin{aligned}\limsup_n \frac{1}{n}\max\{X_1(\omega),\dots,X_n(\omega)\} &\leqslant \limsup_n \frac{M(\omega,\varepsilon)}{n}+\varepsilon\\ &= \lim_n \frac{M(\omega,\varepsilon)}{n}+\varepsilon=\varepsilon.\end{aligned}$$

Now we take the limit as $\varepsilon \to 0$:

$$\limsup_n \frac{1}{n}\max\{X_1(\omega),\dots,X_n(\omega)\} \leqslant \lim_{e\to 0}\varepsilon = 0$$

Thus, using the fact that X_i are positive, we conclude

$$\lim_{n\to\infty}\frac{1}{n}\max\{X_1(\omega),\dots,X_n(\omega)\}=0.$$

Since this holds for $\forall\, \omega \in \{B_n^x \text{ ev.}\}$ and $\mathbb{P}(\{B_n^c \text{ ev.}\})=1$ we have

$$\frac{1}{n}\max\{X_1,\dots,X_n\}\to 0 \qquad \text{a.s.}$$

Exercise 6

[redacted]

Guess: S_n diverges to $+\infty$ because the random variables X_i are such that $\{X_i > \delta\}$ has positive probability. We claim

$$\limsup_n\{X_i \geqslant \delta\} \subset \{\lim_{n\to\infty} S_n = \infty\} \tag{❀}$$

Let's prove it.

$$\begin{aligned}\omega \in \limsup_i\{X_i\geqslant\delta\} &= \bigcap_{i\geqslant 1}\bigcup_{m\geqslant i}\{X_m\geqslant\delta\}\\ &=\{X_i\geqslant\delta \;\text{ i.o.}\}.\end{aligned}$$

This means that $\exists$ a subsequence i_k such that $X_{i_k}(\omega)\geqslant\delta$. Summing over k we get

$$\sum_{k=1}^{\infty}X_{i_k}(\omega)\geqslant\sum_{k=1}^{\infty}X_i(\omega)=\infty.$$

So taking the limit we obtain

$$\begin{aligned}\lim_{n\to\infty}S_n^{(\omega)} &= \lim_{n\to\infty}\sum_{i=1}^{n}X_i(\omega)\\ &=\sum_{i=1}^{\infty}X_i(\omega)\\ &\geqslant \sum_{k=1}^{\infty}X_{i_k}(\omega)=\infty \implies \lim_{n\to\infty}S_n(\omega)=\infty\end{aligned}$$

hence ❀[4] holds. We now apply Borel-Cantelli 2.

[4]Yes, I have ONLY NOW discovered dingbats. My future endeavours will have more of them, do not worry. Even if they look like shit with this font, on god.

Revise with Kotatsu!

Proposition 1.1.1

"Divergence". Let $(A_n)_n$ be a sequence of pairwise independent events.

$$\text{If } \sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty \text{ then } \underbrace{\sum_{n=1}^{\infty} \mathbb{1}_{A_n}}_{=\mathbb{P}(\{A_n \text{ i.o.}\})} = \infty \quad \text{a.s.}$$

An equivalent formulation of Borel-Cantelli 2 is:

Proposition 1.1.2

Take a sequence of Bernoulli random variables B_n defined as

$$B_n = \begin{cases} 1 & \text{on } A_n \\ 0 & \text{on } A_n^c \end{cases} \qquad (\text{or, more simply, } B_n = \mathbb{1}_{A_n})$$

so that B_n is a Bernoulli random variable and we have

$$\mathbb{E}B_n = \mathbb{P}(A_n).$$

The B_n are pairwise independent.

$$\text{If } \sum_{n=1}^{\infty} \mathbb{E}B_n = \infty \text{ then } \sum_{n=1}^{\infty} B_n = \infty \quad \text{a.s.}$$

Remember that $\limsup_n \{X_n = 1\}$ is the same as saying that $\{X_n = 1 \text{ i.o.}\}$.

In our case we have $B_i = \{X_i \geqslant \delta\}$. We know that $\mathbb{P}(B_i) \geqslant \epsilon$ by assumption, so

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i) = \infty.$$

Moreover, the B_i are independent. Hence we can apply BC-2 and say that

$$\mathbb{P}(\{X_i \geqslant \delta\} \text{ i.o.}) = 1 \iff \mathbb{P}(\limsup_i \{X_i \geqslant \delta\}) = 1.$$

By ❀ we have

$$1 = \mathbb{P}(\limsup\{X_i \geqslant \delta\}) \leqslant \mathbb{P}(\lim_n S_n = \infty)$$

hence

$$\mathbb{P}(\lim_n S_n = \infty) = 1$$

which means that we have almost sure divergence of the sum.

1.1.4 Exercise class 4

Revise with Kotatsu!

Hi bitch. Remember all the types of convergence? Me neither!

☞ **Convergence in L^p**:

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

if $X_n, X \in L^p$ and

$$\mathbb{E}|X_n - X|^p \to 0 \text{ as } n \to \infty.$$

A necessary (but not sufficient...) condition for convergence in L^p is that

$$\mathbb{E}|X_n|^p \to \mathbb{E}|X|^p.$$

☞ **Convergence in probability**:

$$X_n \xrightarrow{\mathbb{P}} X$$

if $\forall\, \varepsilon > 0$ we have

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \qquad \text{as } n \to \infty.$$

☞ **Almost sure convergence**:

$$X_n \xrightarrow{\text{a.s.}} X$$

if $\exists \Omega' \subset \Omega$ with $\mathbb{P}(\Omega') = 1$ such that

$$X_n(\omega) \to X(\omega) \qquad \forall\, \omega \in \Omega'.$$

☞ **Convergence in distribution**:

$$X_n \xrightarrow{\text{d}} X$$

if $F_n(x) \to F(x)$ for any point x such that $F(x)$ is continuous. Here F_n and F denote the cumulative distribution functions of X_n and X.

✎ **Link between different convergence modes**:

$$\begin{array}{ccccc} X_n \xrightarrow{L^p} X & \Longrightarrow & X_n \xrightarrow{\mathbb{P}} X & \Longleftarrow & X_n \xrightarrow{\text{a.s.}} X \\ & & \Downarrow & & \\ & & X_n \xrightarrow{\text{d}} X. & & \end{array}$$

✎ **Properties:**

➠ if $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} Y$ then $X = Y$ a.s.;

➠ if $X_n \xrightarrow{L^p} X$ and $\lim_n X_n$ exists a.s., then $X_n \xrightarrow{\text{a.s.}} X$.

Exercise 1

[redacted]

I will now stop[5] with the stupid LaTeX symbols.

➶ L^p convergence: note that

$$\mathbb{E}|X_n|^p = \mathbb{E}X_n^p = 1^p \cdot \frac{1}{n} + 0^p \cdot (1 - \frac{1}{n}) = \frac{1}{n} \xrightarrow{n\to\infty} = 0, \qquad \forall\, p > 0.$$

Hence the candidate limit is $X = 0$. Let's check

$$\mathbb{E}|X_n - 0|^p = \mathbb{E}|X_n|^p = \frac{1}{n} \to 0$$

so

$$X_n \xrightarrow{L^p} X, \qquad \forall\, p \geqslant 1.$$

➶ Almost sure convergence: we note that

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty$$

[5]This is a lie.

hence by BC-2 we have that

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

This implies that we cannot have $X_n \xrightarrow{\text{a.s.}} 0$.

Exercise 2

[redacted]

We must figure out the distribution of Y_n. for $x \in (0,1)$ we have:

$$\begin{aligned}
\mathbb{P}(Y_n > x) &= \mathbb{P}(X_1 > x, X_2 > x, \dots, X_n > x) \\
&= \mathbb{P}(X_1 > x) \cdot \mathbb{P}(X_2 > x), \cdot \dots \cdot \mathbb{P}(X_n > x) \\
&= (1-x)^n.
\end{aligned}$$

Therefore,

$$\mathbb{P}(Y_n \leqslant x) = 1 - (1-x)^n = \int_0^x f_{Y_n}(x)\,\mathrm{d}y$$

so taking the derivative we have:

$$f_{Y_n}(y) = \begin{cases} n(1-y)^{n-1} & \text{for } y \in (0,1) \\ 0 & \text{else.} \end{cases}$$

Let's check if $\mathbb{E}|Y_n|^p$ converges for some p. We start with $p = 1$.

Revise with Kotatsu!

Remember integration by parts? Lmao.

$$\int fg' = fg - \int f'g$$

$$\begin{aligned}
\mathbb{E}|Y_n| &= \mathbb{E}Y_n \\
&= \int_0^1 \underbrace{y}_{f}\, \underbrace{n(1-y)^{n-1}}_{g'}\,\mathrm{d}y \\
&= y\left(-(1-y)^n\right)\Big|_0^1 - \int_0^1 -(1-y)^n\,\mathrm{d}y \\
&= 0 + -\frac{(1-y)^{n+1}}{n+1}\Big|_0^1 \\
&= 0 + \frac{1}{1+n}
\end{aligned}$$

Hence $\mathbb{E}|Y_n| \xrightarrow[n\to\infty]{} 0$, so the candidate Y such that $\mathbb{E}|Y_n - Y| \to 0$ is 0. This implies $\mathbb{E}|Y_n - 0| = \mathbb{E}|Y_n| \to 0$ so that we have $Y_n \xrightarrow{L^1} 0$. What about convergence in L^p though? First check a.s. convergence.
By definition of Y_n we have $Y_{n+1}(\omega) \leqslant Y_n(\omega)$ for almost all ω; this means that Y_n is monotone $\mathbb{P}$-almost surely. This implies that Y_n has a limit $\mathbb{P}$-almost surely, say Y (moreover, $Y_n(\omega) \geqslant 0$ hence $Y(\omega) \geqslant 0$ $\mathbb{P}$-a.s.).
By uniqueness of the limit[6] we have that $Y = 0$. Hence we conclude $Y_n \xrightarrow{\text{a.s.}} 0$.
To check for convergence in L^p, if we can apply the dominated convergence theorem then we have:

$$\lim_{n\to\infty} \mathbb{E}|Y_n - Y|^p = \mathbb{E}\left[\lim_{n\to\infty} |Y_n|^p\right] = 0.$$

Here dominated convergence works because $0 \leqslant X_k \leqslant 1$ hence $|Y_n| \leqslant 1$ and $|Y_n|^p \leqslant 1$ where 1 is integrable. Moreover, the a.s. limit $Y = 0$ is also integrable.

[6] $Y_n \xrightarrow{L^1} 0 \implies Y_n \xrightarrow{\mathbb{P}} 0; Y_b \xrightarrow{\text{a.s.}} Y \implies Y = 0\ \mathbb{P} - \text{a.s.}$

Exercise 3

[redacted]

Let's calculate $\mathbb{E}|X_n|^p$ for some $p \geqslant 1$.

$$\mathbb{E}|X_n|^p = n^p \cdot \mathbb{P}\left(\left[0, \frac{1}{n}\right]\right) + 0^p \cdot \mathbb{P}\left(\left[\frac{1}{n}, 1\right]\right) = n^p \cdot \frac{1}{n} = n^{p-1}$$

which diverges for $n \to \infty$ if $p > 1$. Thus there can be no $L^p - convergence$ for $p > 1$.
Let's check a.s. convergence. For any $\omega > 0$ $\exists N = N(\omega)$ s.t. $\forall\, n > N$ then $X_n(\omega) = 0$ (since the interval $\left[0, \frac{1}{n}\right]$ shrinks). Hence $X_n(\omega) \to 0\ \forall\, \omega\ \mathbb{P} -$ a.s. given that $\mathbb{P}([0,1]) = 1$.
⚠ It does not converge for $\omega = 0$ but $\{\omega = 0\}$ has Lebesgue measure zero.
Back to L^p-convergence with $p = 1$: if $p = 1$ then $\mathbb{E}|X_n| = 1$ but by uniqueness of the limit we should have $X = 0$ and $\mathbb{E}|X| = 0 \neq 1$ so we do not have convergence in L^1.

Exercise 4

[redacted]

$$\mathbb{E}|X_n| = \mathbb{E}|X_n| = n^2 \cdot \frac{1}{n^2} + 0 = 1.$$

Let us check whether $X_n \xrightarrow{L^p} 1$.

$$\begin{aligned}\mathbb{E}|X_n - 1| &= |n^2 - 1|\frac{1}{n^2} + |0 - 1|\left(1 - \frac{1}{n^2}\right) \\ &= \left|\frac{n^2-1}{n^2}\right| + \left|\frac{n^2-1}{n^2}\right| \\ &= 2\left|\frac{n^2-1}{n^2}\right| \to 2 \neq 0\end{aligned}$$

so there is no L^1 convergence. Since $L^p \implies L^1$ $p \geqslant 1$ there is no L^p convergence either. What about a.s. convergence?
We cannot argue directly as in Exercise 3, but do so through Borel-Cantelli: we know that $\forall\, \varepsilon > 0$, $\mathbb{P}(|X_n| \geqslant \varepsilon) = \frac{1}{n^2}$, hence

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|X_n| \geqslant \varepsilon\right) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Thus by Borel-Cantelli 1 we have

$$\mathbb{P}\left(\limsup_n \{|X_n| \geqslant \varepsilon\}\right) = 0$$

and, taking the complement,

$$\mathbb{P}\left(\liminf_n \{|X_n| < \varepsilon\}\right) = 1.$$

Thus, $\forall\, \varepsilon > 0\ \exists\, N = N(\varepsilon)$ such that $|X_n(\omega)| < \varepsilon$ for $\mathbb{P}$-almost all ω and this implies $X_n \xrightarrow{\text{a.s.}} 0$.

Exercise 5

[redacted]

Revise with Kotatsu!

The exponential distribution has the following cumulative distribution function:

$$\mathbb{P}(X_n \geqslant x) = \int_0^x \frac{1}{k} e^{-\frac{\lambda}{k}} \, \mathrm{d}z = -e^{-\frac{\lambda}{k}} \Big|_0^x = 1 - e^{-\frac{x}{k}}$$

Let's figure out the distribution of Y_n. We have

$$\begin{aligned}
\mathbb{P}(Y_n > y) &= \mathbb{P}(X_1 > y, X_2 > y, \dots, X_n > y) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i > y) \\
&= \prod_{i=1}^{n} e^{-\frac{y}{i}} \\
&= e^{-y \cdot \sum_{i=1}^{n} \frac{1}{i}}
\end{aligned}$$

hence

$$Y_n \sim \mathsf{Exp}\left(\sum_{i=1}^{n} \frac{1}{i}\right).$$

At the limit we would have $f_Y(y) = \infty \cdot e^{-\infty \cdot y}$ but we guess that $e^{-\infty}$ is stronger than ∞ and the limit is zero. Thus we have, for $\forall\, \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - 0| > \varepsilon) = \lim_{n\to\infty} e^{\varepsilon \cdot \sum_{i=1}^{n} \frac{1}{i}} = 0$$

So by definition $Y_n \xrightarrow{\mathbb{P}} 0$.

Exercise 6

[redacted]

First, we calculate the distribution of Y_n:

$$\mathbb{P}(Y_n \geqslant y) = \begin{cases} 1 & \text{if } Y \geqslant 1 \\ \mathbb{P}(X_n < \log n) & \text{if } \frac{1}{n} \leqslant y < 1 \\ 0 & \text{if } y < \frac{1}{n} \end{cases}$$

where

$$\mathbb{P}(X_n < \log n) = 1 - e^{\lambda \cdot \log n} = 1 - e^{\log(n^{-\lambda})} = 1 - n^{-\lambda}.$$

From this, it looks like $pr(Y_n \leqslant y) \to 1$ as $n \to \infty$. The complement of the cumulative distribution function, which we are interested in to compute the limit, is

$$\begin{aligned}
\mathbb{P}(Y_n > y) &= 1 - \mathbb{P}(Y_n \leqslant y) \\
&= \begin{cases} 0 & \text{if } Y \geqslant 1 \\ n^{-\lambda} & \text{if } \frac{1}{n} \leqslant y < 1 \\ 1 & \text{if } y < \frac{1}{n} \end{cases}
\end{aligned}$$

and it looks like Figure 1.5.

- Convergence in $\mathbb{P}$: we have that $\forall\, \varepsilon > 0$ (since $Y_n \geqslant 0$),

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - 0| > \varepsilon) = \lim_{n\to\infty} \mathbb{P}(Y_n > \varepsilon) = 0$$

which implies

$$Y_n \xrightarrow{\mathbb{P}} 0, \qquad \forall\, \lambda > 0.$$

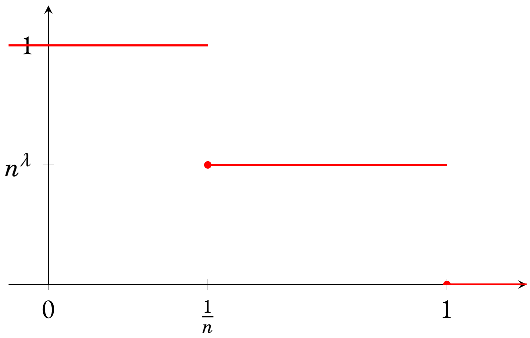


Figure 1.5: It's really just like a cumulative distribution "backwards"...

- Almost sure convergence: We now tackle almost sure convergence and we try to apply Borel-Cantelli. Let $\varepsilon < 1$ and consider

$$\sum_{i=1}^{\infty} \mathbb{P}(Y_n > \varepsilon) = \sum_{n=1}^{N^\star} 1 + \sum_{n=N^\star+1}^{\infty} n^{-\lambda} \leqslant N^\star + \sum_{i=1}^{\infty} n^{-\lambda} < \infty \qquad \text{if } \lambda > 1$$

where $N^\star = \max\left\{n : \varepsilon < \frac{1}{n}\right\}$. Thus, by BC1 (given that $\lambda > 1$),

$$\mathbb{P}\left(\limsup_k \{|Y_n > \varepsilon\}\right) = 0 \iff \mathbb{P}\left(\limsup_k \{|Y_n \leqslant \varepsilon\}\right) = 1$$

i.e., $\forall\, \varepsilon\ \exists N = N(\varepsilon)$ such that $|Y_n(\omega)| \leqslant \varepsilon$ $\mathbb{P}$-almost all ω. This means that

$$Y_n \xrightarrow{\text{a.s.}} 0 \qquad \text{if } \lambda > 1.$$

On the other hand, if $\lambda \leqslant 1$ we instead have

$$\sum_{n=1}^{\infty} \mathbb{P}(Y_n > \varepsilon) = \sum_{n=1}^{N^\star} 1 + \sum_{n=N^\star+1}^{\infty} n^{-\lambda} \leqslant N^\star + \sum_{i=1}^{\infty} n^{-\lambda} = \infty \qquad \text{if } \lambda \leqslant 1.$$

Thus, by BC2, we get (since $X_n \perp\!\!\!\perp \implies Y_n \perp\!\!\!\perp$)

$$\mathbb{P}\left(\limsup_n \{Y_n > \varepsilon\}\right) = 1$$

So the sequence Y_n cannot converge almost surely to 0. By uniqueness f the limit, this implies that $Y_n \stackrel{\text{a.s.}}{\not\longrightarrow} Y$ to any Y.

- Convergence in L^p: let's check the necessary condition

$$\mathbb{E}|Y_n|^p \to \mathbb{E}|Y|^p = 0.$$

Again, by uniqueness of the limit, we must have $Y = 0$ $\mathbb{P}$-a.s. Considering that $X_n \sim \mathsf{Exp}\,(\lambda)$, we have

$$\begin{aligned}
\mathbb{E}|Y_n|^p &= \mathbb{E}Y_n^p \\
&= \frac{1}{n^p}\mathbb{P}(X_n < \log n) + 1 \cdot \mathbb{P}(X_n \geqslant \log n) \\
&= \frac{1}{n^p}\left(1 - e^{\lambda \log n}\right) + e^{-\lambda \log n} \\
&= \frac{1}{n^p}\left(1 - n^{-\lambda}\right) + n^{-\lambda} \to 0 \qquad \text{as } n \to \infty,\ \forall\, \lambda,\ \forall\, \lambda > 0.
\end{aligned}$$

Hence,

$$\mathbb{E}|Y_n - 0|^p = \mathbb{E}|Y_n|^p \to 0$$

that is

$$Y_n \xrightarrow{L^p} 0 \qquad \forall\, p \geqslant 1,\ \forall\, \lambda > 0.$$

1.1.5 Exercise class 5

Revise with Kotatsu!

- **Strong Law Of Large Numbers (SLLN)**: if X_n are pairwise independent and identically distributed as X. If $\mathbb{E}X$ exists ($\pm\infty$ is allowed) then

$$\overline{X}_n \xrightarrow{\text{a.s.}} \mu.$$

- **Weak Law Of Large Numbers (WLLN)**:

  (a) if X_n are pairwise independent and identically distributed as X with $\mathbb{E}X = \mu < \infty$ and $\mathbb{V}\mathrm{ar}\, X = \sigma^2 < \infty$ then

$$\overline{X}_n \to \mu \qquad \text{in } L^2, \text{ in } \mathbb{P} \text{ and a.s.}$$

  (b) If X_n are uncorrelated and $\sum_n \mathbb{V}\mathrm{ar}\left(\frac{X_n}{b_n}\right) < \infty$ for some (b_n) strictly positive and increasing to ∞ then

$$\frac{\sum_{i=1}^n X_i - \mathbb{E}\left(\sum_{i=1}^n X_i\right)}{b_n} \to 0 \qquad \text{in } L^2 \text{ and in } \mathbb{P}.$$

  If the random variables are independent, the convergence holds almost surely.

- **Central Limit Theorem (CLT)** (easy version, whatever this may mean): if X_n are i.i.d. random variables with $\mathbb{E}X_i = \mu < \infty$, $\mathbb{V}\mathrm{ar}\, X = \sigma^2 < \infty$ then

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{d}} \mathrm{N}(0,1).$$

- **Weak Convergence** (for probability measures): a sequence $(\mu_n)_n$ of probability measures converges weakly to μ if for any f bounded and continuous function we have

$$\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu.$$

In this case we write $\mu_n \xrightarrow{\text{weak}} \mu$.

Proposition 1.1.3

$$X_n \xrightarrow{\text{d}} X \iff \mu_n \xrightarrow{\text{weak}} \mu$$

where μ_n is the law of X_n and μ is the law of X.

Proposition 1.1.4

1. $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\text{d}} X$;
2. $X_n \xrightarrow{\text{d}} X$ and $X(\omega) = x$ a.s. $\implies X_n \xrightarrow{\mathbb{P}} X$.

Exercise 1

[redacted]

1. Note that $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^n YX_n = Y\overline{X}_n$. Moreover $\mathbb{E}Z_n = \mathbb{E}Y\mathbb{E}X_n = \mu\mathbb{E}Y < \infty$, but Z_n are not independent: hence we cannot apply SLLN... However, we can apply it to $\overline{X}_n$. Let's guess the limit,

which is $Y\mathbb{E}X$ (because Y does not average). We have:

$$|\overline{Z}_n - Y\mathbb{E}X| = \left| Y \sum_{i=1}^{n} X_i - Y\mathbb{E}X \right| \leqslant |Y| \left| \sum_{i=1}^{n} X_i \mathbb{E}X \right| \leqslant |\overline{X}_n - \mathbb{E}X|.$$

Hence $\forall\, \varepsilon > 0$ and for any $n \geqslant 1$:

$$\left\{ \left| \overline{Z}_n - Y\mathbb{E}X_n \right| \leqslant \varepsilon \right\} \supset \left\{ \left| \overline{X}_n - \mathbb{E}X_n \right| \leqslant \varepsilon \right\}$$

which means

$$\bigcap_{n=k}^{\infty} \left\{ \left| \overline{Z}_n - Y\mathbb{E}X \right| \leqslant \varepsilon \right\} \subset \bigcap_{n=k}^{\infty} \left\{ \left| \overline{X}_n - \mathbb{E}X_n \right| \leqslant \varepsilon \right\}, \qquad \forall\, k$$

and also

$$\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{ \left| \overline{Z}_n - Y\mathbb{E}X \right| \leqslant \varepsilon \right\} \subset \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{ \left| \overline{X}_n - \mathbb{E}X_n \right| \leqslant \varepsilon \right\}, \qquad \forall\, k$$

$$\Updownarrow$$

$$\liminf_n \left\{ \left| \overline{Z}_n - Y\mathbb{E}X \right| \leqslant \varepsilon \right\} \supset \liminf_n \left\{ \left| \overline{X}_n - \mathbb{E}X_n \right| \leqslant \varepsilon \right\}$$

$$\Downarrow$$

$$\mathbb{P}\left(\left\{ \left| \overline{Z}_n - Y\mathbb{E}X \right| \leqslant \varepsilon \right\} \right) \geqslant \underbrace{\mathbb{P}\left(\left\{ \left| \overline{X}_n - \mathbb{E}X_n \right| \leqslant \varepsilon \right\} \right)}_{\to 1 \text{ by SLLN: } \overline{X}_n \to \mathbb{E}X \text{ a.s.}}$$

$$\Downarrow$$

$$\mathbb{P}\left(\liminf_n \left\{ \left| \overline{Z}_n - Y\mathbb{E}X \right| \leqslant \varepsilon \right\} \right) = 1$$

i.e.

$$\overline{Z}_n \xrightarrow{\text{a.s.}} Y\mathbb{E}X.$$

Remark

There is a shorter solution: we know that $X_n \to X$ a.s. and $Y_n \to Y$ a.s. and this implies that

$$X_n Y_n \xrightarrow{\text{a.s.}} XY$$

in the special case where $Y_n \equiv Y$.

2. We know that

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(Y\mathbb{E}X = \mu) \\ &= \mathbb{P}(Y\mathbb{E}X = \mathbb{E}X) \\ &= \mathbb{P}(Y = 1) = p \in (0, 1) \end{aligned}$$

since $\mathbb{E}X \neq 0$.

Exercise 2

[redacted]

Since we know that U_i $\perp\!\!\!\perp$ and that f is Borel-measurable, then $f(U_i)$ are random variables and independent ones. Note that, since U_i are absolutely continuous random variables,

$$\begin{aligned} \mathbb{E}|f(U_i)| &= \int_0^1 |f(u)| \cdot f_{U_i}(u)\, \mathrm{d}u \\ &= \int_0^1 |f(u)| \cdot \underbrace{1}_{\text{p.d.f. of U}} \mathrm{d}u < \infty \end{aligned}$$

since $f \in L^1$. Setting $Y_i = f(U_i)$ we have $I_n = \overline{Y}_n$. By SLLN we have

$$\overline{Y}_n \to \mathbb{E}Y = \int_0^1 f(u)\,\mathrm{d}u.$$

Exercise 3

[redacted]

Let $x \in (0, n)$. We have

$$\begin{aligned}
\mathbb{P}(Z_n > x) &= \mathbb{P}\left(nX_1 > x, nX_2 > x, \dots, nX_n > x\right) \\
&= \mathbb{P}\left(X_1 > \frac{x}{n}\right) \cdot \dots \cdot \mathbb{P}\left(X_n > \frac{x}{n}\right) \\
&= \left(1 - \frac{x}{n}\right) \cdot \dots \cdot \left(1 - \frac{x}{n}\right) \\
&= \left(1 - \frac{x}{n}\right)^n.
\end{aligned}$$

So

$$\mathbb{P}(Z_n \leqslant x) = 1 - \left(1 - \frac{x}{n}\right)^n$$

and for $x \geqslant n$ we have $\mathbb{P}(Z_n \geqslant x) = 1$, for $x \geqslant 0$ we have $\mathbb{P}(Z_n \geqslant x) = 0$.

Revise with Kotatsu!

Remember that

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n.$$

Thus

$$\lim_{n \to \infty} \mathbb{P}(X_n \leqslant x) = \begin{cases} 1 - e^{-x} & \text{for } 0 < x < \infty \\ 0 & \text{for } x \geqslant 0. \end{cases}$$

It follows that

$$Z_n \xrightarrow{\mathrm{d}} Z$$

where $Z \sim \mathrm{Exp}\,(1)$.

Exercise 4

[redacted]

A.s. convergence: if this holds, we also have convergence in probability and in distribution. The distribution of Y_n is

$$\begin{aligned}
\mathbb{P}(Y_n \leqslant y) = \mathbb{P}\left(X_n^{\frac{1}{n}} \leqslant y\right) &= \begin{cases} 0 & \text{if } y < 0 \\ \mathbb{P}(X_n \leqslant y^n) & \text{if } y \geqslant 0. \end{cases} \\
&= \begin{cases} 0 & \text{if } y < 0 \\ y^n \cdot 2^{-n} & \text{if } 0 \leqslant y^n \leqslant 2^n \iff 0 \leqslant y \leqslant 2 \\ 1 & \text{if } y^n > 2^n \iff y > 2. \end{cases}
\end{aligned}$$

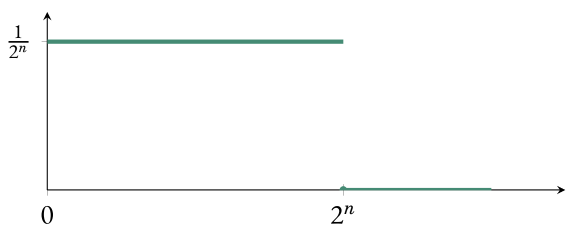


Figure 1.6: Remember how a $\mathrm{U}(0, 2^n)$ uniform distribution is made?

Revise with Kotatsu!

Do you really need to remember that for uniform distributions the probability density function is

$$f(x) = \frac{1}{a-b}?$$

Since $y^n 2^{-n} \longrightarrow 0$ when $y < 2$ and $y^n 2^{-n} \longrightarrow 1$ when $y = 2$ we have:

$$\lim_{n\to\infty} \mathbb{P}(Y_n \leqslant y) = \begin{cases} 0 & \text{if } y < 2 \\ 1 & \text{if } y \geqslant 2. \end{cases}$$

This means that we have convergence in distribution to the degenerate random variable $Y = 2$. Let's check a.s. convergence with BC: for $\varepsilon > 0$ and $\varepsilon < 2$ (because for $\varepsilon \geqslant 2$ the probability is simply 0) we have

$$\begin{aligned}\mathbb{P}(|Y_n - 2| > \varepsilon) = \underbrace{\mathbb{P}(-Y_n + 2 > \varepsilon)}_{Y_n \in [0,2] \text{ since } Y_n = X_n^{\frac{1}{n}} \text{ and } Y_n \in [0,2^n]} &= \mathbb{P}(Y_n < 2 - \varepsilon) \\ &= \left(\frac{2-\varepsilon}{2}\right)^n\end{aligned}$$

hence, since we know that $0 > \varepsilon < 2$ and therefore $\frac{2-\varepsilon}{2} < 1$

$$\sum_{i=1}^{\infty} \mathbb{P}(|Y_n - 2| > \varepsilon) = \sum_{i=1}^{\infty} \left(\frac{2-\varepsilon}{2}\right)^n < +\infty.$$

Thus, by BC-1 we have

$$\mathbb{P}(\limsup_n \{|Y_n - 2| > \varepsilon\}) = 0 \iff \mathbb{P}(\liminf_n \{|Y_n - 2| \geqslant \varepsilon\}) = 1$$

which implies a.s. convergence of Y_n to 2, since $\varepsilon > 0$ was arbitrary.

Exercise 5

[redacted]

Let's chech whether we can apply CLT (ez version). We know that the X_n are independent with $\mathbb{E}X_n = 0$ and

$$\operatorname{Var} X_n = \mathbb{E}X_n^2 = 2 \cdot n^2 \cdot \frac{1}{2n^2} + 2 \cdot 1 \cdot \frac{1}{2}\left(1 - \frac{1}{n^2}\right) = 1 + \left(1 - \frac{1}{n^2}\right)$$

so $\operatorname{Var} X_n$ depends on n and we can't apply CLT! Let's proceed differently. Consider

$$\phi(x) = \begin{cases} 1 & \text{if } x \geqslant 0 \\ -1 & \text{if } x < 0 \end{cases}$$

and notice that

$$Y_n = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} \phi(X_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \phi(X_i)). \qquad (🏃)$$

We study the convergence of the two parts separately. Note that

$$\begin{aligned} \mathbb{P}(X_i \neq \phi(X_i)) &= \mathbb{P}(X_i \neq \pm 1) \\ &= \mathbb{P}(X_i = i \cup X_i = -i) \qquad \text{for } i > 1 \\ &= \mathbb{P}(X_i = i) + \mathbb{P}(X_i = -i) \\ &= \frac{1}{2i^2} + \frac{1}{2i^2} = \frac{1}{i^2}. \end{aligned}$$

Hence we have that

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \neq \phi(X_i)) = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

and by BC-1 we can take the complement $\mathbb{P}(\liminf_i \{X_i = \phi(X_i)\}) = 1$ and thus $\mathbb{P}$-a.a.ω, $\exists N(\omega)$ such that $\forall\, i \geqslant N(\omega), X_i(\Omega) = \phi(X_i(\omega))$ which implies that for a set Ω' of measure 1 we have $\sum_{i=1}^{\infty} X_i - \phi(X_i) < \infty$. Then the second term in 🏃 converges to 0 $\mathbb{P}-$ a.s. (and in distribution) because of the factor $\frac{1}{\sqrt{n}}$ in front of the finite series. For the other terms in 🏃 we have

$$\begin{aligned} \phi(X_i) &\in \{1, -1\} \\ \mathbb{P}(\phi(X_i) = 1) = \mathbb{P}(X_i \geqslant 0) &= \tfrac{1}{2} \qquad \text{i.e. } \phi(X_i) \sim \mathrm{U}(\{-1, 1\}) \\ \mathbb{P}(\phi(X_i) = -1) = \mathbb{P}(X_i < 0) &= \tfrac{1}{2} \end{aligned}$$

so $\phi(X_i)$ are identically distributed random variables with mean 0 and variance

$$\mathbb{V}\mathrm{ar}\, \phi(X_i) = \mathbb{E}\phi(X_i)^2 = 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1.$$

They are independent because X_i are independent and thus, by CLT (ez version) we have

$$\underbrace{\frac{\frac{1}{n}\sum_{i=1}^{n} \phi(X_i) - 0}{\frac{1}{\sqrt{n}}}}_{\xrightarrow{\mathrm{d}} \mathrm{N}(0,1)} = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \phi(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(X_i)$$

hence $Y \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) + 0 = \mathrm{N}(0,1)$ because of the following fact:

Remark

if $X_n \xrightarrow{\mathrm{d}} X, Z_n \xrightarrow{\mathrm{d}} c,\ c \in \mathbb{R}$, then $(X_n, Z_n) \xrightarrow{\mathrm{d}} (X, c)$.

From this fact we know that $X_n + Z_n \xrightarrow{\mathrm{d}} X + c$ and we apply this with

$$X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(X_i) \xrightarrow{\mathrm{d}} \mathrm{N}(0,1) \qquad \text{and} \qquad Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \phi(X_i)) \xrightarrow{\mathrm{d}} 0.$$

Exercise 6

[redacted]

Let $k \in \mathbb{N}$ and select n_k such that $n_k > n_{k-1}$ and

$$\mathbb{P}\left(\left|X_{n_k} - X\right| > \frac{1}{k}\right) \leqslant \frac{1}{k^2}$$

which is always possible since $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) \to 0\ \forall\, \varepsilon > 0$ thanks to the fact that we know about the convergence. Notice, moreover, that $n_k \to \infty$ as $k \to \infty$. We have

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\left|X_{n_k} - X\right| > \frac{1}{k}\right) \leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

so by BC-1 we have

$$\mathbb{P}\left(\limsup_k \left\{\left|X_{n_k} - X\right| > \frac{1}{k}\right\}\right) = 0 \iff \mathbb{P}\left(\liminf_k \left\{\left|X_{n_k} - X\right| \leqslant \frac{1}{k}\right\}\right) = 1$$

which means that $\forall\, \omega \in \left\{\liminf_k \left|X_{n_k} - X\right| \leqslant \frac{1}{k}\right\} = \Omega'$ (which is the set of the ω where the event $\left|X_{n_k} - X\right| \leqslant \frac{1}{k}$ is true as $k \to \infty$) we have

$$\left|X_{n_k}(\omega) - X(\omega)\right| \leqslant \frac{1}{k}$$

i.e.

$$\lim_{k\to\infty} \left|X_{n_k}(\omega) - X(\omega)\right| \leqslant 0 \implies \lim_{k\to\infty} \left|X_{n_k} - X(\omega)\right| = 0$$

and since $\mathbb{P}(\Omega') = 1$ we have a.s. convergence along the subsequence n_k.

Exercise 7

[redacted]

(i) Let $Y_i := f(X_i)$. $(Y_i)_i$ are random variables because f is Borel-measurable and $Y_i \perp\!\!\!\perp$ because $X_i \perp\!\!\!\perp$. Moreover Y_i are identically distributed because

$$\begin{aligned}
\mathbb{P}(Y_i \leqslant y) &= \mathbb{P}(f(X_i) \leqslant y) \\
&= \mathbb{P}(f(X_i) \in (-\infty, y]) \\
&= \mathbb{P}(X_i \in f^{-1}(-\infty, y]) \\
&= \mathbb{P}(X_1 \in f^{-1}(-\infty, y]) \\
&= \mathbb{P}(f(X_1) \in (-\infty, y]) \\
&= \mathbb{P}(Y_1 \leqslant y) \qquad \forall\, i \geqslant 1
\end{aligned}$$

Moreover, since $\mathbb{V}\mathrm{ar}\, Y_i \leqslant c \left\|f\right\|_\infty =: \sigma^2$ (and it is independent of i) then

$$\sum_{i-1}^{k} \mathbb{V}\mathrm{ar}\left(\frac{Y_i}{i}\right) \leqslant \sigma^2 \cdot \sum_{i=1}^{n} \frac{1}{i^2} < \infty$$

So by the second point of the WLLN we have

$$\frac{\sum_{i=1}^n Y_i - \mathbb{E}\left(\sum_{i=1}^n Y_i\right)}{n} \xrightarrow{\text{a.s.}} 0$$

$$\downarrow$$

$$\frac{1}{n}\sum_{i=1}^n f(X_i) - \frac{1}{n}\mathbb{E}[n \cdot Y_i] = \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X_i)]$$

$$\Downarrow$$

$$\frac{1}{n}\sum_{i=1}^n f(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[f(X_1)].$$

(ii) By (i) with $f = \mathbb{1}_A$ we have

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}_A(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[\mathbb{1}_A(X_1)] = \mathbb{P}(X_1 \in A).$$

Remark

Notice that (i) and (ii) provide a Monte-Carlo method to estimate probabilities and expectations.

1.1.6 Exercise class 6

Revise with Kotatsu!

Definition 1.1.4

Conditional expectation: let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space and $\mathcal{F} \subset \mathcal{H}$. Let X be $\mathcal{H}$-measurable with $\mathbb{E}|X| < \infty$. The conditional expectation of X given $\mathcal{F}$, denoted by $\overline{X}$ or $\mathbb{E}(X|\mathcal{F})$ is a random variablesuch that:

- $\overline{X}$ is $\mathcal{F}$-measurable (**measurability**);
- $\mathbb{E}[\mathbb{1}_F \overline{X}] = \mathbb{E}[\mathbb{1}_F X], \ \forall F \in \mathcal{F}$ (**projection**).

This actually defines the conditional expectation of positive random variable first and then uses $X = X^+ - X^-$.

Here are the **main properties of conditional expectation**. Let $X, Y \in L^1(\Omega, \mathcal{H}, \mathbb{P}), \ \mathcal{G} \subset \mathcal{H}, \ \mathcal{F} \subset \mathcal{H}$. Then:

(a) if Y is a version of $\mathbb{E}(X|\mathcal{G})$ then $\mathbb{E}Y = \mathbb{E}X$;

(b) if X is $\mathcal{G}$-measurable then $\mathbb{E}(X|\mathcal{G}) = X, \mathbb{P} -$ a.s.;

(c) $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G}), \mathbb{P} -$ a.s.;

(d) if $X \geqslant 0$ then $\mathbb{E}[X|\mathcal{G}] \geqslant 0$

(e) **tower property**: if $\mathcal{F} \subset \mathcal{G}$ then

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{F}] = \mathbb{E}[\mathbb{E}(X|\mathcal{F})|\mathcal{G}] = \mathbb{E}(X|\mathcal{F});$$

(f) if Y is bounded and $\mathcal{G}$-measurable then

$$\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G});$$

(g) if $\mathcal{F}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$ then

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{F})) = \mathbb{E}(X|\mathcal{G}).$$

Remark

Special case with $\mathcal{G} = \{\emptyset, \Omega\}$: if $X \perp\!\!\!\perp \mathcal{F}$ then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X$.

Conditioning on another random variable:

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y)).$$

Remark

The expectation $\mathbb{E}(X)$ can be viewed as a condition expectation given the trivial σ-algebra $\mathcal{H}_0 = \{\emptyset, \Omega\}$ i.e.

$$\mathbb{E}(X) = \mathbb{E}(X|\mathcal{H}_0).$$

This is useful when applying the tower property, because it gives us

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|\mathcal{G})).$$

Theorem 1.1.2

Suppose X and Y are random variables on $(\Omega, \mathcal{H}, \mathbb{P})$ with values in $(D, \mathcal{D})$, $(E, \mathcal{E})$ respectively. Suppose that the joint probability distribution has the form

$$\pi(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x) k(x, \mathrm{d}y)$$

with some probability kernel form $(D, \mathcal{D})$ to $(E, \mathcal{E})$, i.e. for every f which is $(\mathcal{D} \times \mathcal{E})$-measurable we have

$$\int_{D\times E} f(x,y)\pi(\mathrm{d}x, \mathrm{d}y) = \int_D \left(\int_E f(x,y) k(x, \mathrm{d}y) \right) \mu(dx).$$

Then the kernel L defined by

$$L_\omega(B) = k(X(\omega), B), \qquad \forall\, \omega \in \Omega,\ B \in \mathcal{E}$$

is a version of the conditional distribution of Y given X and for every positive f which is $(\mathcal{D} \times \mathcal{E})$-measurable it holds

$$\mathbb{E}[f(X,Y)|X] = \int_E f(X,y) k(X, \mathrm{d}y).$$

This last equation is also known as **freezing lemma**. It means that I can effectively calculate the conditional expectation given X as if X was a constant (and not a random variable). Cool! Was it so hard to put it in this way?

Exercise 1

[redacted]

Note that $\sigma(X)$ is generated by three disjoint intervals:

$$I_1 := \left[0, \tfrac{1}{3}\right), \quad I_2 := \left[\tfrac{1}{3}, \tfrac{2}{3}\right), I_3 := \quad \left[\tfrac{2}{3}, 1\right].$$

We know that $\overline{Y} := \mathbb{E}(Y|X)$ must be $\sigma(X)$-measurable (by property of measurability) so it must be constant on each interval I_i (since the σ-algebras of constants are constants so any function measurable by a constant must be a constant):

$$\left(\forall\, A \in \mathcal{B}(\mathbb{R}), \qquad \overline{Y}^{-1}(A) \in \sigma(X) = \sigma(I_1, I_2, I_3)\right).$$

Let's denote by a_i the values of $\overline{Y}$ on I_i, that is

$$\overline{Y}(\omega) = \mathbb{E}(Y|X) = \begin{cases} a_1 & \text{on } I_1 \\ a_2 & \text{on } I_2 \\ a_3 & \text{on } I_3. \end{cases}$$

It must be that, for the projection property,

$$\mathbb{E}\left(\mathbb{1}_{I_i}\overline{Y}\right) = \mathbb{E}\left(\mathbb{1}_{I_i} Y\right) \qquad \forall\, I_i \in \sigma(X),$$

that is

$$\int_{I_i} \overline{Y}(\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_{I_i} Y(\omega)\, \mathrm{d}\mathbb{P}(\omega), \qquad \forall\, i = 1, 2, 3. \tag{♣}$$

The left hand side of gives

$$\int_{I_i} \overline{Y}(\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_{I_i} a_i \,\mathrm{d}\mathbb{P} = a_i \mathbb{P}(I_i) = a_i \cdot \frac{1}{3}.$$

The right hand side of gives

$$\begin{aligned}\int_{I_i} Y(\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{I_i} 2\omega^2\,\mathrm{d}\mathbb{P}(\omega)\\ &= \begin{cases} \frac{2}{3}\omega^3\big|_0^{\frac{1}{3}} = \frac{2}{81} & \text{if } i = 1\\ \frac{2}{3}\omega^3\big|_{\frac{1}{3}}^{\frac{2}{3}} = \frac{14}{81} & \text{if } i = 2\\ \frac{2}{3}\omega^3\big|_{\frac{2}{3}}^{1} = \frac{38}{81} & \text{if } i = 3.\end{cases}\end{aligned}$$

Thus we must have

$$\begin{gathered}\tfrac{1}{3}a_1 = \tfrac{2}{81}, \quad \tfrac{1}{3}a_2 = \tfrac{14}{81}, \quad \tfrac{1}{3}a_3 = \tfrac{38}{81}\\ \Downarrow\\ a_1 = \tfrac{2}{27}, \quad a_2 = \tfrac{14}{27}, \quad a_3 = \tfrac{38}{27}\end{gathered}$$

Exercise 2

[redacted]

According to the definition we must show:

[i] $\sigma(Y)$-measurability ($g(X) \in L^1$ by assumption);

[ii] projection.

Let's tackle this one at the time.

[i] For this property it is enough to show that $y \mapsto \varphi(y)$ is measurable. We have $\varphi(y) = \sum_x g(x)\rho_{X|Y}(x|y)$ where $\rho_{X|Y}(x|y)$ depends on $\rho_{X,Y}(x,y)$ and $\rho_Y(y)$ which are both measurable because X and Y are random variables.

[ii] $\forall A \in \sigma(Y)$ we should check whether $\mathbb{E}[\mathbb{1}_A\varphi(Y)] = \mathbb{E}[\mathbb{1}_A g(X)]$. The set $A \in \sigma(Y)$ must be of the form $A = \{\omega : Y(\omega) \in B\}$ for some $B \in \mathcal{F}$. Thus $\mathbb{1}_A(\omega) = \mathbb{1}_B(Y)$. We have

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A\varphi(Y)] &= \int_\Omega \mathbb{1}_A(\omega)\varphi(Y(\omega))\mathbb{P}(\mathrm{d}\omega)\\ &= \int_\Omega \mathbb{1}_B(Y(\omega))\mathbb{P}(\mathrm{d}\omega)\\ &= \sum_{y\in F} \mathbb{1}_B(y)\varphi(y)\rho_Y(y)\text{[7]}\\ &= \sum_{y\in B}\sum_{x\in E} g(x)\rho_{X|Y}(x|y)\rho_Y(y)\\ &= \sum_{y\in B}\sum_{x\in E} g(x)\rho_{X,Y}(x,y).\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_A g(Y)] &= \int_\Omega \mathbb{1}_A(\omega) g(X(\omega)) \mathbb{P}(\mathrm{d}\omega) \\
&= \int_\Omega \mathbb{1}_B(Y(\omega)) g(X(\omega)) \mathbb{P}(\mathrm{d}\omega) \\
&= \sum_{(x,y) \in E \times F} \mathbb{1}_B(y) g(x) \rho_{X,Y}(x,y) \\
&= \sum_{X \in E} \sum_{y \in B} g(x) \rho_{X,Y}(x,y)
\end{aligned}$$

which is the same as in .

Exercise 3

[redacted]

Here we mean that $\mathbb{P}(X = x) = (1-p)p^x,\ x \in \mathbb{N} \cup \{0\}$. This is also known as "modified Geometric". In this case $\mathbb{E}X = \frac{1}{1-p} - 1$. We want to use the result from exercise 2 and hence we compute the joint distribution of (X, Z):

$$\begin{aligned}
\rho_{X,Z}(x,z) &= \mathbb{P}(X = x, Z = z) \\
&= \mathbb{P}(X = x, \min(X,Y) = z) \\
&= \begin{cases} 0 & \text{if } z > x \\ \mathbb{P}(X = x, Y = z) & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ \mathbb{P}(X = x, Y \geqslant x) & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} \mathbb{P}(X = x, Y = z) & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ \mathbb{P}(X = x, Y \geqslant x) & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} \mathbb{P}(X = x, Y = z) & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ \mathbb{P}(X = x, Y \geqslant x) & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} (1-p)p^x(1-q)q^z & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ (1-p)p^x \sum_{k=1}^{\infty}(1-q)q^k & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else.} \end{cases}
\end{aligned}$$

Revise with Kotatsu!

Remember the geometric series:

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$
$$\sum_{k=0}^{N} q^k = \frac{1-q^{N+1}}{1-q}.$$

[7] In the sum only the terms with $\rho_Y(y) > 0$ give a contribution.

In this case we have

$$
\begin{aligned}
\mathbb{P}(Y \geqslant x) &= \sum_{k=x}^{\infty} q^k \\
&= -\sum_{k=0}^{x-1}(1-q)q^k + \sum_{k=0}^{\infty}(1-q)q^k \\
&= -(1-q)\cdot\frac{1-q^x}{1-q} + (1-q)\frac{1}{1-q} \\
&= 1-(1-q^x) = q^x. \qquad\qquad (🍕)
\end{aligned}
$$

We compute now the marginal distribution of Z. Let $z \geqslant 0, z \in \mathbb{N}$.

$$
\begin{aligned}
\rho_Z(z) &= \mathbb{P}(Z = z) \\
&= \mathbb{P}(\min(X,Y) = z) \\
&= \mathbb{P}(min(X,Y) \geqslant z) - \mathbb{P}(\min(X,Y \geqslant z+1)) \\
&= \mathbb{P}(X \geqslant z)\mathbb{P}(Y \geqslant z) - \mathbb{P}(X \geqslant z+1)\mathbb{P}(Y \geqslant z+1) \\
&= p^z q^z - p^{z+1}q^{z+1} \qquad \text{due to 🎸} \\
&= p^z q^z(1-pq) \qquad\qquad (🎵)
\end{aligned}
$$

hence $Z \sim \text{Geom}\,(1-pq)$. We can thus define the conditional probability $\rho_{X|Z}$ as

$$
\begin{aligned}
\rho_{X|Z} &= \begin{cases} \frac{\rho_{X,Z}(x,z)}{\rho_Z(z)} & \text{if } \rho_Z(z) > 0 \\ 0 & \text{if } \rho_Z(z) = 0 \end{cases} \\
\text{due to 🎸 and 🎵} &= \begin{cases} \frac{(1-p)p^x(1-q)\cancel{q^z}}{p^z\cancel{q^z}(1-pq)} & \text{if } z < x,\ \ z,x \in \mathbb{N}\cup\{0\} \\ \frac{(1-p)\cancel{p^x q^x}}{\cancel{p^x q^x}(1-pq)} & \text{if } z = x,\ \ z,x \in \mathbb{N}\cup\{0\} \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} \frac{(1-p)(1-q)p^{x-z}}{1-pq} & \text{if } z < x,\ \ z,x \in \mathbb{N}\cup\{0\} \\ \frac{(1-p)}{1-pq} & \text{if } z = x,\ \ z,x \in \mathbb{N}\cup\{0\} \\ 0 & \text{else.} \end{cases} \qquad (🐾)
\end{aligned}
$$

Using 🐾 and the result of exercise 2 with $g(x) = x$ we define

$$
\begin{aligned}
\varphi(z) &= \sum_{x=0}^{\infty} x\rho_{X|Z}(x|z) \\
&= \sum_{x=0}^{\infty}(x-z+z)\rho_{X|Z}(x|z) \\
&= z\sum_{x=0}^{\infty}\rho_{X|Z}(x|z) + \sum_{x=0}^{\infty}(x-z)\rho_{X|Z}(x|y) \\
&= z + \sum_{x=z}^{\infty}(x-z)\rho_{X|Z}(x|y) \\
&= z + 0 + \sum_{x=z+1}^{\infty}(x-z)\cdot\frac{(1-p)(1-q)p^{x-z}}{1-pq} \\
&= z + \frac{(1-p)(1-q)}{1-pq}\sum_{x=z+1}^{\infty}\underbrace{(x-z)}_{=y}p^{x-z} \\
&= z + \frac{(1-p)(1-q)}{1-pq}\sum_{y=1}^{\infty} y p^y \\
&= z + \frac{1-q}{1-pq}\underbrace{\sum_{y=1}^{\infty} ypY(1-p)}_{\mathbb{E}V \text{ where } V \sim \text{Geom}\,(1-p) \text{ so } \mathbb{E}V = \frac{1}{1-p}-1} \\
&= z + \frac{1-q}{1-pq}\cdot\frac{p}{1-p}.
\end{aligned}
$$

So a version of the conditional expectation $\mathbb{E}(X|Z)$ is

$$\mathbb{E}(X|Z) = Z + \frac{(1-q)p}{(1-pq)(1-p)}.$$

Exercise 4

[redacted]

Idk this had no solution.

Exercise 5

[redacted]

Take $A \in \sigma(N)$ arbitrary. Note that the events $\{N = n\}_{n\in\mathbb{N}}$ for a partition of Ω so

$$Y = \sum_{n=1}^{\infty} \mathbb{1}_{\{N=n\}} = \sum_{n=1}^{\infty}\sum_{k=1}^{n} X_k \mathbb{1}_{\{N=n\}}.$$

Hence

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_A Y\right] &= \mathbb{E}\left[\mathbb{1}_A \sum_{n=1}^{\infty}\sum_{k=1}^{n} X_k \mathbb{1}_{\{N=n\}}\right] \\
\text{monotone convergence} &= \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{1}_A \sum_{k=1}^{n} X_k \mathbb{1}_{\{N=n\}}\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left[\sum_{k=1}^{n} X_k\right] \cdot \mathbb{E}\left[\mathbb{1}_A \mathbb{1}_{\{N=n\}}\right] \\
&= \sum_{n=1}^{\infty} n \cdot \mathbb{E}X_1 \cdot \mathbb{E}\left[\mathbb{1}_A \mathbb{1}_{\{N=n\}}\right] \\
&= \mathbb{E}X_1 \cdot \mathbb{E}\left[\sum_{n=1}^{\infty} n \cdot \mathbb{1}_{\{N=n\}} \cdot \mathbb{1}_A\right] \\
&= \mathbb{E}X_1 \cdot \mathbb{E}\left[N \mathbb{1}_A\right] \\
&= \mathbb{E}\left[\mathbb{E}[X_1] \cdot N \mathbb{1}_A\right]. \qquad (📖)
\end{aligned}$$

Hence, since $\mathbb{E}(X_1) \cdot N$ is $\mathcal{F}(N)$-measurable, we have that the second point of definition and expectation is satisfied for

$$\mathbb{E}(Y|N) = \mathbb{E}(X_1) \cdot N$$

which is then a version of the conditional expectation. Notice that we should also check that $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ which is true by similar computations as in 📖 with $A = \Omega$ so

$$\mathbb{E}[|Y|] = \mathbb{E}(Y) = \mathbb{E}(\mathbb{E}X_1 \cdot N) = \mathbb{E}X_1 \cdot \mathbb{E}N < \infty$$

since $X_1, N \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Exercise 6

[redacted]

We know that

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY|X)) \underset{X\in\sigma(X)}{=} \mathbb{E}(X\mathbb{E}(Y|X)) = \underbrace{\mathbb{E}(X\mathbb{E}Y)}_{\text{by assumption}} = \mathbb{E}X\mathbb{E}Y.$$

Thus $\mathbb{C}\mathrm{ov}\,(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = 0$. Assume now $\mathbb{C}\mathrm{ov}\,(X,Y) = 0$. To show that $\mathbb{E}(Y|X) \neq \mathbb{E}Y$ in general, we find a counterexample. Consider

$$U \sim \mathsf{U}\,(\{(0,1),(-1,-1),(1,-1)\}) \qquad U = (X,Y).$$

Then $X \sim \mathsf{U}\,(\{0,-1,1\})$ which means $\mathbb{E}X = 0$ and

$$XY = \begin{cases} 0 & \text{with prob. } \frac{1}{3} \\ 1 & \text{with prob. } \frac{1}{3} \\ -1 & \text{with prob. } \frac{1}{3}. \end{cases}$$

Hence $XY \sim \mathsf{U}\,(\{0,-1,1\})$ which means $\mathbb{E}(XY) = 0$. We then have $\mathbb{C}\mathrm{ov}\,(XY) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0-0 = 0$. However, $\mathbb{E}(Y|X=0) = 1$ but

$$Y = \begin{cases} 1 & \text{with prob. } \frac{1}{3} \\ -1 & \text{with prob. } \frac{2}{3} \end{cases} \implies \mathbb{E}Y = \frac{1}{3} - \frac{2}{3} = -\frac{1}{3} \neq 1.$$

Exercise 7

[redacted]

We want to apply the freezing lemma:

Theorem 1.1.3

Let (X,Y) be $\mathbb{R}$-valued random variables such that their joint distribution is

$$\pi(\mathrm{d}x,\mathrm{d}y) = \mu(\mathrm{d}x)\underbrace{k(x,\mathrm{d}y)}_{\text{prob. kernel}}.$$

Then $\forall$ positive measurable f we have that $\int_{\mathbb{R}} f(X,Y)k(X,\mathrm{d}y)$ is a version of conditional expectation $\mathbb{E}[f(X,Y)|X]$.

We want to apply this shit to (X,Y) and $f(X,Y) = \mathbb{1}_{\{X+Y\leqslant t\}}$ for some $t \in \mathbb{R}$ given. We have by independence $\pi(\mathrm{d}x,\mathrm{d}y) = \mu_X(\mathrm{d}x)\mu_Y(\mathrm{d}y)$ and $\mathbb{1}_{\{X+Y\leqslant t\}} \geqslant 0$ and measurable. Thus a version of the conditional expectation $\mathbb{E}\left(\mathbb{1}_{\{X+Y\leqslant t\}}|X\right)$ is given by

$$\int_{\mathbb{R}} \mathbb{1}_{\{X+y\geqslant t\}}(y)\mu_Y(\mathrm{d}y).$$

Let's work out the integral (with X replaced by x):

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{1}_{\{x+y\geqslant t\}}(y)\mu_Y(\mathrm{d}y) &= \int_{\mathbb{R}} \mathbb{1}_{\{y\geqslant t-x\}}(y)\mu_Y(\mathrm{d}y) \\ &= \int_{-\infty}^{t-x} \mu_Y(\mathrm{d}y) \\ &= F_Y(t-x). \end{aligned}$$

Hence a version of the conditional expectation is given by $F_Y(t-x)$. In the special case when $X \sim \mathsf{U}\,(0,1)$ we get $\mu(\mathrm{d}x) = \mathrm{d}x$ on (0,1). Then

$$\begin{aligned}
F_{X+Y}(t) &= \mathbb{P}(X+Y \leqslant t) \\
&= \mathbb{E}\left[\mathbb{1}_{\{X+Y\leqslant t\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\{X+Y\leqslant t\}}|X\right]\right] \\
&= \mathbb{E}\left[F_Y(t-X)\right] \\
&= \int_{\mathbb{R}} F_Y(t-x)\mu_X(\mathrm{d}x) \\
&= \int_{\mathbb{R}} F_Y(t-x)\,\mathrm{d}x.
\end{aligned}$$

1.1.7 Exercise class 7

Revise with Kotatsu!

Definition 1.1.5

Sub/Super Martingale: a real-valued stochastic process $(X_t)_{t\in T}$ is a $\mathcal{F}$-submartingale/supermartingale/martingale if:

1. X is $\mathcal{F}$-adapted;
2. X_t is integrable $\forall\, t \in T$, i.e. $\mathbb{E}|X_t| < \infty$;
3. $\mathbb{E}[X_t|\mathcal{F}_s] \geqslant / \leqslant / = X_s,\ \forall\, s \geqslant t,\ t, s \in T$.

For the discrete time processes we have $T \equiv \mathbb{N}$.

Definition 1.1.6

Stopping time: a random time $T : \Omega \to \mathbb{R}^+$ is a stopping time for a filtration $(\mathcal{F}_t)_{t\in T}$ if $\{T \leqslant t\} \in \mathcal{F}_t\ \forall\, t \in T$.

In the discrete case we have $T : \Omega \to \mathbb{N} \cup \infty$ s.t. $\{T = n\} \in \mathcal{F}_n\ \forall\, n \geqslant 0$.

Theorem 1.1.4

Doob's stopping theorem. Let M be a $\mathcal{F}$-adapted process. The following are equivalent:

- M is a $\mathcal{F}$-submartingale;
- $\forall\, T, S$ stopping times such that $S \geqslant T$ and S, T bounded, M_S and M_T are integrable and $\mathbb{E}(M_T - M_S|\mathcal{F}_S) \geqslant 0$;
- $\forall\, T, S$ stopping times such that $S \geqslant T$ and S, T bounded, M_S and M_T are integrable and $\mathbb{E}(M_T - M_S) \geqslant 0$.

We can replace the submartingale with the martingale by switching $\geqslant$ with $=$. Same for supermartingale (switching with $\leqslant$).

Theorem 1.1.5

Variation/extension of Doob's stopping theorem. Let M be a $\mathcal{F}$-martingale and T a stopping time. Then M_T is integrable and $\mathbb{E}M_T = \mathbb{E}M_0$ if and only if one of the following condition holds:

1. T is bounded;
2. M is bounded and $T < \infty$ $\mathbb{P}-$ a.s.;
3. $\mathbb{E}T < \infty$ and M has bounded increments;
4. M is uniformly integrable.

Exercise 1

[redacted]

Each $\mathcal{G}_n$ is a σ-algebra by definition. We must prove that $\mathcal{F}_n \subseteq \mathcal{G}_{n+1}$. We have:

$$\begin{aligned} C_n \ni \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] &= \left(\frac{2k}{n^{n+1}}, \frac{2(k+1)}{n^{n+1}}\right] \\ &= \underbrace{\underbrace{\left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right]}_{\in C_{n+1}} \cup \underbrace{\left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right]}_{\in C_{n+1}}}_{\in \sigma(C_{n+1}) = \mathcal{G}_{n+1}}. \end{aligned}$$

Since it holds $\forall\, k = 0, \dots, 2^n - 1$ we have $C_n \subset \mathcal{G}_{n+1}$ as wanted. Let's check whether X is a $\mathcal{G}$-martingale.

1. *"Is X $\mathcal{G}$-adapted?"* We need to check that $\forall\, B \in \mathcal{B}(\mathbb{R}),\ X_n^{-1}(B) \in \mathcal{G}_n$. This is like asking whether we have $\sigma(X_n) \subseteq \mathcal{G}_n$.

$$\begin{aligned} \sigma(X_n) &= \{X_n^{-1}(B) : B \in \mathcal{B}(\mathcal{B})\} \\ &\equiv \{\varnothing, \Omega, (0, 2^n], (2^n, 1]\} \subset \mathcal{G}_n. \end{aligned}$$

2. *"Is X integrable?"* We know that

$$\begin{aligned} \mathbb{E}|X_n| &= \int_\Omega |X_n(\omega)| \, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_0^1 |X_n(\omega)| \, \mathrm{d}\omega \\ &= \int_0^{2^{-n}} 2^n \, \mathrm{d}\omega + \int_{2^{-n}}^1 0 \, \mathrm{d}\omega = 1 < \infty. \end{aligned}$$

3. *"Does X have martingale property?"* It is enough to show for n and $n+1$ that $\mathbb{E}(X_{n+1}|\mathcal{G}_n) = X_n$. We have to show that

$$\mathbb{E}(X_n \mathbb{1}_A) = \mathbb{E}(X_{n+1}\mathbb{1}_A) \qquad \forall\, A \in \mathcal{G}_n.$$

Since $\mathcal{G}_n = \sigma(C_n)$ we have that $A = \bigcup_k A_k$ where A_k have the form $A_k = \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ and are thus disjoint. Thus

$$\begin{aligned} \mathbb{E}(X_{n+1}\mathbb{1}_A) &= \mathbb{E}(X_{n+1}\mathbb{1}_{\bigcup_k A_k}) \\ &= \sum_k \mathbb{E}(X_{n+1}\mathbb{1}_{A_k}) \end{aligned}$$

and similarly

$$\mathbb{E}(X_n \mathbb{1}_A) = \sum_k \mathbb{E}(X_n \mathbb{1}_{A_k}).$$

We then just have to check that

$$\mathbb{E}(X_n \mathbb{1}_{A_k}) = \mathbb{E}(X_{n+1}\mathbb{1}_{A_l}) \qquad \forall\, A_k \in C_n \tag{$\#$}$$

(which means for $k = 0, 1, \dots, 2^n - 1$). Notice that if $k \geqslant 1$ then $\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] \cap \left(0, \frac{1}{2^n}\right] = \varnothing$ and $\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] \cap \left(0, \frac{1}{2^{n+1}}\right] = \varnothing$. Thus $X_n \mathbb{1}_{A_k} = 0$ and $X_{n+1}\mathbb{1}_{A_k} = 0$, so $\#$ holds.
If $k = 0$ then

$$X_n \mathbb{1}_{A_0} = 2^n \cdot \mathbb{1}_{(0,2^{-n})} \cdot \mathbb{1}_{(0,2^{-n})} = 2^n \cdot \mathbb{1}_{(0,2^{-n})} = X_n$$

and
$$X_{n+1}\mathbb{1}_{A_0} = 2^{n+}\mathbb{1}_{(0,2^{-n-1})}\mathbb{1}_{(0,2^{-n})} = 2^{n+1}\mathbb{1}_{(0,2^{-n-1})} = X_{n+1}.$$
Moreover, $\mathbb{E}X_n = \mathbb{E}N_{n+1} = 1$. Hence ✠ holds also for $k = 0$.

Exercise 2

[redacted]

By the supermartingale property we know that $\mathbb{E}(X_{n+k}|\mathcal{G}_n) \leqslant X_n$ so for any $A \in \mathcal{G}_n$ we have

$$\mathbb{E}(X_{n+k}\mathbb{1}_A) = \mathbb{E}\left[(\mathbb{E}X_{n+k}|\mathcal{G}_n)\mathbb{1}_A\right] \leqslant \mathbb{E}(X_n\mathbb{1}_A). \tag{✣}$$

We can use tower property because if a random variable is $\mathcal{G}$-adapted and positive then it is $\mathcal{G}$-measurable. Moreover, since $X_n \geqslant 0 \forall\, n$ and also $\mathbb{1}_A \geqslant 0$ we have

$$0 \leqslant \mathbb{E}(X_{n+k}\mathbb{1}_A). \tag{✣✣}$$

Putting ✣ and ✣✣ together we get

$$0 \leqslant \mathbb{E}(X_{n+k}\mathbb{1}_A) \leqslant \mathbb{E}(X_n\mathbb{1}_A)$$

and choosing $A = \{X_n = 0\} \in \mathcal{G}_n$ we get

$$0 \leqslant \mathbb{E}(X_{n+k}\mathbb{1}_{\{X_n=0\}}) \leqslant \mathbb{E}(\underbrace{X_n\,\mathbb{1}_{\{X_n=0\}}}_{\text{this is always 0...}}) = 0$$

which implies $X_{n+k}\mathbb{1}_{\{X_n=0\}} = 0$ a.s. since $X_{n+k} \geqslant 0$.

Exercise 3

[redacted]

We need to check the three properties.

1. Check that $Z_n \in m\mathcal{G}_n$ (= it is measurable with respect to $\mathcal{G}_n$): this is true because $X_j \in m\mathcal{G}_n \;\forall\, j \leqslant n$, thus $S_n \in m\mathcal{G}_n$ and also $S_n^2 \in m\mathcal{G}_n$. Finally $\mathbb{E}\left(X_{j+1}^2|\mathcal{G}_j\right)$ is $\mathcal{G}_j$-measurable, so its sum up to $n-1$ is $\mathcal{G}_{n-1}$-measurable.

2. Check integrability:

$$\begin{aligned}
\mathbb{E}|Z_n| &\leqslant \mathbb{E}|S_n|^2 + \sum_{j=0}^{n-1}\mathbb{E}\left(\left|\mathbb{E}\left(X_{j+1}^2|\mathcal{G}_j\right)\right|\right) \\
&\leqslant \mathbb{E}\left(\left|\sum_{j=0}^{n}X_j\right|^2\right) + \sum_{j=0}^{n-1}\mathbb{E}\left(X_{j+1}^2\right) \\
&\leqslant n\sum_{j=0}^{n}\mathbb{E}\left|X_j\right|^2 + \sum_{j=0}^{n-1}\mathbb{E}\left|X_{j+i}\right|^2 < \infty \qquad \text{since } X_j \in L^2(\Omega).
\end{aligned}$$

3. Check martingale property:

$$\begin{aligned}
Z_{n+1} &= S_{n+1}^2 - \sum_{j=0}^{n} \mathbb{E}\left(X_{j+1}^2|\mathcal{G}_n\right) \\
&= S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 - \sum_{j=0}^{n-1} \mathbb{E}\left(X_{j+1}^2|\mathcal{G}_n\right) - \mathbb{E}(X_{n+1}^2|\mathcal{G}_n) \\
&= Z_n + 2X_{n+1}S_n + X_{n+1}^2 - \mathbb{E}\left(X_{n+1}^2|\mathcal{G}_n\right).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}\left(Z_{n+1}|\mathcal{G}_n\right) &= \mathbb{E}(Z_n|\mathcal{G}_n) + 2\mathbb{E}\left(X_{n+1}S_n|\mathcal{G}_n\right) + \cancel{\mathbb{E}\left(X_{n+1}^2|\mathcal{G}_n\right)} - \cancel{\mathbb{E}\left(\mathbb{E}\left(X_{n+1}^2|\mathcal{G}_n\right)|\mathcal{G}_n\right)} \\
&= Z_n + 2S_n \underbrace{\mathbb{E}\left(X_{n+1}|\mathcal{G}_n\right)}_{=0 \text{ by assumption}} \\
&= Z_n.
\end{aligned}$$

Exercise 4

[redacted]

1. We must check the three conditions:
   (a) *measurability*: $X_j \in m\mathcal{G}_n, \forall\, j \leqslant n$, hence $S_n \in m\mathcal{G}_n$;
   (b) *integrability*:
   $$\mathbb{E}|S_n| = \mathbb{E}\left|\sum_{j=1}^{n} X_j\right| \leqslant \mathbb{E}\sum_{j=1}^{n} |X_j| = n\mathbb{E}\,|X_j| = n < \infty.$$
   (c) *martingale property*:
   $$\begin{aligned}
   \mathbb{E}\left(S_{n+1}|\mathcal{G}_n\right) &= \mathbb{E}\left(\sum_{j=1}^{n+1} X_j|\mathcal{G}_n\right) \\
   &= \mathbb{E}\left(X_{n+1} + S_n|\mathcal{G}_n\right) \\
   &= \mathbb{E}\left(X_{n+1}|\mathcal{G}_n\right) + \mathbb{E}\left(S_n|\mathcal{G}_n\right) \\
   &= \mathbb{E}\left(X_{n+1}\right) + S_n \qquad \text{since } X_{n+1} \perp\!\!\!\perp \mathcal{G}_n \text{ and } S_n \in m\mathcal{G}_n \\
   &= 0 + S_n.
   \end{aligned}$$
2. We have clearly $T_i : \Omega \to \mathbb{N}$ and they are measurable (hence random variables) because S_n, S_{n-1} are random variables. Let's check the condition $\{T_i = n\} \in \mathcal{G}_n,\ \forall\, n \geqslant 0$. Remember that this is equivalent to check $\{T_i \leqslant n\} \in \mathcal{G}_n,\ \forall\, n \geqslant 0$.
   $$\begin{aligned}
   \{T_1 \leqslant n\} &= \left\{\omega : \inf\{m \geqslant 1 : \underbrace{S_m - S_{m-1}}_{X_m} = 1\} \leqslant n\right\} \\
   &= \left\{\omega : \min\{m \geqslant 1 : X_m = 1\} \leqslant n\right\} \\
   &= \left\{\omega : X_m(\omega) = 1 \text{ for some } m \leqslant n\right\} \\
   &= \bigcup_{m=1}^{n} \underbrace{\{X_m(\omega) = 1\}}_{\in \mathcal{G}_n,\ \forall\, m \leqslant n} \in \mathcal{G}_n.
   \end{aligned}$$

So T_1 is a stopping time. Now check $\{T_2 \leqslant n\}$.

$$
\begin{aligned}
\{T_2 \leqslant n\} &= \Big\{\omega : S_m(\omega) = 1 \text{ for some } m \leqslant n\Big\} \\
&= \bigcup_{m=1}^{n} \{S_m = 1\} \\
&= \bigcup_{m=1}^{n} \underbrace{\left\{\sum_{j=1}^{n} X_j = 1\right\}}_{\in\, \mathcal{G}_n,\ \forall\, m \leqslant n \text{ because } X_j \,\in\, m\mathcal{G}_n \text{ for } j \leqslant m \leqslant n.} \in \mathcal{G}_n.
\end{aligned}
$$

So T_2 is a $\mathcal{G}$-stopping time.

3. Let's calculate $\mathbb{E}S_{T_2}$ first. We have $S_{T_2} = 1$ a.s., hence $\mathbb{E}\left(S_{T_2}\right) = 1$. Notice that $\mathbb{E}(S_0) = 0 \neq \mathbb{E}(S_{T_2})$, so the variation/extension of the Doob's theorem can't apply. This means that T_2 must be unbounded... Let's calculate $\mathbb{E}S_{T_1}$ now. Notive that T_1 is unbounded hence Doob's theorem does not apply. Let's check whether S_{T_1} is integrable. Since $\{\{T_1 = n\}, n \in \mathbb{N}\}$ is a partition of Ω we have:

$$
\begin{aligned}
\mathbb{E}\left|S_{T_1}\right| &= \sum_{n=1}^{\infty} \mathbb{E}\left(|S_n|\, \mathbb{1}_{\{T_1=n\}}\right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left[\left|\sum_{j=1}^{n} X_j\right| \mathbb{1}_{\{T_1=n\}}\right] \\
&= \sum_{n=1}^{\infty}\sum_{j=1}^{n} \mathbb{E}\left[\left|X_j\right| \mathbb{1}_{\{T_1=n\}}\right] \\
&= \sum_{j=1}^{\infty}\sum_{n=j}^{\infty} \mathbb{E}\left[\left|X_j\right| \mathbb{1}_{\{T_1=n\}}\right] \\
&= \sum_{j=1}^{\infty} \mathbb{E}\left[\left|X_j\right| \mathbb{1}_{\{T_1\geqslant j\}}\right] \\
&= \mathbb{E}|X_1| \sum_{j=1}^{\infty} \mathbb{P}(\{T_1 \geqslant j\}) \\
&= \mathbb{E}|X_1| \cdot \mathbb{E}(T_1).
\end{aligned}
$$

The last equation comes form the fact that X_j are i.i.d., that $X_j \in m\mathcal{G}_j$ and that $\{T_1 \geqslant j\} = \{T_1 \leqslant j-1\}^c \in \mathcal{G}_{j-1}$, so we know that $X_j \perp\!\!\!\perp \{T_i \geqslant j\}$ (for $X \geqslant 0$, $\mathbb{E}X = \int_0^\infty \mathbb{P}(X \geqslant x)\,\mathrm{d}x = \sum_{k=0}^\infty \mathbb{P}(X \geqslant x))$[8]. We have $\mathbb{P}(T_1 = k) = \frac{1}{2} \cdot \frac{1}{2^{k-1}} = \frac{1}{2^k}$, hence

$$
\mathbb{E}T_1 = \sum_{k=1}^{\infty} k\frac{1}{2^k} < \infty \implies S_{T_1} \in L^1(\Omega).
$$

Doing the same computations without absolute value we get

$$
\mathbb{E}(S_{T_1}) \underbrace{=}_{\text{before that we put } \leqslant \text{ due to } |\sum X_j| \leqslant \sum |X_j|} \mathbb{E}(X_1)\mathbb{E}(T_1) = 0 \cdot \mathbb{E}(T_1) = 0.
$$

There is an alternative (shorter) solution to compute $\mathbb{E}S_{T_2}$, using the extended Doob's Theorem since $T_2 < \infty\, \mathbb{P} -$ a.s.. Given that $T_2 \sim \mathsf{Geom}\left(\frac{1}{2}\right)$ hence

$$
\begin{cases} \mathbb{E}T_2 < \infty \\ T_2 \geqslant 0 \end{cases} \implies T_2 < +\infty \mathbb{P} - \text{ a.s.}
$$

Moreover, the increments of the martingale S are

$$
S_n - S_{n-1} = X_n \sim \mathsf{Be}\left(\frac{1}{2}\right)
$$

hence $|X_n(\omega)| \leqslant 1\, \forall\, \omega \in \Omega$. This tells us that S_{T_2} is integrable and $\mathbb{E}\left(S_{T_2}\right) = \mathbb{E}(S_0) = 0$.

[8]What the fuck?

Exercise 5

[redacted]

1. (a) Measurability is obvious by definition of $\mathcal{G}_n$.

   (b) To check for integrability:

$$\mathbb{E}|Y_n| = \mathbb{E}(Y_n) = \mathbb{E}\left(p^{-n}\mathbb{1}_{\{X\geqslant n\}}\right) = p^{-n}\mathbb{P}(X \geqslant n) = p^{-n}p^n = 1.$$

   (c) Martingale property: $\mathbb{E}(Y_{n+1}|\mathcal{G}_n = Y_n$. We must show that, $\forall\, G \in \mathcal{G}$,

$$\mathbb{E}\left(\mathbb{1}_G Y_{n+1}\right) = \mathbb{E}\left(\mathbb{1}_G Y_n\right).$$

Since we know that $\mathcal{G}_n = \sigma(Y_1, \ldots, Y_n)$, then

$$\begin{aligned} Y_n &= 0 &\iff&& X &< n \\ Y_n &= p^{-n} &\iff&& X &\geqslant n. \end{aligned}$$

So G is of the form

$$G = \{Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n\}.$$

If $Y_n \neq 0$ then we have $y_1, y_2, \ldots, y_{n-1} \neq 0$ because $X \geqslant n > n-1 > \ldots > 1$. If $y_n = 0 \iff X < n$ then

$$X > n+1 \iff Y_{n+1} = 0.$$

This also means=

$$\mathbb{E}\left(\mathbb{1}_G Y_n\right) = \begin{cases} 0 & \text{if } y_n = 0 \\ p^{-n} & \text{if } y_n \neq 0 \end{cases}$$

$$\mathbb{E}\left(\mathbb{1}_G Y_{n+1}\right) = \begin{cases} 0 & \text{if } y_n = 0 \text{ (because of)} \\ ? & \text{if } y_n \neq 0. \end{cases}$$

There are two possible values for Y_{n+1}:

$$\begin{aligned} Y_{n+1} &= 0 &\iff&& X &< n+1 \\ Y_{n+1} &= p^{-n-1} &\iff&& X &\geqslant n+1. \end{aligned}$$

If $y_n \neq 0$ we know that $X \geqslant n$. So we get

$$\begin{aligned} \mathbb{1}_G Y_{n+1} &= 0 &\iff&& X &< n+1 && \text{given that } X \geqslant n \\ \mathbb{1}_G Y_{n+1} &= p^{-n-1} &\iff&& X &\geqslant n+1 && \text{given that } X \geqslant n \end{aligned}$$

that is

$$\begin{aligned} \mathbb{P}(\mathbb{1}_G Y_{n+1} = 0) &= \mathbb{P}(X < n+1 | X \geqslant n) \\ &= \frac{\mathbb{P}(X < n+1, X \geqslant n)}{\mathbb{P}(X \geqslant n)} \\ &= \frac{\mathbb{P}(X = n) - \mathbb{P}(X \geqslant n+1)}{\mathbb{P}(X \geqslant n)} \\ &= \frac{p^n - p^{n+1}}{p^n} \\ &= 1 - p \end{aligned}$$

and

$$
\begin{aligned}
\mathbb{P}(\mathbb{1}_G Y_{n+1} = p^{n-1}) &= \mathbb{P}(X \geqslant n+1 | X \geqslant n) \\
&= \frac{\mathbb{P}(X \geqslant n+1, X \geqslant n)}{\mathbb{P}(X \geqslant n)} \\
&= \frac{\mathbb{P}(X \geqslant n+1)}{\mathbb{P}(X \geqslant n)} \\
&= \frac{p^{n+1}}{p^n} \\
&= p.
\end{aligned}
$$

Thus if $y_n \neq 0$ we get

$$
\begin{aligned}
\mathbb{E}(\mathbb{1}_G Y_{n+1}) &= 0 \cdot \mathbb{P}(\mathbb{1}_G Y_{n+1} = 0) + p^{-n-1}\mathbb{P}(\mathbb{1}_G Y_{n+1} = p^{-n-1}) \\
&= 0 + p^{-n-1} \cdot p = p^{-n}
\end{aligned}
$$

and plugging this into we get . Thus holds and $(Y_n)_n$ is a martingale with respect to $\mathcal{G}$.

2. We have

$$
\begin{aligned}
T &= \min\{n \geqslant 0 : Y_n = 0\} \\
&= \min\{n \geqslant 0 : \mathbb{1}_{\{X \geqslant n\}} = 0\} \\
&= X + 1.
\end{aligned}
$$

Note that the indicator $\mathbb{1}_{\{X \geqslant n\}}$ is 1 if $X \geqslant n$ and when $X < n$ it switches to 0. $X \geqslant n$ means $X = n - 1 \iff n = X + 1$. Thus the conditions for stopping time reduce to

$$
\{T \leqslant n\} = \{X + 1 \leqslant n\} \in \mathcal{G}_n
$$

and

$$
\{X + 1 \leqslant n\}^n = \{X + 1 > n\} = \{X > n + 1\} = \underbrace{\{X \leqslant n\} \in \mathcal{G}_n}_{\text{since } Y_n = p^{-n}\mathbb{1}_{\{X \geqslant n\}}}
$$

so $\{T \leqslant n\} \in \mathcal{G}_n$, hence it is a stopping time.

3. Consider

$$
\begin{aligned}
\mathbb{E}X &= \sum_{n \geqslant 1} \mathbb{P}(X \geqslant n) \\
&= \sum_{n \geqslant 1} p^n \\
&= -1 + \sum_{n \geqslant 0} p^n \\
&= -1 + \frac{1}{1-p}
\end{aligned}
$$

We know that

$$
\begin{aligned}
\mathbb{E}|T| = \mathbb{E}|X + 1| \leqslant 1 + \mathbb{E}|X| = 1 + \mathbb{E}X \\
&= 1 + \sum_{n=0}^{\infty} n \cdot \mathbb{P}(X = x) \\
&= 1 + \sum_{n=0}^{\infty} n(\mathbb{P}(X \geqslant n) - \mathbb{P}(X \geqslant n+1)) \\
&= 1 + \sum_{n=0}^{\infty} (p^n - p^{n+1}).
\end{aligned}
$$

Compute the members one by one.

$$
\sum_{n=0}^{\infty} p^n = \frac{1}{1-p} \qquad \text{so} \qquad \sum_{n=0}^{\infty} np^{n-1} = \frac{1}{(1-p)^2} \qquad \text{for } p < 1
$$

so

$$\sum_{n=0}^{\infty} np^n = p \sum_{n=0}^{\infty} np^{n-1} = \frac{p}{(1-p)^2}$$

and

$$\sum_{n=0}^{\infty} np^{n+1} = p^2 \sum_{n=0}^{\infty} np^{n-1} = \frac{p^2}{(1-p)^2}.$$

This gives us

$$\begin{aligned}
&= 1 + \frac{p}{(1-p)^2} - \frac{p^2}{(1-p)^2} \\
&= \frac{(1-p)^2 + p - p^2}{(1-p)^2} \\
&= \frac{1 + \cancel{p^2} - 2p + p - \cancel{p^2}}{(1-p)^2} \\
&= \frac{1-p}{(1-p)^2} \\
&= \frac{1}{1-p} < \infty.
\end{aligned}$$

Remark

To compute $\mathbb{E}(X)$ we can also use the formula for positive random variables:

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X \geqslant n) \qquad \text{or} \qquad \mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X \geqslant x)\,\mathrm{d}x$$

and get the same result $\mathbb{E}(X) = \frac{1}{1-p} - 1$. Take

$$A := \{(x, \omega) : 0 \leqslant x \leqslant X(\omega)\}.$$

Then

$$\begin{aligned}
\int_{\Omega} \int_0^{\infty} \mathbb{1}_A(x, \omega)\,\mathrm{d}x\,\mathrm{d}\mathbb{P} &= \int_0^{\infty} \int_{\omega} \mathbb{1}_A(x, \omega)\,\mathrm{d}\mathbb{P}\,\mathrm{d}x \\
&= \int_0^{\infty} \mathbb{P}(X \geqslant x)\,\mathrm{d}x.
\end{aligned}$$

Exercise 6

[redacted]

1. (a) Measurability: obvious by definition.
   (b) Integrability: obvious by definition.
   (c) Martingale property:

$$\begin{aligned}
\mathbb{E}[Y_{n+1}|\mathcal{G}_n] &= \mathbb{E}[\mathbb{E}(X|\mathcal{G}_{n+1})|\mathcal{G}_n] \\
&= \mathbb{E}[X|\mathcal{G}_n] = Y_n.
\end{aligned}$$

2.

$$
\begin{aligned}
\mathbb{E}|Y_T| &= \mathbb{E}\left|\sum_{n=0}^{\infty} Y_n \mathbb{1}_{\{T=n\}}\right| \\
&\leqslant \sum_{n=0}^{\infty} \mathbb{E}\left(\mathbb{E}(|X||\mathcal{G}_n)\mathbb{1}_{\{T=n\}}\right) \\
\mathbb{1}_{\{T=n\}} \in m\mathcal{G}_n \longrightarrow &= \sum_{n=0}^{\infty} \mathbb{E}\left[\mathbb{E}(|X|\mathbb{1}_{\{T=n\}}|\mathcal{G}_n)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[|X|\mathbb{1}_{\{T=n\}}\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} |X|\mathbb{1}_{\{T=n\}}\right] \\
&= \mathbb{E}|X| < +\infty.
\end{aligned}
$$

Exercise 7

[redacted]

(a) Measurability: follows from the measurability of X_i and the fact that Y_n is a product of X_i with $i \leqslant n$.

(b) Integrability:

$$
\mathbb{E}|Y_n| = \mathbb{E}\left|\prod_{i=1}^{n} X_i\right| = \prod_{i=1}^{n} \mathbb{E}|X_i| = 1 < \infty.
$$

(c) Martingale probability:

$$
\begin{aligned}
\mathbb{E}(Y_{n+1}|\mathcal{G}_n) &= \mathbb{E}\left(\prod_{i=1}^{n+1} X_i|\mathcal{G}_n\right) \\
&= \mathbb{E}(X_{n+1} \cdot Y_n|\mathcal{G}_n) \\
Y_n \in m\mathcal{G}_n \longrightarrow &= Y_n\mathbb{E}(X_{n+1}|\mathcal{G}_n) \\
X_{n+1} \perp\!\!\!\perp \mathcal{G}_n \longrightarrow &= Y_n\mathbb{E}(X_{n+1}) = Y_n.
\end{aligned}
$$

1.1.8 Exercise class 8

Revise with Kotatsu!

Theorem 1.1.6

Martingale convergence theorem. Let X be a submartingale. Suppose that $(X_n)_n$ is bounded in L^1, uniformly in n (that is, $\sup_n \mathbb{E}|X_n| < \infty$). Then X_n converges a.s. to an integrable random variable.

Remark

$$
\sup_n \mathbb{E}|X_n| < \infty \iff \sup_n \mathbb{E}X_n^+ < \infty.
$$

Definition 1.1.7

A sequence of random variables $(X_n)_n$ is said to be **uniformly integrable** if

$$
\lim_{b\to\infty} \sup_n \mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>b\}}\right] = 0.
$$

If a random variable is uniformly integrable then it is L^1-bounded. If it is L^p-bounded and $p > 1$ then it is uniformly integrable.

Theorem 1.1.7

Martingale convergence theorem (equivalence). Let X be a submartingale. Then X converges a.s. and in L^1 to an integrable random variable if and only if it is uniformly integrable.

Exercise 1

[redacted]

If $\lim_n S_n$ exists and it is finite a.s. it should be absolutely summable so it must be that $|X_i| \to 0$ a.s. as $i \to \infty$. This however is impossible because X_i are i.i.d.. So S_n does not convergence a.s..
If $\sup_n \mathbb{E}|S_n| < \infty$ then by the martingale convergence theorem (if S_n is a martingale) we would have S_n converging a.s., but this is not the case and thus S_n is not L^1-bounded.

(a) Measurability with respect to natural filtration.

(b) Integrability:

$$\mathbb{E}|S_n| \leqslant \sum_{i=1}^{n} \mathbb{E}|X_i| = n \cdot c < \infty.$$

(c) Martingale property:

$$\begin{aligned}
\mathbb{E}\left[S_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i + X_n\right] \\
&= \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1}|\mathcal{F}_n) \\
&= S_n + \mathbb{E}(X_{n+1}) = S_n,
\end{aligned}$$

Exercise 2

[redacted]

$(X_n)_n$ is a martingale (same as the exercise 1, we used only the fact that ξ_i are independent and $\mathbb{E}\xi_i = 0$). Thus we could apply martingale convergence theorem:

$$\sup_n \mathbb{E}|X_n| = \sup_n \mathbb{E}\left|\sum_{i=1}^{n} \xi_i\right| \leqslant \sup_n \sum_{i=1}^{n} \mathbb{E}|\xi_i| \leqslant \sup_n \sum_{i=1}^{n} \frac{1}{2^n} = 1.$$

Hence $X_n \to X$ a.s. for some $X \in L^1$ (since $(X_n)_n$ is L^1-bounded).

Exercise 3

[redacted]

$$\mathbb{E}\frac{Y_i}{i} = \frac{1}{i}\left(1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2}\right) = 0.$$

Thus X_n is a martingale with respect to its natural filtration, because it is a sum of independent random variables with mean 0 (see exercise 1). Let's check whether $(X_n)_n$ is L^1-bounded.

$$\begin{aligned}
\sup_n \mathbb{E}|X_n| &= \sup_n \mathbb{E}\left|\sum_{i=1}^{n} \frac{Y_i}{i}\right| \\
&\leqslant \sup_n \sum_{i=1}^{n} \frac{1}{i}\mathbb{E}|Y_i| \\
&= \sup_n \sum_{i=1}^{n} \frac{1}{i}\left(1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}\right) \\
&= \sum_{i=1}^{n} \frac{1}{i} = +\infty.
\end{aligned}$$

From this alone we cannot conclude anything. Let's see if $(X_n)_n$ is L^2-bounded. This would imply L^1-boundedness, hence convergence a.s. to a L^1 random variable.

$$\begin{aligned}
\mathbb{E}|X_n|^2 &= \mathbb{E}\left(\sum_{i=1}^{n} \frac{Y_i}{i}\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^{n} \frac{Y_i^2}{i^2} + \sum_{\substack{i,j=1\\ i\neq 1}} \frac{Y_iY_j}{ij}\right) \\
&= \sum_{i=1}^{n} \frac{1}{i^2}\mathbb{E}Y_i^2 + \sum_{\substack{i,j=1\\ i\neq 1}} \frac{1}{ij}\underbrace{\mathbb{E}(Y_iY_j)}_{\underbrace{\mathbb{E}Y_i}_{=0}\underbrace{\mathbb{E}Y_j}_{=0}} \\
&= \sum_{i=1}^{n} \frac{1}{i^2} \cdot 1 + 0.
\end{aligned}$$

So

$$\sup_n \mathbb{E}|X_n|^2 = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

hence $(X_n)_n$ is L^2-bounded $\implies$ it is uniformly integrable $\implies$ converges a.s. to a random variable in L^1 and it converges also in L^1.

Exercise 4

[redacted]

1. We formalise the procedure to construct X_n and write the conditional p.d.f.:

$$\begin{aligned}
f_{X_1}(x) &= \mathbb{1}_{[0,1]}(x) = \tfrac{1}{x_0}\mathbb{1}_{[0,x_0]}(x) \qquad \text{with } x_0 = 1 \\
f_{X_2|X_1=x_1}(x) &= \tfrac{1}{x_1}\mathbb{1}[0,x_1](x) \\
&\vdots \\
f_{X_n|X_{n-1}=x_{n-1},\dots,X_1=x_1}(x) &= f_{X_n|X_{n-1}=x_{n-1}}(x) = \tfrac{1}{x_{n-1}}\mathbb{1}_{[0,x_{n-1}]}(x).
\end{aligned}$$

Thus we have

$$\mathbb{E}(X_n|\mathcal{G}_{n-1}) = \mathbb{E}(X_n|X_{n-1})$$

and

$$\mathbb{E}(X_n|X_{n-1}=x_{n-1}) = \int x f_{X_n|X_{n-1}=x_{n-1}}(x)\,\mathrm{d}x = \frac{1}{x_{n-1}}\int_0^{x_{n-1}} x\,\mathrm{d}x = \frac{1}{x_{n-1}}\cdot\frac{x_{n-1}^2}{2} = \frac{x_{n-1}}{2}.$$

So $\mathbb{E}(X_n|X_{n-1}) = \frac{X_{n-1}}{2}$ by the freezing lemma. This implies that $\frac{X_{n-1}}{2}$ is a version of the conditional expectation $\mathbb{E}(X_n|\mathcal{G}_{n-1})$. This helps because by the tower property

$$\begin{aligned}
\mathbb{E}(X_n) &= \mathbb{E}(\mathbb{E}(X_n|\mathcal{G}_{n-1})) \\
&= \mathbb{E}\left(\frac{X_{n-1}}{2}\right) \\
&= \frac{1}{2}\mathbb{E}(\mathbb{E}(X_{n-1}|\mathcal{G}_{n-2})) \\
&= \frac{1}{2}\mathbb{E}(\frac{X_{n-2}}{2}) \\
\text{by recurrence } \rightarrow \quad & \vdots \\
&= \frac{1}{2^{2-1}}\mathbb{E}X_1 = \frac{1}{2^{n-1}}\cdot\frac{1}{2} = \frac{1}{2^n}.
\end{aligned}$$

2. Given any $(C_n)_n$ we have

$$\mathbb{E}(Y_n|\mathcal{G}_{n-1}2) = C_n\mathbb{E}(X_n|\mathcal{G}_{n-1}) = \frac{C_n}{2}X_{n-1} \stackrel{?}{=} Y_{n-1}.$$

2 For $Y_{n-1} = C_{n-12}X_n$ we must have $C_{n-1} = \frac{C_n}{2}$ i.e. $C_n = 2C_{n-1}$ defined recursively with C_1 initial condition to get

$$C_n = 2C_{n-1} = 2^2C_{n-2} = 2^{n-1}C_1.$$

3. Notice that if $C_1 > 0$ then $Y_n \geqslant 0$, if $C_1 < 0$ then $Y_n \leqslant 0$. Let's consider the case $C_1 > 0$ (the other case is analogous with a minus sign).

$$\mathbb{E}|Y_n| = \mathbb{E}(Y_n) = \mathbb{E}[C_nX_n] = 2^{n-1}C_1\mathbb{E}X_n = 2^{n-1}C_1\cdot\frac{1}{2^n} = \frac{C_1}{2}$$

and this implies

$$\sup_n \mathbb{E}|Y_n| = \frac{C_1}{2} < \infty \qquad \text{(independent of n.)}$$

Hence by the martingale convergence theorem we have $Y_n \rightarrow Y_\infty$ a.s. wth $Y_\infty \in L^1$.

Exercise 5

[redacted]

1. (a) Measurability is obvious.
   (b) Integrability:

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_n \leqslant \mathbb{E}\left(X_{n-1} + \frac{X_{n-1}}{2}\right) \leqslant \cdots \text{const}\cdot\mathbb{E}(X_0) \leqslant \infty.$$

(c) Martingale property: we have to show $\mathbb{E}(X_{n+1}|\mathcal{G}_n) = X_n$ for every $n \in \mathbb{N}$. Notice that $\mathbb{E}(X_{n+1}|\mathcal{G}_n) = \mathbb{E}(X_{n+1}|X_n)$. We want to apply the freezing lemma so we calculate

$$\mathbb{E}(X_{n+1}|X_n = x_n) = \frac{1}{2}\left(\frac{x_n}{2} + \frac{3}{2}x_n\right) = x_n$$

thus by the freezing lemma we have

$$\mathbb{E}(X_{n+1}|X_n) = X_n.$$

2.

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_0 < \infty \implies \sup_n \mathbb{E}|X_n| = \mathbb{E}X_0$$

hence $X_n \to X$ a.s. for some $X \in L^1$. In order to determine X we make the following calculations"

$$X_{n+1} = \begin{cases} \frac{X_n}{2} \\ \frac{X_n}{2} + X_n \end{cases} \implies |X_{n+1} - X_n| = \begin{cases} \left|\frac{X_n}{2} - X_n\right| & = \frac{X_n}{2} \\ \left|\frac{X_n}{2} + X_n - X_n\right| & = \frac{X_n}{2} \end{cases}$$

$$\Downarrow$$

$$2|X_{n+1} - X_n| = X_n$$

Taking the a.s. limit on both sides we get $2|X - X| = X$ which implies $X = 0$.

3. Since $\mathbb{E}[|X_n - 0|] = \mathbb{E}|X_n| = \mathbb{E}X_n \neq 0$ since $X_0 > 0$ a.s. this implies that X_n does not converge in 0 in L^1.

Exercise 6

[redacted]

1. (a) Measurability holds because M_n is a continuous function of $X_1, \dots, X_n$.

   (b) Integrability:

$$\begin{aligned}
\mathbb{E}|M_n| = \mathbb{E}M_n &= \mathbb{E}\left[(2p)^{S_n}(2q)^{n-S_n}\right] \\
&= \mathbb{E}\left[(2p)^{\sum_{i=1}^n X_i}(2q)^{\sum_{i=1}^n (1-X_i)}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^n (2p)^{X_i}(2q)^{1-X_i}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[(2p)^{X_i}(2q)^{1-X_i}\right] \\
&= \left[\frac{1}{2}(2p)^0 \cdot (2q)^{1-0} + \frac{1}{2}(2p)^1(2q)^{1-1}\right]^n \\
&= \left[\frac{1}{2}2q + \frac{1}{2}2p\right]^n \\
&= (p+q)^n \\
&= 1^n = 1 < \infty.
\end{aligned}$$

   (c) Martingale property: setting $Z_i = (2p)^{X_i}(2q)^{1-X_i}$ we have $M_n = \prod_{i=1}^n Z_i$ and Z_i $\perp\!\!\!\perp$. Then

$$\begin{aligned}
\mathbb{E}(M_{n+1}|G_n) &= \mathbb{E}(Z_{n+1}M_n|\mathcal{G}_n) \\
M_n \in m\mathcal{G}_n \to \quad &= M_n\mathbb{E}(Z_{n+1}|\mathcal{G}_n) \\
Z_{n+1} \perp\!\!\!\perp \mathcal{G}_n \to \quad &= M_n\mathbb{E}_{(Z_{n+1})} \\
\mathbb{E}Z_{n+1} = 1 \to \quad &= M_n.
\end{aligned}$$

2.
$$\mathbb{E}|M_n| = \mathbb{E}M_n = \mathbb{E}\prod_{i=1}^{n} Z_i = \prod_{i=1}^{n} \mathbb{E}Z_i = 1$$

thus $\sup_n \mathbb{E}|M_n| = 1$. By martingale convergence theorem $M_n \to M$ a.s. for some $M \in L^1$. To determine M we must distinguish for different values of p.

- If $p = \frac{1}{2}$ then $M_n \equiv 1$ and so $M = 1$.
- If $p \neq \frac{1}{2}$ we note that

$$\begin{aligned}
M_n &= (2p)^{S_n}(2q)^{n-S_n} \\
&= \exp\left\{\log\left[(2p)^{S_n}(2q)^{n-S_n}\right]\right\} \\
&= \exp\left\{\log(2p)^{S_n} + \log(2q)^{n-S_n}\right\} \\
&= \exp\left\{S_n \log(2p) + (n - S_n)\log(2q)\right\} \\
&= \exp\left\{n\left[\frac{S_n}{n}\log(2p) + \left(1 - \frac{S_n}{n}\right)\log(2q)\right]\right\}.
\end{aligned}$$

By SLLN we have $\frac{S_n}{n} \to \mathbb{E}X_1$ a.s. with $\mathbb{E}X_1 = \frac{1}{2}$. Thus setting

$$Y_n = \frac{S_n}{n}\log(2p) + \left(1 - \frac{S_n}{n}\right)\log(2q)$$

we have

$$Y_n \xrightarrow{\text{a.s.}} \frac{1}{2}\log(2p) + \frac{1}{2}\log(2q) = \frac{1}{2}\log(4pq)$$

where

$$4pq = 4p(1-p) < 1 \implies \frac{1}{2}\log(4pq) < 0.$$

We also have $M_n = \mathbb{E}(n \cdot Y_n)$ and by limit of composite functions (on a set of probability 1) we get

$$\lim_{n\to\infty} M_n = e^{-\infty} = 0 \qquad \text{a.s.}$$

which means $M = 0$.

Exercise 7

[redacted]

1. We must check whether $\{T \leqslant n\} \in \mathcal{F}_n,\ \forall n \in \mathbb{N}$, or equivalently if $\{T = n\} \in \mathcal{F}_n,\ \forall n \in \mathbb{N}$.

$$\{T = n\} = \{\underbrace{X_1 \neq Y_i}_{\in \mathcal{F}_1 \subset \mathcal{F}_n}, \underbrace{X_n \neq Y_n}_{\in \mathcal{F}_2 \subset \mathcal{F}_n}, \dots, X_{n-1} \neq Y_{n-1}, \underbrace{X_n = Y_n}_{\in \mathcal{F}_n}\} \in \mathcal{F}_n.$$

Remember that X and Y are $\mathcal{F}$–martingales and hence measurable.

2.
- Measurability:

$$\{T \geqslant n\} = \{T > n-1\} = \{T \leqslant n-1\}^c$$

and by the previous point we know that $\{T \leqslant n-1\} \in \mathcal{F}_{n-1}$ so $\{T \geqslant n\} \in m\mathcal{F}_{n-1}$. Moreover $X_n, Y_n \in m\mathcal{F}_n$ have $Z_n \in m\mathcal{F}_n$.
- Integrability:

$$\mathbb{E}\left[\left|(X_n - Y_n)\mathbb{1}_{\{T \geqslant n\}}\right|\right] \leqslant \mathbb{E}|X_n| + \mathbb{E}|Y_n| < \infty.$$

- Martingale property:

$$\begin{aligned}
\mathbb{E}\left[(X_n - Y_n)\mathbb{1}_{\{T \geqslant n\}}|\mathcal{G}_{n-1}\right] &= \mathbb{1}_{\{T \geqslant n\}}\mathbb{E}\left[X_n - Y_n|\mathcal{G}_{n-1}\right] \\
&= \mathbb{1}_{\{T \geqslant n\}}(X_{n-1} - Y_{n-1}) \\
&= \left(\mathbb{1}_{\{T \geqslant n-1\}} - \mathbb{1}_{T=n-1}\right)(X_{n-1}Y_{n-1}) \\
&= \mathbb{1}_{\{T \geqslant n-1\}}(X_{n-1} - Y_{n-1}) - \underbrace{\mathbb{1}_{\{T=n-1\}}(X_{n-1} - Y_{n-1})}_{=0 \text{ because if } T = n-1 \text{ it means } X_{n-1} = Y_{n-1}} \\
&= \mathbb{1}_{\{T \geqslant n-1\}}(X_{n-1} - Y_{n-1}).
\end{aligned}$$

1.2 Questions for the oral examination

Here I will try and answer to the questions for the anal examination that were on the moodle page for the academic year 2022-2023. I hope that they will be valid for the years to come as well.

1.2.1 Transition kernels: definition, example, and their usage in the extension of measures to product spaces

First of all, the definition:

Definition 1.2.1

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Let K be a mapping from $E \times \mathcal{F}$ into $\overline{\mathbb{R}}_+$. Then K is called **transition kernel** from space $(E, \mathcal{E})$ into space $(F, \mathcal{F})$ if:

- the mapping $x \mapsto K(x, B)$ is $\mathcal{E}$-measurable $\forall\, B \in \mathcal{F}$;
- the mapping $B \mapsto K(x, B)$ (the second mapping of the kernel, the one regarding the set) is a measure $\forall\, x \in E$.

Then the example:

Example 1.2.1

Take ν, a finite measure on $(F, \mathcal{F})$ and take k, a positive function on $(E \times F)$ which is measurable with respect to $\mathcal{E} \otimes \mathcal{F}$, the product σ-algebra. Then, we integrate

$$\int_B \nu(\mathrm{d}y) k(x, y) \qquad \begin{matrix} B \in \mathcal{F} \\ x \in E \end{matrix}$$

We see how this object depends on x and on the choice of B (a function of x and B...). It defines a transition kernel

$$K(x, B) = \int_B \nu(\mathrm{d}y) k(x, y) \qquad \begin{matrix} B \in \mathcal{F} \\ x \in E \end{matrix}$$

from $(E, \mathcal{E})$ into $(F, \mathcal{F})$.

Now the extensions of measures on product spaces;

Theorem 1.2.1

Extension of measures on product spaces.
Let μ be a measure on the measurable space $(E, \mathcal{E})$. Let K be a Σ-finite[a] transition kernel from space $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Then:

1. if we take our function $f(x, y)$, integrate it against our kernel $K(x, \mathrm{d}x)$ over F and then integrate again against measure μ over E, the operation

$$\pi f = \int_E \mu(\mathrm{d}x) \int_F K(x, \mathrm{d}y) f(x, y)$$

defines a measure π on $(E \times F, \mathcal{E} \otimes \mathcal{F})$;

2. if μ is σ-finite and K is σ-bounded then π is σ-finite and it is the unique measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ satisfying

$$\pi(A \times B) = \int_A \mu(\mathrm{d}x) K(x, B) \qquad \forall\, A \in \mathcal{E}, B, \in \mathcal{F}.$$

[a] Erm... what the sigma?

1.2.2 Kolmogorov's 0-1 law: proof and an example of its usage

First, the statemente of the theorem.

Theorem 1.2.2

Kolmogorov's 0-1 law
This is a theorem about independence. Let $\mathcal{G}_1, \mathcal{G}_2, \ldots$ be independent. Then

$$\mathbb{P}(H) = \begin{cases} 0 \\ 1 \end{cases} \qquad \forall H \in \tau.$$

and here's the short proof:

Proof

Start from your independency $(\mathcal{G}_1, \mathcal{G}_2, \ldots)$, then put the last $\mathcal{G}$'s in a partition (which is still an independency) $(\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n, \tau_n)$. Now consider that the tail σ-algebra is a subset of τ_n and this means that $(\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n, \tau)$ is still an independency... But we also know that this is true for every n up to ∞ and so we can extend the independency to a collection of countably many partitions, since $(\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n, \tau)_n$ us a countably infinite sequence of independencies. So

$$(\tau, \mathcal{G}_1, \mathcal{G}_2, \ldots)$$

is an independency. Do another partition (which still gives us an independency) and get the independency

$$(\tau, \tau_0)$$

but then remember that τ_0 is an *union* of the σ-algebras from 1 to ∞ while τ is an *intersection* of all τ_n including τ_0, so we have that $\tau \subset \tau_0$.
Consider now $\mathbb{P}(H \cap G)$ where $H \in \tau$ and $G \in \tau_0$. Due do independency we have that

$$\mathbb{P}(H \cap G) = \mathbb{P}(H)\mathbb{P}(G)$$

but since all elements of τ are also elements of τ_0 we can choose $G \equiv H$ and our equation becomes

$$\mathbb{P}(H \cap G) = \mathbb{P}(H \cap H) = \mathbb{P}(H) = \mathbb{P}(H)^2$$

and the solution to this can only be 0 or 1. □

We still need an example of the application... that would be, for example, the behavior in the limits of a sequence of random variables with independence between $\{X_n\}$. Then if the limit exists it is a constant, since $\lim_n X_n$ belongs to the tail σ-algebra of the sequence. Independency between random variables tells us that we can apply Kolmogorov's 0-1 law to say that $\mathbb{P}(\lim_n X_n = \infty)$ is either 0 or 1... but if we somehow know that the limit exists then it can't be ∞ so its probability of being ∞ is 0 and therefore it is a constant almost surely! This is cool because it tells us that the limit is either infinite or finite. This is useful for sub-martingales.

1.2.3 Almost sure convergence: definition, properties and characterization theorem

Start with the definition:

Definition 1.2.2

A real-valued sequence of random variables $(X_n)_n$ on $(\Omega, \mathcal{H}, \mathbb{P})$ is said to be **almost sure convergent** (a.s. convergent) is the numerical sequence

$$\Big(X_n(\omega)\Big)_n$$

converges for almost all $\omega \in \Omega$.
It is said to converge to X if X is an almost sure real-valued random variable and

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for almost all $\omega \in \Omega$.

The characterization is

Theorem 1.2.3

Characterization of almost sure convergence:
A sequence of real valued random variables $(X_n)_n$ converges to X almost surely if and only if, for every $\varepsilon > 0$,

$$\sum_n i_\varepsilon \circ |X_n - X| < \infty$$

almost surely.

Where i_ε is the indicator function of (ε, ∞):

$$i_\varepsilon(x) = \mathbb{1}_{(\varepsilon,\infty)}(x) = \begin{cases} 1, & x > \varepsilon \\ 0, & x \leqslant \varepsilon. \end{cases}$$

This basically means that in every interval from a certain point ε to ∞ the sum of the differences of X_n and X must not diverge. Here are some properties:

- **Comparison with Other Types of Convergence:**
  - If $\{X_n\}$ converges almost surely to X, then $\{X_n\}$ also converges to X in probability.
  - Almost sure convergence implies convergence in distribution, but not necessarily vice versa.

  Closure Under Linear Operations:
  - **Addition:** If $X_n \to X$ a.s. and $Y_n \to Y$ a.s., then $X_n + Y_n \to X + Y$ a.s.
  - **Multiplication:** If $X_n \to X$ a.s. and $Y_n \to Y$ a.s., then $X_n Y_n \to XY$ a.s. (provided X and Y are bounded or measurable in a compatible way).
- **Countable Additivity:**
  - **Countable Unions:** If $\{X_n\}$ converges almost surely to X, then for any countable collection of events $\{A_i\}$ with $\Pr(A_i) \to 0$, $\Pr(\limsup_{i\to\infty} A_i) = 0$.
- **Uniform Integrability:**
  - **Expectation Convergence:** If $\{X_n\}$ converges almost surely to X and $\{X_n\}$ is uniformly integrable, then $X_n \to X$ in L^1, meaning $\mathbb{E}[|X_n - X|] \to 0$.
- **Continuity from Below and Above:**
  - **From Below:** If $X_n \leqslant X_{n+1}$ a.s. for all n and $X_n \to X$ a.s., then $X_n \leqslant X$ a.s.
  - **From Above:** If $X_n \geqslant X_{n+1}$ a.s. for all n and $X_n \to X$ a.s., then $X_n \geqslant X$ a.s.
- **Interchange of Limits:**
  - **Expectation Limit:** If $X_n \to X$ a.s. and $\{X_n\}$ is uniformly integrable, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.
- **Fatou's Lemma and Monotone Convergence Theorem:**
  - **Fatou's Lemma:** If $X_n \geqslant 0$ almost surely and $X_n \to X$ a.s., then $\mathbb{E}[X] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n]$.
  - **Monotone Convergence Theorem (for non-negative random variables):** If $X_n \to X$ a.s. and $X_n \geqslant 0$ a.s., then $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n]$.

1.2.4 Borel Cantelli lemmas; proof of the divergence part

Enunciate the two lemmas side by side (and with a slightly different formulation).

Theorem 1.2.4

Borel-Cantelli lemmas

a) Let $(B_n)_n$ be a sequence of Bernoulli random variables.

$$\sum_n \mathbb{E}B_n < +\infty \implies \sum_n B_n < +\infty \text{ a.s.}$$

b) If

$$\sum_n \mathbb{E}B_n = \infty \text{ and } (B_n)_n \text{ are pairwise independent}$$

then

$$\sum_n B_n = +\infty \text{ a.s.}$$

Now let's prove the divergent part.

Note the notation!

$$p_n = \mathbb{E}B_n, \quad a_n = \sum_{i=1}^n \mathbb{E}B_i = \sum_{i=1}^n p_i$$

and we get the partial sum and the limits

$$S_n = \sum_{i=1}^n B_i, \quad S = \lim_n S_n$$

Proof

Ok, this is long. We know that, since the random variables are pairwise independent,

$$\operatorname{Var} S_n = \sum_{i=1}^n \operatorname{Var} B_n = \sum_{i=1}^n p_n(1-p_n) \leqslant \sum_{i=1}^n p_n = a_n.$$

Then fix $b \in (0,\infty)$. Since we know that $(a_n)_n$ is increasing towards ∞ then the same can be said of

$$\left(a_n - \sqrt{ba_n}\right)_n$$

since we are subtracting to a_n a quantity that goes towards ∞ more slowly. So the sequence still goes towards infinity and this means that the event $S < \infty$ is basically the limit of the increasing sequence of events

$$\left\{S < a_n - \sqrt{ba_n}\right\}_n.$$

Since $S_n \leqslant S$ we have that

$$\left\{S < a_n - \sqrt{ba_n}\right\} \subset \left\{S_n < a_n - \sqrt{ba_n}\right\}$$

and we also have that

$$\left\{S_n < a_n - \sqrt{ba_n}\right\} \subset \left\{|S_n - a_n| > \sqrt{ba_n}\right\}$$

Since we are dealing with inclusion of events we can look at this from the point of view of probability measures:

$$\mathbb{P}\left(S < a_n - \sqrt{ba_n}\right) \leqslant \mathbb{P}\left(S_n < a_n - \sqrt{ba_n}\right) \leqslant \mathbb{P}\left(|S_n - a_n| > \sqrt{ba_n}\right).$$

We take the $\limsup_n$ for two of these probabilities:

$$\limsup_n \left(\mathbb{P}\left(S < a_n - \sqrt{ba_n}\right)\right) \leqslant \limsup_n \left(\mathbb{P}\left(|S_n - a_n| > \sqrt{ba_n}\right)\right)$$

but the first term of this inequality becomes

$$\begin{aligned}\limsup_n \left(\mathbb{P}\left(S < a_n - \sqrt{ba_n}\right)\right) &= \mathbb{P}\left(\lim_n \left(\mathbb{P}\left(S < a_n - \sqrt{ba_n}\right)\right)\right)\\ &= \mathbb{P}\left(S < \infty\right)\end{aligned}$$

since the quantity $\left\{S < a_n - \sqrt{ba_n}\right\}_n$ goes towards infinity. Now we can rewrite the previous inequality and use Chebyshev's inequality:

$$\begin{aligned}\mathbb{P}\left(S<\infty\right) &\leqslant \limsup_n \left(\mathbb{P}\left(|S_n - a_n| > \sqrt{ba_n}\right)\right)\\ &\leqslant \limsup_n \left(\frac{\operatorname{Var} S_n}{ba_n}\right)\\ &\leqslant \limsup_n \left(\frac{\cancel{a_n}}{b\cancel{a_n}}\right) = \frac{1}{b}.\end{aligned}$$

Since b is arbitrary we can let it go towards infinty and thus obtain that

$$\mathbb{P}\left(S<\infty\right) \leqslant 0$$

But this means that the probability of the complementary event must be 1:

$$\mathbb{P}\left(S=\infty\right) = 1.$$

□

1.2.5 Borel Cantelli lemmas; proof of the convergence part, relations with a.s. convergence and examples of its application

Start by enunciating the theorem.

Theorem 1.2.5

First Borel-Cantelli Lemma Let $(H_n)_n$ be a sequence of events. Then

$$\sum_n \mathbb{P}(H_n) < +\infty \implies \sum_n \mathbb{1}_{H_n} < +\infty \quad \text{a.s.}$$

The proof is no shit:

Proof

Denote $N = \sum \mathbb{1}_{H_n}$ so that $\sum_n \mathbb{P}(H_n) = \sum_n \mathbb{E}\mathbb{1}_{H_n} = \mathbb{E}\sum \mathbb{1}_{H_n} = \mathbb{E}N$. This means that the new claim is

$$\text{“If } \mathbb{E}N < \infty \text{ then } N < \infty\text{”}$$

But this is true because if the expectation is finite so is its random variable. Of course almost surely... □

Now the implications for a.s. convergence:

Proposition 1.2.1

Let

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < +\infty \qquad \forall\, \varepsilon > 0.$$

Then

$$X_n \xrightarrow{\text{a.s.}} X.$$

Proposition 1.2.2

Suppose that there exists a sequence $(\varepsilon_n)_n$ decreasing to 0 such that

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < +\infty.$$

Then

$$X_n \xrightarrow{\text{a.s.}} X.$$

So we don't need ε to be constant, but just to be decreasing to 0.

Proposition 1.2.3

Suppose that there exists a sequence of positive numbers $(\varepsilon_n)_n$ such that

$$\sum_n \varepsilon_n < +\infty, \qquad \sum_n \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty$$

Then X_n converges almost surely.

1.2.6 Convergence in probability: definition, properties and theorem on the metric for convergence in probability

Let's start with the definition:

Definition 1.2.3

Let $(X_n)_n$ be a sequence of real-valued random variables. Then $(X_n)_n$ converges to a further real-valued random variable **in probability** if

$$\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0 \qquad \forall\, \varepsilon > 0.$$

I am not sure about properties, why can't you state things clearly? Fuck off.

Theorem 1.2.6

Characterization theorem for convergence in probability.

i) if $(X_n)_n$ converges to X almost surely, then it converges to X in probability;

ii) if $(X_n)_n$ converges in probability to X, then it has a subsequence converging to the same random variable X almost surely;

iii) if every subsequence of the main sequence has a further subsequence converging to X almost surely, then the main sequence converges to X in probability.

Remark

Convergence in probability is preserved under arithmetic operations.

So, for example if we have $X \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$ then $(X_n + Y_n)_n \xrightarrow{\mathbb{P}} X + Y$.
We will now introduce a metric for convergence in probability: since we are talking about *distance*, we may think[9] that this has something to do with *metric spaces*, where we measure the distance between different objects. Indeed there is a connection between measure spaces and metric spaces...

Let us introduce now a metric for convergence in probability. If we want to calculate a metric between random variables X and Y we can define the following metric:

$$d(X,Y) = \mathbb{E}(|X - Y| \wedge 1).$$

Remark

1. $d(X,Y) = 0 \iff X = Y$ a.s.
2. $d(X,Y) + d(Y,Z) \geqslant d(X,Z)$.

d is a metric on the space of real-valued random variables if X and Y are indentified as the same random variable if $X = Y$ almost surely.

[9] A remark that reeks of overestimation.

Proposition 1.2.4

$$X_n \xrightarrow{\mathbb{P}} X \iff d(X_n, X) \xrightarrow[n\to\infty]{} 0.$$

1.2.7 Uniform integrability: definitions and its consequences

Start with the definition

Definition 1.2.4

A collection of random variables K is said to be **uniformly integrable** if

$$k(b) = \sup_{X\in K} \left(\mathbb{E}|X|\mathbb{1}_{(|X|>b)}\right)$$

goes to 0 as $b \to \infty$.

What about the properties?

1. If K is finite and each $X \in K$ is integrable then K is uniformly integrable.
2. If K is dominated by an integrable random variable Z then it is uniformly integrable.
3. Uniform integrability of a collection K implies the so-called L^1-boundedness, which means

$$k \subset L^1 \quad \text{and} \quad k(0) = \sup_K \mathbb{E}|X| < \infty.$$

   Note that $k(0)$ considers the whole random variable without truncation.
4. We know that uniform integrability implies L^1... but the converse is not true. We can prove it by a counterexample.
5. If K is L^p-bounded with $p > 1$ then it is uniformly integrable with $f(x) = x^p$.
   To prove this we recur to the following proposition:

   Proposition 1.2.5

   Suppose it exists a positive borel function f on $\mathbb{R}_+$ such that

$$\lim_{x\to\infty} \frac{f(x)}{x} = \infty$$

   and write

$$c = \sup_{X\in K} \mathbb{E} f \circ |X|. \tag{$\star$}$$

   If $c \leqslant \infty$ then K is uniformly integrable.

and one last theorem

Theorem 1.2.7

The following are equivalent:

1. K is uniformly integrable;
2. $h(b) = \sup_K \int_b^{+\infty} \mathrm{d}y \mathbb{P}(|X| > y) \xrightarrow[b\to\infty]{} 0$;
3. $\sup_K \mathbb{E} f \circ |X| < +\infty$ for some increasing convex function f on $\mathbb{R}_+$ such that

$$\lim_{x\to\infty} \frac{f(x)}{x} = +\infty.$$

1.2.8 Law of large numbers: proof of the related theorems

What the hell? This is the definition

Theorem 1.2.8

Law of large numbers:
Let $(X_n)_n$ be a sequence of pairwise independent random variables with the same distribution as X. If $\mathbb{E}X$ exists (infinite values are admitted!) then

$$\overline{X}_n \xrightarrow[n]{\text{a.s.}} \mathbb{E}X.$$

To do this we need two theorems with their related proof.

Theorem 1.2.9

Consider a sequence of real valued, pairwise independent random variables $X_1, X_2, \ldots$ with finite common expectation and variance

$$\mathbb{E}X_n = a, \;\; \mathbb{V}\text{ar}\, X_n = b.$$

Then:

1. $\overline{x}_n = \dfrac{\sum_{i=1}^n X_i}{n} \xrightarrow{L^2} a$;
2. $\overline{x}_n \xrightarrow{\mathbb{P}} a$;
3. $\overline{x}_n \xrightarrow{\text{a.s.}} a$.

Proof

1. Define
$$S_n = n\overline{X}_n.$$
Now we have, due to independency,
$$\mathbb{E}S_n = na, \qquad \mathbb{V}\text{ar}\, S_n = nb$$
and, consequently,
$$\mathbb{E}\overline{X}_n = a, \qquad \mathbb{V}\text{ar}\, \overline{X}_n = \frac{b}{n}.$$
To prove L^2-convergence, consider
$$\mathbb{E}|\overline{X}_n - a|^2 = \mathbb{V}\text{ar}\, \overline{X}_n \xrightarrow[b\to\infty]{} 0$$
so $\overline{X}_n \xrightarrow{L^2} a$.
2. This one comes from Markov's inequality.
3. Assume $X_n \geqslant 0$ (without loss of generality as always since we can just divide negative and positive part). Consider a subsequence $N = (n_k)_{k\in\mathbb{N}^*}$ where $n_k = k^2$. Apply Chebyshev's inequality
$$\mathbb{P}(|\overline{X}_n - a| > \varepsilon) \leqslant \frac{b}{\varepsilon^2 k^2}$$
and by summing everything
$$\varepsilon^2 \sum_{n\in N} \mathbb{P}(|\overline{X}_n - a| > \varepsilon) \leqslant \sum_{k=1}^{\infty} \frac{b}{k^2} \leqslant \infty$$
so we can apply borel-cantelli to $\mathbb{P}(|\overline{X}_n - a| > \varepsilon)$ since it is a finite quantity. This means that
$$\overline{X}_n \xrightarrow{\text{a.s.}} a$$

along N. Call Ω_0 the almost sure set on which the convergence take place. Remember that if we have $N = (n_k) \in \mathbb{N}^*$ and $\lim_k \frac{n_{k+1}}{n_k} = r > 0$ and the sequence $(X_n)_n$ converges along N to a then

$$\frac{a}{r} \leqslant \liminf \overline{X}_n \leqslant \limsup \overline{X}_n \leqslant ra.$$

So note that

$$\frac{(k+1)^2}{k^2} \xrightarrow[k\to\infty]{} 1$$

and hence $\forall\, \omega \in \Omega_0$ we have that

$$\lim_{n\to\infty} \overline{X}_n(\omega) = a$$

with probability 1 which means almost sure convergence.

□

Then there is the second statement:

Proposition 1.2.6

Let $(X_n)_n$ be a sequence of positive independent and identically distributed (i.i.d.) random variables with

$$\mathbb{E}X_1 = +\infty.$$

Consider also a further random variable X distributed as X_1 (which means that they also have the same expectation). Then

$$\overline{X}_n \xrightarrow{\text{a.s.}} \infty.$$

Proof

Start by defining $Y_n = X_n \wedge b$ for some $b \in \mathbb{R}$. Then we have $\overline{Y}_n = \frac{\sum_{i=1}^n Y_i}{n}$. Since the random variable is truncated we have

$$\mathbb{E}Y_n = \mathbb{E}\left[X_n \wedge b\right] < \infty$$

but by the previous result we know that

$$Y_n \xrightarrow{\text{a.s.}} \mathbb{E}\left[X_n \wedge b\right].$$

Now consider the fact that $X_n \geqslant Y_n$ for every n, so we expect

$$\liminf_n X_n \geqslant \lim_n \overline{Y}_n = \mathbb{E}\left[X_n \wedge b\right]$$

for every b, even when b tends to ∞ (which causes $\mathbb{E}\left[X_n \wedge b\right]$ to become $\mathbb{E}X_n$). But this means

$$\liminf_n \overline{X} = \infty \text{ a.s.}$$

and since $\overline{X}_n$ is a non-decreasing sequence

$$\overline{X}_n \xrightarrow{\text{a.s.}} \infty.$$

□

1.2.9 Central limit theorem (Lyapunov theorem) Proof of the theorem and of the corresponding lemma (Lindeberg Lemma)

Theorem 1.2.10

Lyapunov Central limit Theorem:

Suppose $\mathbb{E}X_{nj} = 0 \; \forall n, j$ and $\operatorname{Var} Z_n = 1 \; \forall n$ and $\lim_n \sum_n \mathbb{E}|X_{nj}|^3 = 0$. Then

$$Z_n \xrightarrow{\mathrm{d}} Z \sim N(0,1).$$

Here we have replaced the conditions on the random variables with the condition $\lim_n \sum_n \mathbb{E}|X_{nj}|^3 = 0$ which allows us to use the structure of the triangular array. To prove this theorem we need the following lemma:

Lemma 1.2.1

Lindeberg's lemma: Let $(Y_1, Y_2, \dots, Y_k)$ be independent random variables with mean zero and let $S = \sum_{j=1}^k Y_j$. Let us further assume that $\operatorname{Var} S = 1$. Let f be a function which can be differentiated 3 times and let f', f'', f''' be bounded and continuous and such that

$$|f'''| \leqslant c, \qquad c \in \mathbb{R}_+.$$

Then for $Z \sim N(0,1)$

$$|\mathbb{E}f \circ S - \mathbb{E}f \circ Z| \leqslant c \sum_{j=1}^k \mathbb{E}|Y_j|^3$$

Proof

Let $Z_1, \dots, Z_k$ be independent normal random variables with mean $\mathbb{E}Z_j = \mathbb{E}Y_j = 0$ for $j = 1, \dots, k$ and variance $\operatorname{Var} Z_j = \operatorname{Var} Y_j$ for $j = 1, \dots, k$. Then construct

$$T = \sum_{j=1}^k Z_j \sim N(0, \sum_{j=1}^k \operatorname{Var} Z_j = \sum_{j=1}^k \operatorname{Var} Y_j = 1).$$

So we know that T is distributed as Z (they are both $N(0,1)$) so $T \overset{d}{=} Z$ and since we are using the expectation of Z we can replace $\mathbb{E}f \circ Z$ with $\mathbb{E}f \circ T$. So the Linderberg's lemma becomes

$$|\mathbb{E}f \circ S - \mathbb{E}f \circ T| \leqslant c \sum_{j=1}^k \mathbb{E}|Y_j|^3$$

which we want to prove, to exploit the structure of T. Define now the random variables $V_1, V_2, \dots, V_k$ as follows:

$$\begin{array}{lll}
V_1 & \text{s.t.} & S = V_1 + Y_1 \\
V_2 & \text{s.t.} & V_1 + Z_1 = V_2 + Y_2 \\
 & & \vdots \\
V_j & \text{s.t.} & V_j + Z_j = V_{j+1} + Y_{j+1}, \quad 1 \leqslant j < k \\
 & & \vdots \\
V_k & \text{s.t.} & V_k + Z_k = T.
\end{array}$$

Note that

$$\begin{aligned}
V_1 &= Y_2 + Y_3 + \dots + Y_k \\
V_2 &= Z_1 + Y_3 + \dots + Y_k \\
V_3 &= Z_1 + Z_2 + Y_4 + \dots + Y_k
\end{aligned}$$

so the Y get replaced by the Z one at the time in the V. We can now focus on the following expression:

$$\begin{aligned}
f \circ S - f \circ T &= f(V_1 + Y_1) - f(V_k - Z_k) \\
&= f(V_1 + Y_1) + f(V_2 + Y_2) - \underbrace{f(V_2 + Y_2)}_{f(V_1 + Z_1)} + f(V_3 + Y_3) - \underbrace{f(V_2 + Y_2)}_{f(V_2 + Z_2)} + \\
&\quad + \dots + f(V_k + Z_k) \\
&= \sum_{j=1}^k f(V_j + Y_j) - \sum_{j=1}^k f(V_j + Z_j).
\end{aligned}$$

Now take the expectation and the absolute value:

$$
\begin{aligned}
|\mathbb{E}f \circ s - \mathbb{E}f \circ t| &= \left|\sum_{j=1}^{k} \mathbb{E}f(V_j + Y_j) - \sum_{j=1}^{k} \mathbb{E}f(V_j + Z_j)\right| \\
&\leqslant \sum_{j=1}^{k} \left|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)\right|
\end{aligned}
$$

and now we only need to prove that

$$\left|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)\right| \leqslant c\mathbb{E}|Y_j|^3.$$

Let's write the Taylor formula for this function:

$$f(v + x) = f(v) + f'(v)x + \frac{1}{2}f''(v)x^2 + R_2(v, x)$$

where

$$
\begin{aligned}
R_2(v, x) &= \frac{1}{2}\int_v^{v+x} (v + x - t)_2 f''(t)\,\mathrm{d}t \\
&\leqslant \frac{1}{2}c\int_v^{v+x} (v + x - t)_2\,\mathrm{d}t = c\frac{x^3}{6}
\end{aligned}
$$

so that

$$|R_2(v, x)| \leqslant \frac{c}{6}|x|^3.$$

We now have

$$
\begin{aligned}
f(V_j + Y_j) &= f(V_j) + f'(V_j)Y_j + \frac{1}{2}f''(V_j)Y_j^2 + R_2(V_j, Y_j) \\
f(V_j + Z_j) &= f(V_j) + f'(V_j)Z_j + \frac{1}{2}f''(V_j)Z_j^2 + R_2(V_j, Z_j)
\end{aligned}
$$

and subtract side by side:

$$f(V_j + Y_j) - f(V_j + Z_j) = (Y_j - Z_j)f'(V_j) + \frac{1}{2}f''(V_j)(Y_j^2 - Z_j^2) + R_2(V_j, Y_j) - R_2(V_j, Z_j).$$

Now take the expectation

$$
\begin{aligned}
\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j) &= \frac{1}{2}\mathbb{E}f''(V_j)\underbrace{(\mathbb{E}Y_j^2 - \mathbb{E}Z_j^2)}_{=0 \text{ since they have same variance}} + \mathbb{E}\left[R_2(V_j, Y_j) - R_2(V_j, Z_j)\right] \\
&= \mathbb{E}\left[R_2(V_j, Y_j) - R_2(V_j, Z_j)\right].
\end{aligned}
$$

Now take the absolute value

$$
\begin{aligned}
\left|\mathbb{E}\left[R_2(V_j, Y_j) - R_2(V_j, Z_j)\right]\right| &\leqslant \mathbb{E}|\left[R_2(V_j, Y_j)\right]| + \mathbb{E}\left[|R_2(V_j, Z_j)|\right] \\
&\leqslant \frac{c}{6}\left(\mathbb{E}|Y|^3 + \mathbb{E}|Z|^3\right)
\end{aligned}
$$

so that

$$|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \leqslant \frac{c}{6}\left(\mathbb{E}|Y|^3 + \mathbb{E}|Z|^3\right).$$

Recall that $Z_j \sim N(0, b^2)$ where $b^2 = \mathbb{E}Y_j^2$. We know that

$$\mathbb{E}|Z_j|^3 = b^3\sqrt{\frac{8}{\pi}} \leqslant 2b^3$$

and we also have

$$b = \left(\mathbb{E}Y_j^2\right)^{\frac{1}{2}} \leqslant \left(\mathbb{E}|Y_j|^3\right)^{\frac{1}{3}}$$

Because L^2 norm is less or equal than L^3 norm (revise inclusions in L^p-spaces for different values of p). But this last inequality is equivalent to

$$b^3 \leqslant \mathbb{E}|Y_j|^3$$

which leads to

$$\mathbb{E}\left|Z_j\right|^3 \leqslant 2b^3 \leqslant 2\mathbb{E}\left|Y_j\right|^3.$$

Finally we get

$$\begin{aligned} \left|\mathbb{E}f(V_j+Y_j) - \mathbb{E}f(V_j+Z_j)\right| &\leqslant \frac{c}{6}\left(\mathbb{E}\left|Y_j\right|^3\right) \\ &= \frac{c}{6}3\mathbb{E}\left|Y_3\right|^3 \\ &= \frac{c}{2}\mathbb{E}\left|Y_3\right|^3 \leqslant c\mathbb{E}\left|Y_3\right|^3. \end{aligned}$$

□

This was horrible, horrible. Truly an horrible experience and honestly useless proof. And we still have to prove Lyapunov's theorem.

Proof

Lyapunov's CLT. Recall

$$Z_n = \sum_j X_{nj} \qquad Z \sim N(0,1).$$

We are interested in evaluating the characteristic function.

$$\begin{aligned} e^{irZ_n} &= \cos rZ_n + i\sin rZ_n \\ e^{irZ} &= \cos rZ + i\sin rZ. \end{aligned}$$

Consider now

$$\begin{aligned} \left|\mathbb{E}e^{irZ_n} - \mathbb{E}e^{irZ}\right| &= \left|(\mathbb{E}\cos rZ_n - \mathbb{E}\cos rZ) + i\,(\mathbb{E}\sin rZ_n - \mathbb{E}\sin rZ)\right| \\ &\leqslant \left|\mathbb{E}\cos rZ_n - \mathbb{E}\cos rZ\right| + \underbrace{|i|}_{1}\left|\mathbb{E}\sin rZ_n - \mathbb{E}\sin rZ\right|. \end{aligned}$$

By applying the above lemma we obtain

$$\left|\mathbb{E}e^{irZ_n} - \mathbb{E}e^{irZ}\right| \leqslant \sum_j |r|^3\mathbb{E}|X_{nj}|^3 + \sum_j |r|^3\mathbb{E}|X_{nj}|^3$$

and this is possible since both sine and cosine are differentiable three times and they are bounded by 1 (so our $c = 1$). Now, according to the hypotheses of Lyapunov's theorem, we need to take the limit considering the hypothesis that $\lim_{n\to\infty}\sum_j \mathbb{E}|X_{nj}|^3 = c$ and we obtain the claim. This theorem applies to all triangular arrays which include the one in the CLT. □

1.2.10 Definition of conditional expectation and its main properties

First of all, remember that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)}\int_A \mathbb{1}_B \,\mathrm{d}\mathbb{P}$$

Evaluate our best estimate for $X(\omega)$, which may very well be the average of X over H:

$$\frac{1}{\mathbb{P}(H)}\int_H X(\omega)\mathbb{P}(\mathrm{d}\omega) = \frac{1}{\mathbb{P}(H)}\mathbb{E}X\mathbb{1}_H = \mathbb{E}_H X.$$

If $\mathbb{P}(H) = 0$ then we allow any value for $\mathbb{E}_H X$.

Note the notation!

We usually denote that quantity as $\mathbb{E}(X|H)$ but Çinlar denotes it as $\mathbb{E}_H X$.

Remark

The quantity $\mathbb{E}_H X$ is called **conditional expectation of X given the event H**.

Consider the following facts:

1. $\overline{X} \in \mathcal{F}$: it is measurable with respect to $\mathcal{F}$ and this is clear from the definition ?? of the random variable $\overline{X}$;
2. we know that for each $V \in \mathcal{F}_+$ we have that

$$\mathbb{E}VX = \mathbb{E}\overline{X}$$

which is called **projection property**.

Definition 1.2.5

Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$. The conditional expectation $\mathbb{E}_{\mathcal{F}}X$ of X given $\mathcal{F}$ is defined in two steps:

(a) for $X \in \mathcal{H}_+$ (positive random variables) it is any random variable $\overline{X}$ satisfying:

   (a) measurability ($\overline{X} \in \mathcal{F}_+$);
   (b) projection property ($\mathbb{E}VX = \mathbb{E}V\overline{X} \qquad \forall V \in \mathcal{F}_+$).

(b) for arbitrary $X \in \mathcal{H}$, if $\mathbb{E}X$ exists, we define

$$\mathbb{E}_{\mathcal{F}}X = \mathbb{E}_{\mathcal{F}}X^+ - \mathbb{E}_{\mathcal{F}}X^-.$$

Otherwise, if $\mathbb{E}X^+ = \mathbb{E}X^- = \infty$, then $\mathbb{E}_{\mathcal{F}}$ is left undefined.

more properties:

1. **monotonicity**:
$$X \leqslant Y \implies \mathbb{E}_{\mathcal{F}}X \leqslant \mathbb{E}_{\mathcal{F}}Y;$$

2. **linearity**:
$$\mathbb{E}_{\mathcal{F}}(aX + bY + c) = a\mathbb{E}_{\mathcal{F}}X + b\mathbb{E}_{\mathcal{F}}Y + c;$$

3. **monotone convergence theorem**:
$$(X_n)_n \text{ s.t. } X_n \geqslant 0 \forall n, X_n \nearrow X \implies \mathbb{E}_{\mathcal{F}}X_n \nearrow \mathbb{E}_{\mathcal{F}}X;$$

4. **Fatou's lemma**:
$$X \geqslant 0 \implies \mathbb{E}_{\mathcal{F}} \liminf X_n \leqslant \mathbb{E}_{\mathcal{F}}X_n;$$

5. **Dominated convergence theorem**:
$$(X_n)_n \text{ a.s. } X_n \to X, |X_n| \leqslant Y, Y \text{ integrable} \implies \mathbb{E}_{\mathcal{F}}X_n \to \mathbb{E}_{\mathcal{F}}X;$$

6. **Jensen's inequality**:
$$f \text{ convex} \implies \mathbb{E}_{\mathcal{F}}f(x) \leqslant f(\mathbb{E}_{\mathcal{F}}X).$$

1.2.11 Existence and uniqueness of conditional expectation (for L1 random variables)

Theorem 1.2.11

Let $X \in \mathcal{H}$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{H}$. Then the conditional expectation $\mathbb{E}_{\mathcal{F}}X$ exists and it is unique up to equivalence.

Proof

$\forall H \in \mathcal{F}$ on the measurable space $(\Omega, \mathcal{F})$ consider the restriction of $\mathbb{P}$ on $\mathcal{F}$. Consider now

$$\mathbb{Q}(H) = \int_H \mathbb{P}(\mathrm{d}\omega)X(\omega)$$

where $\mathbb{P}$ is a probability measure and $\mathbb{Q}$ is a measure which is absolutely continuous with respect to $\mathbb{P}$. In this measurable space random variables are functions, so we can apply Radon-Nikodyn theorem. So it exists a random variable $\overline{X} \in \mathcal{F}_+$ such that

$$\int_\Omega \mathbb{Q}(\mathrm{d}\omega)V(\omega) = \int_\Omega \mathbb{P}(\mathrm{d}\omega)\overline{X}(\omega)V(\omega) \qquad \forall V \in \mathcal{F}_+$$

so the projection property is satisfied and $\overline{X}$ is a version of $\mathbb{E}_\mathcal{F} X$.
About the uniqueness: Let $\overline{X}$ and $\overline{\overline{X}}$ be versions of $\mathbb{E}_\mathcal{F} X,\ X \geqslant 0$.

1. Both $\overline{X}$ and $\overline{\overline{X}}$ are $\mathcal{F}^+$-measurable;
2. $\mathbb{E}VX = \mathbb{E}V\overline{X} = \mathbb{E}V\overline{\overline{X}}$ for every $V \in \mathcal{F}_+$. Hence

$$\overline{X} = \overline{\overline{X}} \text{ a.s.}$$

   Conversely, if $\mathbb{E}_\mathcal{F} X = \overline{X}$ and $\overline{\overline{X}} \in \mathcal{F}_+$ and $\overline{X} = \overline{\overline{X}}$ a.s. then $\overline{\overline{X}}$ satisfies the projection property $\mathbb{E}XV = \mathbb{E}\overline{\overline{X}}V$ (i.e. $\overline{\overline{X}}$ is a version of $\mathbb{E}_\mathcal{F} X$).

□

1.2.12 Existence and uniqueness of conditional expectation (for L2 random variables, using orthogonal projection)

Theorem 1.2.12

$\forall X \in L^2(\mathcal{H})$ there exists a unique (up to equivalence) $\overline{X} \in L^2(\mathcal{H})$ such that

$$\mathbb{E}|X - \overline{X}|^2 = \inf_{Y \in L^2(\mathcal{F})} \mathbb{E}|X - Y|^2.$$

Furthermore, $X - \overline{X}$ is orthogonal to $L^2(\mathcal{F})$, i.e.

$$\mathbb{E}V(X - \overline{X}) = 0 \qquad \forall V \in L^2(\mathcal{F})$$

Note that $L^2(\mathcal{H})$ is a complete Hilbert space in which the inner product of X and Y is given by $\mathbb{E}XY$. $\overline{X}$ is the orthogonal projection of the vector X onto the subspace $L^2(\mathcal{F})$ and the decomposition

$$X = \overline{X} + \widetilde{X}$$

holds.

Proof

Let's write the L^2-norm of X calling it $\|X\|$:

$$\|X\| = \|X\|_2 = \sqrt{\mathbb{E}X^2}.$$

Fix $X \in L^2(\mathcal{H})$. Define

$$\delta = \inf_{Y \in L^2(\mathcal{F})} \|X - Y\|.$$

Let $(Y_n)_n \subset L^2(\mathcal{F})$ such that $\delta_n = \|X - Y_n\| \xrightarrow[x \to \infty]{} 0$. Let us prove that $(Y_n)_n$ is a Cauchy sequence for the $L^2(\mathcal{F})$-convergence.

$$|Y_n - Y_m|^2 = 2|X - Y_m|^2 - 4|\underbrace{x - \frac{1}{2}(Y_n + Y_m)}_{\in L^2(\mathcal{F})}|^2.$$

Take the expectation on both sides:

$$\mathbb{E}|Y_n - Y_m|^2 \leqslant 2\delta_m^2 + 2\delta_n^2 - 4\delta^2.$$

Now we take the limit for n and m and what we get is

$$\lim_{m,n\to\infty} \mathbb{E}|Y_n - Y_m|^2 \leqslant 0.$$

Hence it is true that $(Y_n)_n$ is Cauchy and this means that there exists a $\overline{X} \in L^2(\mathcal{F})$ such that $\left|Y_n - \overline{X}\right| \xrightarrow[n\to\infty]{} 0$. Note that $\overline{X}$ is unique up to equivalence (by definition of L^2-norm). Note also

$$\overline{X} \in L^2(\mathcal{F}) \implies \left|X - \overline{X}\right| \geqslant \delta.$$

Now, by Minkowski's inequality we can write that

$$\left|X - \overline{X}\right| \leqslant \|X - Y_n\| + \left|Y_n - \overline{X}\right| \xrightarrow[n\to\infty]{} \delta + 0 = \delta.$$

We have thus

$$\left|X - \overline{X}\right| = \delta.$$

For $V \in L^2(\mathcal{F})$ and $a \in \mathbb{R}$, since $\mathbb{E}|X - \overline{X}|^2 = \delta$ then we have that

$$\begin{aligned} a^2\mathbb{E}V^2 - 2a\mathbb{E}V(X - \overline{X}) + \delta^2 &= \left|aV - (X - \overline{X})\right|^2 \\ &= \|X - (\underbrace{aV + \overline{X}}_{\in L^2(\mathcal{F})})\|^2 = \delta^2 \end{aligned}$$

And therefore

$$a^2\mathbb{E}V^2 - 2a\mathbb{E}V(X - \overline{X}) \leqslant 0 \qquad \forall\, a \in \mathbb{R}$$

which is impossible unless

$$\mathbb{E}V(X - \overline{X}) = 0.$$

□

1.2.13 Filtration, adaptedness; example of non adapted process

Let T be a subset of $\mathbb{R}$. Let $\mathcal{F}_t$ be a sub-σ-algebra of $\mathcal{H}$ $\forall\, t \in T$. The family $\mathcal{F} = (\mathcal{F}_t)_{t\in T}$ is called **filtration** if $\mathcal{F}_s \subset \mathcal{F}_t$ for every $s < t$. Take, for example, $\mathcal{F}_1$: it is much smaller than $\mathcal{H}$ and if we say that a random variable is measurable with respect to $\mathcal{F}_1$ it means that we do not have much knowledge, but if we expand to $\mathcal{F}_2$ we are able to gain more knowledge about the random variable. If we interpret the index set T as time we may actually think about filtrations as our knowledge of the phenomenon as time passes.

Definition 1.2.6

Given $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$, a stochastic process $\{X_t\}_{t\in\mathbb{T}}$ is **adapted** to $\mathcal{F}$ if X_t is $\mathcal{F}_t$-neasurable for each $t \in \mathbb{T}$ and $\mathcal{E}$-measurable.

As an example consider "the secretary problem": in this case we must start from the filtration and then understand the problem. Here i have N candidates for a position; a candidate disregarded after the interview is lost. The interviewer wants to hire exactly 1 candidate and each candidate has different abilities and the interviewer knows only the relative ability of those already interviewed so far. Our goal is to maximizing the probability of hiring the best one. We have three questions:

1. what is Ω?

2. what is the filtration $\mathcal{F}$ for this experiment?

3. what process should we use?

In this case $\Omega = N!$ permutations of the ranking of the candidates (the order in which they show up) and the filtration is the information earned from interview up to time t (that is the ranking of the candidates up to time t). But what is the process that I should use? Consider the sequence

$$V_1, V_2, \dots \qquad \{V_i\}_{i \geqslant 1}$$

with $V_i = 1$ if and only if the best candidate is the i-th candidate and $V_i = 0$ otherwise. Could this process $\{V_i\}_{i\geqslant 1}$ be used? No, because V is not adapted to $\mathcal{F}$... because to understand if i-th candidate is te best we need to compare it to the other candidates, including the ones that didn't show up yet! But then how can we get an adapted process? Let us consider the expectation

$$U_n = \mathbb{E}\left[V_n | \mathcal{F}_n\right]$$

What do we know about the measurability of U_n? We know that it is for sure $\mathcal{F}_n$-measurable. This trick gives us a simple way to build an adapted process. So now we will have: $U_n = 0$ if the candidate is not the best up to n and $U_n = 1$ otherwise. More specifically, we will have

$$\begin{aligned} U_n &= 1 \cdot \substack{\text{proability that the best candidate} \\ \text{is among the first } n} + 0 \cdot \substack{\text{proability that the best candidate} \\ \text{is not among the first } n} \\ &= 1 \cdot \frac{n}{N} + 0 \cdot \frac{N-n}{N}. \\ &= \frac{n}{N} \end{aligned}$$

This is a quantity that I can measure and it is therefore adapted.

1.2.14 Random times and Stopping Times: examples of random times that are not stopping times and examples of stopping times

Definition 1.2.7

$$T : \Omega \mapsto \overline{\mathbb{T}}$$

is called **stopping time** of $\mathcal{F}$ if

$$\{T \leqslant t\} \in \mathcal{F}_t \qquad \forall\, t \in \mathbb{T}.$$

Stopping time have an alternative definition:

Definition 1.2.8

T is a stopping time if the process

$$Z_t = \mathbb{1}_{\{T \leqslant t\}} \qquad t \in \mathbb{T}$$

is adapted to F.

Starting from the previously deifned random walk, consider the quantity

$$M_n = \min(S_1, \dots, S_n)$$

And the random time

$$T_2 = \min\{n : S_n \geqslant M_m + 2\}$$

is a stoppimg time. On the contrary,

$$T_3 = \begin{cases} \max\{n < 50 : S_n = 7\} & \text{if not empty} \\ 50 & \text{otherwise} \end{cases}$$

is not a stopping time. Why? Because I have to wait until $n = 50$ to answer the question.

1.2.15 Stopped filtration: definition and examples

Definition 1.2.9

Let $\mathcal{F}$ be a filtration on $\overline{\mathbb{T}}$ and let T be a stopping time on $\mathcal{F}$. We call **past until T** the σ-algebra $\mathcal{F}_t \subset \mathcal{F}_\infty \subset \mathcal{H}$ such that:

$$\mathcal{F}_T = \left\{H \in \mathcal{H} : H \cap \{T < t\} \in \mathcal{F}_t, \forall t \in \overline{\mathbb{T}}\right\}.$$

This means that $\mathcal{F}_T$, which represents the evolution within T, contains all the events H such that they are before the stopping time in all the filtrations. This is selecting the events H that happen within the time T.

Remark

If we fix a $T \equiv t$ then $\mathcal{F}_T \equiv \mathcal{F}_t$ (it's the normal filtration at time t).

If T is a stopping time of $\mathcal{F}$ then $\{T \leqslant r\}$ belongs to $\mathcal{F}_T$, $\forall r \geqslant 0$:

$$\{T \leqslant r\} \cap \{T \leqslant t\} = \{T \leqslant \min(T, r)\} \in \mathcal{F}_t$$

So T is $\mathcal{F}_T$-measurable.

Remark

$\mathcal{F}_t$ can be read as the collection of all $\mathcal{F}_t$-measurable random variables V: the value of $V(\omega)$ can be told by the time $T(\omega)$: we can read the value of $V(\omega)$ before the ringing of an alarm.

Imagine a factory. The production is blocked when

a) the temperature of the room is above a certain threshold;

b) the machinery has not been cleaned for more than 12 hours.

Alternatively, imagine that the sales of shares happens:

a) when the price is above a fixed value;

b) when the increase of the price is smaller than a fixed value.

1.2.16 Discrete stopping times associated to continuous stopping times

Definition 1.2.10

Let

$$d_n(t) = \begin{cases} \frac{k+1}{2^n} & \text{if } \frac{k}{2^n} \leqslant t < \frac{k+1}{2^n} \\ +\infty & \text{if } t = \infty. \end{cases}$$

We got a function

$$d_n : \overline{\mathbb{R}}_+ \mapsto \overline{\mathbb{R}}_+.$$

Observe that $d_1 \geqslant d_2 \geqslant d_3 \geqslant \dots$. What are the other properties of this function?

1. it is a step function;
2. it is right-continuous;
3. $d_n(t) > t$ (the diagonal line);
4. $\lim_{n\to\infty} d_n(t) = t$.

Now we apply this simple function to our stopping times so that we get the following proposition.

Proposition 1.2.7

Let $\mathcal{F}$ be a filtration on $\overline{\mathbb{R}}_+$ and let T be a stopping time. Let

$$T_n = d_n \circ T.$$

Then $\{T_n\}_n$ is a sequence of discrete stopping times of $\mathcal{F}$ decreasing to T.

Of course T_n is a foretold time by T, because if I know T I also know the value of T_n.

Proof

First of all, fix n. We know that:

- T_n is a measurable function of T so it is for sure $\mathcal{F}_n$-measurable;
- $d_n(t) \geqslant t$ for every $t < \infty$ and $d_n(\infty) = \infty$, so if we apply d_n to T we get $d_n(T) = T_n \geqslant T$;
- we know that T_n is a foretold time by T;
- we know that T_n is a stopping time as well;
- we know that T_n is discrete.

But we also know that $d_n(t) \searrow t$ as $n \to \infty$ and this means that $d_n(T) = T_n \searrow T$. □

We have thus switched from a continuous stopping time to a sequence of discrete stopping time.

1.2.17 Conditional expectation at stopping time

Can I use T fixed as a deterministic time t and thus get $\mathbb{E}(X|\mathcal{F}_t)$? Is this a special case or a different thing altogether? Well yes, because t *is* a stopping time being $\mathcal{F}_t$-measurable! $\mathbb{E}(X|\mathcal{F}_t)$ is a special case of $\mathbb{E}(X|\mathcal{F}_T)$. This becomes clear when we see the properties of stopped expectation. For $\forall\, X, Y, Z$ being positive random variables and for $\forall\, S, T$ stopping times of $\mathcal{F}$ the following properties hold:

- **defining property**:
$$\mathbb{E}_T X = Y \iff V \in \mathcal{F}_T \text{ and } \mathbb{E}VX = \mathbb{E}VY;$$
- **unconditioning**:
$$\mathbb{E}\mathbb{E}_T X = \mathbb{E}X;$$
- **repeated conditioning**:
$$\mathbb{E}_S\mathbb{E}_T X = \mathbb{E}_{\min\{S,T\}} X = \mathbb{E}_{S\wedge T} X;$$
- **conditional determinism**:
$$\mathbb{E}_T(X + YZ) = X + Y\mathbb{E}_T(Z)$$
  if $X, Y \in \mathcal{F}_T$.

We need to prove the third property.

Proof

Notice that if $S \leqslant T$ then $\mathcal{F}_S \subset \mathcal{F}_T$ and since the poorer σ-algebra always wins we have
$$\mathbb{E}_S\mathbb{E}_T = \mathbb{E}_S = \mathbb{E}_{S\wedge T}.$$
But if S and T are arbitrary we can apply the same result to $S \wedge T$ since for sure $S \wedge T \leqslant T$:
$$\mathbb{E}_{S\wedge T}\mathbb{E}_T = \mathbb{E}_{S\wedge T}. \tag{$\bigstar$}$$
Remember that $Y = \mathbb{E}_T X$. We can now write the statement of the proof in a different way:
$$\mathbb{E}_S Y = \mathbb{E}_{S\wedge T} X.$$
and using property ★ we get
$$\begin{aligned}\mathbb{E}_S Y &= \mathbb{E}_{S\wedge T} \\ &= \mathbb{E}_{S\wedge T} X.\end{aligned}$$
This is what we need to prove now. Let's go through the usual two steps.

1. Note that $\mathcal{F}_{S\wedge T} \subset \mathcal{F}_S$. This means that $\mathbb{E}_{S\wedge T}Y$ is in $\mathcal{F}_S$ and therefore $\mathbb{E}_{S\wedge T}Y$ has the measure property and it is a candidate to be a version of $\mathbb{E}_S Y$. We must now check whether this candidate satisfies the defining property of stopped expectation.

2. Check the defining property:
$$\mathbb{E}VY = \mathbb{E}V\mathbb{E}_{S\wedge T}Y$$
for each positive $V \in \mathcal{F}_S$. We start by fixing such $V \in \mathcal{F}_S$ positive. We proved that
$$V\mathbb{1}_{(S\leqslant T)}Y = \mathbb{E}V\mathbb{1}_{(S\leqslant T)}\mathbb{E}_{S\wedge T}Y.$$
Now we observe that $Y \in \mathcal{F}_T$ by definition. Notice that $Y\mathbb{1}_{(T<S)} \in \mathcal{F}_{S\wedge T}$. We can apply this fact to the defining property and obtain
$$\mathbb{E}VY\mathbb{1}_{(T<S)} = \mathbb{E}V\mathbb{E}_{S\wedge T}Y\,\underbrace{\mathbb{1}_{(T<S)}}_{\mathcal{F}_{S\wedge T}\text{-measurable}}$$
and by conditional determinism we get
$$\begin{aligned}\mathbb{E}VY\mathbb{1}_{(T<S)} &= \mathbb{E}V\mathbb{1}_{(T<S)}\\ &= \mathbb{E}V\mathbb{1}_{(T<S)}\mathbb{E}_{S\wedge T}Y.\end{aligned}$$
By putting this together with , we get
$$\mathbb{E}VY = \mathbb{E}V\mathbb{E}_{S\wedge T}Y.$$

□

1.2.18 Comparing stopping times: some interesting features

Uuuuh I'm not really sure what she meant by this. Of course god forbid a fucking SDS students has their life made easy once. I think she means this:

Theorem 1.2.13

1. if $S \leqslant T \implies \mathcal{F}_S \subset \mathcal{F}_T$;
2. $\mathcal{F}_{\min\{S,T\}} = \mathcal{F}_S \cap \mathcal{F}_T$;
3. if $V \in \mathcal{F}$, the following processes are in $\mathcal{F}_{\min\{S,T\}}$:
$$\underset{V\mathbb{1}_{S\leqslant T},}{(a)} \quad \underset{V\mathbb{1}_{S=T},}{(b)} \quad \underset{V\mathbb{1}_{S<T}.}{(c)}$$

Proof

1. Imagine $\{S \leqslant T\}$. Now consider the event $\{S \leqslant n\}$. Since S happens before T then we have that $\{T \leqslant n\} \subset \{S \leqslant n\}$. We know that if $A \in \mathcal{F}_S$ then
$$A \cap \{T \leqslant n\} = A \cap \underbrace{\{S \leqslant n\}}_{\in\mathcal{F}_n} \cap \underbrace{\{T \leqslant n\}}_{\in\mathcal{F}_n}$$
which means that $A \cap \{T \leqslant n\} \in \mathcal{F}_n$ and by definition of $\mathcal{F}_T$ this correponds to saying
$$\mathcal{F}_S \subset \mathcal{F}_T.$$

2. (a) Let us first prove that $\mathcal{F}_{\min\{S,T\}} \subset \mathcal{F}_S \cup \mathcal{F}_T$. Note that $\min\{S,T\}$ is dominated by $\mathcal{F}_s$ and $\mathcal{F}_T$, which means that $\mathcal{F}_{\min\{S,T\}}$ is included both in $\mathcal{F}_S$ and $\mathcal{F}_T$. By point 1 we have
$$\mathcal{F}_{\min\{S,T\}} \subset \mathcal{F}_S \cap \mathcal{F}_T.$$
(b) It remains to prove that $\mathcal{F}_{\min\{S,T\}} \supset \mathcal{F}_S \cup \mathcal{F}_T$. Let $H \in \mathcal{F}_S \cap \mathcal{F}_T$. This means that
$$H \cap \{S \leqslant T\} \in \mathcal{F}_{\min\{S,T\}}$$

by part the of this theorem, since $H \in \mathcal{F}_S$. But we also know that

$$H \cap \{T \leqslant S\} \in \mathcal{F}_{\min\{S,T\}}$$

by part 3 of this theorem since also $H \in \mathcal{F}$. But this means that

$$H = \{H \cap \{S \leqslant T\}\} \cup \{H \cap \{S \leqslant T\}\}$$

belongs to $\mathcal{F}_{\min\{S,T\}}$ and so

$$\mathcal{F}_{\{S \cap T\}} \subset \mathcal{F}_{\min\{S,T\}}.$$

3. (a) $V\mathbb{1}_{\{S \leqslant T\}} \in \mathcal{F}_{\min\{S,T\}}$. To prove this we use the following theorem:

Theorem 1.2.14

Let T be a stopping time of $\mathcal{F}$. Then

$$\mathcal{F}_T = \{X_T : X \in \mathcal{F}\}$$

with X in $\mathcal{F}$.

This baiscally means that $\mathcal{F}_t$ becomes the values X_T of X in $\mathcal{F}$ at time T. Let us consider $X_t = V\mathbb{1}_{\{S \leqslant t\}}$ with $t = \min\{S,T\}$. We know that $\min\{S,T\}$ is a stopping time of $\mathcal{F}_{\min\{S,T\}}$ and by the theorem above

$$X_{\min\{S,T\}} \in \mathcal{F}_{\min\{S,T\}}$$

or, alternatively,

$$\mathbb{1}_{\{S \leqslant T\}} \in \mathcal{F}_{\min\{S,T\}}.$$

(b) Let $V = 2 \implies \{S \leqslant T\} \in \mathcal{F}_{\min\{S,T\}}$ and by symmetry $\{T \leqslant S\} \in \mathcal{F}_{\min\{S,T\}}$. This means that

$$\{S = T\} = \{S \leqslant T\} \cap \{T \leqslant S\} \in \mathcal{F}_{\min\{S,T\}}.$$

Furthermore

$$\{S < T\} = \{S \leqslant T\} \setminus \{S = T\}$$

Which implies memebership of $\mathcal{F}_{\min\{S,T\}}$.

(c) Consider the fact already proven that $V\mathbb{1}_{\{S \leqslant T\}} \in \mathcal{F}_{\min\{S,T\}}$. Multiplying by $\mathbb{1}_{\{S=T\}}$ or $\mathbb{1}_{\{S<T\}}$ does not alter the membership in $\mathcal{F}_{\min\{S,T\}}$.

□

1.2.19 Functions of stopping times that are (or are not) stopping times

Same, I think she means this: are we sure that $S \wedge T$ and $S \vee T$ are stopping times?

Proof

We know that $\{S \leqslant t\}$ and $\{T \leqslant t\}$ belong to $\mathcal{F}_t$ for every t, since they are stopping times. Consider now $\{\min\{(S,T) \leqslant t\}\}$ and $\{\max\{(S,T) \leqslant t\}\}$. We have

$$\{\min\{(S,T) \leqslant t\}\} = \underbrace{\underbrace{\{S \leqslant t\}}_{\in\mathcal{F}_t} \cup \underbrace{\{T \leqslant t\}}_{\in\mathcal{F}_t}}_{\in\mathcal{F}_t}$$

and

$$\{\max\{(S,T) \leqslant t\}\} = \underbrace{\underbrace{\{S \leqslant t\}}_{\in\mathcal{F}_t} \cap \underbrace{\{T \leqslant t\}}_{\in\mathcal{F}_t}}_{\in\mathcal{F}_t}.$$

□

1.2.20 Foretold times and their relationship with stopping times; examples of foretold times.

Definition 1.2.11

Let

S	be a stopping time
T	be a random time $T > S$ but whose value can be told by time S.

We say that T is **foretold by S** (and it is a stopping time).

Example 1.2.2

Consider S and $T = 2S$. Surprise surprise, T is a foretold time.

Remark

At time T we know S: We know that when the alarm rings within three seconds the machinery will stop, so S is actually $\mathcal{F}_T$-measurable. We say that **S is foretold by T**.

1.2.21 Examples of Martingales: Markov Chains

Consider a stochastic process $\{X_n\}_{n\geqslant 1}$. All the possible outcomes are on the same probability space so $(E_n, \mathcal{E}_n) = (E, \mathcal{E})$. Define a Markov kernel on $(E, \mathcal{E})$ such that $(E, \mathcal{E}) \mapsto (E, \mathcal{E})$ and such that

$$K_n(x_0, x_1, \ldots, x_n; A) = \mathbb{P}(x_n, A) \quad \forall\, n \in \mathbb{N},\ \forall\, (x_0, x_1, \ldots, x_n) \in E, A \in \mathcal{E}.$$

The probability that governs the evolution of the process only depends on the last state. If $x_{n+1} = j$ and $x_n = i$ then $\mathbb{P}(i, j) = p_{ij}$ and the process $X = \{X_n\}_{n\geqslant 0}$ is said to be a Markov chain over $(\Omega, \mathcal{H}, \mathbb{P})$ with state spaces $(E, \mathcal{E})$, initial distribution μ and transition kernel P where

$$P = (p_{ij})_{i,j}$$

Now consider a Markov chain with transition probability matrix P such that

$$P \cdot \underbrace{f}_{\text{eigenvector}} = \underbrace{\lambda}_{\text{eigenvalue}} \cdot f.$$

This means that we have the following system of equations:

$$\begin{cases} \sum_j p_{ij} f(j) = \lambda f(i) & i = 1 \\ \vdots & i = 2 \\ \vdots & \vdots \end{cases}$$

We can clearly write these expressions in form of expectation:

$$\mathbb{E}[f(X_{n+1})|X_n = i] = \lambda f(i).$$

We can also write it without fixing i:

$$\mathbb{E}[f(X_{n+1})|X_n] = \lambda f(X_n). \tag{$\bullet$}$$

Look at the last equation: it is not too different from the martingale property. Consider the ratio $\frac{f(X_n)}{\lambda^n}$ and the filtration generated by X $\sigma(X_0, X_1, \ldots, X_n)$. Thanks to the relation given by the equation ($\bullet$) it is a martingale with respect to $\sigma(X_0, X_1, \ldots, X_n)$.

Proof

Consider the expectation

$$\begin{aligned}\mathbb{E}\left[\frac{f(X_{n+1})}{\lambda^{n+1}}\middle|X_0,\dots,X_n\right] &= \frac{1}{\lambda^n}\frac{1}{\lambda}\underbrace{\mathbb{E}\left[f(X_{n+1}|X_n)\right]}_{\lambda f(X_n)} \\ &= \frac{1}{\lambda^n}\frac{1}{\cancel{\lambda}}\cancel{\lambda}f(X_n) \\ &= \frac{1}{\lambda^n}f(X_n) \\ &= Y_n\end{aligned}$$

where the first equality is thanks to Markov property. This is a general property for Markov chains. An example of Markov process is the Branching process. □

1.2.22 Definition of martingale, sub-martingale and super-martingale and their properties

Definition 1.2.12

A real-valued process

$$X = (X_t)_{t\in\mathbb{T}}$$

is called a **$\mathcal{F}$-martingale** if:

1. it is adapted to $\mathcal{F}$;
2. it is integrable for each $t \in \mathbb{T}$;
3. $\mathbb{E}(X_t - X_s|\mathcal{F}_s) = 0 \quad \forall s < t$.

If $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \geqslant 0 \quad \forall s < t$ then the process is called **$\mathcal{F}$-submartingale** and if $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \leqslant 0 \quad \forall s < t$ it is called **$\mathcal{F}$-supermartingale**.

Now the properties:

1. Consider X and Y being $\mathcal{F}$-sub-martingales and $a, b \in \mathbb{R}^+$. Then

$$aX + bY$$

is a $\mathcal{F}$-sub-martingale.

2. Consider X, Y being two $\mathcal{F}$-sub-martingales. Then

$$\max\{X, Y\}$$

is a $\mathcal{F}$-sub-martingale.

Proof

Consider $\mathbb{E}[\max\{X_n, Y_n\}|\mathcal{F}_{n-1}]$. Remember that we have to prove:

- measurability property;
- integrability property;
- martingale property.

Often times measurability and integrability are implied in the form of the function. Since we are using the maximum between X and Y we know that

$$\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] \geqslant \mathbb{E}[X_n|\mathcal{F}_{n-1}\}]$$
$$\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] \geqslant \mathbb{E}[Y_n|\mathcal{F}_{n-1}\}].$$

But since X and Y are martingales we know that $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geqslant X_{n-1}$ and $\mathbb{E}[Y_n|\mathcal{F}_{n-1}] \geqslant Y_{n-1}$; hence we get that

$$\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] \geqslant \max\{X_{n-1}, Y_{n-1}\}$$

so the martingale property is satisfied. □

3. Consider two $\mathcal{F}$-super-martingales X_n, Y_n. Then

$$\min\{X_n, Y_n\}$$

is a $\mathcal{F}$-super-martingale.

Proof

Consider $\max\{-X_n, -Y_n\}$. We know that $-X_n$ is a sub-martingale and so is $-Y_n$. So as proved above $\max\{-X_n, -Y_n\}$ is a sub-martingale. But $\max\{-X_n, -Y_n\} = -\min\{X_n, Y_n\}$ so

$$-\min\{X_n, Y_n\} \text{ is a sub-martingale} \iff \min\{X_n, Y_n\} \text{ is a super-martingale.}$$

□

4. Consider a function f that is convex on $\mathbb{R}$. If X is a $\mathcal{F}$-martingale and $f \circ X$ is integrable then $f \circ X$ is a $\mathcal{F}$-sub-martingale.

Proof

We have $s < t$ and we want to study $\mathbb{E}\left[f(X_t)|\mathcal{F}_s\right]$. By Jensens's inequality we have

$$\mathbb{E}\left[f(X_t)|\mathcal{F}_s\right] \geqslant f\Big[\underbrace{\mathbb{E}(X_t|\mathcal{F}_s)}_{X_s}\Big]$$
$$= f(X_s).$$

□

Remark

Some examples of positive functions of X are $X^+, X^-, |X|$. If X is martingale, these are sub-martingales. $|X|^p$ is a sub-martingale if X is a martingale and $\mathbb{E}|X|^p < \infty$.

5. if f is convex and increasing and X is a $\mathcal{F}$-sub-martingale with $f(X_t)$ integrable $\forall\, t$ then $f(X_t)$ is again a $\mathcal{F}$-sub-martingale.

1.2.23 Definition of sub/super martingale and examples

Definition 1.2.13

A real-valued process

$$X = (X_t)_{t\in\mathbb{T}}$$

is called a **$\mathcal{F}$-martingale** if:

1. it is adapted to $\mathcal{F}$;
2. it is integrable for each $t \in \mathbb{T}$;
3. $\mathbb{E}(X_t - X_s|\mathcal{F}_s) = 0 \quad \forall\, s < t$.

If $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \geqslant 0 \quad \forall\, s < t$ then the process is called **$\mathcal{F}$-submartingale** and if $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \leqslant 0 \quad \forall\, s < t$ it is called **$\mathcal{F}$-supermaringlae**.

1. Consider X and Y being $\mathcal{F}$-sub-martingales and $a, b \in \mathbb{R}^+$. Then

$$aX + bY$$

is a $\mathcal{F}$-sub-martingale.

2. Consider X, Y being two $\mathcal{F}$-sub-martingales. Then

$$\max\{X, Y\}$$

is a $\mathcal{F}$-sub-martingale.

3. Consider two $\mathcal{F}$-super-martingales X_n, Y_n. Then

$$\min\{X_n, Y_n\}$$

is a $\mathcal{F}$-super-martingale.

4. Consider a function f that is convex on $\mathbb{R}$. If X is a $\mathcal{F}$-martingale and $f \circ X$ is integrable then $f \circ X$ is a $\mathcal{F}$-sub-martingale.

Remark

Some examples of positive functions of X are $X^+, X^-, |X|$. If X is martingale, these are sub-martingales. $|X|^p$ is a sub-martingale if X is a martingale and $\mathbb{E}|X|^p < \infty$.

5. if f is convex and increasing and X is a $\mathcal{F}$-sub-martingale with $f(X_t)$ integrable $\forall\, t$ then $f(X_t)$ is again a $\mathcal{F}$-sub-martingale.

1.2.24 Maximum, minimum of martingales

Not so sure what it is being asked here.

Example 1.2.3

We are still with our financebros. Suppose that (M_n) is the price of an asset at time n. We want to buy when the price is below a at time S_i and sell when it is above b at time T_i. In $(0, n]$ we have $U_n(a, b)$ cycles of buying and selling so our strategy could consists in holding a number F_n of shares during period $(m - 1, m]$. This means introducing

$$F_m = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k, T_k]}$$

with $F_0 = 0$. We can thus trace the evolution of our capital with the discrete-time integral:

$$X = \int F \, \mathrm{d}M$$

And the profit during $(0, n]$will be $X - X_0$.
The profit is at least

$$(b - a)U_n(a, b).$$

Proposition 1.2.8

If M is a sub-martingale with respect to its natural filtration then

$$(b - a)\mathbb{E}U_n(a, b) \leqslant \mathbb{E}\left[(M_n - a)^+ - (M_0 - a)^+\right].$$

We wanted to find a bound for our profit, but our profit is a stochastic quantity: so it's only natural to think about the expectation to give a bound to the number of expected up/downcrossing.
Observe that the number of upcrossings does not depend on the value of T_0 that we fix.

Proof

Choose $a = 0$. Consider hence the process $(M-a)^+$ that is sub-martingale (if M is a sub-martingale). Take $M \geqslant 0$ and let

$$F_n = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k,T_k]}(n)$$

and consider

$$X = \int F \,\mathrm{d}M$$

like in our example. We know that F is predictable since by definition $F_{k+1} \in \mathcal{F}_k$. Consider the expectation of the increment

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] = \mathbb{E}\left[F_{k+1}(M_{k+1} - M_k)|\mathcal{F}_k\right].$$

But since F_{k+1} is predictable we can take it out the expectation:

$$F_{k+1}\mathbb{E}\left[M_{k+1} - M_k|\mathcal{F}_k\right].$$

But since F_{k+1} is an indicator we know it is $\leqslant 1$:

$$\mathbb{E}[X_{k+1} - X_k|\mathcal{F}_k] \leqslant \mathbb{E}\left[M_{k+1} - M_k|\mathcal{F}_k\right].$$

Now take the expectation of both sides:

$$\mathbb{E}[X_{k+1} - X_k] \leqslant \mathbb{E}\left[M_{k+1} - M_k\right].$$

If we sum these inequalities over k we get:

$$\mathbb{E}\left[X_n - X_0\right] \leqslant \mathbb{E}\left[M_n - M_0\right].$$

So we now get that

$$bU_n(a,n) \leqslant \mathbb{E}\left[X_n - X_0\right] \leqslant \mathbb{E}\left[M_n - M_0\right]$$

but given that $a = 0$ we get that

$$bU_n(0,b) \leqslant \mathbb{E}[M_n - M_0].$$

Clearly we have to take the positive part. □

This characterizes our martingale and its boundedness. The number of oscillations of a sub-martingale is bounded! The next question is: can we say anything about the behaviour of maximum/minimum of a martingale or sub-martingale?

Remark

Consider a sequence $\{X_n\}$ of independent random variables with $\mathbb{E}X_n = 0$, $S_n = \sum X_i$. We proved that

$$a^2\mathbb{P}(\max_{k\leqslant n}|S_k| > a) \leqslant \operatorname{Var} S_n$$

and we called this **Kolmogorov's inequality**.

What we are doing here is considering a random walk whose jumps have 0 mean. We wonder wether it is above the level a as seen in figure 1.7.

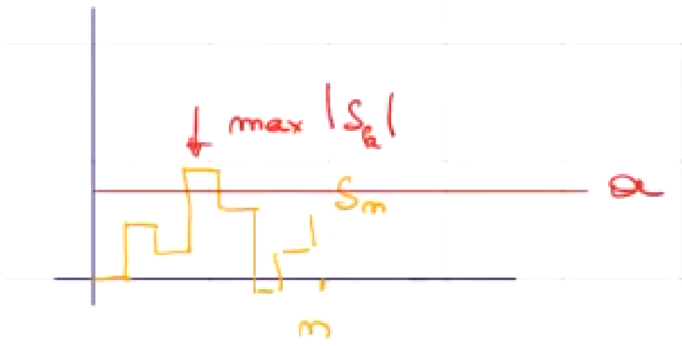


Figure 1.7: The maximum of the random walk.

About this, we know that $\{S_n\}$ is a $\mathcal{F}$-martingale but in proving the Kolmogorov's inequaility we never talked about the martingale property! We could improve this inequality using the **Doob's martingale inequality**. The problem is that we can prove very general inequalities that hold true for any random variables but specifying more characteristic we can obtain stricter bounds. In this framework let's define

$$\begin{aligned} M_n^\star &= \max_{k \leqslant n} M_k \\ m_n^\star &= \min_{k \leqslant n} M_k \end{aligned}$$

as current maximum and current minimum of M.

Theorem 1.2.15

Take M as a sub-martingale. For $b > 0$ it holds:

1. $b\mathbb{P}(M_n^\star \geqslant b) \leqslant \mathbb{E}\left[M_n \mathbb{1}_{\{M_n^\star \geqslant b\}}\right] \leqslant \mathbb{E}\left[M_n^+\right]$;
2. $b\mathbb{P}(m_n^\star \leqslant -b) \leqslant -\mathbb{E}M_0 + \mathbb{E}\left[M_n \mathbb{1}_{\{m_n^\star \geqslant b\}}\right] \leqslant \mathbb{E}M_n^+ - \mathbb{E}M_0$.

So we can further bound the result looking into the property of M_n.

Example 1.2.4

Now need to define the brownian motion (or Weiner process).

Definition 1.2.14

A real-valued stochastic process $B = (B_t)_{t \geqslant 0}$ is called **brownian motion** if:

1. the index set is $\mathbb{R}^+$;
2. $B_0(\omega) = 0$ for almost all ω;
3. $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ are independent for $\forall\, 0 = t_0 < t_1 < \dots < t_n < \infty$;
4. $B_t - B_s \sim B_{t+b} - B_{s+b}$ for every $0 \leqslant s < t < \infty \quad \forall n > -s$;
5. $B_t - B_s \sim N(0, t-s)$;
6. $t \mapsto B_t(\omega)$ are continuous for every ω.

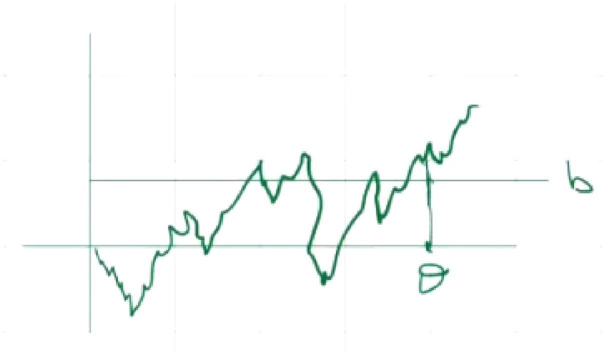
The brownian motion is in itself a martingale, but there is a class of martingales strictly related to it:

$$M_t = \exp\left\{\lambda B_t - \frac{1}{2}\lambda^2 t\right\}, \qquad t \in \mathbb{R}^+.$$

Now we can get to the actual example: if B is a brownian motion, we have

$$\mathbb{P}(\sup_{0 \leqslant t \leqslant \theta} B_t \geqslant b) \leqslant \exp\left\{-\frac{b^2}{s\theta}\right\}.$$

The sample paths of brownian motions are extremely irregular. We are asking with which probability the max of our process will be over b at time θ.



I am basically asking if the maximum attained is above or below b. How can we use the Doob's inequality?

Proof

Consider

$$\mathbb{P}(\sup B_t \geqslant b) = \mathbb{P}(\sup e^{\lambda B_t} \geqslant e^{\lambda b})$$

and using Doob's inequality

$$\begin{aligned}\mathbb{P}(\sup B_t \geqslant b) &= \mathbb{P}(\sup e^{\lambda B_t} \geqslant e^{\lambda b}) \\ &\leqslant \frac{\overbrace{\mathbb{E}\left[e^{\lambda B_\theta - \frac{\lambda^2}{2}\theta}\right]}^{\text{martingale}}}{e^{\lambda b} e^{-\frac{\lambda^2}{2}\theta}} \\ &= e^{-\lambda b + \frac{\lambda^2}{2}\theta} \qquad \forall \lambda > 0\end{aligned}$$

□

1.2.25 Predictable processes and Doob's decomposition

Definition 1.2.15

The process $F = (F_n)_{n \geqslant 1}$ is **$\mathcal{F}$-predictable** if $F_0 \in \mathcal{F}_0$ and $\mathcal{F}_{n+1} \in \mathcal{F}_n$, for $\forall n \in \mathbb{N}$.

This means that the available information up to n is "enough" to have a bet in the period $n+1$. Some predictable processes are:

- any deterministic processes;
- consider two stopping times S, T of $\mathcal{F}$ and let $S \leqslant T$. Consider the random variable V in $\mathcal{F}_n$. Then

$$\overset{(1)}{V\mathbb{1}_{(S,T]}} \quad \overset{(2)}{V\mathbb{1}_{(S,\infty]}} \quad \overset{(3)}{\mathbb{1}_{(S,T]}} \quad \overset{(4)}{\mathbb{1}_{(0,T]}}$$

  are predictable processes.

  Proof

  (2) Consider $F = V\mathbb{1}_{(S,\infty]}$, so that $F_n = \mathbb{1}_{(S,\infty]}(n)$. Consider

$$\begin{aligned}F_{n+1} &= V\mathbb{1}_{(S,\infty]}(n+1) \\ &= V\mathbb{1}_{\{S<n+1\}} \\ &= V\mathbb{1}_{\{S\leqslant n\}} \in \mathcal{F}_n\end{aligned}$$

so the process is $\mathcal{F}$-measurable and predictable.

(1) Since we know by hypothesis that $S \leqslant T$ then $V \in \mathcal{F}_S \subset \mathcal{F}_T$. This means that $V \in \mathcal{F}_T$. Hence, of a consequence,

$$V\mathbb{1}_{(T,\infty]} - V\mathbb{1}_{(S,\infty]} = V\mathbb{1}_{(S,T]}$$

is predictable.

(3) Take $V = 1$.

(4) Take $T = \infty, V = 1$. $\mathbb{1}_{(S,\infty]}$ is predictable. But then

$$\mathbb{1}_{[0,S]} = \mathbb{1} - \mathbb{1}_{(S,\infty]}$$

so it is predictable.

□

Theorem 1.2.16

Doob's decomposition.
X is a stochastic process which is adapted to $\mathcal{F}$ and integrable. Then

1. it can be decomposed as
$$X_n = X_0 + M_n + A_n, \qquad n \in \mathbb{N}$$
where:
   - M_n is a $\mathcal{F}$-martingale with $M_0 = 0$;
   - A_n is a predictable process with $A_0 = 0$.
2. The decomposition is unique up to equivalence.
3. If X_n is a sub-martingale then $\{A_n\}_{n\geqslant 0}$ increasing, while if X_n is a super-martingale then $\{A_n\}_{n\geqslant 0}$ is decreasing.

Proof

Put $A_0 = M_0 = 0$. Define M and A through their increments:

$$\begin{aligned} A_{n+1} - A_n &= \underbrace{\mathbb{E}\Big[X_{n+1} - X_n|\mathcal{F}_n\Big]}_{\mathcal{F}_n-\text{meas.}} \\ M_{n+1} - M_n &= (X_{n+1} - X_n) - (A_{n+1} - A_n). \end{aligned}$$

If we look at these quantities we see that A is predictable and M is martingale. Imagine now there is another decomposition: let

$$X = X_0 + M' + A'$$

be another decomposition. We must have

$$\cancel{X_0} + M' + A' = \cancel{X_0} + M + A \iff A - A' = M - M' = B.$$

Now B is a process and it is predictable and martingale (because it is the difference between two martingales). Since B is predictable *and* a martingale we have

$$\begin{aligned} B_{n+1} - B_n &\underset{\text{predictable}}{=} \mathbb{E}[B_{n+1} - B_n|\mathcal{F}_n] \\ &\underset{\text{martingale}}{=} 0 \\ &\implies B_{n+1} = B_n = B_0 \text{ a.s.}, A = A' \text{ a.s.}, M = M' \text{ a.s.} \end{aligned}$$

If X is a sub-martingale then we have

$$\mathbb{E}[X_{n+1} - X_{n|\mathcal{F}_n}] \geqslant 0$$

and this means that we have $A_{n+1} \geqslant A_n$ is increasing. □

1.2.26 Doob's stopping theorem

For a martingale we know

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s] = 0.$$

The question is: is this true also if s and t are substituted by stopping times $S, T, S \leqslant T$?

Theorem 1.2.17

Let M be adapted to $\mathcal{F}$. Then the following are equivalent:

1. M is a submartingale;
2. for every bounded stopping time $S \leqslant T$ the random variables M_S and M_T are integrable and
$$\mathbb{E}[M_T - M_S|\mathcal{F}_S] \geqslant 0;$$
3. for each pair of bounded stopping times the random variables M_S and M_T are integrable and
$$\mathbb{E}[M_T - M_S] \geqslant 0.$$

Remark

If M is a martingale the theorem can be read in a different way:

$$\mathbb{E}M_T = \mathbb{E}M_S = \mathbb{E}M_0.$$

Previously, $\mathbb{E}M_n = \mathbb{E}M_{n-1} = \mathbb{E}M_0$.

Proof

To prove the theorem we need to show that from condition 1 follows 2 from which follows 3 from which follows 1.

$1 \to 2$ by hypothesis M is a sub-martingale and our thesis is that if $S(\omega) < T(\omega) < n$ (because we asked for bounded times) then:

(a) M_S and M_T are integrable;

(b) $\mathbb{E}[M_T - M_S|\mathcal{F}_S] \geqslant 0$.

We know that S and T are bounded by n. Let V be a positive bounded random variable and define $F = V\mathbb{1}_{(S,T]}$ and use it in the discrete time integral:

$$X_n = \underbrace{M_0 F_0}_{X_0} + (M_1 - M_0)\underbrace{F_1}_{V\mathbb{1}_{\{1\in(S,T]\}}} + \ldots + (M_n - M_{n-1})F_n$$

So we have

$$X_n - X_0 = V(M_T - M_S)$$

So X_n is a sub-martingale. Take $V = 1$ and $S = 0$: now we have

$$\underset{\text{int.}}{\underset{\downarrow}{X_n}} - \underset{\text{int.}}{\underset{\downarrow}{X_0}} = M_T \implies M_T \text{ is integrable.}$$

Now take $V = 1$ and $T = n$ so that we get

$$\underset{\text{int.}}{\underset{\downarrow}{X_n}} - \underset{\text{int.}}{\underset{\downarrow}{X_0}} = \underset{\text{int.}}{\underset{\downarrow}{M_n}} - M_S \implies M_S \text{ is integrable.}$$

We recall that $V \in \mathcal{F}_S$ and we use the defining property for $\mathbb{E}(\cdot|\mathcal{F}_S)$. So we can write

$$\begin{aligned}\mathbb{E}V\mathbb{E}(M_T - M_S|\mathcal{F}_S) &\underset{\text{def. prop.}}{=} \mathbb{E}V(M_T - M_S) \\ &\underset{\text{discr. time int.}}{=} \mathbb{E}[X_n - X_0] \\ &\underset{\text{proved above}}{\geqslant} 0\end{aligned}$$

and this is true $\forall V > 0, V < b, V \in \mathcal{F}_s$. Hence

$$\mathbb{E}(M_T - M_S|\mathcal{F}_S) \geqslant 0$$

So $1 \to 2$.

$2 \to 3$ We can use the tower rule. Take the expectation of point 2:

$$\mathbb{E}[\mathbb{E}(M_T - M_S|\mathcal{F}_S)] = \mathbb{E}[M_T - M_S] \geqslant 0.$$

$3 \to 1$ Let 3 hold, so that $\mathbb{E}[M_T - M_S] \geqslant 0$. Choose $T = n$ and $S = 0$. Then M_n is integrable. Move to adaptness: this holds by hypothesis. Move to the martingale inequality:

$$\mathbb{E}[M_n - M_m|\mathcal{F}_m] \geqslant 0.$$

Note that this is equivalent to prove

$$\mathbb{E}\mathbb{1}_H\mathbb{E}[M_n - M_m|\mathcal{F}_m] \geqslant 0 \qquad H \in \mathcal{F}_m, 0 \leqslant m \leqslant n.$$

Fix H, m, n and define

$$S(\omega) = m \qquad T(\omega) = n\mathbb{1}_H(\omega) + m\mathbb{1}_{\{\Omega\setminus H\}}(\omega)$$

The indicators are non-zero in complementary instances. Notice that:

(a) S is a fixed time so it is a stopping time;
(b) $S \leqslant T \leqslant n$ by definition of S and T because the indicators are non-zero in complementary instances;
(c) $T \geqslant S$ is a foretold time by $S = m$;
(d) $H \in \mathcal{F}_S$ by definition.

So we can write $M_T - M_S = \mathbb{1}_H(M_n - M_m)$ where $M_T - M_S \geqslant 0$ by hypothesis. This means that we have

$$\mathbb{E}[\underbrace{\mathbb{1}_H\mathbb{E}(M_n - M_m|\mathcal{F}_m)]}_{\mathbb{E}[M_n - M_m|\mathcal{F}_m]\geqslant 0} \geqslant 0$$

□

1.2.27 Upcrossing inequality

I believe that what is being asked is

Proposition 1.2.9

If M is a sub-martingale with respect to its natural filtration then

$$(b-a)\mathbb{E}U_n(a,b) \leqslant \mathbb{E}\left[(M_n - a)^+ - (M_0 - a)^+\right].$$

We wanted to find a bound for our profit, but our profit is a stochastic quantity: so it's only natural to think about the expectation to give a bound to the number of expected up/downcrossing.
Observe that the number of upcrossings does not depend on the value of T_0 that we fix.

Proof

Choose $a = 0$. Consider hence the process $(M - a)^+$ that is sub-martingale (if M is a sub-martingale). Take $M \geqslant 0$ and let

$$F_n = \mathbb{1}_{(S_k, T_k]}(n)$$

and consider

$$X = \int F \,\mathrm{d}M$$

like in our example. We know that F is predictable since by definition $F_{k+1} \in \mathcal{F}_k$. Consider the expectation of the increment

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] = \mathbb{E}\left[F_{k+1}(M_{k+1} - M_k) | \mathcal{F}_k\right].$$

But since F_{k+1} is predictable we can take it out the expectation:

$$F_{k+1}\mathbb{E}\left[M_{k+1} - M_k | \mathcal{F}_k\right].$$

But since F_{k+1} is an indicator we know it is $\leqslant 1$:

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] \leqslant \mathbb{E}\left[M_{k+1} - M_k | \mathcal{F}_k\right].$$

Now take the expectation of both sides:

$$\mathbb{E}[X_{k+1} - X_k] \leqslant \mathbb{E}\left[M_{k+1} - M_k\right].$$

If we sum these inequalities over k we get:

$$\mathbb{E}\left[X_n - X_0\right] \leqslant \mathbb{E}\left[M_n - M_0\right].$$

So we now get that

$$bU_n(a, n) \leqslant \mathbb{E}\left[X_n - X_0\right] \leqslant \mathbb{E}\left[M_n - M_0\right]$$

but given that $a = 0$ we get that

$$bU_n(0, b) \leqslant \mathbb{E}[M_n - M_0].$$

Clearly we have to take the positive part. □

But I also found this

Proposition 1.2.10

Suppose that $\{X = X_t : t \in [0, \infty)\}$ satisfies the basic assumptions with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \in [0, \infty)\}$ and let $a, b \in \mathbb{R}$ with $a < b$. Let $U_t = u_t(a, b, X)$ the random number of upcrossings of $[a, b]$ by X up to time $t \in [0, \infty)$.

1. if X is a super-martingale relative to $\mathcal{F}$ then

$$\mathbb{E}(U_t) \leqslant \frac{1}{b-a}\mathbb{E}\left[(X_t - a)^-\right] \leqslant \frac{1}{b-a}\left[\mathbb{E}(X_t^-) + |a|\right] \leqslant \frac{1}{b-a}\left[\mathbb{E}(|X_t|) + |a|\right];$$

2. if X is a sub-martingale relative to $\mathcal{F}$ then

$$\mathbb{E}(U_t) \leqslant \frac{1}{b-a}\mathbb{E}\left[(X_t - a)^+\right] \leqslant \frac{1}{b-a}\left[\mathbb{E}(X_t^+) + |a|\right] \leqslant \frac{1}{b-a}\left[\mathbb{E}(|X_t|) + |a|\right];$$

1.2.28 Doob's decomposition

Theorem 1.2.18

Doob's decomposition.
X is a stochastic process which is adapted to $\mathcal{F}$ and integrable. Then

1. it can be decomposed as
$$X_n = X_0 + M_n + A_n, \qquad n \in \mathbb{N}$$
where:
   - M_n is a $\mathcal{F}$-martingale with $M_0 = 0$;
   - A_n is a predictable process with $A_0 = 0$.
2. The decomposition is unique up to equivalence.
3. If X_n is a sub-martingale then $\{A_n\}_{n\geqslant 0}$ increasing, while if X_n is a super-martingale then $\{A_n\}_{n\geqslant 0}$ is decreasing.

Proof

Put $A_0 = M_0 = 0$. Define M and A through their increments:

$$\begin{aligned} A_{n+1} - A_n &= \underbrace{\mathbb{E}\Big[X_{n+1} - X_n | \mathcal{F}_n\Big]}_{\mathcal{F}_n-\text{meas.}} \\ M_{n+1} - M_n &= (X_{n+1} - X_n) - (A_{n+1} - A_n). \end{aligned}$$

If we look at these quantities we see that A is predictable and M is martingale. Imagine now there is another decomposition: let
$$X = X_0 + M' + A'$$
be another decomposition. We must have
$$\cancel{X_0} + M' + A' = \cancel{X_0} + M + A \iff A - A' = M - M' = B.$$

Now B is a process and it is predictable and martingale (because it is the difference between two martingales). Since B is predictable *and* a martingale we have

$$\begin{aligned} B_{n+1} - B_n &\underset{\text{predictable}}{=} \mathbb{E}[B_{n+1} - B_n | \mathcal{F}_n] \\ &\underset{\text{martingale}}{=} 0 \\ &\implies B_{n+1} = B_n = B_0 \;\text{ a.s.}, A = A' \;\text{ a.s.}, M = M' \text{ a.s.} \end{aligned}$$

If X is a sub-martingale then we have
$$\mathbb{E}[X_{n+1} - X_{n|\mathcal{F}_n}] \geqslant 0$$
and this means that we have $A_{n+1} \geqslant A_n$ is increasing. □

1.2.29 Stochastic integral in discrete time

Let us consider two real-valued processes $M = (M_n)_n$ and $F = (F_n)_n$ and let us define
$$X_n = F_0 M_0 + (M_1 - M_0)F_1 + \ldots + (M_n - M_{n-1})F_n.$$
We say that $\{X_n\}$ is the integral of F with respect to M and we write
$$X_n = \int F \,\mathrm{d}M$$
where $\mathrm{d}M$ is a random signed measure. Remember the Lebesgue-Stieltjes integral? Me neither, but as long as M has bounded variation this is a Lebesgue-Stieltjes integral. So a little explanation is due since I actually

never saw a Stieltjes integral.

Theorem 1.2.19

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 1.2.5

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n=1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^{n}(M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k, but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

1. We have M being a martingale and $F_0, F_1, \ldots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \ldots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to $\mathcal{F}$.

2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b. This implies

$$|X_n| < b(|M_0| + |M_1 + M_0| + \ldots + |M_n - M_{n-1}|)$$

Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n]$$

since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$F_n \underbrace{\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

□

1.2.30 Stochastic integral and its application

Let us consider two real-valued processes $M = (M_n)_n$ and $F = (F_n)_n$ and let us define

$$X_n = F_0 M_0 + (M_1 - M_0)F_1 + \ldots + (M_n - M_{n-1})F_n.$$

We say that $\{X_n\}$ is the integral of F with respect to M and we write

$$X_n = \int F \, \mathrm{d}M$$

where $\mathrm{d}M$ is a random signed measure. Remember the Lebesgue-Stieltjes integral? Me neither, but as long as M has bounded variation this is a Lebesgue-Stieltjes integral. So a little explanation is due since I actually never saw a Stieltjes integral.

Theorem 1.2.20

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 1.2.6

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n=1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^{n}(M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k, but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

1. We have M being a martingale and $F_0, F_1, \ldots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \ldots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to $\mathcal{F}$.

2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b. This implies

$$|X_n| < b(|M_0| + |M_1 + M_0| + \ldots + |M_n - M_{n-1}|)$$

   Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n)|\mathcal{F}_n]$$

   since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$F_n \underbrace{\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

$\square$

and for the applications:

Definition 1.2.16

Define $M = (M_n)_{n\in\mathbb{N}}$ as a process and let T be a random time with values on $\overline{\mathbb{N}}$. The process

$$X_n(\omega) = M_{n\wedge T}(\omega) = \begin{cases} M_n(\omega) & n < T(\omega) \\ M_T(\omega) & n > T(\omega) \end{cases}$$

(where $n \wedge T$ is a truncated random time) is called **M stopped at T**.

As a consequence X is exactly the discrete time integral if $F = \mathbb{1}_{[0,T]}$:

$$X_n = \underbrace{M_0F_0}_{0} + (M_1 + M_0)\mathbb{1}_{[0,T]}(1) + \ldots + (M_n - M_{n-1})\mathbb{1}_{[0,T]}(n).$$

The indicators only select the current time interval. If this is the case we can observe that $F_{[0,T]}$ is bounded, positive and predictable. Hence if M is a martingale the theorem applies with this special choice of M and we can write the result as a different theorem:

Theorem 1.2.21

Let T be a stopping time and let X be the process M stopped at T. If M is a martingale then so is X (the same holds for sub-martingales and super-martingales).

So we cannot determine a policy based on stopping times that can change the nature of our martingale. In the remote case in which you are interested in this you can read Williams - Introduction to martingales.

1.2.31 Impossibility to win against the system: related theorems and examples

Theorem 1.2.22

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 1.2.7

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n=1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^{n}(M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k, but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

1. We have M being a martingale and $F_0, F_1, \ldots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \ldots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to $\mathcal{F}$.

2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b. This implies

$$|X_n| < b(|M_0| + |M_1 + M_0| + \ldots + |M_n - M_{n-1}|)$$

Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n]$$

since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$F_n \underbrace{\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

□

We know that using a policy that it is predictable it is impossible to beat the system. Finance bros try to overcome this possibility using stopping times. If my policy is not only based on a predictable process but I add the randomness of the time in which I decide to sell or buy can I break the curse of martingales and make shareholders want to suck my dick? Well, it depends.

1.2.32 Predictable and adapted processes: definition and examples

Definition 1.2.17

The process $F = (F_n)_{n \geqslant 1}$ is **$\mathcal{F}$-predictable** if $F_0 \in \mathcal{F}_0$ and $\mathcal{F}_{n+1} \in \mathcal{F}_n$, for $\forall\, n \in \mathbb{N}$.

This means that the available information up to n is "enough" to have a bet in the period $n+1$. Some predictable processes are:

- any deterministic processes;
- consider two stopping times S, T of $\mathcal{F}$ and let $S \leqslant T$. Consider the random variable V in $\mathcal{F}_n$. Then

$$\underset{(1)}{V\mathbb{1}_{(S,T]}} \quad \underset{(2)}{V\mathbb{1}_{(S,\infty]}} \quad \underset{(3)}{\mathbb{1}_{(S,T]}} \quad \underset{(4)}{\mathbb{1}_{(0,T]}}$$

are predictable processes.

Proof

(2) Consider $F = V\mathbb{1}_{(S,\infty]}$, so that $F_n = \mathbb{1}_{(S,\infty]}(n)$. Consider

$$\begin{aligned} F_{n+1} &= V\mathbb{1}_{(S,\infty]}(n+1) \\ &= V\mathbb{1}_{\{S<n+1\}} \\ &= V\mathbb{1}_{\{S\leqslant n\}} \in \mathcal{F}_n \end{aligned}$$

so the process is $\mathcal{F}$-measurable and predictable.

(1) Since we know by hypothesis that $S \leqslant T$ then $V \in \mathcal{F}_S \subset \mathcal{F}_T$. This means that $V \in \mathcal{F}_T$. Hence, of a consequence,

$$V\mathbb{1}_{(T,\infty]} - V\mathbb{1}_{(S,\infty]} = V\mathbb{1}_{(S,T]}$$

is predictable.

(3) Take $V = 1$.

(4) Take $T = \infty$, $V = 1$. $\mathbb{1}_{(S,\infty]}$ is predictable. But then

$$\mathbb{1}_{[0,S]} = \mathbb{1} - \mathbb{1}_{(S,\infty]}$$

so it is predictable. □

Example of adapted process:

Example 1.2.8

This is called "the secretary problem": in this case we must start from the filtration and then understand the problem. Here i have N candidates for a position; a candidate disregarded after the interview is lost. The interviewer wants to hire exactly 1 candidate and each candidate has different abilities and the interviewer knows only the relative ability of those already interviewed so far. Our goal is to maximizing the probability of hiring the best one. We have three questions:

1. what is Ω?
2. what is the filtration $\mathcal{F}$ for this experiment?
3. what process should we use?

In this case $\Omega = N!$ permutations of the ranking of the candidates (the order in which they show up) and the filtration is the information earned from interview up to time t (that is the ranking of the candidates up to time t). But what is the process that I should use? Consider the sequence

$$V_1, V_2, \dots \qquad \{V_i\}_{i\geqslant 1}$$

with $V_i = 1$ if and only if the best candidate is the i-th candidate and $V_i = 0$ otherwise. Could this process $\{V_i\}_{i\geqslant 1}$ be used? No, because V is not adapted to $\mathcal{F}$... because to understand if i-th candidate is te best we need to compare it to the other candidates, including the ones that didn't show up yet! But then how can we get an adapted process? Let us consider the expectation

$$U_n = \mathbb{E}\left[V_n | \mathcal{F}_n\right]$$

What do we know about the measurability of U_n? We know that it is for sure $\mathcal{F}_n$-measurable. This trick gives us a simple way to build an adapted process. So now we will have: $U_n = 0$ if the candidate is not the best up to n and $U_n = 1$ otherwise. More specifically, we will have

$$\begin{aligned} U_n &= 1 \cdot \begin{array}{c}\scriptstyle\text{proability that the best candidate}\\ \scriptstyle\text{is among the first } n\end{array} + 0 \cdot \begin{array}{c}\scriptstyle\text{proability that the best candidate}\\ \scriptstyle\text{is not among the first } n\end{array} \\ &= 1 \cdot \frac{n}{N} + 0 \cdot \frac{N-n}{N}. \\ &= \frac{n}{N} \end{aligned}$$

This is a quantity that I can measure and it is therefore adapted.

1.2.33 Poisson process and its martingale property

Consider $\mathbb{R}_+$ as our index set. $\mathcal{F}$ is our filtration and we consider the counting process $N = (N_t)_{y\geqslant 0}$: this counts the number of events up to time t, it has unit jumps and any path starts from 0 so that $N_0(\omega) = 0$, it is increasing and it is right continuous.

Definition 1.2.18

The counting process N is said to be a **Poisson process** with rate λ with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$ and

$$\mathbb{E}[f(N_{t+s} - N_s)|\mathcal{F}_s] = \sum_{k=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!} f(k) \qquad \forall\, s,t \in \mathbb{R}_+, \forall \text{ positive } f \mapsto \mathbb{N}.$$

Theorem 1.2.23

Let N be a counting process. It is a Poisson process with rate λ with respect to $\mathcal{F}$ if and only if:

$$M_t = N_t - \lambda t$$

is a $\mathcal{F}$-martingale.

We only prove that M_t is a martingale if N_t is Poisson.

Proof

We know that

$$\begin{aligned}
\mathbb{E}[M_t|\mathcal{F}_s] &= \mathbb{E}[M_t - M_s + M_s|\mathcal{F}_s] \\
&= \mathbb{E}[M_t - M_s|\mathcal{F}_s] + M_s \\
&= \mathbb{E}[M_t - M_s] + M_s \\
&= \mathbb{E}[N_t - N_s + \lambda t + \lambda s] + M_s \\
&= \underbrace{\mathbb{E}[N_t - N_s]}_{\cancel{\lambda(t-s)}} - \cancel{\lambda(t-s)} + M_s \\
&= M_s
\end{aligned}$$

$\square$

1.2.34 Stopped processes and their properties

Example 1.2.9

Some stopping times:

1. The first time that $X(\omega) \in H \in \Omega$;

$$T(\omega) = \begin{cases} \inf\{n \in \mathbb{N} : X_n(\omega) \in H\} & \\ +\infty & \text{if } X_n(\omega) \notin H \ \forall n \end{cases}$$

So

$$\{T \leqslant n\} = \bigcup_{k=0}^{n} \{X_k \in H\}.$$

2. Consider i.i.d. random variables $X_1, X_2, \ldots$. Consider the probabilities

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$$

and the random walk

$$S_n = \sum_{1}^{n} X_i.$$

Let's define

$$T_1 = \begin{cases} \min\{n < 50 : S_n = 3\} & \\ 50 & \text{otherwise.} \end{cases}$$

This is a stopping time because I can write $\{T_1 \leqslant n\}$ as

$$\underbrace{\bigcup_{k=1}^{n} \underbrace{\{S_k = 3\}}_{\mathcal{F}_k\text{-measurable}}}_{\mathcal{F}_n\text{-measurbale}} \qquad n < 50.$$

Moreover for $n = 50$ we have $T_1 \in \mathcal{F}_{50}$.

3. Starting from the previously deifned random walk, consider the quantity

$$M_n = \min(S_1, \dots, S_n)$$

And the random time

$$T_2 = \min\{n : S_n \geqslant M_m + 2\}$$

is a stoppimg time. On the contrary,

$$T_3 = \begin{cases} \max\{n < 50 : S_n = 7\} & \text{if not empty} \\ 50 & \text{otherwise} \end{cases}$$

is not a stopping time. Why? Because I have to wait until $n = 50$ to answer the question.

Consider the random times on $\mathbb{R}_+$

$$0 < T_1 < T_2 < \dots$$

With $\lim_{n\to\infty} T_n = +\infty$. Define the process $\{N_t\}$ as

$$N_t := \sum \mathbb{1}_{[0,t]}(T_n).$$

This is called **counting process**. It is a basic count of the number of events happened up to time n. N_t is increasing, right continuous and incresases by unitary jumps. Moreover, $N_0 = 0$, $N_t < \infty$ for $t \in \mathbb{R}_+$. Of course $\lim_{t\to\infty} N_t = \infty$. Counting processes generate their natural filtration $\mathcal{F}$.

Some problems require to stop the observation at a stopping time (because we don't care anymore[10])... So we don't actually need the whole knowledge of the complete filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geqslant 0}$. The problem is that as we said before the stopping time is a random time. So what we need is the *information known up to time T $\mathcal{F}_T$*, basically the σ-field that is the filtration at time T.

Example 1.2.10

Truncated stopping time: let T be a stopping time (for example, the time at which we sell certain shares) and that we want a finite horizon for this decision. In this case the quantity of interest is

$$S = T \wedge n = \min\{T, n\}$$

where n could be some sort of time horizon.
Imagine that two cyclists participate to a race. Their children will have their snack when both the parents will arrive to the finish line. How long will the children wait for their snack? We can think about the following stopping times:

$$\begin{aligned} T &: \text{ time employed by the first cyclist} \\ S &: \text{ time employed by the second cyclist} \\ U &: \max\{S, T\}. \end{aligned}$$

The waiting time for the children will be U.

1.2.35 Important inequalities for sub-martingales

The problem is that we can prove very general inequalities that hold true for any random variables but specifying more characteristic we can obtain stricter bounds. In this framework let's define

$$\begin{aligned} M_n^\star &= \max_{k\leqslant n} M_k \\ m_n^\star &= \min_{k\leqslant n} M_k \end{aligned}$$

as current maximum and current minimum of M.

[10] Assuming we ever did.

Theorem 1.2.24

Take M as a sub-martingale. For $b > 0$ it holds:

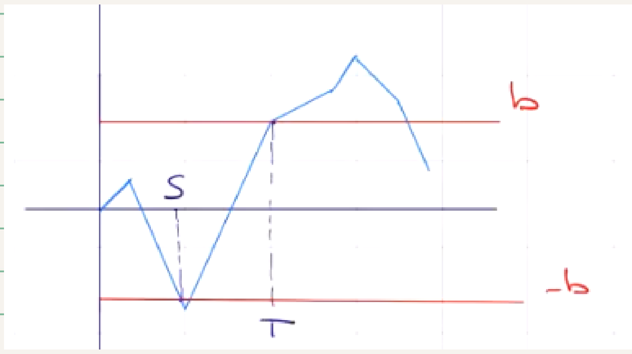
1. $b\mathbb{P}(M_n^\star \geqslant b) \leqslant \mathbb{E}\left[M_n \mathbb{1}_{\{M_n^\star \geqslant b\}}\right] \leqslant \mathbb{E}\left[M_n^+\right]$;
2. $b\mathbb{P}(m_n^\star \leqslant -b) \leqslant -\mathbb{E}M_0 + \mathbb{E}\left[M_n \mathbb{1}_{\{m_n^\star \geqslant b\}}\right] \leqslant \mathbb{E}M_n^+ - \mathbb{E}M_0$.

So we can further bound the result looking into the property of M_n.

Proof

We introduce 2 stopping times:

$$\begin{aligned} T &= \inf\{n \geqslant 0 : M_n \geqslant b\} \\ S &= \inf\{n \geqslant 0 : M_n \leqslant -b\}. \end{aligned}$$



Consider the current maximum and minimum above the level b:

$$\{M_n^\star \geqslant b\} = \{T \leqslant n\} \qquad \{m_n^\star < -b\} = \{S \leqslant n\}.$$

Fix b and n. When $\{T \leqslant n\}$ we have

$$M_{T\wedge n} = M_T \geqslant b.$$

Multiply by $\mathbb{1}_{\{T\leqslant n\}}$:

$$b\mathbb{1}_{\{T\leqslant n\}} \leqslant M_{T\wedge n}\mathbb{1}_{\{T\leqslant n\}}.$$

Using Doob's stopping theorem we know that

$$\mathbb{E}\left[M_T - M_S | \mathcal{F}_S\right] \geqslant 0$$

and using $S = T \wedge n$ and $T = n$ we obtain

$$\mathbb{E}\left[M_n | \mathcal{F}_{T\wedge n}\right] > M_{T\wedge n}.$$

So summing it all up:

$$\begin{aligned} b\mathbb{1}_{\{T\leqslant n\}} &\leqslant M_{T\wedge n}\mathbb{1}_{\{T\leqslant n\}} \\ &\leqslant \mathbb{1}_{\{T\leqslant n\}}\mathbb{E}\left[M_n | \mathcal{F}_{T\wedge n}\right] \\ &= \mathbb{E}\left[M_n \mathbb{1}_{\{T\leqslant n\}} | \mathcal{F}_{T\wedge n}\right]. \end{aligned}$$

Now take the expectation:

$$\begin{aligned} b\mathbb{P}(T \leqslant n) &= b\mathbb{P}(M_n^\star \geqslant b) \\ &\leqslant \underbrace{\mathbb{E}\left[M_n \mathbb{1}_{\{T\leqslant n\}}\right]}_{\mathbb{E}\left[M_n \mathbb{1}_{\{M_n^\star \geqslant b\}}\right]} \\ &\leqslant \mathbb{E}(M_n^+). \end{aligned}$$

For the minimum we work on $\{S \leqslant n\}$ and we get

$$\begin{aligned} M_{S\wedge n} &= M_S \mathbb{1}_{\{S\leqslant n\}} + M_S \mathbb{1}_{\{S>n\}} \\ &\leqslant -b\mathbb{1}_{\{S\leqslant n\}} + M_n \mathbb{1}_{S\geqslant n}. \end{aligned}$$

Now take the expectation:

$$\mathbb{E}M_{S\wedge n} \leqslant -b\mathbb{P}(m_n^\star \leqslant -b) + \mathbb{E}\left[M_n \mathbb{1}_{\{S>n\}}\right].$$

So that we get

$$\mathbb{P}(m_n^\star \leqslant -b) \leqslant -\mathbb{E}M_{S\wedge n} + \mathbb{E}[M_n \mathbb{1}_{\{S>n\}}].$$

Now use Doob's stopping theorem with $T = 0$ and $S = S \wedge n$ so that $\mathbb{E}M_0 \leqslant \mathbb{E}M_{\{S\wedge n\}}$. This gets us our result:

$$\begin{aligned} b\mathbb{P}(m_n^\star - b) &\leqslant -\mathbb{E}M_0 + \mathbb{E}\left[M_n \mathbb{1}_{\{m_n^\star\}}\right] \\ &\leqslant \mathbb{E}M_n^+ - \mathbb{E}M_0. \end{aligned}$$

□

1.2.36 Convergence theorems for sub-martingales

Theorem 1.2.25

Let X be a sub-martingale. If (and note that is a sufficient condition)

$$\sup_n \mathbb{E}X_n^+ < \infty$$

Then

1. $\{X_n\}$ converges a.s.;
2. $\{X_n\}$ converges to an integrable random variable.

Proof

We prove the theorem by contradiction. Pick an outcome ω and suppose that $\{X_n(\omega)\}$ is a numerical sequence that has not a limit. But if it doesn't have a limit, then

$$\exists \inf\lim \neq \sup\lim \qquad \inf\lim < \sup\lim.$$

So there exist at least 2 rationals $a < b$ such that

$$\inf\lim < a < b < \sup\lim.$$

The sequence $\{X_n(\omega)\}$ crosses (a, b) ∞ many times. Now take the union over rational a and b, $a < b$ of the sets

$$\{U(a, b) = \infty\}$$

with $U(a, b) = \lim_{n\to\infty} U_n(a, b)$. Our aim is now to show that $U(a, b) \leqslant \infty$ almost surely to get a contradiction.
Fix a, b. We know that $U_n(a, b)$ is increasing with n. Now consider

$$\begin{aligned} (b-a)\mathbb{E}U(a, b) &= (b-a)\mathbb{E}\lim U_n(a, b) \\ &\underset{\text{monotone conv.}}{=} (b-a)\lim \mathbb{E}U_n(a, b) \\ &\underset{\text{upcross inequalities}}{\leqslant} \sup \mathbb{E}(X_n - a)^+ \\ &\leqslant \sup \mathbb{E}X_n^+ + |a|. \end{aligned}$$

So this means that $\mathbb{E}U(a, b) < \infty$. But this is a contradiction, so it exists a limit $X_n = X_\infty$ a.s..

Now consider the second part of the theorem:

$$\begin{aligned}\mathbb{E}|X_\infty| &= \mathbb{E}\liminf |X_n| \\ &\underset{\text{Fatou's lemma}}{\leqslant} \liminf \mathbb{E}|X_n| \\ &\leqslant 2\sup \mathbb{E}X_n^+ - \mathbb{E}X_0 \leqslant \infty\end{aligned}$$

so the limit is integrable. □

1.2.37 Uniform integrability and its consequences on convergence of martingales

We will need:

1. a collection $\mathcal{K}$ of real random variables is said to be uniformly integrable if

$$k(b) = \sup_{X\in\mathcal{K}} \mathbb{E}|X|\mathbb{1}_{\{X>b\}} \xrightarrow[b\to\infty]{} 0.$$

2. If $\mathcal{K}$ is dominated by an integrable random variable Z then it is uniformly integrable.
3. uniform integrability implies L^1-boundedness but not the converse.
4. If $\mathcal{K}$ is L^p-bounded for some $p > 1$ then it is uniformly integrable.

Lemma 1.2.2

Let Z be an integrable random variable. Then

$$\mathcal{K} = \{X : X = \mathbb{E}(Z|\mathcal{G})\}$$

for some sub-σ-algebra $\mathcal{G}$ of $\mathcal{H}$ is uniformly integrable.

Proposition 1.2.11

Let Z be an integrable random variable. Define

$$X_t = \mathbb{E}(Z|\mathcal{F}_t) \qquad t \in \mathbb{T}.$$

This means that $\{X_t\}$ is a uniformly integrable $\mathcal{F}$-martingale.

Theorem 1.2.26

Let $\{X_n\}$ be a sequence of real-valued random variables. The following are equivalent:

1. it converges in L^1;
2. it converges in probability and it is uniformly integrable.

We can now prove the theorem about the convergence of sub-martingales.

Theorem 1.2.27

Let X be a sub-martingale. We have that X converges almost surely and in L^1 if and only if it is uniformly integrable. Moreover, if it is so, setting

$$X_\infty = \lim X_n$$

extends X to a sub-martingale

$$\overline{X} = (X_n)_{n\in\overline{\mathbb{N}}}.$$

We only prove the first part of the theorem.

Proof

Necessity. If X converges in L^1 by the theorem above it is uniformly integrable.
Sufficiency. If X is uniformly integral then it is L^1-bounded for the property above. So our previous theorem holds and the martingale converges almost surely with X_∞ integrable. Furthermore, for the property above, it also converges in L^1. □

1.2.38 Features of the sample paths of a submartingale (or martingale)

Remark

If M is a martingale then $|M|^p$ is a sub-martingale for $p \geqslant 1$. If $M_n \in L^p \forall n$ we can apply Doob's inequality.

Corollary

If M is martingale in L^p for some $p \geqslant 1$ then for $b > 0$ we have that

$$b^p \mathbb{P}(\max_{k \leqslant n} |M_k| > b) \leqslant \mathbb{E}|M_n|^p.$$

There are also other bounds:

- $b\mathbb{P}(\max_{k\leqslant n} |M_k > b) \leqslant 2\mathbb{E}M_n^+ - 3M_0$;
- Doob's norm inequality: if M is a martingale in L^p, $p \geqslant 1$ and q is the exponent conjugate to p ($\frac{1}{p} + \frac{1}{q} = 1$) then
$$\mathbb{E} \max_{k \leqslant n} |M_k|^p \leqslant q^p \mathbb{E}|M_n|^p.$$
- Consider L^2-bounded martingales characterized by final coordinate X with $\mathbb{V}\mathrm{ar}\, X = \sigma^2$ (that is I am fixing the variance of the last value I consider). We want to assess the variability of this process.

Theorem 1.2.28

Dubin & Schwartz 1998: it holds

1. $\mathbb{E}\left[\max_{0\leqslant T\leqslant t} M_T\right] \leqslant \sigma$;
2. $\mathbb{E}\left[\max_{0\leqslant T\leqslant t} |M_T|\right] \leqslant \sigma\sqrt{2}$.

Moreover, there exist suitable martingales for which this bound is attained and is scrict.

1.2.39 Uniform integrability and its role in convergence problems

Theorem 1.2.29

A process $M = (M_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale if and only if for some integrable random variable Z

$$M_n = \mathbb{E}\left[Z|\mathcal{F}_n\right] \qquad n \in \mathbb{N}. \tag{$\bullet$}$$

If so it converges almost surely and in L^1 to the integrable random variable

$$M_\infty = \mathbb{E}[Z|\mathcal{F}_\infty].$$

Corollary

For every integrable random variable Z we have

$$\mathbb{E}(Z|\mathcal{F}_n) \xrightarrow{\text{a.s.}} \xrightarrow{L^1} \mathbb{E}(Z|\mathcal{F}_\infty).$$

Theorem 1.2.30

Let Z be an integrable random variable and let

$$M_n = \mathbb{E}(Z|\mathcal{F}_n)_{n\in\overline{\mathbb{N}}}.$$

For every stopping time T define

$$M_T = \mathbb{E}[Z|\mathcal{F}_T]$$

and for arbitrary stopping times S and T we get

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_{S\wedge T}.$$

This lets us rethink Doob's theorem.

Theorem 1.2.31

If S and T are arbitrary stopping times such that $S \leqslant T$ then

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S$$

for an uniformly integrable martingale.

The dominated convergence theorem requires adaptness.

Theorem 1.2.32

Hunt's dominated convergence theorem.
Let $\{X_n\}$ be dominated by an integrable random variable and suppose that exists

$$X_\infty = \lim X_n$$

So $(\mathbb{E}_n X_n)_n$ converges to $\mathbb{E}X_\infty$ almost surely and in L^1.

1.2.40 Stopped martingales

Definition 1.2.19

Define $M = (M_n)_{n\in\mathbb{N}}$ as a process and let T be a random time with values on $\overline{\mathbb{N}}$. The process

$$X_n(\omega) = M_{n\wedge T}(\omega) = \begin{cases} M_n(\omega) & n < T(\omega) \\ M_T(\omega) & n > T(\omega) \end{cases}$$

(where $n \wedge T$ is a truncated random time) is called **M stopped at T**.

As a consequence X is exactly the discrete time integral if $F = \mathbb{1}_{[0,T]}$:

$$X_n = \underbrace{M_0 F_0}_{0} + (M_1 + M_0)\mathbb{1}_{[0,T]}(1) + \ldots + (M_n - M_{n-1})\mathbb{1}_{[0,T]}(n).$$

The indicators only select the current time interval. If this is the case we can observe that $F_{[0,T]}$ is bounded, positive and predictable. Hence if M is a martingale the theorem applies with this special choice of M and we can write the result as a different theorem:

Theorem 1.2.33

Let T be a stopping time and let X be the process M stopped at T. If M is a martingale then so is X (the same holds for sub-martingales and super-martingales).

So we cannot determine a policy based on stopping times that can change the nature of our martingale. In the remote case in which you are interested in this you can read Williams - Introduction to martingales.
A further theorem about this is **Doob's stopping theorem**. For a martingale we know

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s] = 0.$$

The question is: is this true also if s and t are substituted by stopping times $S, T, S \leqslant T$?

Theorem 1.2.34

Let M be adapted to $\mathcal{F}$. Then the following are equivalent:

1. M is a submartingale;
2. for every bounded stopping time $S \leqslant T$ the random variables M_S and M_T are integrable and
$$\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geqslant 0;$$
3. for each pair of bounded stopping times the random variables M_S and M_T are integrable and
$$\mathbb{E}[M_T - M_S] \geqslant 0.$$

Remark

If M is a martingale the theorem can be read in a different way:

$$\mathbb{E}M_T = \mathbb{E}M_S = \mathbb{E}M_0.$$

Previously, $\mathbb{E}M_n = \mathbb{E}M_{n-1} = \mathbb{E}M_0$.