Probability Theory

Lecturer
Prof. Chiranjib Mukherjee

Notes
Josia Pietsch

Version
git: e291ad9
compiled: October 21, 2023 19:25

Contents

These are my notes on the lecture Probability Theory, taught by Prof. Chiranjib Mukherjee in the summer term 2023 at the University Münster.

Warning 0.1. This is not an official script. The official lecture notes can be found on Learnweb.

These notes contain errors almost surely. If you find some of them or want to improve something, please send me a message:
`notes_probability_theory@jrpie.de`.

Topics of this lecture

(1) Limit theorems: Laws of large numbers and the central limit theorem for i.i.d. sequences,

(2) Conditional expectation and conditional probabilities,

(3) Martingales,

(4) Markov chains.

This notes follow the way the material was presented in the lecture rather closely. Additions (e.g. from exercise sheets) and slight modifications have been marked with †.

Prerequisites

[Lecture 1, 2023-04-04]

First, let us recall some basic definitions:

Definition 0.2. A **probability space** is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, such that

- $\Omega \neq \emptyset$,
- $\mathcal{F}$ is a σ-algebra over Ω, i.e. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ and
  - $\emptyset, \Omega \in \mathcal{F}$,
  - $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,
  - $A_1, A_2, \ldots \in \mathcal{F} \implies \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

  The elements of $\mathcal{F}$ are called **events**.
- $\mathbb{P}$ is a **probability measure**, i.e. $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ such that
  - $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$,
  - $\mathbb{P}\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ for mutually disjoint $A_n \in \mathcal{F}$.

Definition† 0.2.1. Let X be a random variable and $k \in \mathbb{N}$. Then the k-th **moment** of X is defined as $\mathbb{E}[X^k]$.

Definition 0.3. A **random variable** $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function, i.e. for all $B \in \mathcal{B}(\mathbb{R})$ we have $X^{-1}(B) \in \mathcal{F}$. (Equivalently $X^{-1}((a, b]) \in \mathcal{F}$ for all $a < b \in \mathbb{R}$).

Definition 0.4. $F : \mathbb{R} \to \mathbb{R}_+$ is a **distribution function** iff

- F is monotone non-decreasing,
- F is right-continuous,
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Fact 0.4.2. Let $\mathbb{P}$ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then $F(x) := \mathbb{P}((-\infty, x])$ is a probability distribution function. (See lemma 2.4.2 in the lecture notes of Stochastik)

The converse to this fact is also true:

Theorem 0.5 (Kolmogorov's existence theorem / basic existence theorem of probability theory)**.** Let $\mathcal{F}(\mathbb{R})$ be the set of all distribution functions on $\mathbb{R}$ and let $\mathcal{M}(\mathbb{R})$ be the set of all probability measures on $\mathbb{R}$. Then there is a one-to-one correspondence between $\mathcal{F}(\mathbb{R})$ and $\mathcal{M}(\mathbb{R})$ given by

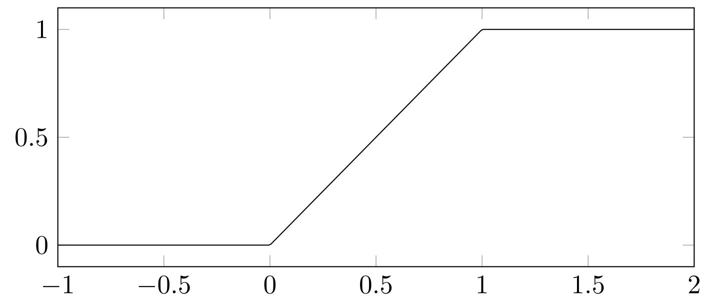
$$\begin{aligned} \mathcal{M}(\mathbb{R}) &\longrightarrow \mathcal{F}(\mathbb{R}) \\ \mathbb{P} &\longmapsto \begin{pmatrix} \mathbb{R} & \longrightarrow & \mathbb{R}_+ \\ x & \longmapsto & \mathbb{P}((-\infty, x]). \end{pmatrix} \end{aligned}$$

Proof. See theorem 2.4.3 in Stochastik. □

Example 0.6 (Some important probability distribution functions)**.**

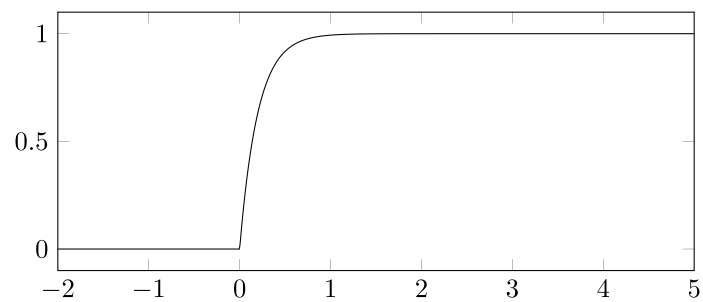
(1) **Uniform distribution** on $[0, 1]$:

$$F(x) = \begin{cases} 0 & x \in (-\infty, 0], \\ x & x \in (0, 1], \\ 1 & x \in (1, \infty). \end{cases}$$



(2) **Exponential distribution**:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geqslant 0, \\ 0 & x < 0. \end{cases}$$

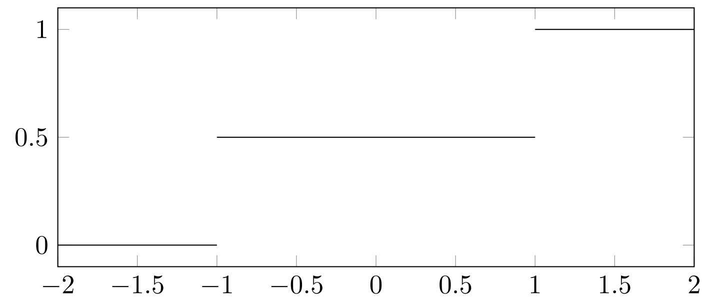


(3) **Gaussian distribution**:

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

(4) $\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = \frac{1}{2}$:

$$F(x) = \begin{cases} 0 & x \in (-\infty, -1), \\ \frac{1}{2} & x \in [-1, 1), \\ 1 & x \in [1, \infty). \end{cases}$$



This section provides a short recap of things that should be known from the lecture on stochastic.

0.1 Notions of Convergence

Definition† 0.6.3. Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $X, X_1, X_2, \ldots$ be random variables.

- We say that X_n converges to X **almost surely** ($X_n \xrightarrow{\text{a.s.}} X$) iff

$$\mathbb{P}(\{\omega | X_n(\omega) \to X(\omega)\}) = 1.$$

- We say that X_n converges to X **in probability** ($X_n \xrightarrow{\mathbb{P}} X$) iff

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$$

  for all $\varepsilon > 0$.

- We say that X_n converges to X **in the p-th mean** ($X_n \xrightarrow{L^p} X$) iff

$$\mathbb{E}[|X_n - X|^p] \xrightarrow{n\to\infty} 0.$$

- We say that X_n converges to X **in distribution**[a] ($X_n \xrightarrow{\text{d}} X$) iff for every continuous, bounded $f : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[f(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[f(X)].$$

[a] This notion of convergence was actually defined during the course of the lecture, but has been added here for completeness; see Definition 2.9.

Theorem† 0.6.4. Let X be a random variable and $X_n, n \in \mathbb{N}$ a sequence of random variables. Let $1 \leqslant p < q < \infty$. Then

$$\begin{array}{ccccc} & & & & X_n \xrightarrow{L^q} X \\ & & & & \Big\Downarrow \\ X_n \xrightarrow{\text{a.s.}} X & & & & X_n \xrightarrow{L^p} X \\ & \searrow\!\!\!\searrow & & \swarrow\!\!\!\swarrow & \\ & & X_n \xrightarrow{\mathbb{P}} X & & \\ & & \Big\Downarrow & & \\ & & X_n \xrightarrow{\text{d}} X & & \end{array}$$

and none of the other implications hold (apart from the transitive closure).

Proof of Theorem† 0.6.4.

Claim 0.6.4.1. $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{\mathbb{P}} X.$

Subproof. Let $\Omega_0 := \{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}$. Fix some $\varepsilon > 0$ and consider $A_n := \bigcup_{m \geqslant n} \{\omega \in \Omega : |X_m(\omega) - X(\omega)| > \varepsilon\}$. Then $A_n \supseteq A_{n+1} \supseteq \ldots$ Define $A := \bigcap_{n \in \mathbb{N}} A_n$. Then $\mathbb{P}[A_n] \xrightarrow{n \to \infty} \mathbb{P}[A]$. Since $X_n \xrightarrow{a.s.} X$ we have that

$$\forall \omega \in \Omega_0.\ \exists n \in \mathbb{N}.\ \forall m \geqslant n.\ |X_m(\omega) - X(\omega)| < \varepsilon.$$

We have $A \subseteq \Omega_0^c$, hence $\mathbb{P}[A_n] \to 0$. Thus

$$\mathbb{P}[\{\omega \in \Omega |\ |X_n(\omega) - X(\omega)| > \varepsilon\}] < \mathbb{P}[A_n] \to 0.$$

■

Claim 0.6.4.2. *Let* $1 \leqslant p < q < \infty$. *Then* $X_n \xrightarrow{L^q} X \implies X_n \xrightarrow{L^p} X$.

Subproof. Take r such that $\frac{1}{p} = \frac{1}{q} + \frac{1}{r}$. We have

$$\begin{aligned} \|X_n - X\|_{L^p} &= \|1 \cdot (X_n - X)\|_{L^p} \\ &\overset{\text{Hölder}}{\leqslant} \|1\|_{L^r} \|X_n - X\|_{L^q} \\ &= \|X_n - X\|_{L^q} \end{aligned}$$

Hence $\mathbb{E}[|X_n - X|^q] \xrightarrow{n \to \infty} 0 \implies \mathbb{E}[|X_n - X|^p] \xrightarrow{n \to \infty} 0.$ ■

Claim 0.6.4.3. $X_n \xrightarrow{L^1} X \implies X_n \xrightarrow{\mathbb{P}} X$.

Subproof. Suppose $\mathbb{E}[|X_n - X|] \to 0$. Then for every $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}[|X_n - X| \geqslant \varepsilon] &\overset{\text{Markov}}{\leqslant} \frac{\mathbb{E}[|X_n - X|]}{\varepsilon} \\ &\xrightarrow{n\to\infty} 0, \end{aligned}$$

hence $X_n \xrightarrow{\mathbb{P}} X$. ■

Claim 0.6.4.4. $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{d} X$.

Subproof. Let F be the distribution function of X and $(F_n)_n$ the distribution functions of $(X_n)_n$. By Theorem 2.13 it suffices to show that $F_n(t) \to F(t)$ for all continuity points t of F. Let t be a continuity point of F. Take some $\varepsilon > 0$. Then there exists $\delta > 0$ such that $|F(t) - F(t')| < \frac{\varepsilon}{2}$ for all t' with $|t - t'| \leqslant \delta$. For all n large enough, we have $\mathbb{P}[|X_n - X| > \delta] < \frac{\varepsilon}{2}$. It is

$$\begin{aligned} |F_n(t) - F(t)| &= |\mathbb{P}[X_n \leqslant t] - F(t)| \\ &\leqslant \max(|\frac{\varepsilon}{2} + \mathbb{P}[X \leqslant t + \delta] - F(t)|, |\mathbb{P}[X \leqslant t - \delta] - F(t)|) \\ &\leqslant \max(|\frac{\varepsilon}{2} + F(t+\delta) - F(t)|, |F(t-\delta) - F(t)|) \\ &\leqslant \varepsilon, \end{aligned}$$

hence $F_n(t) \to F(t)$. ■

Claim 0.6.4.5. $X_n \xrightarrow{\mathbb{P}} X \not\Longrightarrow X_n \xrightarrow{L^1} X$.[1]

Subproof. Take $([0,1], \mathcal{B}([0,1]), \lambda)$ and define $X_n := n\mathbb{1}_{[0,\frac{1}{n}]}$. We have $\mathbb{P}[|X_n| > \varepsilon] = \frac{1}{n}$ for n large enough.

However $\mathbb{E}[|X_n|] = 1$. ■

Claim 0.6.4.6. $X_n \xrightarrow{a.s.} X \not\Longrightarrow X_n \xrightarrow{L^1} X$.

Subproof. We can use the same counterexample as in Claim 0.6.4.5

$\mathbb{P}[\lim_{n\to\infty} X_n = 0] \geqslant \mathbb{P}[X_n = 0] = 1 - \frac{1}{n} \to 0$. We have already seen, that X_n does not converge in L_1. ■

Claim 0.6.4.7. $X_n \xrightarrow{L^1} X \not\Longrightarrow X_n \xrightarrow{a.s.} X$.

[1] Note that the implication holds under certain assumptions, see Theorem 4.24.

Subproof. Take $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]), \mathbb{P} = \lambda$. Define $A_n := [j2^{-k}, (j+1)2^{-k}]$ where $n = 2^k + j$. We have

$$\mathbb{E}[|X_n|] = \int_\Omega |X_n| \, \mathrm{d}\mathbb{P} = \frac{1}{2^k} \to 0.$$

However X_n does not converge a.s. as for all $\omega \in [0,1]$ the sequence $X_n(\omega)$ takes the values 0 and 1 infinitely often. ■

Claim 0.6.4.8. $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{\mathbb{P}} X$.

Subproof. Note that $X_n \xrightarrow{\mathrm{d}} X$ only makes a statement about the distributions of X and $X_1, X_2, \ldots$ For example, take some $p \in (0,1)$ and let $X, X_1, X_2, \ldots$ be i.i.d. with $X \sim \mathrm{Bin}(1,p)$. Trivially $X_n \xrightarrow{\mathrm{d}} X$. However

$$\mathbb{P}[|X_n - X| = 1] = \mathbb{P}[X_n = 0]\mathbb{P}[X = 1] + \mathbb{P}[X_n = 1]\mathbb{P}[X = 0] = 2p(1-p).$$

■

Claim 0.6.4.9. *Let* $1 \leqslant p < q < \infty$*. Then* $X_n \xrightarrow{L^p} X \not\Rightarrow X_n \xrightarrow{L^q} X$.

Subproof. Consider $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P} = \lambda \upharpoonright [0,1]$ and $X_n(\omega) = \frac{1}{n\sqrt[q]{\omega}}$. Then $\|X_0(\omega)\|_{L^p} < \infty$, since $p < q$. Thus $X_n \xrightarrow{L_p} 0$. However $\|X_n(\omega)\|_{L^q} = \infty$ for all n. ■

□

Theorem 0.7 (Bounded convergence theorem)**.** Suppose that $X_n \xrightarrow{\mathbb{P}} X$ and there exists some K such that $|X_n| \leqslant K$ for all n. Then $X_n \xrightarrow{L^1} X$.

Proof. Note that $|X| \leqslant K$ a.s. since

$$\mathbb{P}[|X| \geqslant K + \varepsilon] \leqslant \mathbb{P}[|X_n - X| > \varepsilon] \xrightarrow{n\to\infty} 0.$$

Hence

$$\begin{aligned}\int |X_n - X| \, \mathrm{d}\mathbb{P} &\leqslant \int_{|X_n - X| \geqslant \varepsilon} |X_n - X| \, \mathrm{d}\mathbb{P} + \varepsilon \\ &\leqslant 2K\mathbb{P}[|X_n - X| \geqslant \varepsilon] + \varepsilon.\end{aligned}$$

□

0.2 Some Facts from Measure Theory

Fact† **0.7.5** (Finite measures are regular, Exercise 3.1)**.** Let μ be a finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then for all $\varepsilon > 0$, there exists a compact set $K \in \mathcal{B}(\mathbb{R})$ such that $\mu(K) > \mu(\mathbb{R}) - \varepsilon$.

Proof. We have $[-k, k] \uparrow \mathbb{R}$, hence $\mu([-k, k]) \uparrow \mu(\mathbb{R})$. □

Theorem† **0.7.6** (Change of variables formula)**.** Let X be a random variable with a continuous density f, and let $g : \mathbb{R} \to \mathbb{R}$ be continuous such that $g(X)$ is integrable. Then

$$\mathbb{E}[g(X)] = \int g \circ X \, \mathrm{d}\mathbb{P} = \int_{-\infty}^{\infty} g(y) f(y) \lambda(\mathrm{d}y) = \int_{-\infty}^{\infty} g(y) f(y) \, \mathrm{d}y.$$

Theorem† **0.7.7** (Riemann-Lebesgue)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be integrable. Then

$$\lim_{n \to \infty} \int_{\mathbb{R}} f(x) \cos(nx) \lambda(\mathrm{d}x) = 0.$$

Theorem† **0.7.8** (Fubini-Tonelli)**.** Let $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i), i \in \{0, 1\}$ be probability spaces and $\Omega := \Omega_0 \otimes \Omega_1$, $\mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2$, $\mathbb{P} := \mathbb{P}_0 \otimes \mathbb{P}_1$. Let $f \geqslant 0$ be $(\Omega, \mathcal{F})$-measurable, then

$$\Omega_0 \ni x \mapsto \int_{\Omega_2} f(x, y) \mathbb{P}_2(\mathrm{d}y)$$

and

$$\Omega_1 \ni y \mapsto \int_{\Omega_1} f(x, y) \mathbb{P}_1(\mathrm{d}x)$$

are measurable, and

$$\int f \, \mathrm{d}\mathbb{P} = \int_{\Omega_1} \int_{\Omega_2} f(x, y) \mathbb{P}_2(\mathrm{d}y) \mathbb{P}_1(\mathrm{d}x)(\mathrm{d}x) = \int_{\Omega_2} \int_{\Omega_1} f(x, y) \mathbb{P}_1(\mathrm{d}x) \mathbb{P}_2(\mathrm{d}y).$$

0.3 Inequalities

This is taken from section 6.1 of the notes on Stochastik.

Theorem 0.8 (Markov's inequality)**.** Let X be a random variable and $a > 0$. Then

$$\mathbb{P}[|X| \geqslant a] \leqslant \frac{\mathbb{E}[|X|]}{a}.$$

Proof. We have

$$\mathbb{E}[|X|] \geqslant \int_{|X| \geqslant a} |X| \, \mathrm{d}\mathbb{P} = a \int_{|X| \geqslant a} \mathrm{d}\mathbb{P} = a\mathbb{P}[|X| \geqslant a].$$

□

Theorem 0.9 (Chebyshev's inequality)**.** Let X be a random variable and $a > 0$. Then

$$\mathbb{P}[|X - \mathbb{E}(X)| \geqslant a] \leqslant \frac{\operatorname{Var}(X)}{a^2}.$$

Proof. We have

$$\mathbb{P}[|X - \mathbb{E}(X)| \geqslant a] = \mathbb{P}[|X - \mathbb{E}(X)|^2 \geqslant a^2] \overset{\text{Markov}}{\leqslant} \frac{\mathbb{E}[|X - \mathbb{E}(X)|^2]}{a^2}.$$

□

How do we prove that something happens almost surely? The first thing that should come to mind is:

Lemma 0.10 (Borel-Cantelli)**.** If we have a sequence of events $(A_n)_{n \geqslant 1}$ such that $\sum_{n \geqslant 1} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}[A_n \text{for infinitely many } n] = 0$ (more precisely: $\mathbb{P}[\limsup_{n \to \infty} A_n] = 0$).

For independent events A_n the converse holds as well.

[Lecture 2, 2023-04-11]

1 Independence and Product Measures

In order to define the notion of independence, we first need to construct product measures.

The finite case of a product is straightforward:

Theorem 1.1. Product measure (finite) Let $(\Omega_1, \mathcal{F}, \mathbb{P})$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces. Let $\Omega := \Omega_1 \times \Omega_2$ and $R := \{A_1 \times A_2 | A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$.

Let $\mathcal{F}$ be $\sigma(R)$ (the sigma algebra generated by R). Then there exists a unique probability measure $\mathbb{P}$ on Ω such that for every rectangle $R = A_1 \times A_2 \in \mathcal{R}$, $\mathbb{P}(A_1 \times A_2) = \mathbb{P}(A_1) \times \mathbb{P}(A_2)$.

Proof. See Theorem 5.1.1 in the lecture notes on Stochastik. □

We now want to construct a product measure for infinite products.

Definition 1.2 (Independence)**.** A collection $X_1, X_2, \dots, X_n$ of random variables are called **mutually independent** if

$$\forall a_1, \dots, a_n \in \mathbb{R} : \mathbb{P}[X_1 \leqslant a_1, \dots, x_n \leqslant a_n] = \prod_{i=1}^{n} \mathbb{P}[X_i \leqslant a_i]$$

This is equivalent to

$$\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}) : \mathbb{P}[X_1 \in B_1, \dots, X_n \in B_n] = \prod_{i=1}^{n} \mathbb{P}[X_i \in B_i]$$

Example 1.3. Suppose we throw a dice twice. Let $A \coloneqq \{\text{first throw even}\}$, $B \coloneqq \{\text{second throw even}\}$ and $C \coloneqq \{\text{sum even}\}$.

It is easy the see, that the random variables are pairwise independent, but not mutually independent.

Definition 1.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a random variable. Then $\mathbb{Q}(\cdot) \coloneqq \mathbb{P}[X \in \cdot]$ is called the **distribution** of X under $\mathbb{P}$.

Let $X_1, \dots, X_n$ be random variables and $\mathbb{Q}^{\otimes}(\cdot) \coloneqq \mathbb{P}[(X_1, \dots, X_n) \in \cdot]$ their **joint distribution**. Then $\mathbb{Q}^{\otimes}$ is a probability measure on $\mathbb{R}^n$.

The definition of mutual independence can be rephrased as follows:

Fact 1.4.9. $X_1, \dots, X_n$ are mutually independent iff $\mathbb{Q}^{\otimes} = \mathbb{Q}_1 \otimes \dots \otimes \mathbb{Q}_n$, where $\mathbb{Q}_i$ is the distribution of X_i. In this setting, $\mathbb{Q}_i$ is called the **marginal distribution** of X_i.

By constructing an infinite product, we can thus extend the notion of independence to an infinite number of random variables.

Goal. *Can we construct infinitely many independent random variables?*

Definition 1.5 (Consistent family of random variables)**.** Let $\mathbb{P}_n, n \in \mathbb{N}$ be a family of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The family is called **consistent** if $\mathbb{P}_{n+1}[B_1 \times B_2 \times \dots \times B_n \times \mathbb{R}] = \mathbb{P}_n[B_1 \times \dots \times B_n]$ for all $n \in \mathbb{N}, B_i \in B(\mathbb{R})$.

Theorem 1.6 (Kolmogorov extension / consistency theorem)**.** [a]

Let $\mathbb{P}_n, n \in \mathbb{N}$ be probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ which are **consis-**

tent, then there exists a unique probability measure $\mathbb{P}^{\otimes}$ on $(\mathbb{R}^\infty, B(R^\infty))$ (where $B(R^\infty)$ has to be defined), such that

$$\forall n \in \mathbb{N}, B_1, \ldots, B_n \in B(\mathbb{R}) : \mathbb{P}^{\otimes}[\mathcal{X} : X_i \in B_i \forall 1 \leqslant i \leqslant n] = \mathbb{P}_n[B_1 \times \ldots \times B_n]$$

[a] Informally: "Probability measures are determined by finite-dimensional marginals (as long as these marginals are nice)"

Remark 1.6.10. Kolmogorov's theorem can be strengthened to the case of arbitrary index sets. However this requires a different notion of consistency.

Example 1.7 (A consistent family)**.** Let $F_1, \ldots, F_n$ be probability distribution functions and let $\mathbb{P}_n$ be the probability measure on $\mathbb{R}^n$ defined by

$$\mathbb{P}_n[(a_1, b_1] \times \ldots (a_n, b_n]] := (F_1(b_1) - F_1(a_1)) \cdot \ldots \cdot (F_n(b_n) - F_n(a_n)).$$

It is easy to see that each $\mathbb{P}_n$ is a probability measure.

Define $X_i(\omega) = \omega_i$ where $\omega = (\omega_1, .., \omega_n)$. Then $X_1, \ldots, X_n$ are mutually independent with F_i being the distribution function of X_i. In the case of $F_1 = \ldots = F_n$, then $X_1, \ldots, X_n$ are i.i.d.

[Lecture 3, 2023-04-13]

Notation 1.7.11. Let $\mathcal{B}_n$ denote $\mathcal{B}(\mathbb{R}^n)$.

Goal. *Suppose we have a probability measure μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for each $n \in \mathbb{N}$. We want to show that there exists a unique probability measure $\mathbb{P}^{\otimes}$ on $(\mathbb{R}^\infty, \mathcal{B}_\infty)$ (where the σ-algebra $\mathcal{B}_\infty$ still needs to be defined), such that*

$$\mathbb{P}^{\otimes}\left(\prod_{n \in \mathbb{N}} B_n\right) = \prod_{n \in \mathbb{N}} \mu_n(B_n)$$

for all $\{B_n\}_{n \in \mathbb{N}}$, $B_n \in \mathcal{B}_1$.

Remark 1.7.12. $\prod_{n \in \mathbb{N}} \mu_n(B_n)$ converges, since $0 \leqslant \mu_n(B_n) \leqslant 1$ for all n.

First we need to define $\mathcal{B}_\infty$. This σ-algebra must contain all "boxes" $\prod_{n \in \mathbb{N}} B_n$ for $B_i \in \mathcal{B}_1$. We simply take the smallest σ-algebra with this property:

Definition 1.8.

$$\mathcal{B}_\infty := \sigma\left(\left\{\prod_{n \in \mathbb{N}} B_n : \forall n.\ B_n \in \mathcal{B}(\mathbb{R})\right\}\right).$$

Question 1.8.13. What is there in $\mathcal{B}_\infty$? Can we identify sets in $\mathcal{B}_\infty$ for which we can define the desired product measure easily?

Let $\mathcal{F}_n := \{C \times \mathbb{R}^\infty | C \in \mathcal{B}_n\}$. It is easy to see that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ and using that $\mathcal{B}_n$ is a σ-algebra, we can show that $\mathcal{F}_n$ is also a σ-algebra. Now, for any $C \subseteq \mathbb{R}^n$ let $C^* := C \times \mathbb{R}^\infty$. Note that $C \in \mathcal{B}_n \implies C^* \in \mathcal{F}_n$. Thus $\mathcal{F}_n = \{C^* : C \in \mathcal{B}_n\}$. Define $\lambda_n : \mathcal{F}_n :\to [0,1]$ by $\lambda_n(C^*) := (\mu_1 \otimes \ldots \otimes \mu_n)(C)$. It is easy to see that $\lambda_{n+1}|_{\mathcal{F}_n} = \lambda_n$, i.e. the λ_n form a consistent family.

Recall the following theorem from measure theory:

Theorem 1.9 (Caratheodory's extension theorem)**.** Suppose $\mathcal{A}$ is an algebra (i.e. closed under finite union) und $\Omega \neq \emptyset$. Suppose $\mathbb{P}$ is countably additive on $\mathcal{A}$ (i.e. if $(A_n)_n$ are pairwise disjoint and $\bigcup_{n\in\mathbb{N}} A_n \subseteq \mathcal{A}$ then $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$). Then $\mathbb{P}$ extends uniquely to a probability measure on $(\Omega, \mathcal{F})$, where $\mathcal{F} = \sigma(\mathcal{A})$.

Proof. See theorem 2.3.3 in Stochastik. □

Define $\mathcal{F} = \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$. Then $\mathcal{F}$ is an algebra. We'll show that if we define $\lambda : \mathcal{F} \to [0,1]$ with $\lambda(A) = \lambda_n(A)$ for any n where this is well defined, then λ is countably additive on $\mathcal{F}$. Using Theorem 1.9, λ will extend uniquely to a probability measure on $\sigma(\mathcal{F})$.

We want to prove:

Claim 1. $\sigma(\mathcal{F}) = \mathcal{B}_\infty$.

Claim 2. *λ is countably additive on $\mathcal{F}$.*

Proof of Claim 1. Consider an infinite dimensional box $\prod_{n\in\mathbb{N}} B_n$. We have

$$\left(\prod_{n=1}^{N} B_n\right)^* \in \mathcal{F}_n \subseteq \mathcal{F}$$

thus

$$\prod_{n\in\mathbb{N}} B_n = \bigcap_{N\in\mathbb{N}} \left(\prod_{n=1}^{N} B_n\right)^* \in \sigma(\mathcal{F}).$$

Since $\sigma(\mathcal{F})$ is a σ-algebra, $\mathcal{B}_\infty \subseteq \sigma(\mathcal{F})$. This proves "$\supseteq$". For the other direction we'll show $\mathcal{F}_n \subseteq \mathcal{B}_\infty$ for all n. Let $\mathcal{C} := \{Q \in \mathcal{B}_n | Q^* \in \mathcal{B}_\infty\}$. For $B_1, \ldots, B_n \in \mathcal{B}$, $B_1 \times \ldots \times B_n \in \mathcal{B}_n$ and $(B_1 \times \ldots \times B_n)^* \in \mathcal{B}_\infty$. We have $B_1 \times \ldots \times B_n \in \mathcal{C}$. And $\mathcal{C}$ is a σ-algebra, because:

- $\mathcal{B}_n$ is a σ-algebra
- $\mathcal{B}_\infty$ is a σ-algebra,
- $\emptyset^* = \emptyset$, $(\mathbb{R}^n \setminus Q)^* = \mathbb{R}^\infty \setminus Q^*$, $\bigcup_{i\in I} Q_i^* = \left(\bigcup_{i\in I} Q_i\right)^*$.

Since $\mathcal{C} \subseteq \mathcal{B}_n$ is a σ-algebra and contains all rectangles, it holds that $\mathcal{C} = \mathcal{B}_n$. Hence $\mathcal{F}_n \subseteq \mathcal{B}_\infty$ for all n, thus $\mathcal{F} \subseteq \mathcal{B}_\infty$. Since $\mathcal{B}_\infty$ is a σ-algebra, $\sigma(\mathcal{F}) \subseteq \mathcal{B}_\infty$. □

For the proof of Claim 2, we are going to use the following:

Fact 1.9.14. Suppose $\mathcal{A}$ is an algebra on $\Omega \neq \varnothing$, and suppose $\mathbb{P} : \mathcal{A} \to [0,1]$ is a finitely additive probability measure. Suppose whenever $\{B_n\}_n$ is a sequence of sets from $\mathcal{A}$ decreasing to $\varnothing$ it is the case that $\mathbb{P}(B_n) \to 0$. Then $\mathbb{P}$ must be countably additive.

Proof. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of disjoint, measurable sets with $A := \bigcup_n A_n \in \mathcal{A}$. Let $A_n' := A \setminus \bigcup_{i=1}^n A_i$. Then we have $\mathbb{P}[A] = \mathbb{P}[A_n'] + \sum_{i=1}^n \mathbb{P}[A_i]$ for all n. Thus

$$\mathbb{P}[A] - \lim_{n\to\infty} \mathbb{P}[A_n'] = \lim_{n\to\infty} \sum_{i=1}^n \mathbb{P}[A_i].$$

Since $\bigcap_{n\in\mathbb{N}} A_n' = \varnothing$, we have $\lim\limits_{n\to\infty} \mathbb{P}[A_n'] = 0$, hence

$$\mathbb{P}\left[\bigcup_{i\in\mathbb{N}} A_i\right] = \mathbb{P}[A] = \sum_{i\in\mathbb{N}} \mathbb{P}[A_i].$$

□

Proof of Claim 2. Let us prove that λ is finitely additive. We have $\lambda(\mathbb{R}^\infty) = \lambda_1(\mathbb{R}^\infty) = 1$ and $\lambda(\varnothing) = \lambda_1(\varnothing) = 0$. Suppose that $A_1, A_2 \in \mathcal{F}$ are disjoint. Then pick some n such that $A_1, A_2 \in \mathcal{F}_n$. Take $C_1, C_2 \in \mathcal{B}_n$ such that $C_1^* = A_1$ and $C_2^* = A_2$. Then C_1 and C_2 are disjoint and $A_1 \cup A_2 = (C_1 \cup C_2)^*$. Hence

$$\lambda(A_1 \cup A_2) = \lambda_n(A_1 \cup A_2) = (\mu_1 \otimes \ldots \otimes \mu_n)(C_1 \cup C_2) = \lambda_n(C_1) + \lambda_n(C_2)$$

by the definition of the finite product measure.

[Lecture 4,]

To finish the proof of Claim 2, we need the following:

Fact 1.9.15. Suppose $\{x_k^{(n)}\}_{n\in\mathbb{N}}$ is a bounded sequence of real numbers for each $k \in \mathbb{N}$. Then there exists a strictly increasing sequence of natural number $\{n_i\}_{i\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$ the series $\{x_k^{(n_i)}\}_{i\in\mathbb{N}}$ converges.

Proof. We'll use a diagonalization argument. For $S \subseteq \mathbb{N}$ infinite, we say that a sequence of real number, $(x_n)_{n\in\mathbb{N}}$, **converges along S**, if

$$\lim_{\substack{n\to\infty \\ n\in S}} x_n$$

exists.

Let S_1 be such that $\{x_1^{(n)}\}_{n\in\mathbb{N}}$ converges along S_1. Such an S_1 exists by Bolzano-Weierstraß. We proceed recursively. Suppose we have already chosen $S_1, \ldots, S_{k-1}$. Consider $\{x_k^{(n)}\}_{n\in S_{k-1}}$. By Bolzano-Weierstraß, there exists $S_k \subseteq S_{k-1}$ such that $\{x_k^{(n)}\}_{n\in S_{k-1}}$ converges along S_k. For an infinite subset $T \subseteq \mathbb{N}$ and $\nu \in \mathbb{N}$ let $\#\nu(T)$ denote the ν-th smallest element of T. Let

$$S := \{\#\nu(S_k) : k \in \mathbb{N}\}.$$

Since $S_{k+1} \subseteq S_k$, we have $\#(k+1)(S_{k+1}) > \#k(S_{k+1}) \geqslant \#k(S_k)$. Hence S is infinite. Each $\{x_k^{(n)}\}_{n\in\mathbb{N}}$ converges along S, since all but finitely many elements of S belong to S_k. □

Lemma 1.10. Let $\{K_n\}_{n\in\mathbb{N}}$ be a sequence of compact sets $K_n \subseteq \mathbb{R}^{l_n}$ for some l_n. Suppose for all n

$$\bigcap_{i=1}^{n} K_i^* \neq \varnothing.$$

Then

$$\bigcap_{i\in\mathbb{N}} K_i^* \neq \varnothing.$$

Proof of Lemma 1.10. We know from analysis that if $\{K_n\}_{n\in\mathbb{N}}$ is a sequence of compact sets such that the intersection of finitely many of them is non-empty, then

$$\bigcap_{n\in\mathbb{N}} K_n \neq \varnothing.$$

Here, different K_n may have different dimensions l_n, but we can view them as subsets of $\mathbb{R}^\infty$ by applying $*$. For each n, choose $x^{(n)} \in \bigcap_{i=1}^n K_i^*$. We can assume $x_k^{(n)} = 0$ for $k > \max\{l_1, \ldots, l_n\}$. For all $k \in \mathbb{N}$ we will show that $\{x_k^{(n)}\}$ is bounded.

- Case 1: Suppose every $l_n \leqslant k$. Then $\{x_k^{(n)}\}_n$ only contains zeros.
- Case 2: Suppose some $l_{n_0} \geqslant k$. Let Z be the projection of $K_{n_0} \subseteq \mathbb{R}^{l_{n_0}}$ onto its k-th component. Z is a compact subset of $\mathbb{R}$. Hence it is bounded. For all $n \geqslant n_0$, we have $x^{(n)} \in K_{n_0}^*$ and $x_k^{(n)} \in Z$, so $\{x_k^{(n)}\}_n$ is bounded.

By Fact 1.9.15, there is an infinite set $S \subseteq \mathbb{N}$, such that $\{x_k^{(n)}\}_{n\in S}$ converges for every k. Let $x_k := \lim_{\substack{n\to\infty \\ n\in S}} x_k^{(n)}$. Now let $x = (x_1, x_2, \ldots) \in \mathbb{R}^\infty$.

Claim 1.10.1. $x \in \bigcap_{i\in\mathbb{N}} K_i^*$.

Subproof. Consider $x^{(n)}$ for $n > i$ and $n \in S$. Then $(x_1^{(n)}, \ldots, x_{l_i}^{(n)}) \in K_i$ and

$$\lim_{\substack{n\to\infty \\ n\in S}} (x_1^{(n)}, \ldots, x_{l_i}^{(n)}) = (x_1, \ldots, x_{l_i}).$$

Since K_i is compact, it follows that $x \in K_i^*$. ■

□

Continuation of proof of Claim 2. In order to apply Fact 1.9.14, we need the following:

Claim 2.3. *For any sequence $B_n \in \mathcal{F}$ with $B_n \xrightarrow{n\to\infty} \varnothing$ we have $\lambda(B_n) \xrightarrow{n\to\infty} 0$.*

Subproof. Suppose that $B_1^* \supseteq B_2^* \supseteq \ldots$ is a decreasing sequence such that $\lim_{n\to\infty} \lambda(B_n^*) = \varepsilon > 0$. For each n, let l_n be such that $B_n \in \mathcal{B}_{l_n}$. By regularity of Borel probability measures, given $\varepsilon > 0$, there exists a compact set $L_n \subseteq B_n$, such that

$$(\mu_1 \otimes \ldots \otimes \mu_n)(B_n \backslash L_n) < \frac{\varepsilon}{2^{n+1}}$$

We have

$$B_n^* \backslash \bigcap_{k=1}^n L_k^* \subseteq \bigcup_{k=1}^n (B_k^* \backslash L_k^*)\,.$$

Hence

$$\begin{aligned}
\lambda\left(B_n^* \backslash \bigcap_{k=1}^n L_k^*\right) &\leqslant \lambda\left(\bigcup_{k=1}^n B_k^* \backslash L_k^*\right) \\
&\leqslant \sum_{k=1}^n \lambda(B_k^* \backslash L_k^*) \\
&\leqslant \sum_{k=1}^n \frac{\varepsilon}{2^{k+1}} \\
&\leqslant \frac{\varepsilon}{2}.
\end{aligned}$$

By our assumption, $\lambda(B_n^*) \downarrow \varepsilon > 0$. Hence $\lambda(B_n^*) \geqslant \varepsilon$ for all n. Thus

$$\lambda\left(\bigcap_{k=1}^n L_k^*\right) \geqslant \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

In particular, for all n

$$\bigcap_{k=1}^n L_k^* \neq \varnothing.$$

By Lemma 1.10, it follows that

$$\bigcap_{k\in\mathbb{N}} L_k^* \neq \varnothing.$$

Since

$$\bigcap_{k\in\mathbb{N}} B_k^* \supseteq \bigcap_{k\in\mathbb{N}} L_k^*,$$

we have $\bigcap_{k\in\mathbb{N}} B_k^* \neq \varnothing$. ■

□

The measure λ is as desired: For all $n \in \mathbb{N}$ take some $B_n \in \mathcal{B}_1$ and let $C_n := \prod_{i=1}^n B_i$. Then $C_n^* \downarrow \prod_{i=1}^\infty B_i$, hence

$$\begin{aligned}
\lambda\left(\prod_{i=1}^\infty B_i\right) &\overset{\text{continuity}}{=} \lim_{N\to\infty} \lambda(C_N^*) \\
&= \lim_{N\to\infty} \lambda_N(C_N^*) \\
&= \lim_{N\to\infty} \prod_{n=1}^N \mu_n(B_n) \\
&= \prod_{n\in\mathbb{N}} \mu_n(B_n).
\end{aligned}$$

For the definition of λ as well as the proof of Claim 2 we have only used that $(\lambda_n)_{n\in\mathbb{N}}$ is a consistent family. Hence we have in fact shown Theorem 1.6.

[Lecture 5, 2023-04-21]

1.1 The Laws of Large Numbers

We want to show laws of large numbers: The LHS is random and represents "sane" averaging. The RHS is constant, which we can explicitly compute from the distribution of the RHS.

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ once and for all.

Theorem 1.11. Let $X_1, X_2, \ldots$ be i.i.d. random variables on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $m = \mathbb{E}[X_i] < \infty$ and $\sigma^2 = \mathrm{Var}(X_i) = \mathbb{E}[(X_i - \mathbb{E}(X_i))^2] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 < \infty$.

Then

(a) $\frac{X_1+\ldots+X_n}{n} \xrightarrow{n\to\infty} m$ in probability (**weak law of large numbers**, WLLN),

(b) $\frac{X_1+\ldots+X_n}{n} \xrightarrow{n\to\infty} m$ almost surely (**strong law of large numbers**, SLLN).

Proof of Theorem 1.11. (a) Given $\varepsilon > 0$, we need to show that

$$\mathbb{P}\left[\left|\frac{X_1+\ldots+X_n}{n} - m\right| > \varepsilon\right] \xrightarrow{n\to 0} 0.$$

Let $S_n := X_1 + \ldots + X_n$. Then $\mathbb{E}[S_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = nm$. We have

$$\begin{aligned}\mathbb{P}\left[\left|\frac{X_1 + \ldots + X_n}{n} - m\right| > \varepsilon\right] &= \mathbb{P}\left[\left|\frac{S_n}{n} - m\right| > \varepsilon\right] \\ &\overset{\text{Chebyshev}}{\leqslant} \frac{\operatorname{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{1}{n}\frac{\operatorname{Var}(X_1)}{\varepsilon^2} \xrightarrow{n\to\infty} 0\end{aligned}$$

since

$$\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\operatorname{Var}(S_n) = \frac{1}{n^2} n \operatorname{Var}(X_i).$$

□

For the proof of (b) we need the following general result:

Theorem 1.12. Let $X_1, X_2, \ldots$ be independent (but not necessarily identically distributed) random variables with $\mathbb{E}[X_i] = 0$ for all i and

$$\sum_{i=1}^{n} \operatorname{Var}(X_i) < \infty.$$

Then $\sum_{n\geqslant 1} X_n$ converges almost surely.

We'll prove this later Move proof

Question 1.12.16. Does the converse hold? I.e. does $\sum_{n\geqslant 1} X_n < \infty$ a.s. then $\sum_{n\geqslant 1} \operatorname{Var}(X_n) < \infty$.

This does not hold. Consider the following:

Example 1.13. Let $X_1, X_2, \ldots$ be independent random variables, where X_n has distribution $\frac{1}{n^2}\delta_n + \frac{1}{n^2}\delta_{-n} + (1 - \frac{2}{n^2})\delta_0$. We have $\mathbb{P}[X_n \neq 0] = \frac{2}{n^2}$. Since this is summable, Borel-Cantelli (0.10) yields

$$\mathbb{P}[X_n \neq 0 \text{ for infinitely many } n] = 0.$$

In particular, X_n is summable almost surely. However $\operatorname{Var}(X_n) = 2$ is not summable.

[Lecture 6,]

Continuation of proof of Theorem 1.11. We want to deduce the SLLN (Theorem 1.11) from Theorem 1.12. W.l.o.g. let us assume that $\mathbb{E}[X_i] = 0$ (otherwise define $X_i' := X_i - \mathbb{E}[X_i]$). We will show that $\frac{S_n}{n} \xrightarrow{a.s.} 0$. Define $Y_i := \frac{X_i}{i}$. Then the Y_i are independent and we have $\mathbb{E}[Y_i] = 0$ and $\operatorname{Var}(Y_i) = \frac{\sigma^2}{i^2}$. Thus $\sum_{i=1}^{\infty} \operatorname{Var}(Y_i) < \infty$. From Theorem 1.12 we obtain that $\sum_{i=1}^{\infty} Y_i$ converges a.s.

Claim 1.11.3. *Let (a_n) be a sequence in $\mathbb{R}$ such that $\sum_{n=1}^{\infty} \frac{a_n}{n}$ converges, then* $\frac{a_1+\ldots+a_n}{n} \to 0$.

Subproof. Let $S_m := \sum_{n=1}^{\infty} \frac{a_n}{n}$. By assumption, there exists $S \in \mathbb{R}$ such that $S_m \xrightarrow{m\to\infty} S$. Note that $j \cdot (S_j - S_{j-1}) = a_j$. Define $S_0 := 0$. Then

$$\begin{aligned} a_1 + \ldots + a_n &= (S_1 - S_0) + 2(S_2 - S_1) + \ldots + n(S_n - S_{n-1}) \\ &= nS_n - (S_1 + S_2 + \ldots + S_{n-1}). \end{aligned}$$

Thus

$$\begin{aligned} \frac{a_1 + \ldots + a_n}{n} &= S_n - \frac{S_1 + \ldots + S_{n-1}}{n} \\ &= \underbrace{S_n}_{\to S} - \underbrace{\left(\frac{n-1}{n}\right)}_{\to 1} \cdot \underbrace{\frac{S_1 + \ldots + S_{n-1}}{n-1}}_{\to S} \\ &\to 0, \end{aligned}$$

where we have used

Fact 1.13.17.

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} S_i$$

.

■

The SLLN follows from the claim. □

In order to prove Theorem 1.12, we need the following:

Theorem 1.14 (Kolmogorov's inequality)**.** If $X_1, \ldots, X_n$ are independent with $\mathbb{E}[X_i] = 0$ and $\operatorname{Var}(X_i) = \sigma_i^2$, then

$$\mathbb{P}\left[\max_{1\leqslant i\leqslant n} \left|\sum_{j=1}^{i} X_j\right| > \varepsilon\right] \leqslant \frac{1}{\varepsilon^2} \sum_{i=1}^{m} \sigma_i^2.$$

Proof. Let

$$\begin{aligned} A_1 &:= \{\omega : |X_1(\omega)| > \varepsilon\}, \\ A_2 &:= \{\omega : |X_1(\omega)| \leqslant \varepsilon, |X_1(\omega) + X_2(\omega)| > \varepsilon\}, \\ &\ldots \\ A_i &:= \{\omega : |X_1(\omega)| \leqslant \varepsilon, |X_1(\omega) + X_2(\omega)| \leqslant \varepsilon, \ldots, |X_1(\omega) + \ldots + X_{i-1}(\omega)| \leqslant \varepsilon, \\ &\qquad |X_1(\omega) + \ldots + X_i(\omega)| > \varepsilon\}. \end{aligned}$$

It is clear, that the A_i are disjoint. We are interested in $\bigcup_{1\leqslant i\leqslant n} A_i$.

We have

$$
\begin{aligned}
&\int_{A_i} (\underbrace{X_1 + \ldots + X_i}_{C} + \underbrace{X_{i+1} + \ldots + X_n}_{D})^2 \,\mathrm{d}\mathbb{P} \\
&= \int_{A_i} C^2 \,\mathrm{d}\mathbb{P} + \underbrace{\int_{A_i} D^2 \,\mathrm{d}\mathbb{P}}_{\geqslant 0} + 2\int_{A_i} CD \,\mathrm{d}\mathbb{P} \\
&\geqslant \int_{A_i} \underbrace{C^2}_{\geqslant \varepsilon^2} \,\mathrm{d}\mathbb{P} + 2\int \underbrace{\mathbb{1}_{A_i}(X_1 + \ldots + X_i)}_{E} \underbrace{(X_{i+1} + \ldots + X_n)}_{D} \,\mathrm{d}\mathbb{P} \\
&\geqslant \int_{A_i} \varepsilon^2 \,\mathrm{d}\mathbb{P},
\end{aligned}
$$

since by the independence of E and D, and $\mathbb{E}(X_{i+1}) = \ldots = \mathbb{E}(X_n) = 0$ we have $\int DE \,\mathrm{d}\mathbb{P} = 0$.

Hence

$$
\mathbb{P}(A_i) \leqslant \frac{1}{\varepsilon^2} \int_{A_i} (X_1 + \ldots + X_n)^2 \,\mathrm{d}\mathbb{P}.
$$

Since the A_i are disjoint, we obtain

$$
\begin{aligned}
\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) &\leqslant \frac{1}{\varepsilon^2} \int_{\bigcup_{i \in \mathbb{N}} A_i} (X_1 + \ldots + X_n)^2 \,\mathrm{d}\mathbb{P} \\
&\leqslant \frac{1}{\varepsilon^2} \int_{\Omega} (X_1 + \ldots + X_n)^2 \,\mathrm{d}\mathbb{P} \\
&\overset{\text{independence}}{=} \frac{1}{\varepsilon^2} (\mathbb{E}[X_1^2] + \ldots + \mathbb{E}[X_n^2]) \\
&\overset{\mathbb{E}[X_i]=0}{=} \frac{1}{\varepsilon^2} \left(\operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n)\right).
\end{aligned}
$$

□

Proof of Theorem 1.12. Let $S_n := x_1 + \ldots + x_n$. We'll show that $\{S_n(\omega)\}_{n \in \mathbb{N}}$ is a Cauchy sequence for almost every ω.

Let

$$
a_m(\omega) := \sup_{k \in \mathbb{N}} \{|S_{m+k}(\omega) - S_m(\omega)|\}
$$

and

$$
a(\omega) := \inf_{m \in \mathbb{N}} a_m(\omega).
$$

Then $\{S_n(\omega)\}_{n \in \mathbb{R}}$ is a Cauchy sequence iff $a(\omega) = 0$.

We want to show that $\mathbb{P}[a(\omega) > 0] = 0$. For this, it suffices to show that

$\mathbb{P}[a(\omega) > \varepsilon] = 0$ for all $\varepsilon > 0$. For a fixed $\varepsilon > 0$, we obtain:

$$\begin{aligned}\mathbb{P}[a_m > \varepsilon] &= \mathbb{P}[\sup_{k\in\mathbb{N}} |S_{m+k} - S_m| > \varepsilon] \\ &= \lim_{l\to\infty} \mathbb{P}[\underbrace{\sup_{k\leqslant l} |S_{m+k} - S_m| > \varepsilon}_{=:B_l \uparrow B := \{\sup_{k\in\mathbb{N}} |S_{m+k} - S_m| > \varepsilon\}}]\end{aligned}$$

Now,

$$\begin{aligned} & \max\{|S_{m+1} - S_m|, |S_{m+2} - S_m|, \ldots, |S_{m+l} - S_m|\} \\ = \; & \max\{|X_{m+1}|, |X_{m+1} + X_{m+2}|, \ldots, |X_{m+1} + X_{m+2} + \ldots + X_{m+l}|\} \\ \overset{\text{Kolmogorov}}{\leqslant} \; & \frac{1}{\varepsilon^2} \sum_{i=m}^{l} \mathrm{Var}(X_i) \\ \leqslant \; & \frac{1}{\varepsilon^2} \sum_{i=m}^{\infty} \mathrm{Var}(X_i) \xrightarrow{m\to\infty} 0, \end{aligned}$$

since by our assumption, $\sum_{n\in\mathbb{N}} \mathrm{Var}(X_i) < \infty$.

Hence

$$\mathbb{P}[a_m > \varepsilon] \xrightarrow{m\to\infty} 0.$$

It follows that $\mathbb{P}[a > \varepsilon] = 0$, as claimed. □

1.1.1 Application: Renewal Theorem

Theorem 1.15 (Renewal theorem)**.** Let $X_1, X_2, \ldots$ i.i.d. random variables with $X_i \geqslant 0$, $\mathbb{E}[X_i] = m > 0$. The X_i model waiting times. Let $S_n := \sum_{i=1}^{n} X_i$. For all $t > 0$ let

$$N_t := \sup\{n : S_n \leqslant t\}.$$

Then $\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{m}$ as $t \to \infty$.

The X_i can be thought of as waiting times. S_i models how long you have to wait for i events to occur.

Proof. By SLLN, $\frac{S_n}{n} \xrightarrow{a.s.} m$ as $n \to \infty$. Note that

$$N_t \uparrow \infty \text{ a.s. as } t \to \infty, \tag{1}$$

since $\{N_t \geqslant n\} = \{X_1 + \ldots + X_n \leqslant t\}$.

Claim 1. $\mathbb{P}[\frac{S_n}{n} \xrightarrow{n\to\infty} m \wedge N_t \xrightarrow{t\to\infty} \infty] = 1$.

Subproof. Let $A := \{\omega : \frac{S_n(\omega)}{n} \xrightarrow{n\to\infty} m\}$ and $B := \{\omega : N_t(\omega) \xrightarrow{t\to\infty} \infty\}$. By the SLLN we have $\mathbb{P}(A^C) = 0$ and by (1) it holds that $\mathbb{P}(B^C) = 0$. ■

Equivalently, $\mathbb{P}\left[\frac{S_{N_t}}{N_t} \xrightarrow{t\to\infty} m \wedge \frac{S_{N_t+1}}{N_t+1} \xrightarrow{t\to\infty} m\right] = 1$.

By definition, we have $S_{N_t} \leqslant t \leqslant S_{N_t+1}$. Thus

$$\frac{S_{N_t}}{N_t} \leqslant \frac{t}{N_t} \leqslant \frac{S_{N_t+1}}{N_t} \leqslant \frac{S_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{N_t}.$$

Hence $\frac{t}{N_t} \to m$.

□

[Lecture 7,]

Goal. *We want to drop our assumptions on finite mean or variance and say something about the behaviour of $\sum_{n\geqslant 1} X_n$ when the X_n are independent.*

Theorem 1.16 (Kolmogorov's three-series theorem)**.** Let X_n be a family of independent random variables.

(a) Suppose for some $C \geqslant 0$, the following three series of numbers converge:

- $\sum_{n\geqslant 1} \mathbb{P}(|X_n| > C)$,
- $\sum_{n\geqslant 1} \underbrace{\int_{|X_n|\leqslant C} X_n \,\mathrm{d}\mathbb{P}}_{\text{truncated mean}}$,
- $\sum_{n\geqslant 1} \underbrace{\int_{|X_n|\leqslant C} X_n^2 \,\mathrm{d}\mathbb{P} - \left(\int_{|X_n|\leqslant C} X_n \,\mathrm{d}\mathbb{P}\right)^2}_{\text{truncated variance}}$.

Then $\sum_{n\geqslant 1} X_n$ converges almost surely.

(b) Suppose $\sum_{n\geqslant 1} X_n$ converges almost surely. Then all three series above converge for every $C > 0$.

For the proof we'll need a slight generalization of Theorem 1.12:

Theorem 1.17. Let $\{X_n\}_n$ be independent and **uniformly bounded** (i.e. $\exists M < \infty : \sup_n \sup_\omega |X_n(\omega)| \leqslant M$). Then $\sum_{n\geqslant 1} X_n$ converges almost surely $\iff$ $\sum_{n\geqslant 1} \mathbb{E}(X_n)$ and $\sum_{n\geqslant 1} \mathrm{Var}(X_n)$ converge.

Proof of Theorem 1.16. Assume, that we have already proved Theorem 1.17. We prove part (a) first. Put $Y_n = X_n \cdot \mathbb{1}_{\{|X_n|\leqslant C\}}$. Since the X_n are independent, the Y_n are independent as well. Furthermore, the Y_n are uniformly bounded. By our assumption, the series $\sum_{n\geqslant 1} \int_{|X_n|\leqslant C} X_n \,\mathrm{d}\mathbb{P} = \sum_{n\geqslant 1} \mathbb{E}[Y_n]$ and $\sum_{n\geqslant 1} \int_{|X_n|\leqslant C} X_n^2 \,\mathrm{d}\mathbb{P} - \left(\int_{|X_n|\leqslant C} X_n \,\mathrm{d}\mathbb{P}\right)^2 = \sum_{n\geqslant 1} \mathrm{Var}(Y_n)$ converges. By Theorem 1.17 it follows that $\sum_{n\geqslant 1} Y_n < \infty$ almost surely. Let $A_n := \{\omega :$

$|X_n(\omega)| > C\}$. Since $\sum_{n\geqslant 1} \mathbb{P}(A_n) < \infty$ by assumption, Borel-Cantelli (0.10) yields $\mathbb{P}[\text{infinitely many } A_n \text{ occur}] = 0$.

For the proof of (b), suppose $\sum_{n\geqslant 1} X_n(\omega) < \infty$ for almost every ω. Fix an arbitrary $C > 0$. Define

$$Y_n(\omega) := \begin{cases} X_n(\omega) & \text{if} |X_n(\omega)| \leqslant C, \\ C & \text{if } |X_n(\omega)| > C. \end{cases}$$

Then the Y_n are independent and $\sum_{n\geqslant 1} Y_n(\omega) < \infty$ almost surely and the Y_n are uniformly bounded. By Theorem 1.17 $\sum_{n\geqslant 1} \mathbb{E}[Y_n]$ and $\sum_{n\geqslant 1} \operatorname{Var}(Y_n)$ converge. Define

$$Z_n(\omega) := \begin{cases} X_n(\omega) & \text{if } |X_n| \leqslant C, \\ -C & \text{if } |X_n| > C. \end{cases}$$

Then the Z_n are independent, uniformly bounded and $\sum_{n\geqslant 1} Z_n(\omega) < \infty$ almost surely. By Theorem 1.17 we have $\sum_{n\geqslant 1} \mathbb{E}(Z_n) < \infty$ and $\sum_{n\geqslant 1} \operatorname{Var}(Z_n) < \infty$.

We have

$$\begin{aligned} \mathbb{E}(Y_n) &= \int_{|X_n|\leqslant C} X_n \, \mathrm{d}\mathbb{P} + C\mathbb{P}(|X_n| \geqslant C), \\ \mathbb{E}(Z_n) &= \int_{|X_n|\leqslant C} X_n \, \mathrm{d}\mathbb{P} - C\mathbb{P}(|X_n| \geqslant C). \end{aligned}$$

Since $\mathbb{E}(Y_n) + \mathbb{E}(Z_n) = 2\int_{|X_n|\leqslant C} X_n \, \mathrm{d}\mathbb{P}$ the second series converges, and since $\mathbb{E}(Y_n) - \mathbb{E}(Z_n)$ converges, the first series converges. For the third series, we look at $\sum_{n\geqslant 1} \operatorname{Var}(Y_n)$ and $\sum_{n\geqslant 1} \operatorname{Var}(Z_n)$ to conclude that this series converges as well. ☐

Recall Theorem 1.12. We will see, that the converse of Theorem 1.12 is true if the X_n are uniformly bounded. More formally:

Theorem 1.18 (Theorem 5)**.** Let X_n be a series of independent variables with mean 0, that are uniformly bounded. If $\sum_{n\geqslant 1} X_n < \infty$ almost surely, then $\sum_{n\geqslant 1} \operatorname{Var}(X_n) < \infty$.

Proof of Theorem 1.17. Assume we have proven Theorem 1.18.

" $\Longleftarrow$ " Assume $\{X_n\}$ are independent, uniformly bounded and $\sum_{n\geqslant 1} \mathbb{E}(X_n) < \infty$ as well as $\sum_{n\geqslant 1} \operatorname{Var}(X_n) < \infty$. We need to show that $\sum_{n\geqslant 1} X_n < \infty$ a.s. Let $Y_n := X_n - \mathbb{E}(X_n)$. Then the Y_n are independent, $\mathbb{E}(Y_n) = 0$ and $\operatorname{Var}(Y_n) = \operatorname{Var}(X_n)$. By Theorem 1.12 $\sum_{n\geqslant 1} Y_n < \infty$ a.s. Thus $\sum_{n\geqslant 1} X_n < \infty$ a.s.

" $\Longrightarrow$ " We assume that $\{X_n\}$ are independent, uniformly bounded and $\sum_{n\geqslant 1} X_n(\omega) < \infty$ a.s. We have to show that $\sum_{n\geqslant 1} \mathbb{E}(X_n) < \infty$ and $\sum_{n\geqslant 1} \operatorname{Var}(X_n) < \infty$.

Consider the product space $(\Omega, \mathcal{F}, \mathbb{P}) \otimes (\Omega, \mathcal{F}, \mathbb{P})$. On this product space, we define $Y_n((\omega, \omega')) := X_n(\omega)$ and $Z_n((\omega, \omega')) := X_n(\omega')$.

Claim 1.17.1. *For every fixed n, Y_n and Z_n are independent.*

Subproof. This is obvious, but we will prove it carefully here.

$$
\begin{aligned}
&(\mathbb{P}\otimes\mathbb{P})[Y_n \in (a,b), Z_n \in (a',b')] \\
&= (\mathbb{P}\otimes\mathbb{P})\left((\omega,\omega') : X_n(\omega) \in (a,b) \wedge X_n(\omega') \in (a',b')\right) \\
&= (\mathbb{P}\otimes\mathbb{P})(A \times A') \text{where } A := X_n^{-1}((a,b)) \text{ and } A' := X_n^{-1}\left((a',b')\right) \\
&= \mathbb{P}(A)\mathbb{P}(A')
\end{aligned}
$$

■

Now $\mathbb{E}[Y_n - Z_n] = 0$ (by definition) and $\operatorname{Var}(Y_n - Z_n) = 2\operatorname{Var}(X_n)$. Obviously, $(Y_n - Z_n)_{n\geqslant 1}$ is also uniformly bounded.

Claim 1.17.2. $\sum_{n\geqslant 1}(Y_n - Z_n) < \infty$ *almost surely on* $(\Omega\otimes\Omega, \mathcal{F}\otimes\mathcal{F}, \mathbb{P}\otimes\mathbb{P})$.

Subproof. Suppose $\Omega_0 = \{\omega : \sum_{n\geqslant 1} X_n(\omega) < \infty\}$. Then $\mathbb{P}(\Omega_0) = 1$. Thus $(\mathbb{P}\otimes\mathbb{P})(\Omega_0\otimes\Omega_0) = 1$. Furthermore $\sum_{n\geqslant 1}(Y_n(\omega,\omega') - Z_n(\omega,\omega')) = \sum_{n\geqslant 1}(X_n(\omega) - X_n(\omega'))$. Thus $\sum_{n\geqslant 1}(Y_n(\omega,\omega') - Z_n(\omega,\omega')) < \infty$ a.s. on $\Omega_0\otimes\Omega_0$. ■

By Theorem 1.18, $\sum_n \operatorname{Var}(X_n) = \frac{1}{2}\sum_{n\geqslant 1}\operatorname{Var}(Y_n - Z_n) < \infty$ a.s. Define $U_n := X_n - \mathbb{E}(X_n)$. Then $\mathbb{E}(U_n) = 0$ and the U_n are independent and uniformly bounded. We have $\sum_n \operatorname{Var}(U_n) = \sum_n \operatorname{Var}(X_n) < \infty$. Thus $\sum_n U_n$ converges a.s. by Theorem 1.12. Since by assumption $\sum_n X_n < \infty$ a.s., it follows that $\sum_n \mathbb{E}(X_n) < \infty$. □

> **Remark 1.18.18.** In the proof of Theorem 1.17 " $\Longleftarrow$ " is just a trivial application of Theorem 1.12 and uniform boundedness was not used. The idea of " $\Longrightarrow$ " will lead to coupling.

A proof of Theorem 1.18 can be found in the notes. TODO: copy from official notes

> **Example 1.19** (Application of Theorem 1.17)**.** The series $\sum_n \frac{1}{n^{\frac{1}{2}+\varepsilon}}$ does not converge for $\varepsilon < \frac{1}{2}$. However
>
> $$\sum_n X_n \frac{1}{n^{\frac{1}{2}+\varepsilon}}$$
>
> where $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = \frac{1}{2}$ converges almost surely for all $\varepsilon > 0$. And
>
> $$\sum_n X_n \frac{1}{n^{\frac{1}{2}-\varepsilon}}$$
>
> does not converge.

[Lecture 8, 2023-05-02]

1.2 Kolmogorov's 0-1-law

Some classes of events always have probability 0 or 1. One example of such a 0-1-law is the Borel-Cantelli Lemma and its inverse statement.

We now want to look at events that capture certain aspects of long term behaviour of sequences of random variables.

Definition 1.20. Let $X_n, n \in \mathbb{N}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{T}_i := \sigma(X_i, X_{i+1}, \ldots)$ be the σ-algebra generated by $X_i, X_{i+1}, \ldots$. Then the **tail-σ-algebra** is defined as

$$\mathcal{T} := \bigcap_{i \in \mathbb{N}} \mathcal{T}_i.$$

The events $A \in \mathcal{T} \subseteq \mathcal{F}$ are called **tail events**.

Remark 1.20.19. (i) Since intersections of arbitrarily many σ-algebras is again a σ-algebra, $\mathcal{T}$ is indeed a σ-algebra.

(ii) We have

$$\mathcal{T} = \{A \in \mathcal{F} \mid \forall i \ \exists B \in \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}} : A = \{\omega | (X_i(\omega), X_{i+1}(\omega), \ldots) \in B\}\}.$$

Example 1.21 (What are tail events?)**.** Let $X_n, n \in \mathbb{N}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then

(i) $\{\omega | \sum_{n \in \mathbb{N}} X_n(\omega) \text{ converges}\}$ is a tail event, since for all $\omega \in \Omega$ we have

$$\begin{aligned} & \sum_{i=1}^{\infty} X_i(\omega) \text{ converges} \\ \iff & \sum_{i=2}^{\infty} X_i(\omega) \text{ converges} \\ \iff & \ldots \\ \iff & \sum_{i=k}^{\infty} X_i(\omega) \text{ converges.} \end{aligned}$$

(Since the X_i are independent, the convergence of $\sum_{n \in \mathbb{N}} X_n$ is not influenced by $X_1, \ldots, X_k$ for any k.)

(ii) $\{\omega | \sum_{n \in \mathbb{N}} X_n(\omega) = c\}$ for some $c \in \mathbb{R}$ is not a tail event, because $\sum_{n \in \mathbb{N}} X_n$ depends on X_1.

(iii) $\{\omega | \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) = c\}$ is a tail event, since

$$c = \lim_{n\to\infty}\sum_{i=1}^n X_i = \underbrace{\lim_{n\to\infty}\frac{1}{n}X_1}_{=0} + \lim_{n\to\infty}\frac{1}{n}\sum_{i=2}^n X_i = \ldots = \lim_{n\to\infty}\frac{1}{n}\sum_{i=k}^n X_i.$$

So $\mathcal{T}$ includes all long term behaviour of $X_n, n \in \mathbb{N}$, which does not depend on the realisation of the first k random variables for any $k \in \mathbb{N}$.

Theorem 1.22 (Kolmogorov's 0-1 law)**.** Let $X_n, n \in \mathbb{N}$ be a sequence of independent random variables and let $\mathcal{T}$ denote their tail-σ-algebra. Then $\mathcal{T}$ is **$\mathbb{P}$-trivial**, i.e. $\mathbb{P}[A] \in \{0,1\}$ for all $A \in \mathcal{T}$.

Idea. *The idea behind proving, that a $\mathcal{T}$ is $\mathbb{P}$-trivial is to show that for any $A, B \in \mathcal{F}$ we have*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Taking $A = B$, it follows that $\mathbb{P}[A] = \mathbb{P}[A]^2$, hence $\mathbb{P}[A] \in \{0,1\}$.

Proof of Theorem 1.22. Let $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$ and remember that $\mathcal{T}_n = \sigma(X_n, X_{n+1}, \ldots)$. The proof rests on two claims:

Claim 1.22.1. *For all $n \geqslant 1$, $A \in \mathcal{F}_n$ and $B \in \mathcal{T}_{n+1}$ we have $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.*

Subproof. This follows from the independence of the X_i. It is

$$\sigma(X_1, \ldots, X_n) = \sigma\left(\underbrace{\{X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n) | B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})\}}_{=:\mathcal{A}}\right).$$

$\mathcal{A}$ is a semi-algebra, since

(i) $\emptyset, \Omega \in \mathcal{A}$,

(ii) $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$,

(iii) for $A \in \mathcal{A}$, $A^c = \bigsqcup_{i=1}^n A_i$ for disjoint sets $A_1, \ldots, A_n \in \mathcal{A}$.

Hence it suffices to show the claim for sets $A \in \mathcal{A}$. Similarly

$$\sigma(\mathcal{T}_{n+1}) = \sigma\left(\underbrace{\{X_{n+1}^{-1}(M_1) \cap \ldots \cap X_{n+k}^{-1}(M_k) | k \in \mathbb{N}, M_1, \ldots, M_k \in \mathcal{B}(\mathbb{R})\}}_{=:\mathcal{B}}\right).$$

Again, $\mathcal{B}$ is closed under intersection.

So let $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B)$$

by the independence of $\{X_1, \ldots, X_{n+k}\}$, and since A only depends on $\{X_1, \ldots, X_n\}$ and B only on $\{X_{n+1}, \ldots, X_{n+k}\}$. ■

Claim 1.22.2. $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$ *is an algebra and*

$$\sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\right) = \sigma(X_1, X_2, \ldots) = \mathcal{T}_1.$$

Subproof. "$\supseteq$ " If $A_n \in \sigma(X_n)$, then $A_n \in \mathcal{F}_n$. Hence $A_n \in \bigcup_{n \in \mathbb{N}} \mathcal{F}_n$.

Since $\sigma(X_1, X_2, \ldots)$ is generated by $\{A_n \in \sigma(X_n) : n \in \mathbb{N}\}$, this also means $\sigma(X_1, X_2, \ldots) \subseteq \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\right)$.

"$\subseteq$ " Since $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, obviously $\mathcal{F}_n \subseteq \sigma(X_1, X_2 \ldots)$ for all n. It follows that $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n \subseteq \sigma(X_1, X_2, \ldots)$. Hence $\sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\right) \subseteq \sigma(X_1, X_2, \ldots)$. ■

Now let $T \in \mathcal{T}$. Then $T \in \mathcal{T}_{n+1}$ for any n. Hence $\mathbb{P}[A \cap T] = \mathbb{P}[A]\mathbb{P}[T]$ for all $A \in \mathcal{F}_n$ by the first claim.

It follows that the same folds for all $A \in \bigcup_{n \in \mathbb{N}} \mathcal{F}_n$, hence for all $A \in \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\right)$, and by the second claim for all $A \in \sigma(X_1, X_2, \ldots) = \mathcal{T}_1$. But since $T \in \mathcal{T}$, in particular $T \in \mathcal{T}_1$, so by choosing $A = T$, we get

$$\mathbb{P}[T] = \mathbb{P}[T \cap T] = \mathbb{P}[T]^2$$

hence $\mathbb{P}[T] \in \{0, 1\}$. □

Fact† 1.22.20 (Exercise 5.2 (b))**.** Any random variable measurable with respect to a $\mathbb{P}$-trivial σ-algebra is a.s. a constant.

[Lecture 9,]

1.2.1 Application: Percolation

We will now discuss another application of Kolmogorov's 0-1 Law (1.22), percolation.

Definition 1.23 (**Percolation**)**.** Consider the graph with nodes $\mathbb{Z}^d$, $d \geqslant 2$, where edges from the lattice are added with probability p. The added edges are called **open**; all other edges are called **closed**.

More formally, we consider

- $\Omega = \{0, 1\}^{\mathbb{E}_d}$, where $\mathbb{E}_d$ are all edges in $\mathbb{Z}^d$,
- $\mathcal{F} :=$ product σ-algebra,

- $\mathbb{P} := \left(p \underbrace{\delta_{\{1\}}}_{\text{edge is open}} + (1-p) \underbrace{\delta_{\{0\}}}_{\text{edge is absent closed}} \right)^{\otimes \mathbb{E}_d}$.

Question 1.23.21. Starting at the origin, what is the probability, that there exists an infinite path (without moving backwards)?

Definition 1.24. An **infinite path** consists of an infinite sequence of distinct points $x_0, x_1, \ldots$ such that x_n is connected to x_{n+1}, i.e. the edge $\{x_n, x_{n+1}\}$ is open.

Let $C_\infty := \{\omega | \text{an infinite path exists}\}$.

Exercise. *Show that changing the presence / absence of finitely many edges does not change the existence of an infinite path. Therefore C_∞ is an element of the tail σ-algebra. Hence $\mathbb{P}(C_\infty) \in \{0, 1\}$.*

Obviously, $\mathbb{P}(C_\infty)$ is monotonic with respect to p. For $d = 2$ it is known that $p = \frac{1}{2}$ is the critical value. For $d > 2$ this is unknown.

We'll get back to percolation later.

2 Characteristic Functions, Weak Convergence and the Central Limit Theorem

So far we have dealt with the average behaviour,

$$\frac{\overbrace{X_1 + \ldots + X_n}^{\text{i.i.d.}}}{n} \to \mathbb{E}(X_1).$$

We now want to understand fluctuations from the average behaviour, i.e.

$$X_1 + \ldots + X_n - n \cdot \mathbb{E}(X_1).$$

The question is, what happens on other timescales than n? An example is

$$\frac{X_1 + \ldots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}} \xrightarrow{n \to \infty} \mathcal{N}(0, \operatorname{Var}(X_i)) \tag{2}$$

Why is $\sqrt{n}$ the right order? Handwavey argument:

Suppose $X_1, X_2, \ldots$ are i.i.d. with $X_1 \sim \mathcal{N}(0, 1)$. The mean of the l.h.s. is 0 and for the variance we get

$$\begin{aligned} \operatorname{Var}(\frac{X_1 + \ldots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}}) &= \operatorname{Var}\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right) \\ &= \frac{1}{n}\left(\operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n)\right) = 1 \end{aligned}$$

For the r.h.s. we get a mean of 0 and a variance of 1. So, to determine what (2) could mean, it is necessary that $\sqrt{n}$ is the right scaling. To make (2) precise, we need another notion of convergence. This will be the weakest notion of convergence, hence it is called **weak convergence**. This notion of convergence will be defined in terms of characteristic functions of Fourier transforms.

2.1 Convolutions†

Definition† 2.0.22 (Convolution)**.** Let μ and ν be probability measures on $\mathbb{R}^d$. Then the **convolution** of μ and ν, $\mu * \nu$, is the probability measure on $\mathbb{R}^d$ defined by

$$(\mu * \nu)(A) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_A(x+y)\mu(\mathrm{d}x)\nu(\mathrm{d}y).$$

Fact 2.0.23. If μ and ν have Lebesgue densities f_μ and f_ν, then the convolution has Lebesgue density

$$f_{\mu * \nu}(x) = \int_{\mathbb{R}^d} f_\mu(x-t) f_\nu(t) \lambda^d(\mathrm{d}t).$$

Fact† 2.0.24 (Exercise 6.1)**.** If $X_1, X_2, \ldots$ are independent with distributions $X_1 \sim \mu_1$, $X_2 \sim \mu_2, \ldots$, then $X_1 + \ldots + X_n$ has distribution

$$\mu_1 * \mu_2 * \ldots * \mu_n.$$

TODO

2.2 Characteristic Functions and Fourier Transform

Definition 2.1. Consider $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$. The **characteristic function** of $\mathbb{P}$ is defined as

$$\begin{aligned} \varphi_{\mathbb{P}} : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \int_{\mathbb{R}} e^{\mathbf{i}tx}\mathbb{P}(\mathrm{d}x). \end{aligned}$$

Abuse of Notation 2.1.25. $\varphi_{\mathbb{P}}(t)$ will often be abbreviated as $\varphi(t)$.

We have

$$\varphi(t) = \int_{\mathbb{R}} \cos(tx)\mathbb{P}(dx) + \mathbf{i}\int_{\mathbb{R}} \sin(tx)\mathbb{P}(dx).$$

- Since $|e^{\mathbf{i}tx}| \leqslant 1$ the function $\varphi(\cdot)$ is always defined.
- We have $\varphi(0) = 1$.
- $|\varphi(t)| \leqslant \int_{\mathbb{R}} |e^{\mathbf{i}tx}|\mathbb{P}(dx) = 1$.

Fact† 2.1.26. Let X, Y be independent random variables and $a, b \in \mathbb{R}$. Then

- $\varphi_{aX+b}(t) = e^{\mathbf{i}tb}\varphi_X(\frac{t}{a})$,
- $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$.

Proof. We have

$$\begin{aligned}\varphi_{aX+b}(t) &= \mathbb{E}[e^{\mathbf{i}t(aX+b)}] \\ &= e^{\mathbf{i}tb}\mathbb{E}[e^{\mathbf{i}taX}] \\ &= e^{\mathbf{i}tb}\varphi_X(\frac{t}{a}).\end{aligned}$$

Furthermore

$$\begin{aligned}\varphi_{X+Y}(t) &= \mathbb{E}[e^{\mathbf{i}t(X+Y)}] \\ &= \mathbb{E}[e^{\mathbf{i}tX}]\mathbb{E}[e^{\mathbf{i}tY}] \\ &= \varphi_X(t)\varphi_Y(t).\end{aligned}$$

□

Remark 2.1.27. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is an arbitrary probability space and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable. Then we can define

$$\varphi_X(t) := \mathbb{E}[e^{\mathbf{i}tX}] = \int e^{\mathbf{i}tX(\omega)}\mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{R}} e^{\mathbf{i}tx}\mu(dx) = \varphi_\mu(t),$$

where $\mu = \mathbb{P} \circ X^{-1}$.

Theorem 2.2 (Inversion formula)**.** Let $(\Omega, \mathcal{B}(\mathbb{R}), \mathbb{P})$ be a probability space. Let F be the distribution function of $\mathbb{P}$ (i.e. $F(x) = \mathbb{P}((-\infty, x])$ for all $x \in \mathbb{R}$). Then for every $a < b$ we have

$$\frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\mathbf{i}tb} - e^{-\mathbf{i}ta}}{-\mathbf{i}t}\varphi(t)dt \qquad (3)$$

where $F(b-)$ is the left limit.

We will prove this later.

Theorem 2.3 (Uniqueness theorem)**.** Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then $\varphi_{\mathbb{P}} = \varphi_{\mathbb{Q}} \implies \mathbb{P} = \mathbb{Q}$.

Therefore, probability measures are uniquely determined by their characteristic functions. Moreover, (3) gives a representation of $\mathbb{P}$ (via F) from φ.

Proof of Theorem 2.3. Assume that we have already shown the Inversion Formula (2.2). Suppose that F and G are the distribution functions of $\mathbb{P}$ and $\mathbb{Q}$. Let $a, b \in \mathbb{R}$ with $a < b$. Assume that a and b are continuity points of both F and G. By the Inversion Formula (2.2) we have

$$F(b) - F(a) = G(b) - G(a)$$

Since F and G are monotonic, Equation 4 holds for all $a < b$ outside a countable set.

Take a_n outside this countable set, such that $a_n \searrow -\infty$. Then, Equation 4 implies that $F(b) - F(a_n) = G(b) - G(a_n)$ hence $F(b) = G(b)$. Since F and G are right-continuous, it follows that $F = G$. $\square$

[Lecture 10, 2023-05-09]

First, we will prove some of the most important facts about Fourier transforms.

We consider $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Notation 2.3.28. By $M_1(\mathbb{R})$ we denote the set of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

For all $\mathbb{P} \in M_1(\mathbb{R})$ we define $\varphi_{\mathbb{P}}(t) = \int_{\mathbb{R}} e^{\mathbf{i}tx}\mathbb{P}(\mathrm{d}x)$. If $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable, we write $\varphi_X(t) := \mathbb{E}[e^{\mathbf{i}tX}] = \varphi_\mu(t)$, where $\mu = \mathbb{P}X^{-1}$.

Proof of Theorem 2.2. We will prove that the limit in the RHS of Equation 3 exists and is equal to the LHS. Note that the term on the RHS is integrable, as

$$\lim_{t \to 0} \frac{e^{-\mathbf{i}tb} - e^{-\mathbf{i}ta}}{-\mathbf{i}t}\varphi(t) = a - b$$

and note that $\varphi(0) = 1$ and $|\varphi(t)| \leqslant 1$.

We have

$$\begin{aligned}
& \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \int_{\mathbb{R}} \frac{e^{-\mathbf{i}tb} - e^{-\mathbf{i}ta}}{-\mathbf{i}t} e^{\mathbf{i}tx} \,\mathrm{d}t \mathbb{P}(\mathrm{d}x) \\
\overset{\text{Fubini}}{=} & \lim_{T\to\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{-\mathbf{i}tb} - e^{-\mathbf{i}ta}}{-\mathbf{i}t} e^{\mathbf{i}tx} \,\mathrm{d}t \mathbb{P}(\mathrm{d}x) \\
= & \lim_{T\to\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{\mathbf{i}t(b-x)} - e^{\mathbf{i}t(x-a)}}{-\mathbf{i}t} \,\mathrm{d}t \mathbb{P}(\mathrm{d}x) \\
= & \lim_{T\to\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \underbrace{\int_{-T}^{T} \left[\frac{\cos(t(x-b)) - \cos(t(x-a))}{-\mathbf{i}t}\right] \mathrm{d}t}_{=0,\text{ as the function is odd}} \,\mathbb{P}(\mathrm{d}x) \\
& + \lim_{T\to\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^{T} \frac{\sin(t(x-b)) - \sin(t(x-a))}{-t} \,\mathrm{d}t \mathbb{P}(\mathrm{d}x) \\
= & \lim_{T\to\infty} \frac{1}{\pi} \int_{\mathbb{R}} \int_{0}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} \,\mathrm{d}t \mathbb{P}(\mathrm{d}x) \\
\overset{(2.3.29),\text{DCT}}{=} & \frac{1}{\pi} \int -\frac{\pi}{2} \mathbb{1}_{x<a} + \frac{\pi}{2} \mathbb{1}_{x>a} - (-\frac{\pi}{2} \mathbb{1}_{x<b} + \frac{\pi}{2} \mathbb{1}_{x>b}) \mathbb{P}(\mathrm{d}x) \\
= & \frac{1}{2} \mathbb{P}(\{a\}) + \frac{1}{2} \mathbb{P}(\{b\}) + \mathbb{P}((a,b)) \\
= & \frac{F(b) + F(b-)}{2} - \frac{F(a) - F(a-)}{2}
\end{aligned}$$

□

Fact 2.3.29.

$$\int_0^\infty \frac{\sin x}{x} \,\mathrm{d}x = \frac{\pi}{2}$$

where the LHS is an improper Riemann-integral. Note that the LHS is not Lebesgue-integrable. It follows that

$$\lim_{T\to\infty} \int_0^T \frac{\sin(t(x-a))}{t} \,\mathrm{d}t = \begin{cases} -\frac{\pi}{2} & \text{if } x < a, \\ 0 & \text{if } x = a, \\ \frac{\pi}{2} & \text{if } x > a. \end{cases}$$

Theorem 2.4. Let $\mathbb{P} \in M_1(\mathbb{R})$ such that $\varphi_{\mathbb{P}} \in L^1(\lambda)$. Then $\mathbb{P}$ has a continuous probability density given by

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\mathbf{i}tx} \varphi_{\mathbb{P}}(t) \,\mathrm{d}t.$$

Example 2.5. • Let $\mathbb{P} = \delta_0$. Then

$$\varphi_{\mathbb{P}}(t) = \int e^{\mathbf{i}tx} \delta_0(\mathrm{d}x) = e^{\mathbf{i}t0} = 1$$

• Let $\mathbb{P} = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$. Then

$$\varphi_{\mathbb{P}}(t) = \frac{1}{2}e^{\mathbf{i}t} + \frac{1}{2}e^{-\mathbf{i}t} = \cos(t)$$

Proof of Theorem 2.4. Let $f(x) \coloneqq \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathbf{i}tx}\varphi(t)\,\mathrm{d}t$.

Claim 2.4.1. *If $x_n \to x$, then $f(x_n) \to f(x)$.*

Subproof. Suppose that $e^{-\mathbf{i}tx_n}\varphi(t) \xrightarrow{n\to\infty} e^{-\mathbf{i}tx}\varphi(t)$ for all t. Since

$$|e^{-\mathbf{i}tx}\varphi(t)| \leqslant |\varphi(t)|$$

and $\varphi \in L^1$, we get $f(x_n) \to f(x)$ by the dominated convergence theorem. ■

We'll show that for all $a < b$ we have

$$\mathbb{P}\left((a,b]\right) = \int_a^b f(x)\,\mathrm{d}x.$$

Let F be the distribution function of $\mathbb{P}$. It is enough to prove Claim 2.3.29 for all continuity points a and b of F. We have

$$\begin{aligned}
\text{RHS} &\overset{\text{Fubini}}{=} \frac{1}{2\pi}\int_{\mathbb{R}}\int_a^b e^{-\mathbf{i}tx}\varphi(t)\,\mathrm{d}x\,\mathrm{d}t \\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\varphi(t)\int_a^b e^{-\mathbf{i}tx}\,\mathrm{d}x\,\mathrm{d}t \\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\varphi(t)\left(\frac{e^{-\mathbf{i}tb} - e^{-\mathbf{i}ta}}{-\mathbf{i}t}\right)\mathrm{d}t \\
&\overset{\text{dominated convergence}}{=} \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\varphi(t)\left(\frac{e^{-\mathbf{i}tb} - e^{-\mathbf{i}ta}}{-\mathbf{i}t}\right)\mathrm{d}t
\end{aligned}$$

By the Inversion Formula (2.2), the RHS is equal to $F(b) - F(a) = \mathbb{P}\left((a,b]\right)$. □

However, Fourier analysis is not only useful for continuous probability density functions:

Theorem 2.6 (Bochner's formula for the mass at a point)**.** Let $\mathbb{P} \in M_1(\lambda)$.

Then

$$\forall x \in \mathbb{R}.\ \mathbb{P}(\{x\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-\mathbf{i}tx} \varphi(t)\,\mathrm{d}t.$$

Proof of Theorem 2.6. We have

$$\begin{aligned}
\text{RHS} &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-\mathbf{i}tx} \int_{\mathbb{R}} e^{\mathbf{i}ty} \mathbb{P}(\mathrm{d}y) \\
&\overset{\text{Fubini}}{=} \lim_{T\to\infty} \frac{1}{2T} \int_{\mathbb{R}} \int_{-T}^{T} e^{-\mathbf{i}t(y-x)}\,\mathrm{d}t\mathbb{P}(\mathrm{d}y) \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{\mathbb{R}} \int_{-T}^{T} \cos(t(y-x)) + \underbrace{\mathbf{i}\sin(t(y-x))}_{\text{odd}}\,\mathrm{d}t\mathbb{P}(\mathrm{d}y) \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{\mathbb{R}} \int_{-T}^{T} \cos(t(y-x))\,\mathrm{d}t\mathbb{P}(\mathrm{d}y) \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{\mathbb{R}} 2T\,\mathrm{sinc}(T(y-x))^{2}\mathbb{P}(\mathrm{d}y) \\
&\overset{\text{DCT}}{=} \int_{\mathbb{R}} \lim_{T\to\infty} \mathrm{sinc}(T(y-x))\mathbb{P}(\mathrm{d}y) \\
&= \mathbb{P}(\{x\}).
\end{aligned}$$

□

Theorem 2.7. Let φ be the characteristic function of $\mathbb{P} \in M_1(\lambda)$. Then

(a) $\varphi(0) = 1$, $|\varphi(t)| \leqslant 1$, $\varphi(-t) = \overline{\varphi(t)}$ and $\varphi(\cdot)$ is continuous.

(b) φ is a **positive definite function**, i.e.

$$\forall t_1, \ldots, t_n \in \mathbb{R}, (c_1, \ldots, c_n) \in \mathbb{C}^n \quad \sum_{j,k=1}^{n} c_j \overline{c_k} \varphi(t_j - t_k) \geqslant 0$$

Equivalently, the matrix $(\varphi(t_j - t_k))_{j,k}$ is positive definite.

Proof of Theorem 2.7. Part (a) is obvious.

[2] $\mathrm{sinc}(x) = \begin{cases} \frac{\sin(x)}{x} & \text{if } x \neq 0, \\ 1 & \text{otherwise.} \end{cases}$

For part (b) we have:

$$\begin{aligned}\sum_{j,k} c_j \overline{c_k} \varphi(t_j - t_k) &= \sum_{j,k} c_j \overline{c_k} \int_{\mathbb{R}} e^{\mathbf{i}(t_j - t_k)x} \mathbb{P}(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \sum_{j,k} c_j \overline{c_k} e^{\mathbf{i}t_j x} \overline{e^{\mathbf{i}t_k x}} \mathbb{P}(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \sum_{j,k} c_j e^{\mathbf{i}t_j x} \overline{c_k e^{\mathbf{i}t_k x}} \mathbb{P}(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \left| \sum_l c_l e^{\mathbf{i}t_l x} \right|^2 \geqslant 0\end{aligned}$$

□

Theorem 2.8 (Bochner's theorem)**.** The converse to Theorem 2.7 holds, i.e. any $\varphi : \mathbb{R} \to \mathbb{C}$ satisfying (a) and (b) of Theorem 2.7 must be the Fourier transform of a probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Unfortunately, we won't prove Bochner's Theorem for Positive Definite Functions (2.8) in this lecture.

Definition 2.9 (Convergence in distribution / weak convergence)**.** We say that $\mathbb{P}_n \in M_1(\mathbb{R})$ **converges weakly** towards $\mathbb{P} \in M_1(\mathbb{R})$ (notation: $\mathbb{P}_n \implies \mathbb{P}$), iff

$$\forall f \in C_b(\mathbb{R}) \int f \,\mathrm{d}\mathbb{P}_n \to \int f \,\mathrm{d}\mathbb{P}.$$

Where

$$C_b(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \text{ continuous and bounded}\}$$

In analysis, this is also known as weak* convergence.

Remark 2.9.30. This notion of convergence makes $M_1(\mathbb{R})$ a separable metric space. We can construct a metric on $M_1(\mathbb{R})$ that turns $M_1(\mathbb{R})$ into a complete and separable metric space:

Consider the sets

$$\{\mathbb{P} \in M_1(\mathbb{R}) : \forall i = 1, \ldots, n \int f \,\mathrm{d}\mathbb{P} - \int f_i \,\mathrm{d}\mathbb{P} < \varepsilon\}$$

for any $f, f_1, \ldots, f_n \in C_b(\mathbb{R})$. These sets form a basis for the topology on $M_1(\mathbb{R})$. More of this will follow later.

Example 2.10.

- Let $\mathbb{P}_n = \delta_{\frac{1}{n}}$. Then $\int f \,\mathrm{d}\mathbb{P}_n = f(\frac{1}{n}) \to f(0) = \int f d\delta_0$ for any continuous, bounded function f. Hence $\mathbb{P}_n \to \delta_0$.
- $\mathbb{P}_n := \delta_n$ does not converge weakly, as for example

$$\int \cos(\pi x) \,\mathrm{d}\mathbb{P}_n(x)$$

does not converge.

- $\mathbb{P}_n := \frac{1}{n}\delta_n + (1 - \frac{1}{n})\delta_0$. Let $f \in C_b(\mathbb{R})$ arbitrary. Then

$$\int f \,\mathrm{d}\mathbb{P}_n = \frac{1}{n}(n) + (1 - \frac{1}{n}) f(0) \to f(0)$$

since f is bounded. Hence $\mathbb{P}_n \implies \delta_0$.

- $\mathbb{P}_n := \frac{1}{\sqrt{2\pi n}} e^{-\frac{x^2}{2n}}$. This "converges" towards the 0-measure, which is not a probability measure. Hence $\mathbb{P}_n$ does not converge weakly. (Exercise)

Definition 2.11. We say that a series of random variables X_n **converges in distribution** to X (notation: $X_n \xrightarrow{\mathrm{d}} X$), iff $\mathbb{P}_n \implies \mathbb{P}$, where $\mathbb{P}_n$ is the distribution of X_n and $\mathbb{P}$ is the distribution of X.

It is easy to see, that this is equivalent to $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in C_b(\mathbb{R})$.

Example 2.12. Let $X_n := \frac{1}{n}$ and F_n the distribution function, i.e. $F_n = \mathbb{1}_{[\frac{1}{n}, \infty)}$. Then $\mathbb{P}_n = \delta_{\frac{1}{n}} \implies \delta_0$ which is the distribution of $X \equiv 0$. But $F_n(0) \not\to F(0)$.

Theorem 2.13. $X_n \xrightarrow{\mathrm{d}} X$ iff $F_n(t) \to F(t)$ for all continuity points t of F.

Theorem 2.14 (Levy's continuity theorem)**.** $X_n \xrightarrow{\mathrm{d}} X$ iff $\varphi_{X_n}(t) \to \varphi(t)$ for all $t \in \mathbb{R}$.

We will assume these two theorems for now and derive the central limit theorem. The theorems will be proved later.

[Lecture 11,]

2.3 The Central Limit Theorem

For $X_1, X_2, \ldots$ i.i.d. we were looking at $S_n := \sum_{i=1}^n X_i$. Then the LLN basically states, that S_n can be approximated by $n\mathbb{E}[X_1]$.

Question 2.14.31. What is the error of this approximation?

We set $\mu := \mathbb{E}[X_1]$ and $\sigma^2 := \mathrm{Var}(X_1) \in (0, \infty)$. We know that $\mathbb{E}[S_n] = n\mu$ and $\mathrm{Var}(S_n) = n\sigma^2$.

The central limit theorem basically states, that the distribution of S_n can be approximated by a normal distribution with mean $n\mu$ and variance $n\sigma^2$, i.e. $S_n \approx n\mu + \sigma\sqrt{n}N$ for $N \sim \mathcal{N}(0,1)$, where $\approx$ is to be made precise.

For intuition, watch https://3blue1brown.com/lessons/clt.

Example 2.15. We throw a fair die $n = 100$ times and denote the sum of the faces by $S_n := X_1 + \ldots + X_n$, where $X_1, \ldots, X_n$ are i.i.d. and uniformly distributed on $\{1, \ldots, 6\}$. Then $\mathbb{E}[S_n] = 350$ and $\sqrt{\mathrm{Var}(S_n)} = \sigma \approx 17.07$.

Missing pictures

Question 2.15.32. Why do statisticians care about σ instead of σ^2?

By definition, $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$, hence $\sqrt{\mathrm{Var}(X)}$ can be interpreted as a distance. One could also define $\mathrm{Var}(X)$ to be $\mathbb{E}[|X - \mathbb{E}(X)|]$ but this is not well behaved.

Example 2.16. Let $X_1, \ldots, X_n$ be i.i.d. and $X_1 \sim \mathrm{Exp}(1)$. We knot that for $n \in \mathbb{N}$, $\mathbb{E}[S_n] = n$ and $\sqrt{\mathrm{Var}(S_n)} = \sqrt{n}$. For $n = 100, 300, 500$, we get the following picture

Missing picture

In order to make things nicer, we do the following:

1. center: $S_n - \mathbb{E}[S_n]$,
2. normalize: $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$.

Then $\mathbb{E}[\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}] = 0$ and $\mathrm{Var}(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}) = 1$.

Theorem 2.17 (Central limit theorem, 1920s, Lindeberg and Levy)**.** Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$.

Let $S_n := \sum_{i=1}^n X_i$. Then

$$\frac{S_n - n\nu}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1),$$

i.e. $\forall x \in \mathbb{R}:$

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right] = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt.$$

We will abbreviate the central limit theorem by **CLT**.

There exists a special case of this theorem, which was proved earlier:

Theorem 2.18 (de-Moivre (1730, $p = 0.5$), Laplace (1812, general p))**.** Let $S_n = \text{Bin}(n,p)$, where $p \in (0,1)$ is constant. Then, for all $x \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant x\right] = \Phi(x).$$

Proof. Let $X_1, X_2, \ldots$ i.i.d. with $X_1 \sim \text{Ber}(p)$. Then $\mathbb{E}[X_1] = p$ and $\text{Var}(X_1) = p(1-p)$. Furthermore $\sum_{i=1}^n X_i \sim \text{Bin}(n,p)$, and the special case follows from Central Limit Theorem (2.17). □

Theorem 2.18 is a useful tool for approximating the Binomial distribution with the normal distribution. If $S_n \sim \text{Bin}(n,p)$ and $[a,b] \subseteq \mathbb{R}$, we have

$$\mathbb{P}[a \leqslant S_n \leqslant b] = \mathbb{P}\left[\frac{a-np}{\sqrt{np(1-p)}} \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant \frac{b-np}{\sqrt{np(1-p)}}\right] \approx \Phi(b') - \Phi(a').$$

Example 2.19. We consider a $n = 40$-times Bernoulli trial with success probability $p = \frac{1}{2}$. Then $0.9597 = \mathbb{P}[S \leqslant 25] \approx \Phi(\frac{5}{\sqrt{10}} \approx 0.9431$.

However, S takes only integer values, which means $\mathbb{P}[S \leqslant 25] = \mathbb{P}[S26]$. With this in mind, a better approximation is

$$\mathbb{P}[S \leqslant 25] = \mathbb{P}[S \leqslant 25.5] \approx \Phi\left(\frac{5.5}{\sqrt{10}}\right) \approx 0.9541.$$

Example 2.20. Consider a particle that start at 0 and moves on the lattice $\mathbb{Z}$. In every step, takes a step $+1$ with probability $\frac{1}{2}$ or -1 with probability $\frac{1}{2}$.

More formally: Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = -1] = \frac{1}{2}$ and consider $S_n := \sum_{i=1}^n X_i$.

Then the Central Limit Theorem (2.17) states, that $S_n \approx \mathcal{N}(0,n)$.

Example 2.21. Consider an election with two candidates A and B. The relative number of votes for A is $p \in (0,1)$ (constatn, but unknown) How many ballots do we need to count to make sure that the probability of erring more than 1% is not bigger than 5%?

Each ballot is a vote for A with probability p. We have $S_n \sim \text{Bin}(n,p)$ and we want to find n such that $\mathbb{P}[|S_n - np| \leqslant 0.01n] \leqslant 0.05$. We have that

$$\begin{aligned}
&\mathbb{P}[|S_n - np| \leqslant 0.01n] \\
&= \mathbb{P}[-0.01n \leqslant S_n - np \leqslant 0.01n] \\
&= \mathbb{P}[-\frac{0.01n}{\sqrt{np(1-p)}} \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant \frac{0.01n}{\sqrt{np(1-p)}} \\
&\approx \Phi(0.01\sqrt{\frac{n}{p(1-p)}}) - \Phi(-0.01\sqrt{\frac{n}{p(1-p)}}) \\
&= 2\Phi(0.01\sqrt{\frac{n}{p(1-p)}}) - 1
\end{aligned}$$

Hence, we want $\Phi(0.01\sqrt{\frac{n}{p(1-p)}}) \approx \frac{1.95}{2}$, i.e. $n = (1.96)^2 100^2 p \cdot (1-p)$ We have $p \cdot (1-p) \leqslant \frac{1}{4}$, thus $n \approx (1.96)^2 \cdot 100^2 \cdot \frac{1}{4} = 9600$ suffices.

[Lecture 12, 2023-05-16]

We now want to prove the Central Limit Theorem (2.17). The plan is to do the following:

1. Identify the characteristic function of a standard normal
2. Show that the characteristic functions of the V_n converge pointwise to that of $\mathcal{N}$.
3. Apply Levy's Continuity Theorem (2.14)

First, we need to prove some properties of characteristic functions.

Lemma 2.22. For every real random variable X, we have

(i) $\varphi_X(0) = 1$ and $|\varphi_X(t)| \leqslant 1$ for all $t \in \mathbb{R}$.

(ii) φ_X is uniformly continuous.

(iii) If $\mathbb{E}[|X|^n] < \infty$ for any $n \in \mathbb{N}$, then φ_X i n-times continuously differentiable and $\mathbb{E}[X^n] = (-\mathbf{i})^n \varphi_X^{(n)}(0)$.

(iv) For independent random variables X and Y, we have

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t).$$

Proof of Lemma 2.22. (i) $\varphi_X(0) = \mathbb{E}[e^{\mathbf{i}0X}] = \mathbb{E}[1] = 1$. For $t \in \mathbb{R}$, we have $|\varphi_X(t)| = |\mathbb{E}[e^{\mathbf{i}tX}]| \overset{\text{Jensen}}{\leqslant} \mathbb{E}[|e^{\mathbf{i}tX}|] = 1$.

(ii) Let $t, h \in \mathbb{R}$. Then

$$\begin{aligned} |\varphi_X(t+h) - \varphi_X(t)| &= |\mathbb{E}[e^{\mathbf{i}(t+h)X} - e^{\mathbf{i}tX}]| \\ &= |\mathbb{E}[e^{\mathbf{i}tX}(e^{\mathbf{i}hX} - 1)]| \\ &\overset{\text{Jensen}}{\leqslant} \mathbb{E}[|e^{\mathbf{i}tX}| \cdot |e^{\mathbf{i}hX} - 1|] \\ &= \mathbb{E}[|e^{\mathbf{i}hX} - 1|] =: g(h) \end{aligned}$$

Hence $\sup_{t \in \mathbb{R}} |\varphi_X(t+h) - \varphi_X(t)| \leqslant g(h)$. We show that $\lim_{h \to 0} g(h) = 0$.

For all $\omega \in \Omega$, we realize

$$\lim_{h \to 0} |e^{\mathbf{i}hX(\omega)} - 1| = 0.$$

Thus $|e^{\mathbf{i}hX} - 1| \xrightarrow{h \to 0} 0$ almost surely. Since also for all $h \in \mathbb{R}$ we have $|e^{\mathbf{i}hX} - 1| \leqslant 2$, it follow that $|e^{\mathbf{i}hX} - 1|$ is dominated for all $h \in \mathbb{R}$. Thus, we can apply the dominated convergence theorem and obtain

$$\lim_{h \to 0} g(h) = \lim_{h \to 0} \mathbb{E}[|e^{\mathbf{i}hX} - 1|] = \mathbb{E}[\lim_{h \to 0} |e^{\mathbf{i}hX} - 1|] = 0.$$

It follows that

$$\lim_{n \to 0} \sup_{t \in \mathbb{R}} |\varphi_X(t+h) - \varphi_X(t)| = 0,$$

which means that φ_X is uniformly continuous.

(iii)

Claim 2.22.1. *For $y \in \mathbb{R}$, we have $|e^{\mathbf{i}y} - 1| \leqslant |y|$.*

Subproof. For $y \geqslant 0$, we have

$$\begin{aligned} |e^{\mathbf{i}y} - 1| &= |\int_0^y \cos(s)\, \mathrm{d}s + \mathbf{i} \int_0^y \sin(s)\, \mathrm{d}s| \\ &= |\int_0^y e^{\mathbf{i}s}\, \mathrm{d}s| \\ &\overset{\text{Jensen}}{\leqslant} \int_0^y |e^{\mathbf{i}s}| ds = y. \end{aligned}$$

For $y < 0$, we have $|e^{\mathbf{i}y} - 1| = |e^{-\mathbf{i}y} - 1|$ and we can apply the above to $-y$. ■

First, we look at $n = 1$. Then $\mathbb{E}[|X|] < \infty$. Consider

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathbb{E}\left[e^{\mathbf{i}tX} \frac{e^{\mathbf{i}hX} - 1}{h}\right].$$

We have $e^z = \sum_{n=0}^{\infty} \frac{z^k}{n!}$. Hence

$$\lim_{n\to\infty} e^{\mathbf{i}tX}\left(\frac{1+\mathbf{i}hX+\frac{(\mathbf{i}hX)^2}{2}+o(h^2)-1}{h}\right) = e^{\mathbf{i}tX}\mathbf{i}X \text{ almost surely.}$$

For arbitrary $h \in \mathbb{R}$, we have

$$\begin{aligned} |e^{\mathbf{i}tX}\frac{e^{\mathbf{i}hX}}{h}| &\leqslant \left|\frac{1}{h}\left(e^{\mathbf{i}hX}-1\right)\right| \\ &\overset{(2.22.1)}{\leqslant} \left|\frac{1}{h}\mathbf{i}hX\right| = |X|. \end{aligned}$$

Thus the dominated convergence theorem can be applied and we obtain

$$\lim_{h\to 0}\frac{\varphi_X(t+h)-\varphi_X(t)}{h} = \lim_{h\to 0}\mathbb{E}\left[e^{\mathbf{i}tX}\left(\frac{e^{\mathbf{i}hX}-1}{h}\right)\right] = \mathbb{E}[e^{\mathbf{i}tX}\mathbf{i}X].$$

It follows that φ_X is differentiable and $\varphi_X(t) = \mathbb{E}[e^{\mathbf{i}tX}\mathbf{i}X]$. For $t = 0$ we get $\varphi_X'(0) = \mathbf{i}\mathbb{E}[X]$, i.e. $\text{-}\mathbf{i}\varphi_X'(0) = \mathbb{E}[X]$.

Adapting the proof of (ii) gives that $\varphi_X'(t)$ is continuous.

Adapting the proof of (iii) gives the statement for arbitrary $n \in \mathbb{N}$.

(iv) Easy exercise.

□

Lemma 2.23. For $X \sim \mathcal{N}(0,1)$, we have $\varphi_X(t) = e^{-\frac{t^2}{2}}$ for all $t \in \mathbb{R}$.

Proof of Lemma 2.23. We have

$$\begin{aligned} \varphi_X(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\mathbf{i}tx}e^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} (\cos(tx)+\mathbf{i}\sin(tx))e^{-\frac{x^2}{2}}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \cos(tx)e^{-\frac{x^2}{2}}\,\mathrm{d}x, \end{aligned}$$

since, as $x \mapsto \sin(tx)$ is odd and $x \mapsto e^{-\frac{x^2}{2}}$ is even, their product is odd, wich gives that the integral is 0.

$$
\begin{aligned}
\varphi_X'(t) &= \mathbb{E}[\mathbf{i}Xe^{\mathbf{i}tX}] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbf{i}x\left(\cos(tx) + \mathbf{i}\sin(tx)\right) e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \left(\mathbf{i} \int_{-\infty}^{\infty} x\cos(tx)\right) e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \left(\underbrace{\mathbf{i} \int_{-\infty}^{\infty} x\cos(tx) e^{-\frac{x^2}{2}} \,\mathrm{d}x}_{=0} + \int_{-\infty}^{\infty} -\sin(tx) e^{-\frac{x^2}{2}} \,\mathrm{d}x \right) \\
&= \int_{-\infty}^{\infty} \underbrace{\sin(tx)}_{y(x)} \underbrace{\frac{1}{\sqrt{2\pi}}(-x)e^{\mathbf{i}\frac{x^2}{2}}}_{f'(x)} \,\mathrm{d}x \\
&= \underbrace{[\sin(tx)\frac{1}{\sqrt{2\pi}e^{-\frac{x^2}{2}}}]_{x=-\infty}^{\infty}}_{=0} - \int_{-\infty}^{\infty} t\cos(tx)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= -t\varphi_X(t)
\end{aligned}
$$

Thus, for all $t \in \mathbb{R}$

$$(\log(\varphi_X(t)))' = \frac{\varphi_X'(t)}{\varphi_X(t)} = -t.$$

Hence there exists $c \in \mathbb{R}$, such that

$$\log(\varphi_X(t)) = -\frac{t^2}{2} + c.$$

Since $\varphi_X(0) = 1$, we obtain $c = 0$. Thus

$$\varphi_X(t) = e^{-\frac{t^2}{2}}.$$

□

Now, we can finally prove the Central Limit Theorem (2.17):

Proof of Theorem 2.17. Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}[X_1] = \mu_1$, $\mathrm{Var}(X_1) = \sigma^2$.

Let

$$Y_i := \frac{X_i - \mu}{\sigma}$$

i.e. we normalize to $\mathbb{E}[Y_1] = 0$ and $\mathrm{Var}(Y_1) = 1$. We need to show that

$$V_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{Y_1 + \ldots + Y_n}{\sqrt{n}} \xrightarrow{\omega, n\to\infty} \mathcal{N}(0,1)$$

Let $t \in \mathbb{R}$. Then

$$\begin{aligned}\varphi_{V_n}(t) &= \mathbb{E}[e^{\mathbf{i}tY_n}] \\ &= \mathbb{E}[e^{\mathbf{i}t\left(\frac{Y_1+\ldots+Y_n}{\sqrt{n}}\right)}] \\ &= \mathbb{E}\left[e^{\mathbf{i}t\frac{Y_1}{\sqrt{n}}}\right] \cdot \ldots \cdot \mathbb{E}\left[e^{\mathbf{i}t\frac{Y_n}{\sqrt{n}}}\right] \\ &= \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n.\end{aligned}$$

where $\varphi(t) := \varphi_{Y_1}(t)$.

We have

$$\begin{aligned}\varphi(s) = \varphi(0) + \varphi'(0)s + \frac{\varphi''(0)}{2}s^2 + o(s^2),\ \text{as } s \to 0 \\ = 1 - \underbrace{\mathbf{i}\mathbb{E}[Y_1]s}_{=0} - \mathbb{E}[Y_1^2]\frac{s^2}{2} + o(s^2),\ \text{as } s \to 0 \\ = 1 - \frac{s^2}{2} + o(s^2),\ \text{as } s \to 0\end{aligned}$$

Setting $s := \frac{t}{\sqrt{n}}$ we obtain

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \quad \text{as } n \to \infty$$

$$\varphi_{V_n}(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}},$$

where we have used the following:

Claim 2.17.1. *For a sequence $a_n, n \in \mathbb{N}$ with $\lim_{n\to\infty} na_n = \lambda$, it holds that $\lim_{n\to\infty}(1+a_n)^n = e^{\lambda}$.*

We have shown that

$$\varphi_n(t) \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}} = \varphi_{\mathcal{N}(0,1)}(t).$$

Using Levy's Continuity Theorem (2.14), we obtain the Central Limit Theorem (2.17). □

Remark 2.23.33. If $X : \Omega \to \mathbb{R}^d$ with distribution ν, we define

$$\begin{aligned}\varphi_X : \mathbb{R}^d &\longrightarrow \mathbb{C} \\ t &\longmapsto \mathbb{E}[e^{\mathbf{i}\langle t, X\rangle}]\end{aligned}$$

where $\langle t, X\rangle := \sum_{i=1}^d t_i X_i$.

Exercise: Find out, which properties also hold for $d > 1$. TODO

[Lecture 13, 2023-05]

We have seen, that if $X_1, X_2, \ldots$ are i.i.d. with $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \operatorname{Var}(X_1)$, then $\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{(d)} \mathcal{N}(0,1)$.

Question 2.23.34. What happens if $X_1, X_2, \ldots$ are independent, but not identically distributed? Do we still have a CLT?

Theorem 2.24 (Lindeberg CLT)**.** Assume $X_1, X_2, \ldots,$ are independent (but not necessarily identically distributed) with $\mu_i = \mathbb{E}[X_i] < \infty$ and $\sigma_i^2 = \operatorname{Var}(X_i) < \infty$. Let $S_n = \sqrt{\sum_{i=1}^n \sigma_i^2}$ and assume that

$$\lim_{n\to\infty} \frac{1}{S_n^2} \sum_{i=1}^n \mathbb{E}\left[(X_i - \mu_i)^2 \mathbb{1}_{|X_i - \mu_i| > \varepsilon S_n}\right] = 0$$

for all $\varepsilon > 0$ (**Lindeberg condition**[a]).

Then the CLT holds, i.e.

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{S_n} \xrightarrow{(d)} \mathcal{N}(0,1).$$

[a] "The truncated variance is negligible compared to the variance."

Theorem 2.25 (Lyapunov condition)**.** Let $X_1, X_2, \ldots$ be independent, $\mu_i = \mathbb{E}[X_i] < \infty$, $\sigma_i^2 = \operatorname{Var}(X_i) < \infty$ and $S_n := \sqrt{\sum_{i=1}^n \sigma_i^2}$. Then, assume that, for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{S_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}[(X_i - \mu_i)^{2+\delta}] = 0$$

(**Lyapunov condition**). Then the CLT holds.

Remark 2.25.35. The Lyapunov condition implies the Lindeberg condition. (Exercise).

We will not prove Lindeberg's CLT (2.24) or Lyapunov's CLT (2.25) in this lecture. However, they are quite important.

We will now sketch the proof of Levy's Continuity Theorem (2.14), details can be found in the notes. TODO: copy from official notes

Definition 2.26. Let $(X_n)_n$ be a sequence of random variables. The distribution of $(X_n)_n$ is called **tight** (dt. "straff"), if

$$\lim_{a \to \infty} \sup_{n \in \mathbb{N}} \mathbb{P}[|X_n| > a] = 0.$$

Example[†] 2.26.36 (Exercise 8.1).

Copy

A generalized version of Levy's Continuity Theorem (2.14) is the following:

Theorem 2.27 (A generalized version of Levy's Continuity Theorem (2.14))**.** Suppose we have random variables $(X_n)_n$ such that $\mathbb{E}[e^{\mathbf{i}tX_n}] \xrightarrow{n\to\infty} \varphi(t)$ for all $t \in \mathbb{R}$ for some function φ on $\mathbb{R}$. Then the following are equivalent:

(a) The distribution of X_n is tight.

(b) $X_n \xrightarrow{(d)} X$ for some real-valued random variable X.

(c) φ is the characteristic function of X.

(d) φ is continuous on all of $\mathbb{R}$.

(e) φ is continuous at 0.

Proof of Theorem 2.27 (Exercise 8.2)

Example 2.28. Let $Z \sim \mathcal{N}(0,1)$ and $X_n := nZ$. We have $\varphi_{X_n}(t) = \mathbb{E}[[e^{\mathbf{i}tX_n}] = e^{-\frac{1}{2}t^2n^2} \xrightarrow{n\to\infty} \mathbb{1}_{\{t=0\}}$. $\mathbb{1}_{\{t=0\}}$ is not continuous at 0. By Theorem 2.27, X_n can not converge to a real-valued random variable.

Exercise: $X_n \xrightarrow{(d)} \overline{X}$, where $\mathbb{P}[\overline{X} = \infty] = \frac{1}{2} = \mathbb{P}[\overline{X} = -\infty]$.

Similar examples are $\mu_n := \delta_n$ and $\mu_n := \frac{1}{2}\delta_n + \frac{1}{2}\delta_{-n}$.

Example 2.29. Suppose that $X_1, X_2, \ldots$ are i.d.d. with $\mathbb{E}[X_1] = 0$. Let $\sigma^2 := \operatorname{Var}(X_i)$. Then the distribution of $\frac{S_n}{\sigma\sqrt{n}}$ is tight:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}^2\right)^2\right] &= \frac{1}{n}\mathbb{E}[(X_1 + \ldots + X_n)^2] \\ &= \sigma^2\end{aligned}$$

For $a > 0$, by Chebyshev's Inequality (0.9), we have

$$\mathbb{P}\left[\left|\frac{S_n}{\sqrt{n}}\right| > a\right] \leqslant \frac{\sigma^2}{a^2} \xrightarrow{a\to\infty} 0.$$

verifying Theorem 2.27.

Example 2.30. Suppose C is a random variable which is **Cauchy distributed**, i.e. C has probability distribution $f_C(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

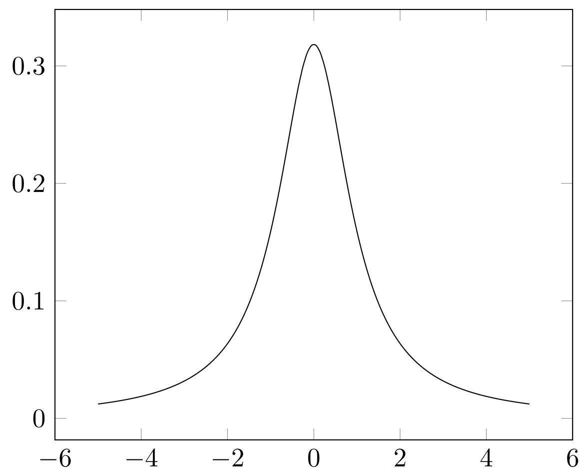


Figure 1: Probability density function of C

We know that $\mathbb{E}[|C|] = \infty$.

We have $\varphi_C(t) = \mathbb{E}[e^{\mathbf{i}tC}] = e^{-|t|}$. Suppose $C_1, C_2, \ldots, C_n$ are i.i.d. Cauchy distributed and let $S_n := C_1 + \ldots + C_n$.

Exercise: $\varphi_{\frac{S_n}{n}}(t) = e^{-|t|} = \varphi_{C_1}(t)$, thus $\frac{S_n}{n} \sim C$.

We will prove Levy's Continuity Theorem (2.14) assuming Theorem 2.13. Theorem 2.13 will be shown in the notes. We will need the following:

TODO: copy from official notes

Lemma 2.31. Given a sequence $(F_n)_n$ of probability distribution functions, there is a subsequence $(F_{n_k})_k$ of F_n and a right continuous, non-decreasing function F, such that $F_{n_k} \to F$ at all continuity points of F. (We do not yet claim, that F is a probability distribution function, as we ignore $\lim\limits_{x\to\infty} F(x)$ and $\lim\limits_{x\to-\infty} F(x)$ for now).

Lemma 2.32. Let $\mu \in M_1(\mathbb{R})$, $A > 0$ and φ the characteristic function of μ. Then $\mu((-A, A)) \geqslant \frac{A}{2}\left|\int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi(t)dt\right| - 1$.

Proof of Lemma 2.32. We have

$$\begin{aligned}\int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi(t)\,\mathrm{d}t &= \int_{-\frac{2}{A}}^{\frac{2}{A}} \int_{\mathbb{R}} e^{\mathbf{i}tx}\mu(\mathrm{d}x)\,\mathrm{d}t \\ &= \int_{\mathbb{R}} \int_{-\frac{2}{A}}^{\frac{2}{A}} e^{\mathbf{i}tx}\,\mathrm{d}t\mu(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \int_{-\frac{2}{A}}^{\frac{2}{A}} \cos(tx)\,\mathrm{d}t\mu(\mathrm{d}x) \\ &= \int_{\mathbb{R}} \frac{2\sin\left(\frac{2x}{A}\right)}{x}\mu(\mathrm{d}x).\end{aligned}$$

Hence

$$\begin{aligned}\frac{A}{2}\left|\int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi(t)\,\mathrm{d}t\right| &= \left|A\int_{\mathbb{R}} \frac{\sin\left(\frac{2x}{A}\right)}{x}\mu(\mathrm{d}t)\right| \\ &= 2\left|\int_{\mathbb{R}} \operatorname{sinc}\left(\frac{2x}{A}\right)\mu(\mathrm{d}t)\right| \\ &\leqslant 2\left[\int_{|x|<A} \underbrace{\left|\operatorname{sinc}\left(\frac{2x}{A}\right)\right|}_{\leqslant 1}\mu(\mathrm{d}x) + \int_{|x|\geqslant A}\left|\operatorname{sinc}\left(\frac{2x}{A}\right)\right|\mu(\mathrm{d}x)\right] \\ &\leqslant 2\left[\mu\left((-A,A)\right) + \frac{A}{2}\int_{|x|\geqslant A}\frac{\sin(2x/A)|}{|x|}\mu(\mathrm{d}x)\right] \\ &\leqslant 2\left[\mu\left((-A,A)\right) + \frac{A}{2}\int_{|x|\geqslant A}\frac{1}{A}\mu(\mathrm{d}x)\right] \\ &\leqslant 2\mu((-A,A)) + \mu((-A,A)^c) \\ &= 1 + \mu((-A,A)).\end{aligned}$$

□

Proof of Theorem 2.14. " $\implies$ " If $\mu_n \implies \mu$, then by definition $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$ for all $f \in C_b$. Since $x \to e^{\mathbf{i}tx}$ is continuous and bounded, it follows that $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$.

" $\impliedby$ "

Claim 2.14.1. *Given $\varepsilon > 0$ there exists $A > 0$ such that $\liminf_n \mu_n\left((-A,A)\right) \geqslant 1 - 2\varepsilon$.*

Proof of Claim 2.14.1. If f is continuous, then

$$\frac{1}{\eta}\int_{x-\eta}^{x+\eta} f(t)dt \xrightarrow{\eta\downarrow 0} f(x).$$

Applying this to φ at $t = 0$, one obtains:

$$\left| \frac{A}{4} \int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi(t) dt - 1 \right| < \frac{\varepsilon}{2} \tag{4}$$

Claim 2.14.1.1. *For n large enough, we have*

$$\left| \frac{A}{4} \int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi_n(t) dt - 1 \right| < \varepsilon. \tag{5}$$

Subproof. Apply dominated convergence. ■

So to prove $\mu_n ((-A, A)) \geqslant 1 - 2\varepsilon$, apply Lemma 2.32. It suffices to show that

$$\frac{A}{2} \left| \int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi_n(t) dt \right| - 1 \geqslant 1 - 2\varepsilon$$

or

$$1 - \frac{A}{4} \left| \int_{-\frac{2}{A}}^{\frac{2}{A}} \varphi_n(t) dt \right| \leqslant \varepsilon,$$

which follows from Equation 5. □

By Lemma 2.31 there exists a right continuous, non-decreasing F and a subsequence $(F_{n_k})_k$ of $(F_n)_n$ where F_n is the probability distribution function of μ_n, such that $F_{n_k}(x) \to F(x)$ for all x where F is continuous.

Claim 2.14.2.

$$\lim_{n \to -\infty} F(x) = 0$$

and

$$\lim_{n \to \infty} F(x) = 1,$$

i.e. F is a probability distribution function.[3]

Subproof. We have

$$\mu_{n_k} ((-\infty, x]) = F_{n_k}(x) \to F(x).$$

Again, given $\varepsilon > 0$, there exists $A > 0$, such that $\mu_{n_k} ((-A, A)) > 1 - 2\varepsilon$ (Claim 2.14.1).

Hence $F(x) \geqslant 1 - 2\varepsilon$ for $x > A$ and $F(x) \leqslant 2\varepsilon$ for $x < -A$. This proves the claim. ■

[3] This does not hold in general!

Since F is a probability distribution function, there exists a probability measure ν on $\mathbb{R}$ such that F is the distribution function of ν. Since $F_{n_k}(x) \to F_n(x)$ at all continuity points x of F, by Theorem 2.13 we obtain that $\mu_{n_k} \overset{k\to\infty}{\implies} \nu$. Hence $\varphi_{\mu_{n_k}}(t) \to \varphi_\nu(t)$, by the other direction of that theorem. But by assumption, $\varphi_{\mu_{n_k}}(\cdot) \to \varphi_n(\cdot)$ so $\varphi_\mu(\cdot) = \varphi_\nu(\cdot)$. By the Uniqueness Theorem (2.3), we get $\mu = \nu$.

We have shown, that $\mu_{n_k} \implies \mu$ along a subsequence. We still need to show that $\mu_n \implies \mu$.

Fact 2.32.37. Suppose a_n is a bounded sequence in $\mathbb{R}$, such that any convergent subsequence converges to $a \in \mathbb{R}$. Then $a_n \to a$.

Assume that μ_n does not converge to μ. By Theorem 2.13, pick a continuity point x_0 of F, such that $F_n(x_0) \not\to F(x_0)$. Pick $\delta > 0$ and a subsequence $F_{n_1}(x_0), F_{n_2}(x_0), \ldots$ which are all outside $(F(x_0) - \delta, F(x_0) + \delta)$. Then $\varphi_{n_1}, \varphi_{n_2}, \ldots \to \varphi$. Now, there exists a further subsequence $G_1, G_2, \ldots$ of F_{n_i}, which converges. $G_1, G_2, \ldots$ is a subsequence of $F_1, F_2, \ldots$. However $G_1, G_2, \ldots$ is not converging to F, as this would fail at x_0. This is a contradiction. □

Proof of Theorem 2.27. □

2.4 Summary

What did we learn:

- How to construct product measures
- WLLN and SLLN
- Kolmogorov's three series theorem
- Fourier transform, weak convergence and CLT

[Lecture 14, 2023-05-25]

3 Conditional Expectation

3.1 Introduction

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two events $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$.

Definition 3.1. The **conditional probability** of A given B is defined as

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Suppose we have two random variables X and Y on Ω, such that X takes distinct values $x_1, x_2, \ldots, x_m$ and Y takes distinct values $y_1, \ldots, y_n$. Then for this case, define the **conditional expectation** of X given $Y = y_j$ as

$$\mathbb{E}[X|Y = y_j] := \sum_{i=1}^{m} x_i \mathbb{P}[X = x_i | Y = y_j].$$

The random variable $Z = \mathbb{E}[X|Y]$ is defined as follows: If $Y(\omega) = y_j$ then

$$Z(\omega) := \underbrace{\mathbb{E}[X|Y = y_j]}_{=: z_j}.$$

Note that $\Omega_j := \{\omega : Y(\omega) = y_j\}$ defines a partition of Ω and on each Ω_j ("the j^{th} Y-atom") Z is constant.

Let $\mathcal{G} := \sigma(Y)$. Then Z is measurable with respect to $\mathcal{G}$. Furthermore

$$\begin{aligned}
\int_{\{Y = y_j\}} Z \, \mathrm{d}\mathbb{P} &= z_j \int_{\{Y = y_j\}} \mathrm{d}\mathbb{P} \\
&= z_j \mathbb{P}[Y = y_j] \\
&= \sum_{i=1}^{m} x_i \mathbb{P}[X = x_i | Y = y_j] \mathbb{P}[Y = y_j] \\
&= \sum_{i=1}^{m} x_i \mathbb{P}[X = x_i, Y = y_j] \\
&= \int_{\{Y = y_j\}} X \, \mathrm{d}\mathbb{P}.
\end{aligned}$$

Hence

$$\int_G Z \, \mathrm{d}\mathbb{P} = \int_G X \, \mathrm{d}\mathbb{P}$$

for all $G \in \mathcal{G}$.

We now want to generalize this to arbitrary random variables.

Theorem 3.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X \in L^1(\mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. Then there exists a random variable Z such that

(a) Z is $\mathcal{G}$-measurable and $Z \in L^1(\mathbb{P})$,

(b) $\int_G Z \, \mathrm{d}\mathbb{P} = \int_G X \, \mathrm{d}\mathbb{P}$ for all $G \in \mathcal{G}$.

Such a Z is unique up to sets of measure 0 and is called the **conditional expectation** of X given the σ-algebra $\mathcal{G}$ and written $Z = \mathbb{E}[X|\mathcal{G}]$.

Remark 3.2.38. Suppose $\mathcal{G} = \{\emptyset, \Omega\}$, then

$$\mathbb{E}[X|\mathcal{G}] = (\omega \mapsto \mathbb{E}[X])$$

is a constant random variable.

Definition 3.3 (Conditional probability)**.** Let $A \subseteq \Omega$ be an event and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. We define the **conditional probability** of A given $\mathcal{G}$ by

$$\mathbb{P}[A|\mathcal{G}] := \mathbb{E}[\mathbb{1}_A|\mathcal{G}].$$

3.2 Existence of Conditional Probability

We will give two different proves of Theorem 3.2. The first one will use orthogonal projections. The second will use the Radon-Nikodym theorem. We'll first do the easy proof, derive some properties and then do the harder proof.

Lemma 3.4. Suppose H is a **Hilbert space**, i.e. H is a vector space with an inner product $\langle \cdot, \cdot \rangle_H$ which defines a norm by $\|x\|_H^2 = \langle x, x \rangle_H$ making H a complete metric space.

For any $x \in H$ and closed, convex subspace $K \subseteq H$, there exists a unique $z \in K$ such that the following equivalent conditions hold:

(a) $\forall y \in K : \langle x - z, y \rangle_H = 0$,

(b) $\forall y \in K : \|z - x\|_H \leqslant \|z - x\|_H$.

Proof. □

TODO: copy from official notes

Proof of Theorem 3.2. Almost sure uniqueness of Z:

Suppose $X \in L^1$ and Z and Z' satisfy (a) and (b). We need to show that $\mathbb{P}[Z \neq Z'] = 0$. By (a), we have $Z, Z' \in L^1(\Omega, \mathcal{G}, \mathbb{P})$. By (b), $\mathbb{E}[(Z - Z')\mathbb{1}_G] = 0$ for all $G \in \mathcal{G}$.

Assume that $\mathbb{P}[Z > Z'] > 0$. Since $\{Z > Z' + \frac{1}{n}\} \uparrow \{Z > Z'\}$, we see that $\mathbb{P}[Z > Z' + \frac{1}{n}] > 0$ for some n. However $\{Z > Z' + \frac{1}{n}\} \in \mathcal{G}$, which is a contradiction, since

$$\mathbb{E}[(Z - Z')\mathbb{1}_{Z - Z' > \frac{1}{n}}] \geqslant \frac{1}{n}\mathbb{P}[Z - Z' > \frac{1}{n}] > 0.$$

Existence of $\mathbb{E}(X|\mathcal{G})$ for $X \in L^2$:

Let $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $K = L^2(\Omega, \mathcal{G}, \mathbb{P})$.

K is closed, since a pointwise limit of $\mathcal{G}$-measurable functions is $\mathcal{G}$ measurable (if it exists). By Lemma 3.4, there exists $z \in K$ such that

$$\mathbb{E}[(X - Z)^2] = \inf\{\mathbb{E}[(X - W)^2] \mid W \in L^2(\mathcal{G})\}$$

and

$$\forall Y \in L^2(\mathcal{G}) : \langle X - Z, Y \rangle = 0. \tag{6}$$

Now, if $G \in \mathcal{G}$, then $Y := \mathbb{1}_G \in L^2(\mathcal{G})$ and by (6) $\mathbb{E}[Z\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$.

Existence of $\mathbb{E}(X|\mathcal{G})$ for $X \in L^1$:

Let $X = X^+ - X^-$. It suffices to show (a) and (b) for X^+. Choose bounded random variables $X_n \geqslant 0$ such that $X_n \uparrow X$. Since each $X_n \in L^2$, we can choose a version Z_n of $\mathbb{E}(X_n|\mathcal{G})$.

Claim 3.2.1. $0 \overset{a.s.}{\leqslant} Z_n \uparrow$.

Subproof. ■

TODO: copy from official notes

Define $Z(\omega) := \limsup_{n\to\infty} Z_n(\omega)$. Then Z is $\mathcal{G}$-measurable and since $Z_n \uparrow Z$, by the Conditional Monotone Converence Theorem (3.10), $\mathbb{E}(Z\mathbb{1}_G) = \mathbb{E}(X\mathbb{1}_G)$ for all $G \in \mathcal{G}$. □

[Lecture 15, 2023-06-06]

3.3 Properties of Conditional Expectation

We want to derive some properties of conditional expectation.

Theorem 3.5 (Law of total expectation)**.**

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X].$$

Proof. Apply (b) from the definition for $G = \Omega \in \mathcal{G}$. □

Theorem 3.6. If X is $\mathcal{G}$-measurable, then $X \overset{\text{a.s.}}{=} \mathbb{E}[X|\mathcal{G}]$.

Proof. Suppose $\mathbb{P}[X \neq Y] > 0$. Without loss of generality $\mathbb{P}[X > Y] > 0$. Hence $\mathbb{P}[X > Y + \frac{1}{n}] > 0$ for some $n \in \mathbb{N}$. Let $A := \{X > Y + \frac{1}{n}\}$. Then

$$\int_A X \,\mathrm{d}\mathbb{P} \geqslant \frac{1}{n}\mathbb{P}(A) + \int_A Y \,\mathrm{d}\mathbb{P},$$

contradicting property (b) from Theorem 3.2. □

Example 3.7. Suppose $X \in L^1(\mathbb{P})$, $\mathcal{G} := \sigma(X)$. Then X is measurable with respect to $\mathcal{G}$. Hence $\mathbb{E}[X|\mathcal{G}] = X$.

Theorem 3.8 (Linearity)**.** For all $a, b \in \mathbb{R}$ we have

$$\mathbb{E}[aX_1 + bX_2|\mathcal{G}] = a\mathbb{E}[X_1|\mathcal{G}] + b\mathbb{E}[X_2|\mathcal{G}].$$

Proof. trivial ☐

add details

Theorem 3.9 (Positivity)**.** If $X \geqslant 0$, then $\mathbb{E}[X|\mathcal{G}] \geqslant 0$ a.s.

Proof. Let W be a version of $\mathbb{E}[X|\mathcal{G}]$. Suppose $\mathbb{P}[W < 0] > 0$. Then

$$G := \{W < -\frac{1}{n}\} \in \mathcal{G}.$$

For some $n \in \mathbb{N}$, we have $\mathbb{P}[G] > 0$. However it follows that

$$\int_G \mathbb{P}[X|\mathcal{G}] \, d\mathbb{P} \leqslant -\frac{1}{n}\mathbb{P}[G] < 0 \leqslant \int_G X \, d\mathbb{P}.$$

☐

Theorem 3.10 (Conditional monotone convergence theorem)**.** Let $X_n, X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $X_n \geqslant 0$ with $X_n \uparrow X$. Then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$.

Proof. Let Z_n be a version of $\mathbb{E}[X_n|Y]$. Since $X_n \geqslant 0$ and $X_n \uparrow$, by the Positivity of Conditional Expectation (3.9), we have

$$\mathbb{E}[X_n|\mathcal{G}] \overset{\text{a.s.}}{\geqslant} 0$$

and

$$\mathbb{E}[X_n|\mathcal{G}] \uparrow \text{a.s.}$$

(consider $X_{n+1} - X_n$).

Define $Z := \limsup_{n\to\infty} Z_n$. Then Z is $\mathcal{G}$-measurable and $Z_n \uparrow Z$ a.s.

Take some $G \in \mathcal{G}$. We know by (b) that $\mathbb{E}[Z_n \mathbb{1}_G] = \mathbb{E}[X_n \mathbb{1}_G]$. The LHS increases to $\mathbb{E}[Z\mathbb{1}_G]$ by the monotone convergence theorem. Again by MCT, $\mathbb{E}[X_n \mathbb{1}_G]$ increases to $\mathbb{E}[X\mathbb{1}_G]$. Hence Z is a version of $\mathbb{E}[X|\mathcal{G}]$. ☐

Theorem 3.11 (Conditional Fatou)**.** Let $X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $X_n \geqslant 0$. Then

$$\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}].$$

Proof. ☐

TODO: copy from official notes

Theorem 3.12 (Conditional dominated convergence theorem)**.** Let $X_n, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that $\sup_n |X_n(\omega)| < Y(\omega)$ a.e. and that X_n converges to a pointwise limit X. Then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ a.e.

Proof. □

TODO: copy from official notes

Recall

Fact 3.12.39 (Jensen's inequality)**.** If $c : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}[|c \circ X|] < \infty$, then $\mathbb{E}[c \circ X] \overset{\text{a.s.}}{\geqslant} c(\mathbb{E}[X])$.

For conditional expectation, we have

Theorem 3.13 (Conditional Jensen's inequality)**.** Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. If $c : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}[|c \circ X|] < \infty$, then $\mathbb{E}[c \circ X|\mathcal{G}] \geqslant c(\mathbb{E}[X|\mathcal{G}])$ a.s.

Fact 3.13.40. If c is convex, then there are two sequences of real numbers $a_n, b_n \in \mathbb{R}$ such that

$$c(x) = \sup_n (a_n x + b_n).$$

Proof of Theorem 3.13. By Fact 3.13.40, $c(x) \geqslant a_n X + b_n$ for all n. Hence

$$\mathbb{E}[c(X)|\mathcal{G}] \geqslant a_n \mathbb{E}[X|\mathcal{G}] + \mathbb{E}[b_n|\mathcal{G}] = a_n \mathbb{E}[X|\mathcal{G}] + b_n \text{ a.s.}$$

for all n. Using that a countable union of sets o f measure zero has measure zero, we conclude that a.s this happens simultaneously for all n. Hence

$$\mathbb{E}[c(X)|\mathcal{G}] \geqslant \sup_n (a_n \mathbb{E}[X|\mathcal{G}] + b_n) \overset{(3.13.40)}{=} c(\mathbb{E}(X|\mathcal{G})).$$

□

Recall

Fact 3.13.41 (Hölder's inequality)**.** Let $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Suppose $X \in L^p(\mathbb{P})$ and $Y \in L^q(\mathbb{P})$. Then

$$\mathbb{E}(XY) \leqslant \underbrace{\mathbb{E}(|X|^p)^{\frac{1}{p}}}_{=:\|X\|_{L^p}} \mathbb{E}(|Y|^q)^{\frac{1}{q}}.$$

Theorem 3.14 (Conditional Hölder's inequality)**.** Let $p, q \geqslant 1$ such that

$\frac{1}{p} + \frac{1}{q} = 1$. Suppose $X \in L^p(\mathbb{P})$ and $Y \in L^q(\mathbb{P})$. Then

$$\mathbb{E}(XY|\mathcal{G}) \leqslant \mathbb{E}(|X|^p|\mathcal{G})^{\frac{1}{p}}\mathbb{E}(|Y|^q|\mathcal{G})^{\frac{1}{q}}.$$

Theorem 3.15 (Tower property)**.** Suppose $\mathcal{F} \supseteq \mathcal{G} \supseteq \mathcal{H}$ are sub-σ-algebras. Then

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}] \mid \mathcal{H}\right] \overset{\text{a.s.}}{=} \mathbb{E}[X|\mathcal{H}].$$

Proof. By definition, $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ is $\mathcal{H}$-measurable. For any $H \in \mathcal{H}$, we have

$$\begin{aligned}\int_H \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\,\mathrm{d}\mathbb{P} &= \int_H \mathbb{E}[X|\mathcal{G}]\,\mathrm{d}\mathbb{P} \\ &= \int_H X\,\mathrm{d}\mathbb{P}.\end{aligned}$$

Hence $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] \overset{\text{a.s.}}{=} \mathbb{E}[X|\mathcal{H}]$. □

Theorem 3.16 (Taking out what is known)**.** If Y is $\mathcal{G}$-measurable and bounded, then

$$\mathbb{E}[YX|\mathcal{G}] \overset{\text{a.s.}}{=} Y\mathbb{E}[X|\mathcal{G}].$$

Proof. Assume w.l.o.g. $X \geqslant 0$. Assume $Y = \mathbb{1}_B$, then Y simple, then take the limit (using that Y is bounded). □ Exercise

Definition 3.17. Let $\mathcal{G}$ and $\mathcal{H}$ be σ-algebras. We call $\mathcal{G}$ and $\mathcal{H}$ **independent**, if $\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}(H)$ for all events $G \in \mathcal{G}$, $H \in \mathcal{H}$.

Theorem 3.18 (Role of independence)**.** Let X be a random variable, and let $\mathcal{G}, \mathcal{H}$ be σ-algebras.

If $\mathcal{H}$ is independent of $\sigma\left(\sigma(X), \mathcal{G}\right)$, then

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] \overset{\text{a.s.}}{=} \mathbb{E}[X|\mathcal{G}].$$

In particular, if X is independent of $\mathcal{G}$, then

$$\mathbb{E}[X|\mathcal{G}] \overset{\text{a.s.}}{=} \mathbb{E}[X].$$

Example 3.19 (Martingale property of the simple random walk)**.** Suppose $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Let $S_n := \sum_{i=1}^n X_i$ be the **simple random walk**. Let $\mathcal{F}$ denote the σ-algebra on the product space. Define $\mathcal{F}_n := \sigma(X_1, \ldots)$. Intuitively, $\mathcal{F}_n$ contains all the information gathered until time n. We have $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \subseteq \dots$

For $\mathbb{E}[S_{n+1}|\mathcal{F}_n]$ we obtain

$$\begin{aligned}\mathbb{E}[S_{n+1}|\mathcal{F}_n] &\overset{\text{Linearity}}{=} \mathbb{E}[S_n|\mathcal{F}_n] + \mathbb{E}[X_{n+1}|\mathcal{F}_n] \\ &\overset{\text{a.s.}}{=} S_n + \mathbb{E}[X_{n+1}|\mathcal{F}_n] \\ &\overset{\text{Independence}}{=} S_n + \mathbb{E}[X_n] \\ &= S_n.\end{aligned}$$

[Lecture 16, 2023-06-13]

Proof of Theorem 3.18. Let $\mathcal{H}$ be independent of $\sigma(\sigma(X), \mathcal{G})$. Then for all $H \in \mathcal{H}$, we have that $\mathbb{1}_H$ and any random variable measurable with respect to either $\sigma(X)$ or $\mathcal{G}$ must be independent.

It suffices to consider the case of $X \geqslant 0$. Let $G \in \mathcal{G}$ and $H \in \mathcal{H}$. By assumption, $X\mathbb{1}_G$ and $\mathbb{1}_H$ are independent. Let $Z := \mathbb{E}[X|\mathcal{G}]$. Then

$$\begin{aligned}\underbrace{\mathbb{E}[X; G \cap H]}_{:= \int_{G \cap H} X \, \mathrm{d}\mathbb{P}} &= \mathbb{E}[(X\mathbb{1}_G)\mathbb{1}_H] \\ &= \mathbb{E}[X\mathbb{1}_G]\mathbb{E}[\mathbb{1}_H] \\ &= \mathbb{E}[Z\mathbb{1}_G]\mathbb{P}(H) \\ &= \mathbb{E}[Z; G \cap H]\end{aligned}$$

The identity above means, that the measures $A \mapsto \mathbb{E}[X; A]$ and $A \mapsto \mathbb{E}[Z; A]$ agree on the σ-algebra $\sigma(\mathcal{G}, \mathcal{H})$ for events of the form $G \cap H$. Since sets of this form generate $\sigma(\mathcal{G}, \mathcal{H})$, these two measures must agree on $\sigma(\mathcal{G}, \mathcal{H})$. The claim of the theorem follows by the uniqueness of conditional expectation.

To deduce the second statement, choose $\mathcal{G} = \{\varnothing, \Omega\}$. □

3.4 The Radon Nikodym Theorem

First, let us recall some basic facts:

Fact 3.19.42. Let $(\Omega, \mathcal{F}, \mu)$ be a **σ-finite measure space**, i.e. Ω can be decomposed into countably many subsets of finite measure. Let $f : \Omega \to [0, \infty)$ be measurable. Define $\nu(A) := \int_A f \, \mathrm{d}\mu$. Then ν is also a σ-finite measure on $(\Omega, \mathcal{F})$. Moreover, ν is finite iff f is integrable.

Application of mct

Note that in this setting, if $\mu(A) = 0$ it follows that $\nu(A) = 0$.

The Radon Nikodym theorem is the converse of that:

Theorem 3.20 (Radon-Nikodym)**.** Let μ and ν be two σ-finite measures on $(\Omega, \mathcal{F})$. Suppose

$$\forall A \in \mathcal{F}.\ \mu(A) = 0 \implies \nu(A) = 0.$$

Then

(1) there exists $Z : \Omega \to [0, \infty)$ measurable, such that

$$\forall A \in \mathcal{F}.\ \nu(A) = \int_A Z \, \mathrm{d}\mu.$$

(2) Such a Z is unique up to equality a.e. (w.r.t. μ).

(3) Z is integrable w.r.t. μ iff ν is a finite measure.

Such a Z is called the **Radon-Nikodym derivative**.

Definition 3.21. Whenever the property $\forall A \in \mathcal{F}, \mu(A) = 0 \implies \nu(A) = 0$ holds for two measures μ and ν, we say that ν is **absolutely continuous** w.r.t. μ. This is written as $\nu \ll \mu$.

Definition† 3.21.43. Two measures μ and ν on a measure space $(\Omega, \mathcal{F})$ are called **singular**, denoted $\mu \perp \nu$, iff there exists $A \in \mathcal{F}$ such that

$$\mu(A) = \nu(A^c) = 0.$$

With the Radon-Nikodym Theorem (3.20) we get a very short proof of the existence of conditional expectation:

Proof (Second proof of Theorem 3.2). Let $(\Omega, \mathcal{F}, \mathbb{P})$ as always, $X \in L^1(\mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$. It suffices to consider the case of $X \geqslant 0$. For all $G \in \mathcal{G}$, define $\nu(G) := \int_G X \, \mathrm{d}\mathbb{P}$. Obviously, $\nu \ll \mathbb{P}$ on $\mathcal{G}$. Then apply the Radon-Nikodym Theorem (3.20). ☐

Proof of Theorem 3.20. We will only sketch the proof. A full proof can be found in the official notes.

Step 1: Uniqueness

TODO: copy from official

Step 2: Reduction to the finite measure case

TODO: copy from official notes

Step 3: Getting hold of Z Assume now that μ and ν are two finite measures. Let

$$\mathcal{C} := \left\{ f : \Omega \to [0, \infty] \middle| \forall A \in \mathcal{F}.\ \int_A f \, \mathrm{d}\mu \leqslant \nu(A) \right\}.$$

We have $\mathcal{C} \neq \varnothing$ since $0 \in \mathcal{C}$. The goal is to find a maximal function Z in $\mathcal{C}$. Obviously its integral will also be maximal.

(a) If $f, g \in \mathcal{C}$, than $f \vee g$ (the pointwise maximum) s also in $\mathcal{C}$.

(b) Suppose $\{f_n\}_{n \geqslant 1}$ is an increasing sequence in $\mathcal{C}$. Let f be the pointwise limit. Then $f \in \mathcal{C}$.

(c) For all $f \in \mathcal{C}$, we have

$$\int_\Omega f \,\mathrm{d}\mu \leqslant \nu(\Omega) < \infty.$$

Define $\alpha := \sup\{\int f \,\mathrm{d}\mu : f \in \mathcal{C}\} \leqslant \nu(\Omega) < \infty$. Let $f_n \in \mathcal{C}, n \in \mathbb{N}$ be a sequence with $\int f_n \,\mathrm{d}\mu \to \alpha$. Define $g_n := \max\{f_1, \ldots, f_n\} \in \mathcal{C}$. Applying (b), we get that the pointwise limit, Z, is an element of $\mathcal{C}$.

Step 4: Showing that our choice of Z works Define $\lambda(A) := \nu(A) - \int_A Z \,\mathrm{d}\mu \geqslant 0$. λ is a measure.

Claim 3.20.1. $\lambda = 0$.

Subproof. Call $G \in \mathcal{F}$ *good* if the following hold:

(i) $\lambda(G) - \frac{1}{k}\mu(G) > 0$.

(ii) $\forall B \subseteq G, B \in \mathcal{F}.\ \lambda(B) - \frac{1}{k}\mu(B) \geqslant 0$.

Suppose we know that for all $A \in \mathcal{F}, k \in \mathbb{N}$ we have $\lambda(A) \leqslant \frac{1}{k}\mu(A)$. Then $\lambda(A) = 0$ since μ is finite.

Assume the claim does not hold. Then there must be some $k \in \mathbb{N}$, $A \in \mathcal{F}$ such that $\lambda(A) - \frac{1}{k}\mu(A) > 0$. Fix this A and k. Then A satisfies condition (i) of being good, but it need not satisfy (ii).

The tricky part is to make A smaller such that it also satisfies (ii). ■

TODO: copy from official notes

□

4 Martingales

4.1 Definition

We have already worked with martingales, but we will define them rigorously now.

> **Definition 4.1** (Filtration)**.** A **filtration** is a sequence $(\mathcal{F}_n)$ of σ-algebras such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \geqslant 1$.

Intuitively, we can think of a $\mathcal{F}_n$ as the set of information we have gathered up to time n. Typically $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for a sequence of random variables.

Definition 4.2. Let $(\mathcal{F}_n)$ be a filtration and $X_1, \ldots, X_n$ be random variables such that $X_i \in L^1(\mathbb{P})$. Then we say that $(X_n)_{n \geqslant 1}$ is an $(\mathcal{F}_n)_n$-**martingale** if the following hold:

- X_n is $\mathcal{F}_n$-measurable for all n.

  (X_n is **adapted to the filtration** $\mathcal{F}_n$).
- $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \overset{\text{a.s.}}{=} X_n$ for all n.

$(X_n)_n$ is called a **submartingale**, if it is adapted to $\mathcal{F}_n$ but

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] \overset{\text{a.s.}}{\geqslant} X_n.$$

It is called a **supermartingale** if it is adapted but $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \overset{\text{a.s.}}{\leqslant} X_n$.

Corollary 4.3. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a convex function such that $f(X_n) \in L^1(\mathbb{P})$. Suppose that $(X_n)_n$ is a martingale[a]. Then $(f(X_n))_n$ is a submartingale. Likewise, if f is concave, then $((f(X_n))_n$ is a supermartingale.

[a] In this form it means, that there is some filtration, that we don't explicitly specify

Proof. Apply Jensen's Inequality (3.13). □

Corollary 4.4. If $(X_n)_n$ is a martingale, then $\mathbb{E}[X_n] = \mathbb{E}[X_0]$.

Example 4.5. The simple random walk:

Let $\xi_1, \xi_2, ..$ iid, $\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = \frac{1}{2}$, $X_n := \xi_1 + \ldots + \xi_n$ and $\mathcal{F}_n := \sigma(\xi_1, \ldots, \xi_n) = \sigma(X_1, \ldots, X_n)$. Then X_n is $\mathcal{F}_n$-measurable. Showing that $(X_n)_n$ is a martingale is left as an exercise.

Example 4.6. See exercise sheet 9.

Copy

[Lecture 17, 2023-06-15]

4.2 Doob's Martingale Convergence Theorem

Definition 4.7 (Stochastic process)**.** A **stochastic process** is a collection of random variables $(X_t)_{t \in T}$ for some index set T. In this lecture we will consider the case $T = \mathbb{N}$.

Definition 4.8 (Previsible process)**.** Consider a filtration $(\mathcal{F}_n)_{n \geqslant 0}$. A stochastic process $(C_n)_{n \geqslant 1}$ is called **previsible**, iff C_n is $\mathcal{F}_{n-1}$-measurable.

Goal. *What about a "gambling strategy"?*

Consider a stochastic process $(X_n)_{n\in\mathbb{N}}$*.*

Note that the increments $X_{n+1} - X_n$ *can be thought of as the win or loss per round of a game. Suppose that there is another stochastic process* $(C_n)_{n\geqslant 1}$ *such that* C_n *is determined by the information gathered up until time* n*, i.e.* C_n *is previsible. Think of* C_n *as our strategy of playing the game. Then* $C_n(X_n - X_{n-1})$ *defines the win in the* n*-th game, while*

$$Y_n := \sum_{j=1}^{n} C_j(X_j - X_{j-1}) \tag{7}$$

defines the cumulative win process.

Lemma 4.9. If $(C_n)_{n\geqslant 1}$ is previsible and $(X_n)_{n\geqslant 0}$ is a martingale and there exists a constant K_n such that $|C_n(\omega)| \leqslant K_n$. Then $(Y_n)_{n\geqslant 1}$ defined in (7) is also a martingale.

Remark 4.9.44. The assumption of K_n being constant can be weakened to $C_n \in L^p(\mathbb{P})$, $X_n \in L^q(\mathbb{P})$ with $\frac{1}{p} + \frac{1}{q} = 1$.

If $C_n \geqslant 0$ the assumption of $(X_n)_{n\geqslant 0}$ being a martingale can be weakened to it being a sub-/supermartingale. Then $(Y_n)_{n\geqslant 1}$ is a sub-/supermartingale as well.

Proof of Lemma 4.9. It is clear that Y_n is $\mathcal{F}_n$-measurable. Suppose that $C_n \in L^p(\mathbb{P})$ and $X_n \in L^q(\mathbb{P})$ for all n. We have

$$\begin{aligned} \|Y_n\|_{L^1} &\leqslant \sum_{i=1}^{n} \|C_i(X_i - X_{i-1})\|_{L^1} \\ &\overset{\text{Hölder}}{\leqslant} \sum_{i=1}^{n} \|C_i\|_{L^p} \|(X_i - X_{i-1})\|_{L^q} \\ &< \infty \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[Y_{n+1} - Y_n|\mathcal{F}_n] &= \mathbb{E}[C_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] \\ &= C_{n+1}(\mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n) \\ &= 0. \end{aligned}$$

□

Suppose we have (X_n) adapted, $X_n \in L^1(\mathbb{P})$, $(C_n)_{n\geqslant 1}$ previsible. We play according to the following principle: Pick two real numbers $a < b$. Wait until $X_n \leqslant a$, then start playing. Stop playing when $X_n \geqslant b$. I.e. define

$$C_1 := 0,$$
$$C_n := \mathbb{1}_{\{C_{n-1}=1\}} \cdot \mathbb{1}_{\{X_{n-1} \leqslant b\}} + \mathbb{1}_{\{C_{n-1}=0\}} \mathbb{1}_{\{X_{n-1} < a\}}. \tag{8}$$

Definition 4.10. Fix $N \in \mathbb{N}$ and let

$$U_N^X([a,b]) := \#\{\text{Upcrossings of } [a,b] \text{ made by } n \mapsto X_n(\omega) \text{ by time } N\},$$

i.e. $U_N([a,b])(\omega)$ is the largest $k \in \mathbb{N}_0$ such that we can find a sequence $0 \leqslant s_1 < t_1 < s_2 < t_2 < \ldots < s_k < t_k \leqslant N$ such that $X_{s_j}(\omega) < a$ and $X_{t_j}(\omega) > b$ for all $1 \leqslant j \leqslant k$.

Clearly $U_N^X([a,b]) \uparrow$ as N increases. It follows that the monotonic limit

$$U_\infty([a,b]) := \lim_{N \to \infty} U_N([a,b])$$

exists pointwise.

Lemma 4.11.

$$\{\omega | \liminf_{N \to \infty} Z_N(\omega) < a < b < \limsup_{N \to \infty} Z_N(\omega)\} \subseteq \{\omega : U_\infty^Z([a,b])(\omega) = \infty\}$$

for every sequence of measurable functions $(Z_n)_{n \geqslant 1}$.

Lemma 4.12. Let $Y_n(\omega) := \sum_{j=1}^n C_j(X_j - X_{j-1})$, where C_n is defined as in (8) Then

$$Y_N \geqslant (b-a)U_N([a,b]) - (X_N - a)^-.$$

Proof. Every upcrossing of $[a,b]$ increases the value of Y by $(b-a)$, while the last interval of play $(X_n - a)^-$ overemphasizes the loss. □

Lemma 4.13. Suppose $(X_n)_n$ is a supermartingale. Then in the above setup

$$(b-a)\mathbb{E}[U_N([a,b])] \leqslant \mathbb{E}[(X_N - a)^-].$$

Proof. Since $C_n \geqslant 0$, by Lemma 4.9 we have that Y_n is a supermartingale. Hence $\mathbb{E}[Y_N] \leqslant \mathbb{E}[Y_1] = 0$. From Lemma 4.12 it follows that

$$(b-a)\mathbb{E}[U_N([a,b])] \leqslant \mathbb{E}[Y_n] + \mathbb{E}[(X_N - a)^-] \leqslant \mathbb{E}[(X_N - a)^-].$$

□

Corollary 4.14. Let $(X_n)_n$ be a **supermartingale bounded in** $L^1(\mathbb{P})$, i.e. $\sup_n \mathbb{E}[|X_n|] < \infty$. Then $(b-a)\mathbb{E}(U_\infty) \leqslant |a| + \sup_n \mathbb{E}(|X_n|)$. In particular, $\mathbb{P}[U_\infty = \infty] = 0$.

Proof. By Lemma 4.13 we have that

$$(b-a)\mathbb{E}[U_N([a,b])] \leqslant \mathbb{E}[|X_N|] + |a| \leqslant \sup_n \mathbb{E}[|X_n|] + |a|.$$

Since $U_N(\cdot) \geqslant 0$ and $U_N(\cdot) \uparrow U_\infty(\cdot)$, by the monotone convergence theorem

$$\mathbb{E}(U_N([a,b])] \uparrow \mathbb{E}[U_\infty([a,b])].$$

□

Let us now consider the case that our process $(X_n)_{n\geqslant 1}$ is a supermartingale bounded in $L^1(\mathbb{P})$.

Theorem 4.15 (Doob's martingale convergence theorem)**.** Any supermartingale bounded in L^1 converges almost surely to a random variable, which is almost surely finite. In particular, any non-negative supermartingale converges a.s. to a finite random variable.

Proof of Theorem 4.15. Let

$$\Lambda := \{\omega | X_n(\omega) \text{ does not converge to anything in } [-\infty, \infty]\}.$$

We have

$$\begin{aligned}\Lambda &= \{\omega | \liminf_N X_N(\omega) < \limsup_N X_N(\omega)\} \\ &= \{\omega | \liminf_N X_N(\omega) < a < b < \limsup_N X_N(\omega)\} \\ &= \bigcup_{a,b\in\mathbb{Q}} \underbrace{\{\omega | \liminf_N X_N(\omega) < a < b < \limsup_N X_N(\omega)\}}_{\Lambda_{a,b}}\end{aligned}$$

We have $\Lambda_{a,b} \subseteq \{\omega : U_\infty([a,b])(\omega) = \infty\}$ by Lemma 4.11. By Lemma 4.13 we have $\mathbb{P}(\Lambda_{a,b}) = 0$, hence $\mathbb{P}(\Lambda) = 0$. Thus there exists a random variable X_∞ such that $X_n \xrightarrow{a.s.} X_\infty$.

Claim 4.15.1. $\mathbb{P}[X_\infty \in \{\pm\infty\}] = 0$.

Subproof. It suffices to show that $\mathbb{E}[|X_\infty|] < \infty$. We have.

$$\begin{aligned}\mathbb{E}[|X_\infty|] &= \mathbb{E}[\liminf_{n\to\infty} |X_n|] \\ &\overset{\text{Fatou}}{\leqslant} \liminf_n \mathbb{E}[|X_n|] \\ &\leqslant \sup_n \mathbb{E}[|X_n|] \\ &< \infty.\end{aligned}$$

■

The second part follows from

Claim 4.15.2. *Any non-negative supermartingale is bounded in L^1.*

Subproof. We need to show $\sup_n \mathbb{E}(|X_n|) < \infty$. Since the supermartingale is non-negative, we have $\mathbb{E}[|X_n|] = \mathbb{E}[X_n]$ and since it is a supermartingale $\mathbb{E}[X_n] \leqslant \mathbb{E}[X_0]$. ■

□

[Lecture 18, 2023-06-20]

Recall our key lemma 4.13 for supermartingales from last time:

$$(b-a)\mathbb{E}[U_N([a,b])] \leqslant \mathbb{E}[(X_n - a)^-].$$

What happens for submartingales? If $(X_n)_{n\in\mathbb{N}}$ is a submartingale, then $(-X_n)_{n\in\mathbb{N}}$ is a supermartingale. Hence the same holds for submartingales, i.e.

Lemma 4.16. A (sub-/super-) martingale bounded in L^1 converges a.s. to a finite limit, which is a.s. finite.

4.3 Doob's L^p Inequality

Question 4.16.45. What about L^p convergence of martingales?

Example 4.17 (A martingale not converging in L^1)**.** Fix $u > 1$ and let $p = \frac{1}{1+u}$. Let $(Z_n)_{n\geqslant 1}$ be i.i.d. ± 1 with $\mathbb{P}[Z_n = 1] = p$.

Let $X_0 = x > 0$ and define $X_{n+1} \coloneqq u^{Z_{n+1}} X_n$.

Then $(X_n)_n$ is a martingale, since

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= X_n\mathbb{E}[u^{Z_{n+1}}]\\ &= X_n\left(p\cdot u + (1-p)\cdot\frac{1}{u}\right)\\ &= X_n\left(\frac{p(u^2-1)+1}{u}\right)\\ &= X_n.\end{aligned}$$

By Doob's Martingale Convergence Theorem (4.15), there exists an a.s. limit X_∞. By the SLLN, we have almost surely

$$\frac{1}{n}\sum_{k=1}^n Z_k \xrightarrow{a.s.} \mathbb{E}[Z_1] = 2p-1.$$

Hence

$$\left(\frac{X_n}{x}\right)^{\frac{1}{n}} = u^{\frac{1}{n}\sum_{k=1}^n Z_k} \xrightarrow{\text{a.s.}} u^{2p-1}.$$

Since $(X_n)_{n\geqslant 0}$ is a martingale, we have $\mathbb{E}[u^{Z_1}] = 1$. Hence $2p-1 < 0$, because $u > 1$. Choose $\varepsilon > 0$ small enough such that $u^{2p-1}(1+\varepsilon) < 1$. Then there exists $N_0(\varepsilon)$ (possibly random) such that for all $n > N_0(\varepsilon)$ almost

$$\left(\frac{X_n}{x}\right)^{\frac{1}{n}} \overset{\text{a.s.}}{\leqslant} u^{2p-1}(1+\varepsilon) \implies x[\underbrace{u^{2p-1}(1+\varepsilon)}_{<1}]^n \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

However, X_n cannot converge to 0 in L^1, as $\mathbb{E}[X_n] = \mathbb{E}[X_0] = x > 0$.

L^2 is nice, since it is a Hilbert space. So we will first consider L^2.

Fact 4.17.46 (Martingale increments are orthogonal in L^2)**.** Let $(X_n)_n$ be a martingale with $X_n \in L^2$ for all n and let $Y_n := X_n - X_{n-1}$ denote the **martingale increments**. Then for all $m \neq n$ we have that

$$\langle Y_m | Y_n\rangle_{L^2} = \mathbb{E}[Y_nY_m] = 0.$$

Proof. As $\mathbb{E}[Y_n^2] = \mathbb{E}[X_n^2] - 2\mathbb{E}[X_nX_{n-1}] + \mathbb{E}[X_{n-1}^2] < \infty$, we have $Y_n \in L^2$. Since $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1}$ a.s., by induction $\mathbb{E}[X_n|\mathcal{F}_k] = X_k$ a.s. for all $k \leqslant n$. In particular $\mathbb{E}[Y_n|\mathcal{F}_k] = 0$ for $k < n$. Suppose that $m < n$. Then

$$\begin{aligned}\mathbb{E}[Y_nY_m] &= \mathbb{E}[\mathbb{E}[Y_nY_m|\mathcal{F}_m]]\\ &= \mathbb{E}[Y_m\mathbb{E}[Y_n|\mathcal{F}_m]]\\ &= 0\end{aligned}$$

□

Fact 4.17.47 (**Parallelogram identity**)**.** Let $X, Y \in L^2$. Then

$$2\mathbb{E}[X^2] + 2\mathbb{E}[Y^2] = \mathbb{E}[(X+Y)^2] + \mathbb{E}[(X-Y)^2].$$

Theorem 4.18. Suppose that $(X_n)_n$ is a martingale bounded in L^2, i.e. $\sup_n \mathbb{E}[X_n^2] < \infty$. Then there is a random variable X_∞ such that

$$X_n \xrightarrow{L^2} X_\infty.$$

Proof. Let $Y_n := X_n - X_{n-1}$ and write

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

We have

$$\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \sum_{j=1}^n \mathbb{E}[Y_j^2]$$

by Fact 4.17.46. In particular,

$$\sup_n \mathbb{E}[X_n^2] < \infty \iff \sum_{j=1}^\infty \mathbb{E}[Y_j^2] < \infty.$$

Since $(X_n)_n$ is bounded in L^2, there exists X_∞ such that $X_n \xrightarrow{\text{a.s.}} X_\infty$ by Doob's Martingale Convergence Theorem (4.15).

It remains to show $X_n \xrightarrow{L^2} X_\infty$. For any $r \in \mathbb{N}$, consider

$$\mathbb{E}[(X_{n+r} - X_n)^2] = \sum_{j=n+1}^{n+r} \mathbb{E}[Y_j^2] \xrightarrow{n\to\infty} 0$$

as a tail of a convergent series.

Hence $(X_n)_n$ is Cauchy, thus it converges in L^2. Since $\mathbb{E}[(X_\infty - X_n)^2]$ converges to the increasing limit

$$\sum_{j \geqslant n+1} \mathbb{E}[Y_j^2] \xrightarrow{n\to\infty} 0$$

we get $\mathbb{E}[(X_\infty - X_n)^2] \xrightarrow{n\to\infty} 0$. □

Now let $p \geqslant 1$ be not necessarily 2. First, we need a very important inequality:

Theorem 4.19 (Doob's L^p inequality)**.** Suppose that $(X_n)_n$ is a martingale or a non-negative submartingale. Let $X_n^* := \max\{|X_1|, |X_2|, \ldots, |X_n|\}$ denote the **running maximum**.

(1) Then

$$\forall \ell > 0.\ \mathbb{P}[X_n^* \geqslant \ell] \leqslant \frac{1}{\ell} \int_{\{X_n^* \geqslant \ell\}} |X_n| \,\mathrm{d}\mathbb{P} \leqslant \frac{1}{\ell} \mathbb{E}[|X_n|].$$

(Doob's L^1 inequality).

(2) Fix $p > 1$. Then

$$\mathbb{E}[(X_n^*)^p] \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

(Doob's L^p inequality).

In order to prove Doob's Martingale Inequalities (4.19), we first need

Lemma 4.20. Let $p > 1$ and X, Y non-negative random variables such that

$$\forall \ell > 0.\ \mathbb{P}[Y \geqslant \ell] \leqslant \frac{1}{\ell} \int_{\{Y \geqslant \ell\}} X \,\mathrm{d}\mathbb{P}$$

Then

$$\mathbb{E}[Y^p] \leqslant \left(\frac{p}{p-1}\right)^p \mathbb{E}[X^p].$$

Proof. First, assume $Y \in L^p$.

Then

$$\begin{aligned}
\|Y\|_{L^p}^p &= \mathbb{E}[Y^p] && (9)\\
&= \int Y(\omega)^p \,\mathrm{d}\mathbb{P}(\omega) && (10)\\
&= \int_\Omega \left(\int_0^{Y(\omega)} p\ell^{p-1} \,\mathrm{d}\ell \right) \mathrm{d}\mathbb{P}(\omega) && (11)\\
&\overset{\text{Fubini}}{=} \int_0^\infty p\ell^{p-1} \underbrace{\int_\Omega \mathbb{1}_{Y \geqslant \ell} \,\mathrm{d}\mathbb{P}}_{\mathbb{P}[Y \geqslant \ell]} \,\mathrm{d}\ell. && (12)
\end{aligned}$$

By the assumption it follows that

$$\begin{aligned}
(12) &\leqslant \int_0^\infty p\ell^{p-2} \int_{\{Y(\omega)\geqslant \ell\}} X(\omega)\mathbb{P}(\mathrm{d}\omega)\,\mathrm{d}\ell \\
&\overset{\text{Fubini}}{=} \int_\Omega X(\omega)\int_0^{Y(\omega)} p\ell^{p-2}\,\mathrm{d}\ell\mathbb{P}(\mathrm{d}\omega) \\
&= \frac{p}{p-1}\int_\omega X(\omega)Y(\omega)^{p-1}\mathbb{P}(\mathrm{d}\omega) \\
&\overset{\text{Hölder}}{\leqslant} \frac{p}{p-1}\|X\|_{L^p}\|Y\|_p^{p-1},
\end{aligned}$$

where the assumption was used to apply Hölder.

Suppose now $Y \notin L^p$. Then look at $Y_M = Y \wedge M$. Apply the above to $Y_M \in L^p$ and use the monotone convergence theorem. □

Proof of Theorem 4.19. Let $E := \{X_n^* \geqslant \ell\} = E_1 \sqcup \ldots \sqcup E_n$ where

$$E_j = \{|X_1| \leqslant \ell, |X_2| \leqslant \ell, \ldots, |X_{j-1}| \leqslant \ell, |X_j| \geqslant \ell\}.$$

Then

$$\mathbb{P}[E_j] \overset{\text{Markov}}{\leqslant} \frac{1}{\ell}\int_{E_j} |X_j|\,\mathrm{d}\mathbb{P} \tag{13}$$

We have that $(|X_n|)_n$ is a submartingale, by Corollary 4.3 in the case of X_n being a martingale and trivially if X_n is non-negative. Hence

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{E_j}(|X_n| - |X_j|)|\mathcal{F}_j] &= \mathbb{1}_{E_j}\mathbb{E}[(|X_n| - |X_j|)|\mathcal{F}_j] \\
&\overset{\text{a.s.}}{\geqslant} 0.
\end{aligned}$$

By the Law of Total Expectation (3.5), it follows that

$$\mathbb{E}[\mathbb{1}_{E_j}(|X_n| - |X_j|)] \geqslant 0. \tag{14}$$

Now

$$\begin{aligned}
\mathbb{P}(E) &= \sum_{j=1}^n \mathbb{P}(E_j) \\
&\overset{(13),(14)}{\leqslant} \frac{1}{\ell}\left(\int_{E_1} |X_n|\,\mathrm{d}\mathbb{P} + \ldots + \int_{E_n} |X_n|\,\mathrm{d}\mathbb{P}\right) \\
&= \frac{1}{\ell}\int_E |X_n|\,\mathrm{d}\mathbb{P}
\end{aligned}$$

This proves the first part.

For the second part, we apply the first part and Lemma 4.20 (choose $Y := X_n^*$). □

Branching process

[Lecture 19, 2023-06-22]

4.4 Uniform Integrability

Example 4.21. Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}$ and $\mathbb{P} = \lambda|_{[0,1]}$. Consider $X_n := n\mathbb{1}_{(0,\frac{1}{n}))}$. We know that $X_n \xrightarrow{n\to\infty} 0$ a.s., however $\mathbb{E}[X_n] = \mathbb{E}[|X_n|] = 1$, hence X_n does not converge in $L^1(\mathbb{P})$.

Let $\mu_n(\cdot) = \mathbb{P}[X_n \in \cdot]$.

Intuitively, for a series that converges in probability, for L^1-convergence to hold we somehow need to make sure that probability measures don't assign mass far away from 0. This will be made precise in the notion of uniform integrability.

Goal. *We want to show that uniform integrability and convergence in probability is equivalent to convergence in* L^1.

Definition 4.22. A sequence of random variables $(X_n)_n$ is called **uniformly integrable** (UI), if

$$\forall \varepsilon > 0.\ \exists K > 0.\ \forall n.\ \mathbb{E}[|X_n|\mathbb{1}_{\{|X_n|>K\}}] < \varepsilon.$$

Similarly, we define uniformly integrable for sets of random variables.

Example 4.23. $X_n := n\mathbb{1}_{(0,\frac{1}{n})}$ is not uniformly integrable.

There is no nice description of uniform integrability. However, some subsets can be easily described, e.g.

Fact 4.23.48. If $(X_n)_{n\geqslant 1}$ is a sequence bounded in $L^{1+\delta}(\mathbb{P})$ for some $\delta > 0$ (i.e. $\sup_n \mathbb{E}[|X_n|^{1+\delta}] < \infty$), then $(X_n)_n$ is uniformly integrable.

Proof. Let $\varepsilon > 0$. Let $p := 1 + \delta > 1$. Choose q such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\mathbb{E}[|X_n|\mathbb{1}_{|X_n|>K}] \leqslant \mathbb{E}[|X_n|^p]^{\frac{1}{p}}\mathbb{P}[|X_n| > k]^{\frac{1}{q}},$$

i.e.

$$\sup_n \mathbb{E}[|X_n|\mathbb{1}_{|X_n|>k}] \leqslant \underbrace{\sup_n \mathbb{E}[|X_n|^p]^{\frac{1}{p}}}_{<\infty} \sup_n \underbrace{\mathbb{P}[|X_n| > K]^{\frac{1}{q}}}_{\leqslant K^{-\frac{1}{q}}\mathbb{E}[|X_n|]^{\frac{1}{q}}}$$

where we have applied Markov's Inequality (0.8).

Since $\sup_n \mathbb{E}[|X_n|^{1+\delta}] < \infty$, we have that $\sup_n \mathbb{E}[|X_n|] < \infty$ by Jensen's Inequality (3.12.39). Hence for K large enough the relevant term is less than ε. $\square$

Fact 4.23.49. If $(X_n)_n$ is uniformly integrable, then $(X_n)_n$ is bounded in L^1.

Proof. Take some $\varepsilon > 0$ and K such that $\sup_n \mathbb{E}[|X_n| \mathbb{1}_{|X_n|>K}] < \varepsilon$. Then $\sup_n \|X_n\|_{L^1} \leqslant K + \varepsilon$. □

Fact 4.23.50. Suppose $Y \in L^1(\mathbb{P})$ and $\sup_n |X_n(\cdot)| \leqslant Y(\cdot)$. Then $(X_n)_n$ is uniformly integrable.

Fact 4.23.51. Let $X \in L^1(\mathbb{P})$.

(a) $\forall \varepsilon > 0.\ \exists \delta > 0.\ \forall F \in \mathcal{F}.\ \mathbb{P}(F) < \delta \implies \int_F |X| \,\mathrm{d}\mathbb{P} < \varepsilon$.

(b) $\forall \varepsilon > 0.\ \exists k \in (0, \infty).\ \int_{|X|>k} |X| \,\mathrm{d}\mathbb{P} < \varepsilon$.

Proof. (a) Suppose not. Then for $\delta = 1, \frac{1}{2}, \frac{1}{2^2}, \ldots$ there exists F_n such that $\mathbb{P}(F_n) < \frac{1}{2^n}$ but $\int_{F_n} |X| \,\mathrm{d}\mathbb{P} \geqslant \varepsilon$.

Since $\sum_n \mathbb{P}(F_n) < \infty$, by Borel-Cantelli (0.10),

$$\mathbb{P}[\underbrace{\limsup_n F_n}_{=:F}] = 0.$$

We have

$$\begin{aligned} \int_F |X| \,\mathrm{d}\mathbb{P} &= \int |X| \mathbb{1}_F \,\mathrm{d}\mathbb{P} \\ &= \int \limsup_n (|X| \mathbb{1}_{F_n}) \,\mathrm{d}\mathbb{P} \\ &\overset{\text{Reverse Fatou}}{\geqslant} \limsup_n \int |X| \mathbb{1}_{F_n} \,\mathrm{d}\mathbb{P} \\ &\geqslant \varepsilon \end{aligned}$$

where the assumption that X is in L^1 was used to apply the reverse of Fatou's lemma.

This yields a contradiction since $\mathbb{P}(F) = 0$.

(b) We want to apply part (a) to $F = \{|X| > k\}$. By Markov's Inequality (0.8), $\mathbb{P}(F) \leqslant \frac{1}{k}\mathbb{E}[|X|]$. Since $\mathbb{E}[|X|] < \infty$, we can choose k large enough to get $\mathbb{P}(F) \leqslant \delta$.

□

Proof of Fact 4.23.50. Fix $\varepsilon > 0$. We have

$$\mathbb{E}[|X_n| \mathbb{1}_{|X_n|>k}] \leqslant \mathbb{E}[|Y| \mathbb{1}_{|Y|>k}] < \varepsilon$$

for k large enough by Fact 4.23.51 (b). $\square$

Fact 4.23.52. Let $X \in L^1(\mathbb{P})$. Then $\mathbb{F} := \{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \subseteq \mathcal{F} \text{ sub-}\sigma\text{-algebra}\}$ is uniformly integrable.

Proof. Fix $\varepsilon > 0$. Choose $\delta > 0$ such that

$$\forall F \in \mathcal{F}.\ \mathbb{P}(F) < \delta \implies \mathbb{E}[|X|\mathbb{1}_F] < \varepsilon. \tag{15}$$

Let $Y = \mathbb{E}[X|\mathcal{G}]$ for some sub-σ-algebra $\mathcal{G}$. Then, by Jensen's Inequality (3.13), $|Y| \leqslant \mathbb{E}[|X||\mathcal{G}]$. Hence $\mathbb{E}[|Y|] \leqslant \mathbb{E}[|X|]$. By Markov's Inequality (0.8), it follows that $\mathbb{P}[|Y| > k] < \delta$ for $k > \frac{\mathbb{E}[|X|]}{\delta}$. Note that $\{|Y| > k\} \in \mathcal{G}$. We have

$$\mathbb{E}[|Y|\mathbb{1}_{\{|Y|>k\}}] < \varepsilon$$

by (15), since $\mathbb{P}[|Y| > k] < \delta$. $\square$

Theorem 4.24. Assume that $X_n \in L^1$ for all n and $X \in L^1$. Then the following are equivalent:

(1) $X_n \to X$ in L^1.

(2) $(X_n)_n$ is uniformly integrable and $X_n \to X$ in probability.

Proof. (2) $\implies$ (1)

Define

$$\varphi(x) := \begin{cases} -k, & x \leqslant -k \\ x, & x \in (-k, k) \\ k, & x \geqslant k. \end{cases}$$

φ is 1-Lipschitz.

We have

$$\int |X_n - X| \, \mathrm{d}\mathbb{P} \leqslant \int |X_n - \varphi(X_n)| \, \mathrm{d}\mathbb{P} + \int |\varphi(X) - X| \, \mathrm{d}\mathbb{P} + \int |\varphi(X_n) - \varphi(X)| \, \mathrm{d}\mathbb{P}$$

We have $\int_{|X_n|>k} \underbrace{|X_n - \varphi(X_n)|}_{\leqslant |X_n| + |\varphi(X_n)| \leqslant 2|X_n|} \mathrm{d}\mathbb{P} \leqslant \varepsilon$ by uniform integrability and Fact 4.23.51 part (b). Similarly $\int_{|X|>k} |X - \varphi(X)| \, \mathrm{d}\mathbb{P} < \varepsilon$.

Since φ is Lipschitz, $X_n \xrightarrow{\mathbb{P}} X \implies \varphi(X_n) \xrightarrow{\mathbb{P}} \varphi(X)$. By the Bounded Convergence Theorem (0.7) $|\varphi(X_n)| \leqslant k \implies \int |\varphi(X_n) - \varphi(X)| \, \mathrm{d}\mathbb{P} \to 0$.

(1) $\implies$ (2)

$X_n \xrightarrow{L^1} X \implies X_n \xrightarrow{\mathbb{P}} X$ by Markov's Inequality (0.8) (see Claim 0.6.4.3).

Fix $\varepsilon > 0$. We have

$$\begin{aligned}\mathbb{E}[|X_n|] &= \mathbb{E}[|X_n - X + X|] \\ &\leqslant \varepsilon + \mathbb{E}[|X|] \\ &< \delta k\end{aligned}$$

for all $\delta > 0$ and suitable k.

Hence $\mathbb{P}[|X_n| > k] < \delta$ by Markov's Inequality (0.8). Then by Fact 4.23.51 part (a) it follows that

$$\int_{|X_n|>k} |X_n| \,\mathrm{d}\mathbb{P} \leqslant \underbrace{\int |X - X_n| \,\mathrm{d}\mathbb{P}}_{<\varepsilon} + \int_{|X_n|>k} |X| \,\mathrm{d}\mathbb{P} \leqslant 2\varepsilon.$$

□

4.5 Martingale Convergence Theorems in $L^p, p \geqslant 1$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ as always and let $(\mathcal{F}_n)_n$ always be a filtration.

Fact 4.24.53. Suppose that $X \in L^p$ for some $p \geqslant 1$.

Then $(\mathbb{E}[X|\mathcal{F}_n])_n$ is an $\mathcal{F}_n$-martingale.

Proof. It is clear that $(\mathbb{E}[X|\mathcal{F}_n])_n$ is adapted to $(\mathcal{F}_n)_n$.

Let $X_n := \mathbb{E}[X|\mathcal{F}_n]$. Consider

$$\begin{aligned}\mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_n] - \mathbb{E}[X|\mathcal{F}_{n-1}]|\mathcal{F}_{n-1}] \\ &= \mathbb{E}[X|\mathcal{F}_{n-1}] - \mathbb{E}[X|\mathcal{F}_{n-1}] \\ &= 0.\end{aligned}$$

□

Theorem 4.25. Let $X \in L^p$ for some $p \geqslant 1$ and $\bigcup_n \mathcal{F}_n \to \mathcal{F} \supseteq \sigma(X)$. Then $X_n := \mathbb{E}[X|\mathcal{F}_n]$ defines a martingale which converges to X in L^p.

Theorem 4.26. Let $p > 1$. Let $(X_n)_n$ be a martingale bounded in L^p. Then there exists a random variable $X \in L^p$, such that $X_n = \mathbb{E}[X|\mathcal{F}_n]$ for all n. In particular, $X_n \xrightarrow{L^p} X$.

[Lecture 20, 2023-06-27]

Proof of Theorem 4.25. By the Tower Property (3.15) it is clear that $(\mathbb{E}[X|\mathcal{F}_n])_n$ is a martingale.

First step: Assume that X is bounded. Then, by Jensen's Inequality (3.13), $|X_n| \leqslant \mathbb{E}[|X||\mathcal{F}_n]$, hence $\sup_{\substack{n\in\mathbb{N}\\ \omega\in\Omega}} |X_n(\omega)| < \infty$. Thus $(X_n)_n$ is a martingale in $L^\infty \subseteq L^2$. By the convergence theorem for martingales in L^2 (Theorem 4.18) there exists a random variable Y, such that $X_n \xrightarrow{L^2} Y$.

Fix $m \in \mathbb{N}$ and $A \in \mathcal{F}_m$. Then

$$\begin{aligned} \int_A Y \,\mathrm{d}\mathbb{P} &= \lim_{n\to\infty} \int_A X_n \,\mathrm{d}\mathbb{P} \\ &= \lim_{n\to\infty} \mathbb{E}[X_n \mathbb{1}_A] \\ &= \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X|\mathcal{F}_n]\mathbb{1}_A] \\ &\overset{A\in\mathcal{F}_n}{=} \lim_{\substack{n\to\infty\\ n\geqslant m}} \mathbb{E}[X\mathbb{1}_A] \end{aligned}$$

Hence $\int_A Y \,\mathrm{d}\mathbb{P} = \int_A X \,\mathrm{d}\mathbb{P}$ for all $m \in \mathbb{N}, A \in \mathcal{F}_m$. Since $\sigma(X) = \bigcup \mathcal{F}_n$ this holds for all $A \in \sigma(X)$. Hence $X = Y$ a.s., so $X_n \xrightarrow{L^2} X$. Since $(X_n)_n$ is uniformly bounded, this also means $X_n \xrightarrow{L^p} X$.

Second step: Now let $X \in L^p$ be general and define

$$X'(\omega) := \begin{cases} X(\omega) & \text{if } |X(\omega)| \leqslant M, \\ 0 & \text{otherwise} \end{cases}$$

for some $M > 0$. Then $X' \in L^\infty$ and

$$\int |X - X'|^p \,\mathrm{d}\mathbb{P} = \int_{\{|X|>M\}} |X|^p \,\mathrm{d}\mathbb{P} \xrightarrow{M\to\infty} 0$$

as $\mathbb{P}$ is regular, i.e. $\forall \varepsilon > 0.\ \exists k.\ \mathbb{P}[|X|^p \in [-k,k]] \geqslant 1 - \varepsilon$.

Take some $\varepsilon > 0$ and M large enough such that

$$\int |X - X'| \,\mathrm{d}\mathbb{P} < \varepsilon.$$

Let $(X'_n)_n$ be the martingale given by $(\mathbb{E}[X'|\mathcal{F}_n])_n$. Then $X'_n \xrightarrow{L^p} X'$ by the first step.

It is

$$\begin{aligned} \|X_n - X'_n\|_{L^p}^p &= \mathbb{E}[\mathbb{E}[X - X'|\mathcal{F}_n]^p] \\ &\overset{\text{Jensen}}{\leqslant} \mathbb{E}[\mathbb{E}[(X - X')^p|\mathcal{F}_n]] \\ &= \|X - X'\|_{L^p}^p \\ &< \varepsilon. \end{aligned}$$

Hence

$$\|X_n - X\|_{L^p} \leqslant \|X_n - X_n'\|_{L^p} + \|X_n' - X'\|_{L^p} + \|X - X'\|_{L^p} \leqslant 3\varepsilon.$$

Thus $X_n \xrightarrow{L^p} X$. ☐

For the proof of Theorem 4.26, we need the following theorem, which we won't prove here:

Theorem 4.27 (Banach Alaoglu)**.** Let X be a normed vector space and X^* its continuous dual. Then the closed unit ball in X^* is compact w.r.t. the weak* topology.

Fact 4.27.54. We have $L^p \cong (L^q)^*$ for $\frac{1}{p} + \frac{1}{q} = 1$ via

$$\begin{aligned} L^p &\longrightarrow (L^q)^* \\ f &\longmapsto (g \mapsto \int gf \, \mathrm{d}\mathbb{P}) \end{aligned}$$

We also have $(L^1)^* \cong L^\infty$, however $(L^\infty)^* \ncong L^1$.

Proof of Theorem 4.26. Since $(X_n)_n$ is bounded in L^p, by Banach Alaoglu (4.27), there exists $X \in L^p$ and a subsequence $(X_{n_k})_k$ such that for all $Y \in L^q$, where as always $\frac{1}{p} + \frac{1}{q} = 1$,

$$\int X_{n_k} Y \, \mathrm{d}\mathbb{P} \to \int XY \, \mathrm{d}\mathbb{P}$$

(Note that this argument does not work for $p = 1$, because $(L^\infty)^* \ncong L^1$).

Let $A \in \mathcal{F}_m$ for some fixed m and choose $Y = \mathbb{1}_A$. Then

$$\begin{aligned} \int_A X \, \mathrm{d}\mathbb{P} &= \lim_{k\to\infty} \int_A X_{n_k} \, \mathrm{d}\mathbb{P} \\ &= \lim_{k\to\infty} \mathbb{E}[X_{n_k} \mathbb{1}_A] \\ &\overset{\text{for } n_k \geqslant m}{=} \mathbb{E}[X_m \mathbb{1}_A]. \end{aligned}$$

Hence $X_n = \mathbb{E}[X|\mathcal{F}_m]$ by the uniqueness of conditional expectation and by Theorem 4.25, we get the convergence. ☐

Example† 4.27.55 (**Branching Process**; Exercise 10.1, 12.4)**.** Let $(Y_{n,k})_{n\in\mathbb{N}_0, k\in\mathbb{N}}$ be i.i.d. with values in $\mathbb{N}_0$ such that $0 < \mathbb{E}[Y_{n,k}] = m < \infty$. Define

$$S_0 := 1, S_n := \sum_{k=1}^{S_{n-1}} Y_{n-1,k}$$

and let $M_n := \frac{S_n}{m^n}$. S_n models the size of a population.

Claim 3. *M_n is a martingale.*

Subproof. We have

$$\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n] = \frac{1}{m^n}\left(\frac{1}{m}\sum_{k=1}^{S_n}\mathbb{E}[X_{n,k}] - S_n\right) = \frac{1}{m^n}(S_n - S_n).$$

■

Claim 4. *$(M_n)_{n\in\mathbb{N}}$ is bounded in L^2 iff $m > 1$.*

TODO

Claim 5. *If $m > 1$ and $M_n \to M_\infty$, then*

$$\operatorname{Var}(M_\infty) = \sigma^2(m(m-1))^{-1}.$$

TODO

4.6 Stopping Times

Definition 4.28 (Stopping time)**.** A random variable $T : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_n, \mathbb{P})$ is called a **stopping time**, if

$$\{T \leqslant n\} \in \mathcal{F}_n$$

for all $n \in \mathbb{N}$. Equivalently, $\{T = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

Example 4.29. A constant random variable $T = c$ is a stopping time.

Example 4.30 (Hitting times)**.** For an adapted process $(X_n)_n$ with values in $\mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, the **hitting time**

$$T := \inf\{n \in \mathbb{N} : X_n \in A\}$$

is a stopping time, as

$$\{T \leqslant n\} = \bigcup_{k=1}^{n}\{X_k \in A\} \in \mathcal{F}_n.$$

However, the last exit time

$$T := \sup\{n \in \mathbb{N} : X_n \in A\}$$

is not a stopping time.

Example 4.31. Consider the simple random walk, i.e. X_n i.i.d. with $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = \frac{1}{2}$. Set $S_n := \sum_{i=1}^{n} X_n$. Then

$$T := \inf\{n \in \mathbb{N} : S_n \geqslant A \vee S_n \leqslant B\}$$

is a stopping time.

Fact 4.31.56. If T_1, T_2 are stopping times with respect to the same filtration, then

- $T_1 + T_2$,
- $\min\{T_1, T_2\}$ and
- $\max\{T_1, T_2\}$

are stopping times.

Warning 4.32. Note that $T_1 - T_2$ is not a stopping time.

Remark 4.32.57. There are two ways to look at the interaction between a stopping time T and a stochastic process $(X_n)_n$:

- The behaviour of X_n until T, i.e.

  $$X^T := (X_{T \wedge n})_{n \in \mathbb{N}}$$

  is called the **stopped process**.
- The value of $(X_n)_n)$ at time T, i.e. looking at X_T.

Example 4.33. If we look at a process

$$S_n = \sum_{i=1}^{n} X_i$$

for some $(X_n)_n$, then

$$S^T = (\sum_{i=1}^{T \wedge n} X_i)_n$$

and

$$S_T = \sum_{i=1}^{T} X_i.$$

Theorem 4.34. If $(X_n)_n$ is a supermartingale and T is a stopping time, then X^T is also a supermartingale, and we have $\mathbb{E}[X_{T\wedge n}] \leqslant \mathbb{E}[X_0]$ for all n. If $(X_n)_n$ is a martingale, then so is X^T and $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$.

Proof. First, we need to show that X^T is adapted. This is clear since

$$\begin{aligned} X_n^T &= X_T \mathbb{1}_{T<n} + X_n \mathbb{1}_{T \geqslant n} \\ &= \sum_{k=1}^{n-1} X_k \mathbb{1}_{T=k} + X_n \mathbb{1}_{T\geqslant n}. \end{aligned}$$

It is also clear that X_n^T is integrable since

$$\mathbb{E}[|X_n^T|] \leqslant \sum_{k=1}^{n} \mathbb{E}[|X_k|] < \infty.$$

We have

$$\begin{aligned} &\mathbb{E}[X_n^T - X_{n-1}^T | \mathcal{F}_{n-1}] \\ &= \mathbb{E}\left[X_n \mathbb{1}_{\{T\geqslant n\}} + \sum_{k=1}^{n-1} X_k \mathbb{1}_{\{T=k\}} - X_{n-1}(\mathbb{1}_{T\geqslant n} + \mathbb{1}_{\{T=n-1\}}) \right. \\ &\quad \left. + \sum_{k=1}^{n-2} X_k \mathbb{1}_{\{T=k\}} \middle| \mathcal{F}_{n-1} \right] \\ &= \mathbb{E}[(X_n - X_{n-1}) \mathbb{1}_{\{T\geqslant n\}} | \mathcal{F}_{n-1}] \\ &= \mathbb{1}_{\{T\geqslant n\}} (\mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}) \begin{cases} \leqslant 0 \\ = 0 \text{ if } (X_n)_n \text{ is a martingale.} \end{cases} \end{aligned}$$

□

Remark 4.34.58. We now want a similar statement for X_T. In the case that $T \leqslant M$ is bounded, we get from the above that

$$\mathbb{E}[X_T] \overset{n\geqslant M}{=} \mathbb{E}[X_n^T] \begin{cases} \leqslant \mathbb{E}[X_0] & \text{supermartingale,} \\ = \mathbb{E}[X_0] & \text{martingale.} \end{cases}$$

However if T is not bounded, this does not hold in general.

Example 4.35. Let $(S_n)_n$ be the simple random walk and take $T := \inf\{n : S_n = 1\}$. Then $\mathbb{P}[T < \infty] = 1$, but

$$1 = \mathbb{E}[S_T] \neq \mathbb{E}[S_0] = 0.$$

Theorem 4.36 (Optional Stopping)**.** Let $(X_n)_n$ be a supermartingale and let T be a stopping time taking values in $\mathbb{N}$.

If one of the following holds

(i) $T \leqslant M$ is bounded,

(ii) $(X_n)_n$ is uniformly bounded and $T < \infty$ a.s.,

(iii) $\mathbb{E}[T] < \infty$ and $|X_n(\omega) - X_{n-1}(\omega)| \leqslant K$ for all $n \in \mathbb{N}, \omega \in \Omega$ and some $K > 0$,

then $\mathbb{E}[X_T] \leqslant \mathbb{E}[X_0]$.

If $(X_n)_n$ even is a martingale, then under the same conditions $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Proof. (i) was already done in Remark 4.34.58.

(ii): Since $(X_n)_n$ is bounded, we get that

$$\begin{aligned}\mathbb{E}[|X_T - X_0|] &\overset{\text{dominated convergence}}{=} \lim_{n \to \infty} \mathbb{E}[|X_{T \wedge n} - X_0|] \\ &\overset{\text{part (i)}}{\leqslant} 0.\end{aligned}$$

(iii): It is

$$\begin{aligned}|X_{T \wedge n} - X_0| &\leqslant |\sum_{k=1}^{T \wedge n} X_k - X_{k-1}| \\ &\leqslant (T \wedge n) \cdot K \\ &\leqslant T \cdot K < \infty.\end{aligned}$$

Hence, we can apply dominated convergence and obtain

$$\mathbb{E}[X_T - X_0] = \lim_{n \to \infty} \mathbb{E}[X_{T \wedge n} - X_0].$$

Thus, we can apply (ii).

The statement about martingales follows from applying this to $(X_n)_n$ and $(-X_n)_n$, which are both supermartingales. □

Remark† 4.36.59. Let $(X_n)_n$ be a supermartingale and T a stopping time. If $(X_n)_n$ itself is not bounded, but T ensures boundedness, i.e. $T < \infty$ a.s. and $(X_{T \wedge n})_n$ is uniformly bounded, the Optional Stopping

Theorem (4.36) can still be applied, as

$$\mathbb{E}[X_T] = \mathbb{E}[X_{T\wedge T}] \overset{\text{Optional Stopping}}{\leqslant} \mathbb{E}[X_{T\wedge 0}] = \mathbb{E}[X_0].$$

[Lecture 21, 2023-06-29]

4.7 An Application of the Optional Stopping Theorem

This is the last lecture relevant for the exam. (Apart from lecture 22 which will be a repetion).

Goal. *We want to see an application of the 4.36.*

Notation 4.36.60. Let E be a complete, separable metric space (e.g. $E = \mathbb{R}$). Suppose that for all $x \in E$ we have a probability measure $\mathbf{P}(x, \mathrm{d}y)$ on E. Such a probability measure is a called a **transition probability measure**.

Example 4.37. $E = \mathbb{R}$,

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

is a transition probability measure.

Example 4.38 (Simple random walk as a transition probability measure)**.** $E = \mathbb{Z}$, $\mathbf{P}(x, \mathrm{d}y)$ assigns mass $\frac{1}{2}$ to $y = x + 1$ and $y = x - 1$.

Definition 4.39. For every bounded, measurable function $f : E \to \mathbb{R}$, $x \in E$ define

$$(\mathbf{P}f)(x) \coloneqq \int_E f(y)\mathbf{P}(x, \mathrm{d}y).$$

This $\mathbf{P}$ is called a **transition operator**.

Fact 4.39.61. If $f \geqslant 0$, then $(\mathbf{P}f)(\cdot) \geqslant 0$.

If $f \equiv 1$, we have $(\mathbf{P}f) \equiv 1$.

Notation 4.39.62. Let $\mathbf{I}$ denote the **identity operator**, i.e.

$$(\mathbf{I}f)(x) = f(x)$$

for all f. Then for a transition operator $\mathbf{P}$ we write

$$\mathbf{L} \coloneqq \mathbf{I} - \mathbf{P}.$$

Goal. *Take $E = \mathbb{R}$. Suppose that $A^c \subseteq \mathbb{R}$ is a bounded domain. Given a bounded function f on $\mathbb{R}$, we want a function u which is bounded, such that $\mathbf{L}u = 0$ on A^c and $u = f$ on A.*

We will show that $u(x) = \mathbb{E}_x[f(X_{T_A})]$ is the unique solution to this problem.

Definition 4.40. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_n, \mathbb{P}_x)$ be a filtered probability space, where for every $x \in \mathbb{R}$, $\mathbb{P}_x$ is a probability measure. Let $\mathbb{E}_x$ denote expectation with respect to $\mathbf{P}(x, \cdot)$. Then $(X_n)_{n\geqslant 0}$ is a **Markov chain** starting at $x \in \mathbb{R}$ with **transition probability** $\mathbf{P}(x, \cdot)$ if

(i) $\mathbb{P}_x[X_0 = x] = 1$,

(ii) for all bounded, measurable $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}_x[f(X_{n+1})|\mathcal{F}_n] \overset{\text{a.s.}}{=} \mathbb{E}_x[f(X_{n+1})|X_n] = \int f(y)\mathbf{P}(X_n, \mathrm{d}y).$$

(Recall $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.)

Example 4.41. Suppose $B \in \mathcal{B}(\mathbb{R})$ and $f = \mathbb{1}_B$. Then the first equality of (ii) simplifies to

$$\mathbb{P}_x[X_{n+1} \in B|\mathcal{F}_n] = \mathbb{P}_x[X_{n+1} \in B|\sigma(X_n)].$$

Example 4.42. Let ξ_i be i.i.d. with$\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = \frac{1}{2}$ and define $X_n := \sum_{i=1}^n \xi_i$.

Intuitively, conditioned on X_n, X_{n+1} should be independent of $\sigma(X_1, \ldots, X_{n-1})$.

Claim. *For a set B, we have*

$$\mathbb{E}[\mathbb{1}_{X_{n+1}\in B}|\sigma(X_1, \ldots, X_n)] = \mathbb{E}[\mathbb{1}_{X_{n+1}\in B}|\sigma(X_n)].$$

Subproof. ■ TODO

New information after this point is not relevant for the exam.

Stopping times and optional stopping are very relevant for the exam, the Markov property is not. No notes will be allowed in the exam. Theorems from the lecture as well as homework exercises might be part of the exam.

5 Markov Chains

Merge this with the end of lecture 21

Goal. *We want to start with the basics of the theory of Markov chains.*

Example 5.1 (Markov chains with two states)**.** Suppose there are two states of a phone line, 0,"free", or 1, "busy". We assume that the state only changes at discrete units of time and model this as a sequence of random variables. Assume

$$\begin{aligned}\mathbb{P}[X_{n+1} = 0 | X_n = 0] &= p\\ \mathbb{P}[X_{n+1} = 0 | X_n = 1] &= (1-p)\\ \mathbb{P}[X_{n+1} = 1 | X_n = 0] &= q\\ \mathbb{P}[X_{n+1} = 1 | X_n = 1] &= (1-q)\end{aligned}$$

for some $p, q \in (0,1)$. We can write this as a matrix

$$P = \begin{pmatrix} p & (1-p) \\ q & (1-q) \end{pmatrix}$$

Note that the rows of this matrix sum up to 1.

Additionally, we make the following assmption: Given that at some time n, the phone is in state $i \in \{0,1\}$, the behavior of the phone after time n does not depend on the way, the phone reached state i.

Question 5.1.63. Suppose $X_0 = 0$. What is the probability, that the phone will be free at times 1&2 and will become busy at time 3, i.e. what is $\mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 1]$?

We have

$$\begin{aligned}&\mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 1]\\ &= \mathbb{P}[X_3 = 0 | X_2 = 0, X_1 = 0]\mathbb{P}[X_2 = 0, X_1 = 0]\\ &= \mathbb{P}[X_3 = 0 | X_2 = 0]\mathbb{P}[X_2 = 0, X_1 = 0]\\ &= \mathbb{P}[X_3 = 0 | X_2 = 0]\mathbb{P}[X_2 = 0 | X_1 = 0]\mathbb{P}[X_1 = 0]\\ &= P_{0,1}P_{0,0}P_{0,0}\end{aligned}$$

Question 5.1.64. Assume $X_0 = 0$. What is $\mathbb{P}[X_3 = 1]$?

For $\{X_3 = 1\}$ to happen, we need to look at the following disjoint events:

$$\mathbb{P}(\{X_3 = 1, X_2 = 0, X_1 = 0\}) = P_{0,1}P_{0,0}^2,$$
$$\mathbb{P}(\{X_3 = 1, X_2 = 0, X_1 = 1\}) = P_{0,1}^2 P_{1,0},$$
$$\mathbb{P}(\{X_3 = 1, X_2 = 1, X_1 = 0\}) = P_{0,0}P_{0,1}P_{1,1},$$
$$\mathbb{P}(\{X_3 = 1, X_2 = 1, X_1 = 1\}) = P_{0,1}P_{1,1}^2.$$

More generally, consider a Matrix $P \in (0,1)^{n\times n}$ whose rows sum up to 1. Then we get a Markov Chain with n states by defining

$$\mathbb{P}[X_{n+1} = i | X_n = j] = P_{i,j}.$$

Definition 5.2. Let E denote a **discrete state space**, usually $E = \{1, \ldots, N\}$ or $E = \mathbb{N}$ or $E = \mathbb{Z}$.

Let α be a probability measure on E. We say that $(p_{i,j})_{i\in E, j\in E}$ is a **transition probability matrix**, if

$$\forall i, j \in E.\ p_{i,j} \geqslant 0 \wedge \forall i \in E \sum_{j\in E} p_{i,j} = 1.$$

Given a triplet (E, α, P), we say that a stochastic process $(X_n)_{n\geqslant 0}$, i.e. $X_n : \Omega \to E$, is a **Markov chain taking values on the state space E with initial distribution α and transition probability matrix P**, if the following conditions hold:

(i) $\mathbb{P}[X_0 = i] = \alpha(i)$ for all $i \in E$,

(ii)

$$\begin{aligned}&\mathbb{P}[X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n]\\ &= \mathbb{P}[X_{n+1} = i_{n+1} | X_n = i_n]\end{aligned}$$

for all $n = 0, \ldots, i_0, \ldots, i_{n+1} \in E$ (provided $\mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \neq 0$).

Fact 5.2.65. For all $n \in \mathbb{N}_0$ and $i_0, \ldots, i_n \in E$, we have

$$\mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = \alpha(i_0) \cdot p_{i_0,i_1} \cdot p_{i_1,i_2} \cdot \ldots \cdot p_{i_{n-1},i_n}.$$

Fact 5.2.66. For all $n \in \mathbb{N}$, $i_n \in E$, we have

$$\mathbb{P}[X_n = i_n] = \sum_{i_0,\ldots,i_{n-1}\in E} \alpha_{i_0} p_{i_0,i_1} \cdot \ldots \cdot p_{i_{n-1},i_n}.$$

Example 5.3 (Simple random walk on $\mathbb{Z}$)**.** Let $E := \mathbb{Z}$, $(\xi_n)_n$ i.i.d. with $\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = \frac{1}{2}$. Let $X_0 = 0, X_n = \xi_1 + \ldots + \xi_n$.

Let $\alpha = \delta_0 \in M_1(\mathbb{Z})$. Consider

$$P = \begin{pmatrix} & \ddots & \ddots & \ddots & & & & & 0 \\ \ldots & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \ldots & & \\ & \ldots & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \ldots & \\ & & \ldots & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \ldots \\ 0 & & & & & \ddots & \ddots & \ddots & \end{pmatrix}$$

Definition 5.4. Let E be a complete, separable metric space, $\alpha \in M_1(E)$. For every $x \in E$, let $\mathbf{P}(x, \cdot)$ be a probability measure on E.[a]

Given the triples $(E, \alpha, \{\mathbf{P}(x, \cdot)\}_{x \in E})$, we say that a stochastic process $(X_n)_{n \geqslant 0}$ is a **Markov chain taking values on E with starting distribution α and transition probability $\{\mathbf{P}(x, \cdot)\}_{x \in E}$** if

(i) $\mathbb{P}[X_0 \in \cdot] = \alpha(\cdot)$,

(ii) For all bounded, measurable $f : E \to \mathbb{R}$,

$$\mathbb{E}[f(X_{n+1}) | \mathcal{F}_n] = \mathbb{E}[f(X_{n+1}) | X_n] = \int_E f(y) \mathbf{P}(X_n, \mathrm{d}y) \text{ a.s.}$$

[a] $\mathbf{P}(x, \cdot)$ corresponds to a row of our matrix in the discrete case

Remark 5.4.67. This agrees with the definition in the discrete case, as all bounded, measurable $f : E \to \mathbb{R}$ can be approximated by simple functions, i.e. (ii) from the discrete case implies (ii) from the general definition.

Notation 5.4.68. If $\{\mathbf{P}(x, \cdot)\}_{x \in E}$ is a transition probability, then for all $f : E \to \mathbb{R}$ bounded and measurable, define $\mathbf{P} : \mathcal{B}_{\text{bdd}}(E) \to \mathcal{B}_{\text{bdd}}$ by

$$(\mathbf{P}f)(x) := \int_E f(y) \mathbf{P}(x, \mathrm{d}y).$$

We get the following fundamental link between martingales and Markov chains:

Theorem 5.5. Suppose $(E, \alpha, \{\mathbf{P}(x, \cdot)\}_{x \in E})$ is given. Then a stochastic process $(X_n)_{n \geqslant 0}$ is a Markov chain iff for every $f : E \to \mathbb{R}$ bounded, measurable,

$$M_n(f) := f(X_n) - f(X_0) - \sum_{j=1}^{n-1} (\mathbf{I} - \mathbf{P}) f(X_j)$$

is a martingale with respect to the canonical filtration of (X_n).

Proof. $\implies$ Fix some bounded, measurable $f : E \to \mathbb{R}$. Then, for all n, $M_n(f)$ is bounded and hence $M_n(f) \in L^1$. $M_n(f)$ is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$.

In order to prove $\mathbb{E}[M_{n+1}(f)|\mathcal{F}_n] = M_n(f)$, it suffices to show $\mathbb{E}[M_{n+1}(f) - M_n(f)|\mathcal{F}_n] = 0$ a.s.

We have

$$\begin{aligned} \mathbb{E}[M_{n+1}(f) - M_n(f)|\mathcal{F}_n] &= \mathbb{E}[f(X_{n+1}|\mathcal{F}_n] - (\mathbf{P}f)(X_n) \\ &\overset{\text{Markov property}}{=} (\mathbf{P}f)(X_n) - (\mathbf{P}f)(X_n) \\ &= 0 \end{aligned}$$

$\impliedby$ Suppose $(M_n(f))_n$ is a martingale for all bounded, measurable f. By the martingale property, we have

$$\begin{aligned} \mathbb{E}[f(X_{n+1})|X_n] &= (\mathbf{P}f)(X_n) \\ &= \int f(y)\mathbf{P}(X_n, \mathrm{d}y) \end{aligned}$$

This proves (ii). ☐

Definition 5.6. Given $\{\mathbf{P}(x,\cdot)\}_{x\in E}$, we say that $f : E \to \mathbb{R}$ is **harmonic**, iff $f(x) = (\mathbf{P}f)(x)$ for all $x \in E$. We call f **super-harmonic**, if $(\mathbf{I}-\mathbf{P})f \geqslant 0$ and **sub-harmonic**, if $(\mathbf{I} - \mathbf{P})f \leqslant 0$.

Corollary 5.7. If f is (sub/super) harmonic, then for every $(E, \{\mathbf{P}(x,\cdot)\}_{x\in E}, \alpha)$ and every Markov chain $(X_n)_{n\geqslant 0}$, we have that $f(X_n)$ is a (sub/super) martingale.

Question 5.7.69. Given a set A and a function f on a superset of A. Find a function u, such that u is harmonic, and $u = f$ on A.

Let $u(x) := \mathbb{E}_x[f(X_{T_A}]$, where $\mathbb{E}_x$ is the expectation with respect to the Markov chain starting in x, and T_A is the stopping time defined by the Markov chain hitting A.

6 Appendix

6.1 List of Distributions

	Symbol	Mass (PMF)	Distribution (CDF)	$\mathbb{E}$	Var	$\varphi_X(t) = \mathbb{E}[e^{\mathbf{i}tX}]$	$M_X(t) = \mathbb{E}[e^{tX}]$
Deterministic	δ_a	$\mathbb{1}_{x=a}$	$\mathbb{1}_{[a,\infty)}$	a	0	$e^{\mathbf{i}ta}$	e^{ta}
Bernoulli	$\mathrm{Bin}(1,p)$						
Binomial	$\mathrm{Bin}(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	$\sum_{j=0}^{\lfloor x \rfloor}\binom{n}{j}p^j(1-p)^{n-j}$	np	$np(1-p)$	$((1-p)+pe^{\mathbf{i}t})^n$	$((1-p)+pe^t)^n$
Geometric	$\mathrm{Geo}(p)$	$(1-p)^{k-1}p$	$1-(1-p)^{\lfloor x \rfloor}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^{\mathbf{i}t}}{1-(1-p)e^{\mathbf{i}t}}$	$\frac{pe^t}{1-(1-p)e^t}$
Poisson	$\mathrm{Poi}(\lambda)$	$\frac{\lambda^k e^{-\lambda}}{k!}$	$e^{-\lambda}\sum_{j=0}^{\lfloor x \rfloor}\frac{\lambda^j}{j!}$	λ	λ	$e^{\lambda(e^{\mathbf{i}t}-1)}$	$e^{\lambda(e^t-1)}$

	Symbol	Density (PDF)	Distribution (CDF)	$\mathbb{E}$	Var	$\varphi_X(t) = \mathbb{E}[e^{\mathbf{i}tX}]$	$M_X(t) = \mathbb{E}[e^{tX}]$
Uniform	$\mathrm{Unif}([a,b])$	$\frac{1}{b-a}\mathbb{1}_{[a,b]}$	$\frac{x-a}{b-a}\mathbb{1}_{[a,b]} + \mathbb{1}_{(b,\infty)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{\mathbf{i}tb}-e^{\mathbf{i}ta}}{\mathbf{i}t(b-a)}$ [4]	$\frac{e^{tb}-e^{ta}}{t(b-a)}$ [5]
Exponential	$\mathrm{Exp}(\lambda)$	$\mathbb{1}_{x\geqslant 0}\lambda e^{-\lambda x}$	$\mathbb{1}_{x\geqslant 0}(1-e^{-\lambda x})$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-\mathbf{i}t}$	$\frac{\lambda}{\lambda-t}, t<\lambda$
Cauchy	$\mathrm{Cauchy}(x_0,\gamma)$	$\frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)}$	$\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}$	n/a	n/a	$e^{x_0\mathbf{i}t-\gamma\lvert t\rvert}$	n/a
Normal	$\mathcal{N}(\mu,\sigma)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\mu-x)^2}{2\sigma^2}}$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	μ	σ^2	$e^{\mathbf{i}\mu t-\frac{\sigma^2 t^2}{2}}$	$e^{\mu t+\frac{\sigma^2 t^2}{2}}$

[4] $\varphi_X(0) = 1$

[5] $M_X(0) = 1$

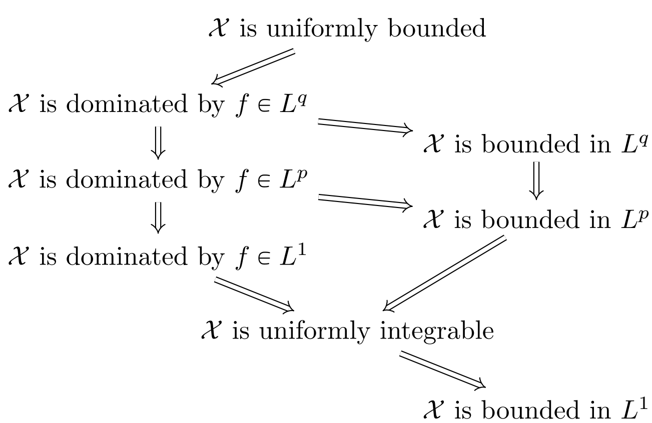
6.2 Notions of boundedness

The following is just a short overview of all the notions of boundedness we used in the lecture.

Definition† 6.0.70 (Boundedness)**.** Let $\mathcal{X}$ be a set of random variables. We say that $\mathcal{X}$ is

- **uniformly bounded** iff
$$\sup_{X \in \mathcal{X}} \sup_{\omega \in \Omega} |X(\omega)| < \infty,$$
- **dominated by** $f \in L^p$ for $p \geqslant 1$ iff
$$\forall X \in \mathcal{X}.\ |X| \leqslant f,$$
- **bounded in** L^p for $p \geqslant 1$ iff
$$\sup_{X \in \mathcal{X}} \|X\|_{L^p} < \infty,$$
- **uniformly integrable** iff
$$\forall \varepsilon > 0.\ \exists K.\ \forall X \in \mathcal{X}.\ \mathbb{E}[|X| \mathbb{1}_{|X| > K}] < \varepsilon.$$

Fact† 6.0.71. Let $\mathcal{X}$ be a set of random variables. Let $1 < p \leqslant q < \infty$ Then the following implications hold:



6.3 Laplace Transforms

Write something about Laplace Transforms

[Lecture 23, 2023-07-06]

6.4 Recap

6.4.1 Construction of iid random variables.

- Definition of a consistent family (Definition 1.5)
- Important construction:

  Consider a distribution function F and define

  $$\prod_{i=1}^{n}(F(b_i) - F(a_i)) =: \mu_n\left((a_1, b_1] \times x \ldots \times x(a_n, b_n]\right).$$

- Examples of consistent and inconsistent families Exercises
- Kolmogorov's consistency theorem (Theorem 1.6)

6.4.2 Limit theorems

- Work with iid. random variables.
- Notions of convergence (Definition† 0.6.3)
- Implications between different notions of convergence (very important) and counter examples. (Theorem† 0.6.4)
- Laws of large numbers: (Theorem 1.11)
  - WLLN: convergence in probability
  - SLLN: weak convergence
- Theorem 1.12 (building block for SLLN): Let (X_n) be independent with mean 0 and $\sum \sigma_n^2 < \infty$, then $\sum X_n$ converges a.s.
  - Counter examples showing that $\Longleftarrow$ does not hold in general are important
  - $\Longleftarrow$ holds for iid. uniformly bounded random variables
  - Application:

    $\sum_{i=1}^{\infty} \frac{(\pm 1)}{n^{\frac{1}{2}+\varepsilon}}$ converges a.s. for all $\varepsilon > 0$.

    $\sum \frac{\pm 1}{n^{\frac{1}{2}-\varepsilon}}$ does not converge a.s. for any $\varepsilon > 0$.
- Kolmogorov's Inequality (1.14)
- Kolmogorov's 0-1 Law (1.22)

  In particular, a series of independent random variables converges with probability 0 or 1.
- Kolmogorov's Three-Series Theorem (1.16)

- What are those 3 series?
- Applications

6.4.2.1 Fourier transform / characteristic functions / weak convegence

- Definition of Fourier transform (Definition 2.1)
- The Fourier transform uniquely determines the probability distribution. It is bounded, so many theorems are easily applicable.
- Uniqueness Theorem (2.3), Inversion Formula (2.2), ...
- Levy's Continuity Theorem (2.14), Theorem 2.27
- Bochner's Theorem for Positive Definite Functions (2.8)
- Bochner's Formula for the Mass at a Point (2.6)
- Related notions TODO
  - Laplace transforms $\mathbb{E}[e^{-\lambda X}]$ for some $\lambda > 0$ (not done in the lecture, but still useful).
  - Moments $\mathbb{E}[X^k]$ (not done in the lecture, but still useful) All moments together uniquely determine the distribution.

Weak convergence

- Definition of weak convergence (Definition 2.9)
- Examples:
  - $(\delta_{\frac{1}{n}})_n$,
  - $(\frac{1}{2}\delta_{-\frac{1}{n}} + \frac{1}{2}\delta_{\frac{1}{n}})_n$,
  - $(\mathcal{N}(0, \frac{1}{n}))_n$,
  - $(\frac{1}{n}\delta_n + (1 - \frac{1}{n})\delta_{\frac{1}{n}})_n$.
- Non-examples: $(\delta_n)_n$
- How does one prove weak convergence? How does one write this down in a clear way?
  - Theorem 2.13,
  - Levy's Continuity Theorem (2.14),
  - Generalization of Levy's continuity theorem 2.27

Convolution

- Definition of convolution.
- $X_i \sim \mu_i$ iid. $\implies X_1 + \ldots + X_n \sim \mu_1 * \ldots * \mu_n$.

Copy from exercise sheet and write a subsection about this

6.4.2.2 CLT

- Statement of the Central Limit Theorem (2.17)
- Several versions:
  - iid,
  - Lindeberg's CLT (2.24),
  - Lyapunov's CLT (2.25)
- How to apply this? Exercises!

6.4.3 Conditional expectation

- Definition and existence of conditional expectation for $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ (Theorem 3.2)
- If $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}[\cdot|\mathcal{G}]$ is the (unique) projection on the closed subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Why is this a closed subspace? Why is the projection orthogonal?
- Radon-Nikodym Theorem (3.20) (Proof not relevant for the exam)
- (Non-)examples of mutually absolutely continuous measures Singularity in this context?

6.4.4 Martingales

- Definition of Martingales (Definition 4.2)
- Doob's convergence theorem (Doob's Martingale Convergence Theorem (4.15)), Upcrossing inequality (Lemma 4.11, Lemma 4.12, Lemma 4.13) (downcrossings for submartingales)
- Examples of Martingales converging a.s. but not in L^1 (Example 4.17)
- Bounded in $L^2 \implies$ convergence in L^2 (Theorem 4.18).
- Martingale increments are orthogonal in L^2! (Fact 4.17.46)
- Doob's (sub-)martingale inequalities (Doob's Martingale Inequalities (4.19)),
- $\mathbb{P}[\sup_{k \leqslant n} M_k \geqslant x] \rightsquigarrow$ Look at martingale inequalities! Estimates might come from Doob's inequalities if $(M_k)_k$ is a (sub-)martingale.
- Doob's L^p convergence theorem (Theorem 4.25, Theorem 4.26).

- Why is $p > 1$ important? **Role of Banach Alaoglu (4.27)**
  - This is an important proof.
- Uniform integrability (Definition 4.22)
- What are stopping times? (Definition 4.28)
- (Non-)examples of stopping times
- **Optional Stopping Theorem (4.36)** - be really comfortable with this.

6.4.5 Markov Chains

- What are Markov chains?
- State space, initial distribution
- Important examples
- **What is the relation between Martingales and Markov chains?** u **harmonic** $\iff Lu = 0$. (sub-/super-) harmonic $u \iff$ for a Markov chain (X_n), $u(X_n)$ is a (sub-/super-)martingale
- Dirichlet problem (Not done in the lecture)
- ... (more in Probability Theory II)

Index