Part II — Probability and Measure

Based on lectures by E. Breuillard

Notes taken by Joseph Tedds using Dexter Chua's header and Gilles Castel's snippets.

Michaelmas 2019

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Measure spaces, σ-algebras, π-systems and uniqueness of extension, statement $\star$ and proof $\star$ of Carathodorys extension theorem. Construction of Lebesgue measure on $\mathbb{R}$. The Borel σ-algebra of $\mathbb{R}$. Existence of non-measurable subsets of $\mathbb{R}$. Lebesgue-Stieltjes measures and probability distribution functions. Independence of events, independence of σ-algebras. The Borel-Cantelli lemmas. Kolmogorovs zero-one law. [6]

Measurable functions, random variables, independence of random variables. Construction of the integral, expectation. Convergence in measure and convergence almost everywhere. Fatous lemma, monotone and dominated convergence, differentiation under the integral sign. Discussion of product measure and statement of Fubinis theorem. [6]

Chebyshevs inequality, tail estimates. Jensens inequality. Completeness of L^p for $1 \leq p \leq \infty$ The Hölder and Minkowski inequalities, uniform integrability. [4]

L^2 as a Hilbert space. Orthogonal projection, relation with elementary conditional probability. Variance and covariance. Gaussian random variables, the multivariate normal distribution. [2]

The strong law of large numbers, proof for independent random variables with bounded fourth moments. Measure preserving transformations, Bernoulli shifts. Statements $\star$ and proofs $\star$ of maximal ergodic theorem and Birkhoffs almost everywhere ergodic theorem, proof of the strong law. [4]

The Fourier transform of a finite measure, characteristic functions, uniqueness and inversion. Weak convergence, statement of Lvys convergence theorem for characteristic functions. The central limit theorem. [2]

Contents

0 Introduction

- Week 1 - Lebesgue measure
- Week 2 - Abstract measure theory
- Week 3 - Integration
- Week 4 - Measure theoretic foundations of probability theory
- Week 5 - Random variables, modes of convergence
- Week 6 - Hilbert spaces, C^p spaces
- Week 7 - Fourier transform, central limit theorem
- Week 8 - Ergodic theory

1 Boolean algebras and Finitely additive measures

Definition (Boolean Algebra)**.** A family of subset of X is called a *Boolean algebra* if it is stable under complementation and finite unions and contains the empty set. In other words

- $\emptyset \in \mathcal{B}$
- $\forall\ A, B \in \mathcal{B}, A^c \in \mathcal{B}$ and $A \cup B \in \mathcal{B}$

Remark. Clearly $\mathcal{B}$ is also stable under finite intersection and difference and symmetric difference i.e.

$$\begin{aligned} A, B \in \mathcal{B} \implies & A \cap B \in \mathcal{B} \\ & A \setminus B \in \mathcal{B} \\ & A \triangle B \in \mathcal{B}. \end{aligned}$$

Example.
- The *trivial* Boolean algebra $\mathcal{B} = \{\emptyset, X\}$
- The *discrete* Boolean algebra $\mathcal{B} = 2^X = \mathcal{P}(X)$, the family of subsets of X
- X topological space, the Boolean algebra of *constructible sets* is the family of all finite unions of *locally closed sets* (locally closes $= U \cap F$, for U open, F, closed).

Definition (Finitely additive measure)**.** A *finitely additive measure* on $(X, \mathcal{B})$ is a function $m : \mathcal{B} \mapsto [0, \infty]$ such that

(i) $m(\emptyset) = 0$

(ii) $m(E \cup F) = m(E) + m(F)$ whenever $E, F \subseteq \mathcal{B}$ are disjoint.

Remark. A finitely additive measure on (X, B) is also

(i) *monotone* - $E \subseteq F$ are in $\mathcal{B}$ then $m(E) \leq m(F)$

(ii) *subadditive* - $\forall\ E, F \in \mathcal{B}, M(E \cup F) \leq m(E) + m(F)$

Example.
(i) $\mathcal{B} = 2^X, m(E) :=$ number of E, is called the *counting measure on X*

(ii) $\mathcal{B} = 2^X$ if $f : X \mapsto [0, \infty]$ a function, $m_f(E) = \sum_{e \in E} f(e)$ is a finitely additive measure on X

(iii) $X = \bigcup_i^n X_i \quad X_i$ pairwise disjoint, let $\mathcal{B}$ be the Boolean algebra generated by this partition. If you assign some weight say $a_i \geq 0$ to each X_i, you can define a finitely additive measure on $\mathcal{B}$

$$m(E) = \sum_{i, X_i \subseteq E} a_i.$$

2 Jordan measure on $\mathbb{R}$

Definition (elementary)**.** A subset $E \subseteq \mathbb{R}^d$ is called *elementary* if it is a finite union of *boxes*. A *box* is a product of finite intervals

$$B = \prod_{i=1}^{d} I_i, \quad I_i = \text{ an interval } \subseteq \mathbb{R}.$$

For instance $(a, b), [a, b], (a, b], [a, b)$.

Denote by $|B|$ the "volume" of a box B.

$$B = \prod_{i=1}^{d} |b_i - a_i| \text{ if } B = \prod_{i=1}^{d} I_i \text{ and } (a_i, b_i) \subseteq I_i \subseteq [a_i, b_i].$$

Proposition. Let B be a box in $\mathbb{R}^d$ and let $\mathcal{E}(B)$ be the family of elementary subsets of B

(i) $\mathcal{E}(B)$ is a Boolean algebra

(ii) Every $E \in \mathcal{E}(B)$ can be written as a finite *disjoint* union of boxes

(iii) If $E \in \mathcal{E}(B)$ is written in 2 ways $E = \bigcup_i^N B_i = \bigcup_j^{N'} B'_j$ with B_i, B'_j pairwise disjoint, then $\sum_{i=1}^N |B_i| = \sum_{i=1}^{N'} |B'_j|$

Proof. When $d = 1$ it is obvious

Exercise. $d > 1$

□

Proposition. We may set $m(E) = \sum_{i=1}^N |B_i|$ whenever E is an elementary set written as $E = \bigcup_i^N B_i$ for B_i pairwise disjoint. Then m is a finitely additive measure on $(B, \mathcal{E}(B))$.

Definition (Jordan measurable set)**.** A subset $A \subseteq \mathbb{R}^d$ is called *Jordan-measurable* if $\forall\ \epsilon > 0\ \exists\ E, F$ elementary sets such that

- $E \subseteq A \subseteq F$
- $m(F \setminus E) < \epsilon$

Definition. If A is Jordan measurable, then set

$$m(A) = \inf\{m(F), A \subseteq F, F \text{ elementary}\}$$

Remark. This implies that

$$m(A) = \sup\{m(F), A \subseteq F, F \text{ elementary}\}$$

indeed,

$$\forall\ \epsilon\ \exists\ E, F\ E \subseteq A \subseteq F : m(F \setminus E) < \epsilon.$$

So $m(E) = m(F) - m(F \setminus E) \geq m(A) - \epsilon$.

Proposition. Let B be a box. The family $J(B)$ of Jordan measurable subsets of B is a Boolean algebra and m is a finitely additive measure on $(B, J(B))$.

Proof. Exercise $\square$

Remark. $A \subseteq [0,1]$ is Jordan measurable $\iff$ 1_A is Riemann-integrable.

Example.

$$f_n(x) = \mathbb{1}_{[0,1] \cap \frac{1}{n!} \in \mathbb{Z}} \ \forall\ x, \quad f_n(x) \to \mathbb{1}_{\mathbb{Q} \cap [0,1]}(x).$$

3 Lebesgue measurable set

Definition (Outer-measure)**.** To a subset E of $\mathbb{R}^d$ we associate its *outer-measure*

$$m^*(E) = \inf \left\{ \sum_{i \geq 1} m(B_i), E \subseteq \bigcup_{i \geq 1} B_i, B_i \text{ boxes} \right\}.$$

Definition (Lebesgue measurable set)**.** A subset $E \subseteq \mathbb{R}^d$ is called *Lebesgue measurable* if

$$\forall\, \varepsilon > 0\ \exists\, C = \bigcup_{i \geq 1} B_i,$$

a countable union of boxes, such that

$$m^*(C \setminus E) < \varepsilon, E \subseteq C.$$

Remark.
- $m^*(E + x) = m^*(E),\ \forall\, E,\ \forall\, x \in \mathbb{R}^d$
- We can take open boxes if we wish
- Jordan measurable sets are Lebesgue measurable

Our main proposition for this section is as follows:

Proposition.
(i) m^* extends to m on Jordan measurable sets.

(ii) The family $\mathcal{L}$ of Lebesgue measurable sets is a Boolean algebra, stable under countable unions.

(iii) m^* is a countably additive measure on $(\mathbb{R}^d, \mathcal{L})$. i.e.

$$m^* \left(\bigcup_{n \geq 1} E_n \right) = \sum_{n \geq 1} m^*(E_n) \text{ for } E_n \text{ pairwise disjoint.}$$

Remark.
- $\mathbb{Q}$ is in $\mathcal{L}$.
- m^* when restricted the family $\mathcal{L}$ is called the Lebesge measure
- Not every subset of $\mathbb{R}^d$ is in $\mathcal{L}$.
- m^* is not finitely additive on all subsets of $\mathbb{R}^d$.

Lemma. m^* is

(i) Monotone i.e. $E \subseteq F \implies m^*(E) \leq m^*(F)$

(ii) Countably subadditive $\forall\, E_n \subseteq \mathbb{R}^d$

$$m^* \left(\bigcup_{n \geq 1} E_n \right) \leq \sum_{n \geq 1} m^*(E_n).$$

Proof. (i) Clear

(ii) By definition of m^*, $\forall\, \varepsilon > 0\ \exists\, C_n = \bigcup_{i \geq 1} B_{n,i}$ a countable union of boxes such that $E_n \subseteq C_n$ and

$$m^*(E_n) + \frac{\varepsilon}{2^n} \geq \sum_{i \geq 1} m(B_{n,i})$$

by definition of m^*. Summing over all n

$$\left(\sum_{n \geq 1} m^*(E_n) \right) + \varepsilon \geq \sum_{n,i} m(B_{n,i})$$

and since $\bigcup E_n \subseteq \bigcup_{n,i} B_{n,i}$ by monotonicity of m^*

$$\sum m^*(E_n) \geq m^* \left(\bigcup_{n \geq 1} E_n \right)$$

□

Remark. It is easy to check (see the example sheet) that a finitely additive measure on a Boolean algebra is countably additive iff it has the "continuity property".

Definition (Continuity property)**.** Let X be a set, $\mathcal{B}$ a Boolean algebra of subsets of X. Let m be a finitely additive measure on X such that $m(X) < \infty$. We say that $(X, \mathcal{B}, m)$ have the *continuity property* if

$$\forall\ E_n \in \mathcal{B}, E_{n+1} \subseteq E_n \text{ and } \bigcap_n E_n = \emptyset \implies \lim_{n \to \infty} m(E_n) = 0.$$

Proposition. The Jordan measure has the continuity property on elementary sets

Proof. Suppose not. We get $E_{n+1} \subseteq E_n, \bigcap_n E_n = \emptyset$ and $m(E_n) \not\to 0, E_n$ elementary. $\exists\ F_n \subseteq E_n$ elementary sets $m(F_n) \geq m(E_n) - \frac{\varepsilon}{2^n}$ and F_n closed. By Heine-Borel, since

$$\bigcap_n F_n = \emptyset \implies \exists\ N < \infty \bigcap_{n=1}^{N} F_n = \emptyset$$

(The F_i are closed and bounded and hence compact; in particular, F_1 is compact. Since the intersection of all the F_i is $\emptyset$ then the open sets $F_1 \setminus F_n \subseteq F_1$ form an open cover of F_1. Since F_1 is compact, there is a finite subcover and in particular $\exists N$ such that $\bigcup_{n=1}^{N} F_1 \setminus F_n = F_1$)

Then,

$$\begin{aligned} m(E_n \setminus (F_1 \cap, \ldots, \cap F_n)) &= m\left(\bigcup_{i=1}^{n} E_n \setminus F_n\right) \\ &\leq \sum_{i=1}^{n} m(E_n \setminus F_i) \\ &\leq \sum_{i=1}^{n} m(E_i \setminus F_i) \\ &\leq \sum_{i=1}^{n} \frac{\varepsilon}{2^i} \leq \varepsilon. \end{aligned}$$

them $m(F_1 \cap \ldots, \cap F_n) \geq m(E_n) - \varepsilon \geq 2\varepsilon - \varepsilon \geq \varepsilon > 0$. For $n = N$ this gives a contradiction □

We can now begin our proof of the main proposition

Proof. (i) To show $m^* = m$ on Jordan measurable sets

- It is clear $m^*(A) \leq m(A)$ by definition
- We need to show converse inequality
- First suppose A is elementary, Pick

$$\varepsilon > 0, A \subseteq \bigcup_{n \geq 1} B_n, m^*(A) + \varepsilon \geq \sum_{n \geq 1} m(B_n).$$

Let $E_n = A \setminus (B_1 \cup \ldots \cup B_n)$ an elementary set.

$$E_{n+1} \subseteq E_n, \bigcap_n E_n = \emptyset \implies m(E_n) \underset{n \to \infty}{\to} 0$$

but

$$\begin{aligned} m(A) &\leq m(A \setminus B_1 \cup \ldots \cup B_n) + m(B_1 \cup \ldots \cup B_n) \\ &\leq \underset{\to 0}{m(E_n)} + \sum_{i=1}^{n} m(B_i). \end{aligned}$$

So $m(A) \leq m^*(A) + \varepsilon$, ε arbitrary $\implies m(A) \leq m^*(A)$

- In general, if A is Jordan measurable, $\forall\, \varepsilon > 0\ \exists\, E$ elementary

$$E \subseteq A, m(A) \leq m(E) + \varepsilon.$$

(ii) We show that $\mathcal{L}$ is stable under countable unions. Let $E = \bigcup_n E_n$ with each $E_n \in \mathcal{L}$ then we need to show $E \in \mathcal{L}$.

$$E_n \in \mathcal{L} \iff \forall\, \epsilon > 0\ \exists\, C_n \text{ a countable union of boxes :}$$

$$m^*(C_n \setminus E_n) < \frac{\epsilon}{2^n}, E_n \subseteq C_n$$

Set $C = \bigcup_n C_n$ is still a countable union of boxes. $E \subseteq C$ and $m^*(C \setminus E) = m^*(\bigcup C_n \setminus \bigcup E_n)$.
m^* is monotone

$$\implies m^*\left(\left\{\bigcup_n C_n\right\} \setminus \left\{\bigcup_n E_n\right\}\right) \leq m^*\left(\bigcup_n (C_n \setminus E_n)\right).$$

m^* is countably subadditive

$$\implies \sum_n m^*(C_n \setminus E_n) \leq \sum \frac{\epsilon}{2^n} \leq \epsilon.$$

So $E \in \mathcal{L}$

□

Example. $E = \mathbb{Q} \cap [0,1]$, $m^*(E) = 0$. Since Q is countable, we can make the boxes singletons which each have measure 0.

Lemma. If $(E_n)_n$ is a family of elementary sets such that $E_{n+1} \subseteq E_n$ then $A = \bigcap_n E_n \in \mathcal{L}$ and $\lim_{n\to\infty} m(E_n) = m^*(A)$

Proof. Note $E_n \setminus A = \bigcup_{i \geq n} E_i \setminus E_{i+1}$. So

$$\begin{aligned} m^*(E_n \setminus A) &\leq \sum_{i\geq n} m^*(E_i \setminus E_{i+1}) (m^* \text{ only subadditive}) \\ &\leq \sum_{i \geq n} m(E_i \setminus E_{i+1}) \\ &= \sum_{i\geq n} m(E_i) - m(E_{i+1}). \end{aligned}$$

We get

$$m^*(E_n \setminus A) \leq m(E_n) - \lim_{i\to\infty} m(E_i) \underset{n\to\infty}{\to} 0,$$

and hence

$$\begin{aligned} &A \in \mathcal{L} \text{ and } m^*(A) \leq m^*(E_n) \leq m^*(E_n \setminus A) + m^*(A) \\ &\implies m(E_n) = m^*(E_n) \to m^*(A). \end{aligned}$$

□

Corollary. (i) Countable intersection of elementary sets are in $\mathcal{L}$,

(ii) Open sets and closed sets in $\mathbb{R}^d$ are in $\mathcal{L}$

Proof. (i) $A = \bigcap E_n \implies A = \bigcap F_n, F_n = E_1 \cap \ldots \cap E_n$, with F_n still elementary and $F_{n+1} \subseteq F_n$ so the lemma shows $A \in \mathcal{L}$.

(ii) Every open subset of $\mathbb{R}^d$ is a countable union of open boxes. So it must be in $\mathcal{L}$ by the proof of stability under countable unions.
If $C \subseteq R^d$ is closed

$$C = \bigcup_{n\geq 1} (C \cap [-n, n]^d).$$

So wlog we may assume that C is bounded say $C \subseteq B$ an open box. $B \setminus C$ is open so $B \setminus C = \bigcup_n B_n, B_n$ open boxes

$$C = B \setminus (B \setminus C) = B \setminus \bigcup_n B_n = \bigcap_n (B \setminus B_n).$$

$B \setminus B_n$ is elementary so by $(i), C \in \mathcal{L}$.

□

Definition (Null set)**.** We say a subset $E \subseteq \mathbb{R}^d$ is called a *null set* if $m^*(E) = 0$

Lemma. Every null set is in $\mathcal{L}$

Proposition. Clear since $\exists\ C = \bigcup B_i, E \subseteq C, \sum m(B_i) < \epsilon$

$$\implies m^*(C \setminus E) \leq m^*(C) \leq \sum m^*(B_i) < \epsilon.$$

Proof. End of proof of main proposition (ii) : to show $E \in \mathcal{L} \implies E^c \in \mathcal{L}$ wlog we can assume that E is bounded i.e. $E \subseteq B = [-k, k]^d$ for some k. Indeed,

$$E^c = \bigcup_{k \geq 1} E^c \cap [-k, k].$$

$$E^c \cap [-k, k]^d = [-k, k]^d \setminus \underset{\in \mathcal{L}}{E} \cap [-k, k]^d.$$

We need to show

$$E \subseteq B, E \in \mathcal{L} \implies B \setminus E \in \mathcal{L}.$$

$$E \subseteq C_n = \bigcup_{i \geq 1} B_{i,n}, \quad m^*(C_n \setminus E) < \frac{1}{n}.$$

We can assume $B_{i,n} \subseteq B$.

$$B \setminus C_n = \bigcap_{i \geq 1} \underbrace{B \setminus B_{i,n}}_{\text{elementary}} \in \mathcal{L},$$

by the corollary (i). Let $F = \bigcup_n B \setminus C_n \in \mathcal{L}$ (as a countable union of sets in $\mathcal{L}$, but $F \subseteq B \setminus E$ and $(B \setminus E) \setminus F$ is null.

$$m^*((B \setminus E) \setminus F) \leq m^*((B \setminus E) \setminus (B \setminus C_n)) = m^*(C_n \setminus E) < \frac{1}{n}.$$

$$\begin{aligned} &\implies (B \setminus E) \setminus F \text{ is null} \\ &\implies B \setminus E = \underset{\in \mathcal{L}}{F} \cup ((B \setminus E) \setminus F) \text{ null so in } \mathcal{L}. \end{aligned}$$

So $B \setminus E \in \mathcal{L}$ □

Proposition. (i) If $E \in \mathcal{L}$, then $\forall\ \epsilon > 0\ \exists\ U$ open, $\exists\ F$ closed such that

$$F \subseteq E \subseteq U \text{ and } m^*(U \setminus F) < \epsilon.$$

(ii) Every $E \in \mathcal{L}$ can be written as

$$E = G \setminus N = F \cup M, \quad F, M \text{ disjoint}.$$

Where G is a countable intersection of open sets, F is a countable union of closed sets and M, N are null.

Proof. Easy $\forall\ \epsilon > 0, E \subseteq U = \bigcup_i B_i$ a countable union of open boxes such that $m^*(E \setminus U_n) < \frac{\epsilon}{2}$ by definition of $E \in \mathcal{L}$. We know $E^c \in \mathcal{L}$ so $E^c \subseteq \Omega = \bigcup_i B_i', m^*(\underbrace{\Omega \setminus E^c}_{E \setminus \Omega^c}) < \frac{\epsilon}{2}$. So $\Omega^c = \bigcap_i B_i^{'c}$

$$\implies \Omega^c \subseteq E \subseteq U, m^*(u \setminus \Omega^c) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

Question. Is every subset of $\mathbb{R}^d$ in $\mathcal{L}$?

The Vitali counter-example is as follows: Let E be a set of representatives of the cosets of $(\mathbb{Q}, +)$ in $(\mathbb{R}, +)$ i,e, $x + \mathbb{Q} \subseteq \mathbb{R}$. For each coset $x + \mathbb{Q}$ pick one element $e \in x + \mathbb{Q}$ such that $e \in [0, 1]$.

$$\implies \forall\ x \in R\ \exists!\ e_x \in E : e_x \in x + \mathbb{Q}.$$

Claim 1 m^* is not finitely additive on all subsets of $\mathbb{R}^d$. Claim 2 E is not in $\mathcal{L}$

Proof. Observe if $r_1, \ldots, r_N$ are N distinct rationals then $(r_i + E)_{1 \leq i \leq N}$ are pairwise disjoint [

$$r_i + e = r_j + f \implies e + \mathbb{Q} = f + \mathbb{Q} \implies e = f, \implies r_i = r_j.$$

] Assume each $r_i \in [0, 1]$, then $r_i + E \subseteq [0, 2]$

$$\implies \bigcup_{i=1}^{N} r_i + E \subseteq [0, 2], m^*(\bigcup_{i=1}^{N} r_i + E) \leq m^*([0, 2]) = 2.$$

If m^* were finitely additive on the family of all subsets of $\mathbb{R}^d$ then

$$m^*(\bigcup_{i=1}^{N} r_i + E) = \sum_{i=1}^{N} m^*(r_i + E) = N m^*(E) \implies m^*(E) = 0.$$

However, $[0, 1] \subseteq \mathbb{R} = \bigcup_{r \in \mathbb{Q}} E + r$.

$$m^* \text{ is countably subadditive } \implies m^*([0, 1]) \leq \sum_{r \in \mathbb{Q}} m^*(r + E).$$

But $m^*(r + E) = m^*(E) = 0$m so $m^*([0, 1]) = 0$ but we know that $m^*([0, 1]) = 1$, contradicting our assumption of finite additivity and non-measurability follows by the same argument. □

4 Abstract measure theory

Definition (σ-algebra)**.** A *σ-algebra* on X is a Boolean algebra which is stable under countable unions.

Definition (Measurable space)**.** A *measurable space* is a pair $(X, \mathcal{A})$ with X a set and $\mathcal{A}$ a σ-algebra of subsets of X.

Definition (Measure)**.** A *measure* on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, +\infty]$ such that

(i) $\mu(\emptyset) = 0$

(ii) It is countably additive, i.e

$$\mu\left(\bigcup_{n\geq 1} E_n\right) = \sum_{n\geq 1} \mu(E_n).$$

if $E_n \in \mathcal{A}$ are disjoint.

Definition (Measure space)**.** The triple $(X, \mathcal{A}, \mu)$ is called a *measure space*.

Example. (i) $\mu = 0$ is always a measure.

(ii) $(\mathbb{R}^d, \mathcal{L}, m)$ is a measure space.

(iii) If $A_0 \in \mathcal{L}$, $m_0(E) = m^*(A_0 \cap E)$, then we get a measure on $(\mathbb{R}^d, \mathcal{L}, m^*)$

(iv) $(X, 2^X, \text{counting measure})$

(v) $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ pick $a_n \geq 0$, $\mu(I) = \sum_{i\in I} a_i$

Proposition. Let $(X, \mathcal{A}, \mu)$ be a measure space

(i) μ is monotone: $\mu(A) \leq \mu(B)$ if $A \subseteq B,\ A, B \in \mathcal{A}$.

(ii) μ is countably subadditive

$$\mu\left(\bigcup_{n\geq 1} E_n\right) \leq \sum_{n\geq 1} \mu(E_n)\ \forall\ E_n \in \mathcal{A}.$$

(iii) Upward monotone convergence. If

$$E_1 \subseteq \ldots \subseteq E_n, E_i \in \mathcal{A}.$$

then

$$\mu(E_n) \underset{n\to\infty}{\to} \mu\left(\bigcup_{n\geq 1} E_n\right).$$

(iv) Downward monotone convergence. If $\mu(E_n) < \infty$ and

$$E_1 \supseteq \ldots \supseteq E_n, E_i \in \mathcal{A}.$$

then

$$\mu(E_n) \to \mu\left(\bigcap E_n\right).$$

Proof. (i) $\mu(B) = \mu(B \setminus A) + \mu(A)$

(ii)

$$\bigcup E_n = \bigcup F_n,\ F_n = E_n \setminus (E_1 \cup \ldots \cup E_{n-1}).$$

$$\mu(\cup E_n) = \mu(\cup F_n) = \sum \mu(F_n) \leq \sum \mu(E_n).$$

(iii) Set $E_0 = \emptyset$ and write

$$\mu(\bigcup_{n=1}^{N} F_n) = \sum_{n=1}^{N} \mu(F_n) = \sum_{n=1}^{N} \mu(E_n) - \mu(E_{n+1}) = \mu(E_N).$$

(iv) Same as (iii) applied to $E_1 \setminus E_n$

□

Caveat :We need $\mu(E_1) < \infty$ in (iv) e.g.

$$E_n = [n, +\infty] \subseteq \mathbb{R}, \bigcap_n E_n = \emptyset \text{ but } \mu(E_n) = \infty.$$

Proposition. If $\mathcal{F}$ is a family of subsets of X, the intersection of all σ-algebras containing $\mathcal{F}$ is a σ-algebra denoted by $\sigma(\mathcal{F})$.

Proof. See the example sheet. □

Example. (i) $X = \bigcup_{i=1}^{N} X_i, \mathcal{F} = \{X_1, \ldots, X_n\}$ the atoms of the partition. $\sigma(\mathcal{F})$ = subsets of X that are finite unions of X_i 's.

(ii) X = countable set $\mathcal{F}$ = singletons, then $\sigma(\mathcal{F}) = 2^X$

Definition (Borel σ-algebra)**.** If X is a topological space, then the σ-algebra generated by open subsets of X is called the *Borel* σ-algebra of X denoted by $\mathcal{B}(X)$

Definition (Borel measure)**.** A measure $(X, \mathcal{B}(X))$ is called a *Borel measure.*

Example. $\mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{L}$

Remark. The Boolean algebra generated by a family $\mathcal{F}$ of subsets of X call it $\beta(\mathcal{F})$: every set in $\beta(\mathcal{F})$ is a finite union of subsets of the form $F_1 \cap \ldots \cap F_n$ where $\forall\ i$ either $F_i \in \mathcal{F}$ or $F_i^c \in \mathcal{F}$.

Definition (σ-finite)**.** A σ-*finite* measure

$$\exists\ X = \bigcup_{n \geq 1} X_n,\ X_n \in \mathcal{F} : \mu(X_n) < \infty.$$

Example. The Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $(\mathbb{R}^d, \mathcal{L})$ are σ-finite.

Question. Given a set X and a Boolean algebra $\mathcal{B}$ of subsets of X and given a finitely additive measure μ on $(X, \mathcal{B})$, can we extend μ to $\sigma(\mathcal{B})$?

Theorem (Caratheodory extension theorem)**.** Let $\mathcal{B}$ be a Boolean algebra on a set X and μ is a finitely additive measure on $(X, \mathcal{B})$ such that

- μ is σ-finite.
- μ has the continuity property.

Then μ extends to a unique measure on $(X, \sigma(\mathcal{B}))$.

Definition (Outer measure)**.** The *outer measure* μ^* is defined on all subsets E of X as

$$\mu^*(E) = \inf\{\sum_{i\geq 1} \mu(E_i), E \subseteq \bigcup_{i\geq 1} B_i B_i \in \mathcal{B}\}.$$

Definition (μ^*-measurable)**.** We say that a subset $E \subseteq X$ is μ^*-*measurable* if

$$\forall\, \epsilon > 0\ \exists\ C = \bigcup_{i\geq 1} B_i, B_i \in \mathcal{B} : E \subseteq C \text{ and } \mu^*(C \setminus E) < \epsilon.$$

The existence part of the Caratheodory's theorem follows from

Proposition. Under the assumptions of the Caratheodory theorem

(i) $\mu^*|_{\mathcal{B}} = \mu$.

(ii) The family $\mathcal{B}^*$ of μ^* measurable sets is a σ-algebra containing $\mathcal{B}$.

(iii) μ^* is countably additive on B^*.

Proof. Already done.

$$\mathcal{B} \leftrightarrow \text{Boolean algebra generated by elementary sets.}$$

$E \subseteq X_n \leftrightarrow$ bounded sets.

□

Remark.
- $\mathcal{B}^*$ contains the μ^*-null sets ($(\mathcal{B}^*, \mu^*)$ is called the completion of $(\mathcal{B}, \mu)$.
- When $X = \mathbb{R}^d, \mathcal{B} =$ Boolean algebra generated by elementary sets and $\mu = m$ Jordan measure, then $\mu^* = m^*, \mathcal{B}^* = \mathcal{L}$ is the Lebesgue measure.

5 Uniqueness of measures

Definition (π-system)**.** Let X be a set. Let $\mathcal{F}$ be a family of subsets of X, $\mathcal{F}$ is called a *π-system* if

- $\emptyset \in \mathcal{F}$.
- $\mathcal{F}$ is stable under finite intersections.

Now we state our main proposition for this section.

Proposition. Let (X, d) be a measurable space. Let μ_1, μ_2 be two measures on it, assume that

- $\mu_1(X) = \mu_2(X) < \infty$
- $\mu_1(F) = \mu_2(F) \ \forall \ F \in \mathcal{F}$

for some π-system $\mathcal{F}$ such that $\sigma(\mathcal{F}) = \mathcal{A}$, then $\mu_1 = \mu_2$.

Lemma. (Dynkin Lemma) $\mathcal{F}$ is a π-system. Let $\mathcal{F} \subseteq \mathcal{G}$ a family of subsets stable under:

- complementation
- disjoint countable unions

Then $\sigma(\mathcal{F}) \subseteq \mathcal{G}$

we use this to prove our main proposition.

Proof. Let $\mathcal{G} = \{A \in \mathcal{A}, \mu_1(A) = \mu_2(A)\}$. Clearly $\mathcal{F} \subseteq \mathcal{G}$, $\mathcal{G}$ is stable under disjoint countable union. $\mathcal{G}$ is stable under complementation since $\mu_1(A^c) = \mu_1(X) - \mu_1(A) = \mu_2(A^c)$ (since $\mu_1(X) < \infty$). By Dynkin's lemma $\mathcal{G} = \sigma(\mathcal{F}) = \mathcal{A}$ so $\mu_1 = \mu_2$. □

Remark. The proposition continues to hold if μ_1 and μ_2 are only assumed σ-finite and $X = \bigcup_{n \geq 1} X_n, X_n \in \mathcal{F}$.

Proof. Dynkin's lemma. Let $\mathcal{M}$ be the smallest family of subsets of X that contains $\mathcal{F}$ and is stable under complementation and disjoint countable union. We need to show that $\mathcal{M}$ is a σ-algebra ($\implies \sigma(\mathcal{F}) \subseteq \mathcal{M}$). It is enough to show that $\mathcal{M}$ is a Boolean algebra and further, that it is stable under finite intersection.

$$\mathcal{M}' = \{A \in \mathcal{M}, A \cap B \in \mathcal{M} \ \forall \ B \in \mathcal{F}\}.$$

We need to show that $\mathcal{M}' = \mathcal{M}$. But $\mathcal{M}'$ is stable under complementation

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

under disjoint countable union, $\mathcal{F} \subseteq \mathcal{M}'$ because $\mathcal{F}$ is a π-system. By minimality of $\mathcal{M} \implies \mathcal{M}' = \mathcal{M}$

$$\mathcal{M}'' = \{A \in \mathcal{M} | A \cap B \in \mathcal{M} \ \forall \ B \in \mathcal{M}\}.$$

To show $\mathcal{M}'' = \mathcal{M}$ again $\mathcal{M}''$ is stable under complementation, countable disjoint union of and $\mathcal{F} \subseteq \mathcal{M}'' \implies \mathcal{M} = \mathcal{M}''$ □

Corollary. The Lebesgue measure is the unique translation invariant Borel measure on $\mathbb{R}^d$ such that $\mu([0,1]^d) = 1$

Proof. $m = m^*$ on $\mathcal{L}$. m^* is translation invariant by definition. To show uniqueness apply the proposition and show that if μ is another such measure $\mu(F) = m(F)\ \forall\ F \subseteq \mathbb{R}^d$, where F is a dyadic box. □

Definition (Dyadic box)**.** A *dyadic box* is a box $\prod_{i=1}^d [a_i, b_i]$ where a_i, b_i are dyadic numbers i.e. of the form $\frac{k}{2^n} k, n \in \mathbb{Z}$.

Remark.

(i) There is no (countably additive) measure μ defined on all subsets of $\mathbb{R}^d$ that is translation invariant with $0 < \mu([0,1]^d) < \infty$ (Vitali example leads to a contradiction).

(ii) However, one can show that there exists infinitely many finitely additive translation invariant measures on $\mathbb{R}^d$ with $\mu([0,1]^d) = 1$

6 Measurable functions

Definition (Measurable function)**.** Let $(X, \mathcal{A})$ be a measurable space. A function $f : X \to \mathbb{R}$ is called $\mathcal{A}$-*measurable* if

$$\forall\, t \in \mathbb{R}\{x \in X | f(x) < t\} \in \mathcal{A}.$$

Remark. (i) f is measurable $\iff$ $f^{-1}(B) \in \mathcal{A}\ \forall\ B$ Borel subsets contained in $\mathbb{R}$.

(ii) If $f : X \to \mathbb{R} \cup \{\pm\infty\}$ we say that f is measurable if additionally $f^{-1}(-\infty)$ and $f^{-1}(\infty)$ are in $\mathcal{A}$.

Definition. Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be measurable spaces. A map $f : X \to Y$ is called measurable if

$$f^{-1}(B) \in \mathcal{A}\ \forall\ B \in \mathcal{B}.$$

Example. (i) A continuous function $f : \mathbb{R}^d \to \mathbb{R}$ is measurable (with respect to $(\mathbb{R}^d, \mathcal{B}(R^d)$).

(ii) $E \subseteq X, E \in \mathcal{A} \iff 1_E$ is $\mathcal{A}$-measurable.

(iii) If $X = \bigcup_{i=1}^N X_i$, with the X_i disjoint, if $\mathcal{A}$ is the Boolean algebra generated by the X_i 's. f is $\mathcal{A}$-measurable $\iff$ f is constant on each X_i.

Proposition. (i) If $f : X \to Y$ and $g : Y \to Z$ are measurable, then $g \circ f : X \to Z$ is also measurable.

(ii) If f, g are measurable functions on $(X, \mathcal{A})$, then so are

$$f + g, fg, \lambda f\ \ \lambda \in \mathbb{R}.$$

(iii) If $(f_n)_{n\geq 1}$ is a sequence of measurable functions on $(X, \mathcal{A})$ then so are

$$\lim_n \sup_n f_n(x), \lim_n \inf_n f_n(x), \sup_n f_n(x), \inf_n f_n(x).$$

Proof. (i) Clear from definition

(ii) Follows from (i) and the fact that

$$\begin{aligned} \mathbb{R}^2 &\to R \\ (x,y) &\mapsto x + y \end{aligned} \qquad \text{and} \qquad \begin{aligned} \mathbb{R}^2 &\to R \\ (x,y) &\mapsto xy \end{aligned}$$

are Borel measurable since they are continuous.

(iii) Straightforward, for example

$$\{x \in X, \lim_n \sup f_n(x) < t\} = \bigcup_{m\geq 1} \bigcup_{k\geq 1} \bigcap_{n\geq k} \{x \in X, f_n(x) < t - \frac{1}{m}\}.$$

$\square$

Caveat. (i) If f is a $\mathcal{A}$-measurable function, then the preimage of a Lebesgue measureable set may not be in $\mathcal{A}$

(ii) The image of a measurable set by a measurable map may not be measurable.

7 Integration

Definition (Simple function)**.** A *simple function* on a measure space $(X, \mathcal{A}, \mu)$ is a function of the form $\sum_{i=1}^N a_i 1_{A_i}, a_i \geq 0, A_i \in \mathcal{A}$. Equivalently f is $\mathcal{A}$-measurable and takes only finitely many non-negative values.

Proposition. If a simple function on $(X, \mathcal{A}, \mu)$ can be written in two ways as $f = \sum_{i=1}^N a_i 1_{A_i} = \sum_{j=1}^M b_j 1_{B_J}$, then

$$\sum_{i=1}^N a_i \mu(A_i) = \sum_{j=1}^M b_j \mu(B_j).$$

This common value is called the μ-integral of f and denoted by $\mu(f)$ on $\int_X f \mathrm{d}\mu$.

Definition. Suppose f is an arbitrarily non-negative $\mathcal{A}$-measurable function. Let

$$\mu(f) := \sup\{\mu(g), g \text{ simple and } g \leq f\}.$$

Remark. (i) When $(X, \mathcal{A}, \mu) = (\mathbb{R}^d, \mathcal{L}, m)$ and if f if Riemann integrable, then

$$\int_{\mathbb{R}^d} f(x)\mathrm{d}x = \mu(f).$$

(ii) The previous definition is consistent with the definition for simple functions.

Proposition. (Positivity of the μ-integral)
Let $f, g \geq 0$ be measurable.

(i) If $f \geq g$ then $\mu(f) \geq \mu(g)$

(ii) If $f \geq g$ and $\mu(f) = \mu(g) < \infty$ then $f = g$ μ-almost everywhere i.e.

$$\mu(\{x \in X | f(x) \neq g(x)\}) = 0.$$

Proof. First assume that f, g are simple

(i) $f - g$ is simple too, we can find a common refinement $f = \sum_{i=1}^N a_i 1_{E_i}, g = \sum_{i=1}^N b_i 1_{E_i}$. $f \geq g$ means $a_i \geq bi$ $\forall$ i so $\mu(f) \geq \mu(g)$. This shows consistency of the two definitions of $\mu(f)$ for f simple.

(ii) $\mu(f) = \mu(g) \iff \sum a_i \mu(E_i) = \sum b_i \mu(E_i)$ but $a_i \geq b_i$ $\forall$ i so in fact $a_i = b_i$ if $\mu(E_i) \neq 0$. So $f = g$ μ-a.e.

Now we tackle the general case. Let $f, g \geq 0$ be measurable functions.

(i) Clear from the definition of $\mu(f)$.

(ii) $A_n := \{x \in X | f(x) - g(x) > \frac{1}{n}\}$. We need to show that $\mu(A_n) = 0 \; \forall \; n \geq 1$. Note that $f - g \geq \frac{1}{n} 1_{A_n}$ but from the definition of the μ-integral $f = g + f - g$ and $\mu(f) \geq \mu(g) + \mu(f - g)$ hence $\mu(f - g) = 0$ so $\mu(A_n) = 0$.

□

Lemma. If $f \geq 0$ and measurable then f is a limit of simple functions $\exists$ $(g_n)_n$ of simple functions such that $g_{n+1} \geq g_n$ and $g_n(x) \to f(x)$ $\forall$ $x \in X$.

Proof.

$$g_n(x) = \frac{1}{2^n}\left\lfloor 2^n \min\{f(x), n\}\right\rfloor.$$

□

Theorem (Monotone convergence theorem)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, let $(f_n)_{n\geq 1}$ be measurable functions such that $f_{n+1} \geq f_n \geq 0$. Then

$$\mu(\lim_{n\to\infty} f_n) = \lim_n \mu(f_n) \in [0, \infty],$$

where the pointwise limit $\lim_{n\to\infty} f_n = f$

Lemma. Let g be a simple function. Then

$$m_g : \mathcal{A} \to [0, \infty], E \mapsto \mu(1_E g)$$

is a measure on $(X, \mathcal{A})$.

Proof. If $g = \sum a_i 1_{A_i}$ so $m_g(E) = \sum_{i=1}^N a_i \mu(A_i \cap E)$ is clearly a measure □

Proof. Of MCT. Let $f := \lim_n f_n$. Then $\mu(f_n) \leq \mu(f_{n+1}) \leq$ by positivity hence $\lim_{n\to\infty} \mu(f_n) \leq \mu(f)$.
Now pick $\varepsilon > 0$, pick a simple function g such that $g \leq f$. Let

$$E_n := \{x \in X | f_n(x) \geq (1-\varepsilon)g(x)\}.$$

Then $X = \bigcup_n E_n$ and $E_n \subseteq E_{n+1}$, so $m_g(E_n) \to m_g(X)$ ($\mu((1-\varepsilon)g1_{E_n}) \to \mu(g)$). So

$$(1-\varepsilon)m_g(E_n) = \mu((1-\varepsilon)g1_{E_n}) \leq \mu(f_n).$$

Where the inequality comes from the definition of E_n. Now we let $n \to \infty$ so

$$(1-\varepsilon)\mu(g) \leq \lim_{n\to\infty} \mu(f_n).$$

This holds $\forall\, \varepsilon > 0\ \forall\, g$ simple $\leq f$. Hence $\mu(f) \leq \lim_{n\to\infty} \mu(f_n)$ □

Lemma. (Fatou's Lemma)
Let $(f_n)_n$ be measurable non-negative functions. Then

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

Proof. Let $g_n = \inf_{k\geq n} f_k$, $g := \liminf_{n\to\infty} f_n$. Note that

- $g_{n+1} \geq g_n \geq 0$.
- $\lim g_n = g$.
- $g_n \leq f_n$.

Hence apply MCT to the (g_n)'s

$$\mu(g) = \lim \mu(g_n) \leq \liminf_n \mu(f_n).$$

□

Remark. We can have a strict inequality e.g. moving bumps $f_n = 1_{[n,n+1]}$ on $(\mathbb{R}, \mathcal{L}, m)$ with $f_n \to 0$ pointwise but $\mu(f_n) = 1$

Definition (μ-integrable)**.** A measurable function on $(X, \mathcal{A}, \mu)$ is said to be *μ-integrable* if $\mu(|f|) < \infty$ and we define $\mu(f) = \mu(f^+) - \mu(f^-)$ where $f^+ = \max\{f, 0\}$, $f^- = (-f)^+$. Note that

$$|f| = f^+ + f^-, f = f^+ - f^-, f^\pm \geq 0.$$

Proposition. (Linearity of μ-integral)
If f, g are μ-integrable and if $\alpha, \beta \in \mathbb{R}$ then $\alpha f + \beta g$ is μ-integrable and in particular

$$\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g).$$

Proof. Reduce to the case where $f, g, \alpha, \beta \geq 0$. If $f, \alpha \geq 0$, $\mu(\alpha f) = \alpha\mu(f)$ which is clear by definition. So now we have left to show that $u(f + g) = \mu(f) + \mu(g)$ if $f, g \geq 0$. This is clear for simple functions and for the general case we use the lemma that approximates f, g by simple functions $f_n \uparrow f, g_n \uparrow g$ and use the MCT. □

Theorem (Lebesgue's dominated convergence theorem)**.** Let f_n, g be μ-integrable functions such that

(i) $|f_n| \leq g$ (domination condition).

(ii) $f(x) := \lim_{n\to\infty} f_n(x)$ exists $\forall\, x \in X$.

Then $\lim_{n\to\infty} \mu(f_n) = \mu(f)$

Proof. Clearly, $|f| \leq g$ from (i) so f is μ-integrable and $g + f_n \geq 0$. Fatou's lemma says that $\mu(g) + \mu(f) = \mu(g + f) \leq \underset{n}{\text{liminf}}\mu(g + f) \leq \mu(g) + \underset{n}{\text{liminf}}\mu(f_n)$. Since $\mu(g) < \infty$ we must have $\mu(f) \leq \liminf \mu(f_n)$. Apply the same argument to $-f_n$ and we obtain $\mu(-f) \leq \liminf_n \mu(-f) = -\limsup_n f(_n)$. So we obtain

$$\mu(f) \leq \liminf \mu(f_n) \leq \limsup \mu(f_n) \leq \mu(f).$$

□

Remark. It is enough to assume that the hypothesis in any of the three main theorems hold μ-almost everywhere. The reason for this is that if

$$E = \{x \in X| \text{ some hypothesis fails}\},$$

e.g. $E = \{x|f_{n+1}(x) < f_n(x)\}$ in MCT then $\mu(E) = 0$ so let $g_n := f_n 1_{E^c}$ then apply the results to g_n.

Corollary. (Exchange between $\int$ and $\sum$)
Let (f_n) be a sequence of measurable functions on X.

(i) If $f \geq 0$ then $\mu(\sum_{n\geq 0} f_n = \sum_{n\geq 0} \mu(f_n)$

(ii) Assume that $\sum_{n\geq 0} |f_n|$ is μ-integrable, then $\sum_{n\geq 0} f_n(x)$ is μ-integrable and $\mu(\sum_{n\geq 0} f_n) = \sum_{n\geq 0} \mu(f_n)$

Proof. (i) Set $G_N = \sum_{n=1}^{N} f_n$ and apply MCT

(ii) Set $g = \sum_{n \geq 0} |f_n|$ and observe that $|g_N| \leq g \ \forall \ N$ so DCT implies that $\mu(\lim g_n) = \lim \mu(g_n)$

□

Corollary. (Differentiation under $\int$)
Let $U \subseteq \mathbb{R}$ be an open interval. Let $f : U \times X \to \mathbb{R}$ be a function such that

(i) $\forall \ t \in U \ x \mapsto f(t,x)$ is μ-integrable

(ii) $\forall \ xinX \ t \mapsto f(t,x)$ is differentiable

(iii) (domination) $\exists \ g : X \to \mathbb{R}$ that is μ-integrable

$$\forall \ t \in U, \ \forall \ x \in X \left| \frac{\partial f}{\partial t}(t,x) \right| \leq g(x).$$

Then $x \mapsto \frac{\partial f}{\partial t}(t,x)$ is μ-integrable and setting

$$F(t) := \int_X f(t,x) \mathrm{d}\mu(x)$$

we have that F is differentiable and

$$F'(t) = \int_X \frac{\partial f}{\partial t}(t,x) \mathrm{d}\mu(x).$$

Proof. Let $h_n > 0$ be a sequence of real numbers $h_n \to 0$. We need to show that

$$\frac{F(t+h+n) - F(t)}{h_n} \underset{n \to \infty}{\to} \int_X \frac{\partial f}{\partial t}(t,x) \mathrm{d}\mu(x).$$

By the mean value theorem $\exists \ \theta \in [t, t+h_n]$ such that

$$g_n(t,x) := \frac{f(t+h_n,x) - f(t,x)}{h_n} = \frac{\partial f}{\partial t}(\theta, x).$$

By (iii) we must have that $|g_n(t,x)| \leq g(x)$. But,

- $g_n(t,x) \underset{n \to \infty}{\to} \frac{\partial f}{\partial t}(t,x)$
- $\mu(g_n(t,\cdot)) = \frac{1}{h_n}(F(t+h_n) - F(t))$

So by DCT we have that

$$\mu(\frac{\partial f}{\partial t}(t,\cdot)) = \lim \mu(g_n(t,\cdot)) = F'(t).$$

□

8 Product measures

Definition (Product σ-algebra)**.** Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. We define the*product σ-algebra* denoted $\mathcal{A} \otimes \mathcal{B}$ as the σ-algebra on $X \times Y$ generated by all $A \times B$ where $A \in \mathcal{A}, B \in \mathcal{B}$

Remark. (i) The family $\mathcal{F} = \{A \times B, A \in \mathcal{A} B \in \mathcal{B}\}$ forms a π-system.

(ii) $\mathcal{A} \otimes \mathcal{B}$ is the smallest σ-algebra on $X \times Y$ that makes the projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ measurable.

(iii) $B(\mathbb{R}^{d_1}) \otimes B(\mathbb{R}^{d_2}) = B(\mathbb{R}^{d_1+d_2})$ but it is not true for the Lebesgue σ-algebra (example sheet) as there are more null sets in the product space.

Lemma. (Slices) Let $E \subseteq X \times Y$ be $\mathcal{A} \otimes \mathcal{B}$ measurable. Then the slices $E_X = \{y \in Y | (x, y) \in E\}$ are in $\mathcal{B}$ $\forall$ $x \in X$.

Proof. Let $\xi = \{E \subseteq X \times Y : E_x \in B \ \forall \ x \in X\}$

- Ξ contains all the product sets $A \times B$. $A \in \mathcal{A}$ $B \in \mathcal{B}$.
- Ξ forms a σ-algebra:
  - $E \in \xi \implies E^c \in \xi$ because $(E^c)_x = (E_x)^c$
  - $E_n \in \xi \implies \bigcup_n E_n \in \xi$ because $(\bigcup_n E_n)_x = \bigcup_n (E_n)_x$

Hence $\mathcal{A} \otimes \mathcal{B} \subseteq \xi$ □

Lemma. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. Let $f : X \times Y \to [0, \infty]$ an $\mathcal{A} \otimes \mathcal{B}$ measurable function. Then

(i) $\forall$ $x \in X$ $\quad y \mapsto f(x, y)$ is $\mathcal{B}$-measurable.

(ii) $x \mapsto \int_Y f(x, y) \mathrm{d}\nu y$ is $\mathcal{A}$-measurable

Proof. (i) When $f = 1_E$ with $E \subseteq \mathcal{A} \otimes \mathcal{B}$ then this is the previous lemma. Hence all linear combinations of $\sum_i a_i 1_{E_i}$ are $\mathcal{B}$-measurable in Y. Here (i) hold for simple functions f. But we can approximate an arbitrary f by simple functions so (i) holds for all f as well.

(ii) We can reduce as in (i) to the case when f if an indicator function $f = 1_E$ for some $E \subseteq \mathcal{A} \otimes \mathcal{B}$.

$$\int_Y f(x, y) \mathrm{d}\nu(y) = \int_Y 1_E(x, y) \mathrm{d}\nu(y) = \nu(E_x).$$

We need to show that $x \mapsto \nu(E_x)$ is $\mathcal{A}$-measurable. We write $Y = \bigcup_{m \geq 0} Y_m$, $Y_m \subseteq Y_{m+1}$ with $\nu(Y_m) < \infty$ (ν is σ-finite). It is enough to show that $\forall$ $m : x \mapsto \nu(E_x \cap Y_m)$ is $\mathcal{A}$-measurable. Set

$$\xi = \{E \in \mathcal{A} \otimes \mathcal{B}, x \mapsto \nu(E_x \cap Y_m)\}$$

is $\mathcal{A}$-measurable $\forall$ m

(a) By (i) ξ contains all product sets $A \times B$ for $A \in \mathcal{A}, B \in \mathcal{B}$ indeed if $E = A \times B$

$$\nu(E_x \cap Y_m) = 1_A(x)\nu(B \cap Y_m)$$

is measurable in X

(b) $E \in \xi \implies E^c \in \xi$ because

$$\nu((E^c)_x \cap Y_m) = \underbrace{\nu(Y_m)}_{<\infty} - \nu(E_X \cap Y_m).$$

(c) $\bigcup_n E_n \in \xi$ if $E_n \in \xi \; \forall \; n$ with the E_n pairwise disjoint.

$$\nu\left(\left(\bigcup E_n\right)_x \cap Y_m\right) = \sum_n \nu\left((E_N)_x \cap Y_m\right).$$

Use Dynkin's lemma then $\xi \supseteq \sigma(\pi$-system of product sets $= \mathcal{A} \otimes \mathcal{B}$

□

Proposition. (Product measure) Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \mu)$ be σ-finite measure spaces. There exists a measure σ on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ such that

$$\sigma(A \times B) = \mu(A)\sigma(B), \;\; \forall \; A \in \mathcal{A}, B \in \mathcal{B}.$$

This measure is unique and we denote it by $\mu \otimes \nu$.

Proof. (Existence)
Set

$$\sigma(E) = \int_X \nu(E_x)\mathrm{d}\;\mu(x),$$

this clearly satisfies our property, it is also clearly σ-additive. If

$$E = \bigcup_n E_n \in \mathcal{A} \otimes \mathcal{B}, \;\; E_x = \bigcup_n (E_n)_x,$$

with the E_n pairwise disjoint. So

$$\begin{aligned}
\sigma(\bigcup_n E_n) &= \int_X \sum \nu((E_n)_x)\mathrm{d}\mu(x) \\
&= \sum_{n \geq 0} \int_X \nu((E_n)_x)\mathrm{d}\mu(x) \\
&= \sum_{n \geq 0} \sigma(E_n).
\end{aligned}$$

Uniqueness: $\mathcal{F} = \{A \times B, A \in \mathcal{A}B \in \mathcal{B}\}$ is a π-system and $\sigma(\mathcal{F}) = \mathcal{A} \otimes \mathcal{B}$ by definition. Dynkin's lemma applies and proves the uniqueness of the measure σ □

Remark. $(X, \mathcal{A}, \mu)$ a measure space and a measurable map $f : X \to Y$ where $(Y, \mathcal{B})$ a measurable space, then the image measure is the measure $f_*\mu$ on $(Y, \mathcal{B})$ defined as

$$f_*\mu(\mathcal{B}) := \mu(f^{-1}\mathcal{B}).$$

For probability measures it is easy to see that the projection maps $\pi_X : X \times Y \to X, \pi_Y : X \times Y \to Y$ then

$$(\pi_X)_*(\mu \otimes \nu) = \mu \;\; (\pi_Y)_*(\mu \otimes \nu) = \nu.$$

Remark. The product operation on σ-algebras and on measures is associative.

$$(\mathcal{A} \otimes \mathcal{B}) \otimes \mathcal{C} = \mathcal{A} \otimes (\mathcal{B} \otimes \mathcal{C}).$$

Similarly,

$$(\mu \otimes \nu) \otimes \pi = \mu \otimes (\nu \otimes \pi).$$

m_d = Lebesgue measure on $\mathbb{R}^d$, $m_d = m_1 \otimes \ldots \otimes m_1$ the d-fold product

Theorem (Fubini-Tonelli theorem)**.** Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \mu)$ be measure spaces.

(i) If $f \geq 0$ measurable on $\mathcal{A} \otimes \mathcal{B}$ then

$$\int_{X \times Y} f \mathrm{d}\mu \otimes \nu = \int_X \left[\int_Y f(x, y) \mathrm{d}\nu(y) \right] \mathrm{d}\mu(x) = \int_Y \left[\int_X f(x, y) \mathrm{d}\mu(x) \right] \mathrm{d}\nu(x).$$

(ii) If f is $\mu \otimes \nu$-integrable then μ almost every x $Y \mapsto f(x, y)$ is ν-integrable and $x \to \int_Y f(x, y) \mathrm{d}\nu(y)$

Proof. – The equation holds if $f = 1_E$ where $E \in \mathcal{A} \otimes \mathcal{B}$ because

$$\mu \otimes \nu(E) = \int_X \nu(E_x) \mathrm{d}\mu(x) = \int_Y \mu(E^y) \mathrm{d}\nu(y).$$

Then the equation holds for simple functions and hence for all $f \geq 0$ by MCT.

– Apply (i) to f^+ and f^- where $f = f^+ - f^-$

□

If f is continuous on $[a, b] \to \mathbb{R}$ then $\int_a^x f(t) \mathrm{d}t = \int_R 1_{[a,x] f \mathrm{d}m}$ and $F'(x) = f(x)$. Later we'll see that if f is m-integrable, then $F'(x)$ exists and is m-a.e. and equals $f(x)$

9 Foundations of probability theory

- Ω the universe of all possible outcomes
- $\mathcal{F}$ the family of *events*
- $\mathbb{P}$ a *probability*

  (i) $\mathbb{P}(A) \in [0,1]$
  (ii) $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$
  (iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if A, B disjoint (this makes $\mathbb{P}$ a finitely additive measure on $(\Omega, \mathcal{F})$).
  (iv) The continuity axiom $A_{n+1} \subseteq A_n, A_n \in \mathcal{F}$ with]

$$\bigcap_n A_n = \emptyset \implies \mathbb{P}(A_n) \underset{n\to\infty}{\to} 0.$$

We can define some equivalent objects in measure theory and probability theory

$$\begin{aligned}
(X, \mathcal{A}, \mu) &\leftrightarrow (\Omega, \mathcal{F}, \mathbb{P}) \\
\text{measurable function } f &\leftrightarrow \text{random variable} X \\
\mu\text{-integral} \mu(f) &\leftrightarrow \text{expectation} \mathbb{E}[X] \\
\mu\text{-a.e.} &\leftrightarrow \text{almost surely} .
\end{aligned}$$

Definition. Every random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ determines a Borel probability measure on $\mathbb{R}$ called its *law* of X (or the distribution of X)

$$\mu_X(A) = \mathbb{P}(X \in A), \ A \text{ a Borel subset of } \mathbb{R},$$

where

$$\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in A\}).$$

μ_x is the image measure of $\mathbb{P}$ under $X : \Omega \to \mathbb{R}$, $\mu_x = X_* \mathbb{P}$

Example. (Archimedes) Let $\Omega = S^2 \subseteq \mathbb{R}_3$ pick a point at random on S^2 project orthogonally on the N-S axis $\pi(x)$. What is the probability that $\pi(x)$ is closer to the origin than to the poles? $\frac{1}{2}$

If $f \geq 0$ a measurable function $f : \mathbb{R} \to \mathbb{R}$ then

$$\mathbb{E}[f(x)] = \int_\Omega f \circ X(\omega) \mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} f(x) \mathrm{d}\mu_x(x).$$

Proposition. For $t \in \mathbb{R}$ let $F_X(t) := \mathbb{P}(X \leq t)$. F_X is called the *distribution function* of X. It is non-decreasing and right-continuous. Moreover, F_X determines μ_X uniquely.

Proof. All clear e.g. right continuity $t_n > t$ $t_n \downarrow t$ $F_X(t_n) \to F_X(t)$ since $\{X \leq t\} = \bigcap_n \{X \leq t_n\}$ so by downward monotone convergence for sets

$$\mathbb{P}\left(\bigcap_n \{X \leq t_n\}\right) = \lim \mathbb{P}(X \leq t_N).$$

uniqueness $F_X(t) \to \mu_X((a,b]) = F_X(b) - F_X(a)$ ∎

Definition (Density)**.** We say that a random variable X has a *density* with respect to the Lebesgue measure if $\exists\ f \geq 0$ measurable such that

$$\mu_X((a,b]) = \int_a^b f(t)\mathrm{d}t (= m(f1_{(a,b]}).$$

f is called the *density* of X (or μ_X)

Example. (i) The uniform distribution on $[0,1]$

$$\mathbb{P}(X \in A) = m(A \cap [0,1]).$$

This has density function $f(x) = 1_{[0,1]}(x)$ on $\mathbb{R}$.

$$F_X(t) = \mu_X((-\infty, t]) = \int_{-\infty}^t f(x)\mathrm{d}x.$$

(ii) The exponential distribution with rate $\lambda > 0$. It has density

$$f(x) = \lambda e^{-\lambda x} 1_{x \geq 0}.$$

$$F(t) = \int_{-\infty}^t f(x)\mathrm{d}x = 1_{t\geq 0}(1 - \exp(-\frac{t}{\lambda}).$$

(iii) The *gaussian distribution* with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$ denoted by $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

(iv) *Dirac mass* δ_m for $m \in \mathbb{R}$

$$\delta_m(A) = 1_{m \in A} A \text{ a Borel set in } \mathbb{R}.$$

$$F_{\delta_m}(t) = 1_{t \geq m}.$$

Proposition. Given any function $F : \mathbb{R} \to [0,1]$ such that F is non-decreasing, right-continuous, with $\lim_{t\to-\infty} F(t) = 0$, $\lim_{t\to\infty} = 1$, then there exists a unique Borel probability measure on $\mathbb{R}$ μ_F such that

$$F(t) = \mu_F((-\infty, t])\ \forall\ t \in \mathbb{R}.$$

μ_F is called the Lebesgue-Stieltjes measure of F and we have

$$\mu_F((a,b]) = F(b) - F(a)\ \forall\ a < b.$$

We define "the inverse" of F $g : (0,1) \to \mathbb{R}$

$$y \mapsto \inf\{t \in \mathbb{R}, F(t) \geq y\}.$$

Lemma. g is non-decreasing and $\forall\ t \in \mathbb{R}$

$$\forall\ y \in (0,1) \quad g(y) \leq t \iff F(t) \geq y.$$

Proof. Let $I_y = \{t \in \mathbb{R}, F(t) \geq y\}$. We claim that $I_y = [g(y), \infty]$

- I_y is an interval: $t < t' \implies t'' \in I_y \;\forall\; t'' \in (t.t')$
- Check $g(y) \in I_y$: Let $t_n \in I_y$

$$t_n \to g(y) \implies F(t_n) \geq y \;\forall\; n.$$

  Since F is right continuous, $F(g(y)) \geq y$

Note that this claim proves the lemma. □

We can now define $\mu := g_* m_{|(0,1)}$. μ is a Borel probability measure and by construction

$$\mu((a,b]) = m(g^{-1}((a,b])) = m((F(a), F(b)]) = F(b) - F(a).$$

Question. How to construct a random variable with prescribed law?

Proposition. Let μ be a Borel probability measure on $\mathbb{R}$. Then there exists a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]), \mathbb{P} = m_{|(0,1)}$ such that $\mu_X = m$.

Proof. Let $F(t) = \mu((-\infty, t])$ and let g be the 'inverse' of F set $X(\omega) = g(\omega)$ □

Definition (Mean)**.** We define

- The *mean* of $X : \mathbb{E}[X]$ which is well defined if $X \geq 0$ or if X is integrable
- The *moment of order k* $\mathbb{E}\left[X^k\right]$ well defined if $|X|^k$ is integrable
- The *variance* of X $\operatorname{var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$ which is well defined if the second moment is finite.

Note that $\operatorname{var} X = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$ and also that if $\mathbb{E}\left[X^2\right] < \infty$, then X is integrable.

Proposition. (Cauchy-Schwarz) If X, Y are random variables with $\mathbb{E}\left[X^2\right], \mathbb{E}\left[Y^2\right] < \infty$ then

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]}.$$

Proof.

$$\mathbb{E}\left[(|X| + \lambda|Y|)^2\right] = \mathbb{E}\left[X^2\right] + 2\lambda\mathbb{E}[|XY|] + \lambda^2\mathbb{E}\left[Y^2\right] \geq 0.$$

Then we conclude in the obvious manner. □

10 Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

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

Definition (Independent).
- A sequence of events $(A_i)_{i \geq 1}$ is called (mutually) *independent* if $\forall\ F \subseteq \mathbb{N}$ a finite set of indices

$$\mathbb{P}\left(\bigcap_{i \in F} A_i\right) = \prod_{i \in F} \mathbb{P}(A_i).$$

- A sequence of sub-σ-algebras $\mathcal{A}_i \subseteq \mathcal{F}$ is *independent* if the above holds $\forall\ A_i \in \mathcal{A}_i$

Definition. Given a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ we denote by $\sigma(X)$ the smallest σ-subalgebra of $\mathcal{F}$ making X measurable. In other words

$$\sigma(X) = \sigma\left(\{X^{-1}A, A \text{ Borel subset } \subseteq \mathbb{R}\}\right) = X^{-1}\left(\mathcal{B}(\mathbb{R})\right).$$

Definition. A sequence $(X_i)_{i \geq 1}$ of random variables is called *independent* if the $(\sigma(X_i))_{i \geq 1}$ is independent.

Proposition. A sequence $(X_i)_{i \geq 1}$ of random variables is independent if and only if

$$\forall\ F = \{i_1, \dots, i_k\} \subseteq \mathbb{N}\ \mu(X_{i_1}, \dots, X_{i_k}) = \mu_{X_1} \otimes \cdots \otimes \mu_{X_k}.$$

Where μ_{X_i} is the law of X_i the Borel probability measure on $\mathbb{R}$, and $\mu_{X_1, \dots, X_k)}$ is the law of $(X_1, \dots, X_k)$ is the Borel probability measure on $\mathbb{R}^k$.

Proof.

$$\mathbb{P}\left(\bigcap_{1 \leq j \leq k} \{X_{i_j} \in A_{i_j}\}\right) = \prod_{1 \leq j \leq k} \mathbb{P}\left(X_{i_j} \in A_{i_j}\right) = \prod_{1 \leq j \leq k} \mu_{X_{i_j}}(A_{i_j}).$$

Where we have $A_{i_j} \in \mathcal{B}(\mathbb{R})$ and $(X_{i_1}, \dots, X_{i_k}) \in \prod_1^k A_j = \mathbb{R}^k$ □

Proposition. If X, Y are two independent random variables, then

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

If either X and $Y \geq 0$ or if X and Y are integrable.

Proof. This is an instance of the Fubini-Tonnelli theorem where one considers $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with $f(x, y) \mapsto xy$ with $\mathbb{R} \times \mathbb{R}$ endowed with $\mu_X \otimes \mu_Y$.

$$\mathbb{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) \mathrm{d}(\mu_X \otimes \mu_Y)$$

and

$$\int_{\mathbb{R}} f(x, y) \mathrm{d}\mu_X(x) = y\mathbb{E}[X].$$

□

Example. (Bernstein's example) An example of 3 random variables, X, Y, Z such that each of the 3 pairs $(X,Y), (X,Z), (Y,Z)$ are independent, but (X,Y,Z) are not mutually independent. X and Y are independent coin flips $X, Y \in \{0,1\}$ with $\mathbb{P}(X=0)=\frac{1}{2}$. Then define $Z = |X-Y| \in \{0,1\}$. Clearly X and Y are independent, and with a little effort you can show the pairwise independence with Z.

$$\mathbb{P}\left((X,Y,Z)=(1,1,0)\right) = \mathbb{P}\left((X,Y)=(1,1)\right) = \frac{1}{4},$$

and so is clearly not independent.

Lemma. (Borel-Cantelli)

(i) Let $(A_n)_{n\geq 1}$ be a sequence of events on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\sum_{n\geq 1}\mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup A_n) = 0$. Where

$$\limsup A_n = \{\omega \in \Omega | \omega \text{ belongs to infinitely many } A_n'\text{s}\}.$$

(ii) If $\sum_{n\geq 1}\mathbb{P}(A_n) = \infty$ and if $(A_n)_{n\geq 1}$ is mutually independent, then $\mathbb{P}(\limsup A_n) = 1$

Proof. (i) $\mathbb{E}\left[\sum_{n\geq 1} 1_{A_n}\right] = \sum_{n\geq 1}\mathbb{E}[1_{A_n}] = \sum_{n\geq 1}\mathbb{P}(A_n) < \infty$ so a.s. $\sum_{n\geq 1} 1_{A_n}(\omega) < \infty$ this means that $\mathbb{P}(\limsup A_n) = 0$

(ii) $(\limsup A_N)^c = \bigcup_N \bigcap_{n\geq N} A_n^c$ so it is enough to show that $\mathbb{P}\left(\bigcap_{n\geq N} A_n^c\right) = 0 \ \forall \ N$.

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{n\geq N} A_n^c\right) \geq \mathbb{P}\left(\bigcap_{N\leq n\leq M} A_n^c\right) &= \prod_{N\leq n\leq M}\mathbb{P}(A_n^c) \\
&= \prod_{N\leq n\leq M}(1-\mathbb{P}(A_n)) \leq \prod_{N\leq n\leq M}\exp(-\mathbb{P}(A_n)) \\
&= \exp - \sum_{n=N}^{M}\mathbb{P}(A_n) \to 0 \text{ as } M\to\infty.
\end{aligned}$$

Where we used the fact that $1-x \leq \exp(-x)$ if $x \in [0,1]$

□

Example. (decimal expansion) Let $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}, m)$. If we pick $\omega \in \Omega$ then

$$\omega = 0.\epsilon_1\epsilon_2\cdot\epsilon_n\cdots \quad \epsilon_i \in \{0,1,\ldots,9\}.$$

Let $X_i(\omega) = \epsilon_i(\omega)$.
Claim $(X_n)_{n\geq 1}$ is an independent sequence of random variables all distributed uniformly on $\{0,1,\ldots,9\}$. We need to show that

$$\mathbb{P}(X_1=\epsilon_1, X_2=\epsilon_2\ldots X_n=\epsilon_n) = \prod_1^n \mathbb{P}(X_i=\epsilon_i).$$

The set $\{X_n=\epsilon\}$ is a union of 10^{n-1} intervals of size 10^{-n}

Remark. $\omega = \sum_{n\geq 1}\frac{X_n(\omega)}{10^n}$

Remark. $(A_n)_{n\geq 1}$ independent $\implies (A_n^c)_{n\geq 1}$ are independent.

Proposition. If $(\nu_i)_{i\geq 1}$ are a sequence of Borel probability measures on $\mathbb{R}$ then $\exists\ (\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $\exists\ (X_i)_{i\geq 1}$ *independent* random variables on Ω such that X_i has law ν_i for each i.

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space

(i) A sequence of random variables $(X_n)_{n\geq 1}$ is called a *random stochastic process.*

(ii) The σ-algebra $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$ is the nth term of the *filtration* $(\mathcal{F}_n)_{n\geq 1}$ associated to the process $(X_n)_{n\geq 1}$ with $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$.

(iii) The σ-algebra $\tau := \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \ldots,)$ is called the *tail σ-algebra* e.g. the event " $\limsup X_n \geq T$" is a tail event.

Theorem (Kolmogorov 0-1 Law)**.** If $(X_n)_{n\geq 1}$ is a sequence of independent random variables then the tail σ-algebra is trivial in the sense that $\forall\ A \in \tau, \mathbb{P}(A) \in \{0, 1\}$.

Proof. Let $A \in \tau$, let $B \in \sigma(X_1, \ldots, X_n)$ then A and B are independent because $A \in \tau \subseteq (X_{n+1, X_{n+2}, \ldots}$ are independent of $\sigma(X_1, \ldots, X_n)$ so $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, this means that the two measures

$$\mu_1 : B \mapsto \mathbb{P}(A \cap B) \quad \mu_2 : B \mapsto \mathbb{P}(A)\mathbb{P}(B)$$

coincide on $\sigma(X_1, \ldots, X_n)$ so they coincide on the π-system $\bigcup_{n\geq 1} \mathcal{F}_n$ so by the Dynkin lemma they coincide on $\sigma\left(\bigcup_{n\geq 1} \mathcal{F}_n\right) = \sigma(X_1, \ldots, X_n, \ldots)$ but $\tau \subseteq \sigma(X_1, \ldots, X_n, \ldots)$ so they coincide on τ. Hence

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mu_1(A) = \mu_2(A) = \mathbb{P}(A)^2 .$$

Hence $\mathbb{P}(A) \in \{0, 1\}$ □

Example. Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables all distributed according to the same law μ (called an i.i.d. process). Assume that $\forall\ T > 0 \mu([T, +\infty]) > 0$ then a.s. $\limsup X_n = +\infty$.

Proof. By the second Borel-Cantelli lemma $A_n = \{X_n \geq T\} \sum_{n\geq 1} \mathbb{P}(A_n) = \infty \cdot \mu([T, +\infty]) = +\infty$. Hence $\mathbb{P}(\limsup A_n) = 1$ where $\limsup A_n = \{\omega \in \Omega, X_n(\omega) \geq T$ for infinitely many $n\}$ so $\mathbb{P}(\limsup X_N \geq T) = 1$ holds $\forall\ T > 0$. Hence $\mathbb{P}(\limsup X_n = \infty) = 1$ □

Theorem (Strong law of large numbers)**.** Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with common law μ. Assume that $\mathbb{E}[|X_1|] = \int |x| \mathrm{d}\mu(x)$ is finite. Then,

$$\frac{X_1 + \cdots + X_n}{n} \underset{n\to\infty}{\to} \mathbb{E}[X_1] \text{ a.s .}$$

We prove under the additional assumption that $\mathbb{E}\left[X_1^4\right] < \infty$

Remark. This then shows $\mathbb{E}[|X_1|], \mathbb{E}\left[|X_1|^2\right], \mathbb{E}\left[|X_1|^3\right]$ are finite by Cauchy-Schwarz wlog we may assume that $\mathbb{E}[X_1] = 0$ otherwise consider $Y_i = X_i - \mathbb{E}[X_i]$

Proof. Let $S_n = \frac{1}{n}\sum_{i=1}^n X_i$, estimate $\mathbb{E}\left[S_n^4\right]$

$$S_n^4 = \frac{1}{n^4}(X_1 + \cdots + X_n)^4 = \frac{1}{n^4}\sum_{1\le i,j,k,l\le n} X_i X_j X_k X_l.$$

Now

$$\begin{aligned}\mathbb{E}\left[S_n^4\right] &= \frac{1}{n^4}\sum \mathbb{E}\left[X_i X_j X_k X_l\right] \\ &= \frac{1}{n^4}\left(\sum_1^n \mathbb{E}\left[X_i^4\right] + 6\sum_{i<j}\mathbb{E}\left[X_i^2 X_j^2\right]\right).\end{aligned}$$

Where Cauchy-Schwarz gives us that $\mathbb{E}\left[X_i^2 X_j^2\right] \le \sqrt{\mathbb{E}\left[X_i^4\right]\mathbb{E}\left[X_j^4\right]} = \mathbb{E}\left[X_i^4\right]$. So,

$$\mathbb{E}\left[S_n^4\right] \le \frac{1}{n^4}(n + \frac{6}{2}n(n-1))\mathbb{E}\left[X_i^4\right] = O\left(\frac{1}{n^2}\right).$$

Hence $\sum_{n\ge1}$ □

11 Convergence of sequences of random variables

Definition (Weak convergence of measures)**.** Let $(\mu_n)_{n\geq 1}$ and μ Borel probability measures on $\mathbb{R}^d$. Say that $(\mu_n)_{n\geq 1}$ converges weakly to μ if for every bounded and continuous function $f : \mathbb{R}^d \to \mathbb{R}$

$$\lim_{n\to\infty} \mu_n(f) = \mu(f).$$

Example. (i) $\mu_n = \delta_{\frac{1}{n}}, \mu_n \to \delta_0$ because f is continous so $f\left(\frac{1}{n}\right) \to f(0)$

(ii) $\mu_n = N(0, \sigma_n^2)$ and $\sigma_n \to 0$ then $\mu_n \to \delta_0$

$$\mu_n(f) = \int \frac{1}{\sqrt{2\pi\sigma_n^2}} f(x) e^{-\frac{x^2}{2\sigma_n^2}} \mathrm{d}x = \int \frac{1}{\sqrt{2\pi}} f(\sigma_n x) e^{-\frac{x^2}{2}} \mathrm{d}x$$

which converges to $f(0)$ by the dominated convergence theorem.

(iii) $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\frac{i}{n}}$ then $\mu_n \to m1_{[0,1]}$ since

$$\mu_n(f) = \frac{1}{n}\sum_{i=1}^n f\left(\frac{i}{n}\right) \underset{n\to\infty}{\to} \int_0^1 f(x)\mathrm{d}x.$$

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, let $(X_n)_{n\geq 1}$ and X be a ($\mathbb{R}^d$-valued) random variables. We say that $(X_n)_{n\geq 1}$ converges to X

(i) *almost surely* (or a.s.) if for $\mathbb{P}$-a.e. $\omega \in \Omega$

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

(ii) *in probability* if

$$\forall\, \varepsilon > 0 \mathbb{P}\left(\|X_n - X\| > \varepsilon\right) \underset{n\to\infty}{\to} 0.$$

(iii) *in law* if

$$\mu_{X_n} \underset{n\to\infty}{\to} \mu_X \text{ weakly.}$$

Proposition. (i) $\implies$ (ii) $\implies$ (iii)

Proof. (i) $\implies$ (ii)

$$\mathbb{P}\left(\|X_n - X\| > \varepsilon\right) = \mathbb{E}\left[1_{\|X_n - X\| > \varepsilon} > \varepsilon^\omega\right],$$

which by dominated convergence theorem tends to zero.
(ii) $\implies$ (i) Let f be a continuous and bounded function on $\mathbb{R}^d$, then we need to show that $\mu_{X_n}(f) \to \mu_X(f)$

$$|\mu_{X_n}(f) - \mu_X(f)| = |\mathbb{E}\left[f(X_n) - f(x)\right]|.$$

Recall that f is uniformly continuous on bounded subsets of $\mathbb{R}^d$ so $\forall\, \varepsilon > 0\, \exists\, \delta > 0$ such that if $\|x\| < \frac{1}{\varepsilon}$ and $\|y - x\| < \delta$ then $|f(x) - f(y)| < \varepsilon$. So

$$\begin{aligned} |\mathbb{E}\left[f(X_n) - f(x)\right]| &\leq \mathbb{E}\left[|f(X_n) - f(X) 1_{\|x\| < \frac{1}{\varepsilon}} 1_{\|X - X_n\| < \delta}\right] + 2\|f\|_\infty \left(\mathbb{P}\left(\|X\| \geq \frac{1}{\varepsilon}\right) + \mathbb{P}\left(\|X - X_n\| \geq \delta\right)\right) \\ &\leq \varepsilon + +2\|f\|_\infty \left(\mathbb{P}\left(\|X\| \geq \frac{1}{\varepsilon}\right) + \mathbb{P}\left(\|X - X_n\| \geq \delta\right)\right). \end{aligned}$$

So

$$\limsup |\mu_{X_n}(f) - \mu_X(f)| \le \varepsilon + 2\|f\|_\infty \mathbb{P}\left(\|X\| \ge \frac{1}{\varepsilon}\right).$$

Let $\varepsilon \to 0$ and conclude that $\limsup |\mu_{X_n}(f) - \mu(f)| = 0$ $\square$

Example. The converses do not hold.
(ii) $\not\Longrightarrow$ (i) - the moving bump . $X_{k,n} = 1_{[\frac{k}{n},\frac{k+1}{n}]} k = 0, \ldots, n-1$ on $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{L}, m)$. Consider $X_{0,n}, X_{1,n}, \ldots, X_{n-1,n}, X_{0,n+1}, X_{1,n+1}$ and get a single sequence $(Y_m)_{m\ge1}$. We claim that $Y_m \to 0$ in probability but not a.s. $\forall\, \omega \in (0,1)\, \lim_{m\to\infty} Y_m(\omega)$ does not exist. But,

$$\mathbb{P}(|Y_m| > \varepsilon) = \mathbb{P}(Y_m = 1)\, \frac{1}{n(m)} \underset{n\to\infty}{\to} 0$$

for $Y_m = X_{k(m),n(m)}$ and $\varepsilon \in (0,1)$.
(iii) $\not\Longrightarrow$ (ii) Pick $(X_n)_{n\ge1}$ some i.i.d. random variables with common distribution μ, take any other random variable with law μ. then $\mu_{X_n} = \mu = \mu_Y$ so $X_n \to Y$ in law but $\mathbb{P}(\|X_n - Y\| > \varepsilon)$ is independent of n so it cannot tend to zero.

Lemma. If $(X_n)_{n\ge1}$ converges in probability to X, then $\exists$ a subsequence $(n_k')_{k\ge1}$ such that $X_{n_k} \overset{\text{a.s.}}{\underset{n\to\infty}{\to}} X$

Proof. By assumption, $\forall\, \varepsilon > 0 \mathbb{P}(\|X_n - X\| > \varepsilon) \underset{n\to\infty}{to} 0$. So $\forall\, k \in \mathbb{N}\, \exists\, n_k$ such that $\mathbb{P}\left(\|X_{n_k} - X\| > \frac{1}{k}\right) \le \frac{1}{2^k}$. Now,

$$\sum_{k\ge0} \mathbb{P}\left(\|X_{n_k} - X\| > \frac{1}{k}\right) \le \sum_{k\ge1} \frac{1}{2^k} < \infty.$$

By the Borelli-Cantelli lemma, $\mathbb{P}\left(\limsup\{\|X_{n_k} - X\| > \frac{1}{k}\}\right) = 0$ where the lim sup is

$$\{\omega \in \Omega : \|X_{n_k}(\omega) - X(\omega)\| > \frac{1}{k} \text{ for infinitely many k}\} \supseteq \{\omega \in \Omega |\limsup_k \|X_{n_k} - X\| \ne 0\}.$$

Hence for a.e. $\omega, X_{n_k}(\omega) \underset{k\to\infty}{\to} X(\omega)$ $\square$

Definition. Let $(X_n)_{n\ge1}$ and X be integrable random variables. We say that $X_n \underset{n\to\infty}{\to} X$ in L^1 if

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

Proposition. If $X_n \to X$ in L^1, then $X_n \to X$ in probability

Proof.

$$\mathbb{P}(\|X_n - X\| > \varepsilon) \le \frac{1}{\varepsilon}\mathbb{E}[\|X_n - X\|] \underset{n\to\infty}{\to} 0.$$

$\square$

Remark. (2 classical probability inequalities)

- Markov inequality - if $X \ge 0$ a random variable and $t \ge 0$, then $\mathbb{P}(X \ge t) \le \frac{1}{t}\mathbb{E}[X]$

Proof.
$$\mathbb{E}[X] \geq \mathbb{E}[X 1_{x\geq t}] \geq t\mathbb{E}[1_{X\geq t}] = t \cdot \mathbb{P}(X \geq t).$$
□

– Chebyshev's inequality - if X any random variable then
$$\mathbb{P}(|X - \mathbb{E}[X]| > t) \leq \frac{1}{t^2} \operatorname{var} X.$$

Proof. Apply Markov's to $Y = |X - \mathbb{E}[X]|^2$ □

Definition (Uniformly integrable)**.** Let $(X_n)_{n\geq 1}$ be a sequence of integrable random variables ($\mathbb{R}^d$-value). We say that $(X_n)_{n\geq 1}$ is *uniformly integrable* (U.I.) if
$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[\|X_n\| 1_{\|X_n\|>M}\right] = 0.$$

Example. (i) If $(X_n)_{n\geq 1}$ are dominated (i.e. $\exists$ an integrable Y such that $\|X_n\| \leq Y \ \forall\ n$) then $(X_n)_{n\geq 1}$ is U.I. Indeed, $\mathbb{E}\left[\|X_n\| 1_{\|X_n\|>M}\right] \leq \mathbb{E}[Y 1_{Y>M}] \to 0$ as $M \to \infty$ by DCT.

(ii) If $(X_n)_{n\geq 1}$ are bounded in L^p for some $p > 1$,
$$\exists\ p > 1, C > 0 \mathbb{E}[\|X_n\|^p] \leq C \ \forall\ n$$
then $(X_n)_{n\geq 1}$ is U.I. Indeed
$$\mathbb{E}\left[\|X_n\| 1_{\|X_n\|>M}\right] \leq \frac{1}{M^{p-1}} \mathbb{E}[\|X_n\|^p] \leq \frac{C}{M^{p-1}} \underset{M\to\infty}{\to} 0.$$

Example. $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{L}, m)$ we take $X_n(\omega) = n 1_{[0,\frac{1}{n}]}$ so $X_n \to 0$ in probability but not in L^1 because $\mathbb{E}[X_n] = 1$

Theorem. Let $(X_n)_{n\geq 1}$ be a sequence of integrable random variables. Let X be another random variable. Then TFAE

(i) X is integrable and $X_n \to X$ in L^1

(ii) $(X_n)_{n\geq 1}$ is U.I. and $X_n \to X$ in probability

Lemma. If Y is an integrable random variable and $(X_n)_{n\geq 1}$ is U.I. , then $(Y + X_n)_{n\geq 1}$ is also U.I.

Proof. $\mathbb{E}\left[\|X_n + Y\| 1_{\|X_n+Y\|\geq M}\right] \leq \mathbb{E}[(\|X_n\| + \|Y\|)] 1_{\|X_n+Y\|\geq M}(1_{\|X_n\|\geq \frac{n}{2}} + 1_{\|X_n\|\leq \frac{M}{2}}$ set $x_n = \|X_n\|, y = \|Y\|$. Observe that
$$\begin{aligned} 1_{\|X_n+Y\|\geq M} \cdot 1_{\|X_n\|\leq \frac{M}{2}} &\leq 1_{\|Y\|\leq \frac{M}{2}} \\ &\leq \mathbb{E}\left[(x_n + y) 1_{x_n \geq \frac{M}{2}}\right] + \mathbb{E}\left[2y 1_{y\geq \frac{M}{2}}\right] \\ &\leq \mathbb{E}\left[2x_n 1_{x_n \geq \frac{M}{2}}\right] + 3\mathbb{E}\left[y 1_{y\geq \frac{M}{2}}\right] \\ &\leq 0 + 0. \end{aligned}$$

Following from U.I. and Y integrable respectively. □

Proof. (i) $\implies$ (ii) We've seen that convergence in L^1 implies convergence in probability. Need only check that $(X_n)_{n\geq 1}$ is U.I. by the lemma it is enough to check that $(X_n - X)_{n\geq 1}$ is U.I. But this is clear as

$$\forall\ M\ \forall\ n\ \mathbb{E}\left[\|X_n - X\|1_{\|X_n - X\|\geq M}\right] \leq \mathbb{E}\left[\|X_n - X\|\right] \underset{n\to\infty}{\to}.$$

(ii) $\implies$ (i) We've seen that as $X_n \to X$ in probability $\exists\ \{n_k\}_K$ a subsequence converging a.s. to X. Let us first show that X is integrable. Fatou's lemma implies

.

$$\begin{aligned}
\mathbb{E}\left[\|X\|1_{\|X\|\geq M}\right] &= \mathbb{E}\left[\lim_k \|X_{n_k}\|1_{\|X_{n_k}\geq M\|}\right] \\
&\leq \liminf_k \mathbb{E}\left[\|X_{n_k}\|1_{\|X_{n_k}\|\geq M}\right] \\
&\leq \limsup_k \mathbb{E}\left[\|X_{n_k}\|1_{\|X_{n_k}\|\geq M}\right] \underset{M\to\infty}{\to}.
\end{aligned}$$

So

$$\mathbb{E}\left[\|X\|\right] \leq \mathbb{E}\left[\|X\|1_{\|X\|<M}\right] + \mathbb{E}\left[\|X\|1_{\|X\|\geq M}\right],$$

hence X is integrable. Now by contradiction if $X_n \not\to X$ in L^1, then $\exists$ subsequence $\{n_k\}_k$ and $\exists\ \varepsilon > 0$ such that $\mathbb{E}\left[\|X_{n_k} - X\|\right] > \varepsilon\ \forall\ k$. But $X_{n_k} \to X$ in probability so up to passing to a finer subsequence if necessary wlog we may assume $X_{n_k} \underset{k\to\infty}{\to} X$ a.s. By the lemma $X_{n_k} - X$ is U.I. Hence $\exists\ M > 0$ such that

$$\limsup_k \mathbb{E}\left[\|X_{n_k} - X\|1_{\|X_{n_k} - X\|>M}\right] < \varepsilon.$$

Then

$$\mathbb{E}\left[\|X_{n_k} - X\|\right] = \mathbb{E}\left[\|X_{n_k} - X\|1_{\|X_{n_k} - X\|\leq M}\right] + \mathbb{E}\left[\|X_{n_k} - X\|1_{\|X_{n_k} - X\|>M}\right].$$

The first term tends to 0 by DCT and the second term is $< \varepsilon$ for k large. Which is a contradiction. $\square$

12 L^p-spaces

Jensen inequality Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be an integrable random variable $X : \Omega : I \subseteq \mathbb{R}$ (I is an open interval). Let ϕ be a convex function on I. Then

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

Definition (Convex)**.** We say that a function ϕ is *convex* if

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y) \ \forall\ x, y \in I \ \forall\ t \in [0, 1].$$

Lemma. ϕ is convex $\iff$ $\exists$ family F of affine functions $\ell(x) = ax + b$ such that $\phi = \sup_{\ell \in F} \ell$

Proof. Clearly a supremum of convex functions is convex. Conversely, if ϕ is convex then we claim that

$$\forall\ x_0 \in I \ \exists\ \ell_{x_0}(x) = a_0 x + b_0 : \phi \geq \ell_{x_0} \text{ on } I \text{ and } \phi(x_0) = \ell_{x_0}(x_0)$$

(this shows what we want with $F = \{\ell_{x_0}, x_0 \in I\}$). Observe that $\forall\ x, y\ x < x_0 < y$

$$\frac{\phi(x_0) - \phi(x)}{x_0 - x} \leq \frac{\phi(y) - \phi(x_0)}{y - x_0}.$$

Set $t = \frac{x_0 - x}{y - x}$ the inequality is then equivalent to $\phi(x_0) \leq t\phi(x) + (1-t)\phi(y)$. Pick

$$a_0 \in \left[\frac{\phi(x_0) - \phi(x)}{x_0 - x}, \frac{\phi(y) - \phi(x_0)}{y - x_0}\right]$$

so this holds $\forall\ x < x_0$ and $y > x_0$ and set $\ell_{x_0}(u) = a_0(u - x_0) + \phi(x_0)$ then we get $\phi \geq \ell_{x_0}$ on I and $\phi(x_0) = \ell_{x_0}(x_0)$. $\square$

Proof. First observe that $\phi(x)^-$ is integrable because $\phi = \sup_{\ell \in F} \ell$ so

$$\phi(x)^- = (-\phi(x))^+ = (\inf_{\ell} -\ell(x))^+ \leq |\ell(x)| \leq a|X| + b.$$

So we can talk about $\mathbb{E}\left[\phi(X)\right] = \mathbb{E}\left[\phi(X)^+\right] - \mathbb{E}\left[\phi(X)^-\right]$. Now $\phi = \sup_{\ell \in F} \ell$

$$\mathbb{E}\left[\phi(X)\right] \geq \mathbb{E}\left[\ell(X)\right] = \ell(\mathbb{E}\left[X\right]) \ \forall\ \ell \in F.$$

Taking supremums we obtain the desired result. $\square$

Definition. Let $(X, \mathcal{A}, \mu)$ be a measure space and $p \in [1, \infty)$. Let $f : X \to \mathbb{R}$ be measurable. We st

$$\|f\|_p = \left(\int_X |f|^p \mathrm{d}\mu\right) \frac{1}{p},$$

and call it the $\mathbb{L}^p$*-norm of* f. When $p = \infty$,

$$\|f\|_\infty = \text{essup}(|f|) = \inf\{t \in \mathbb{R} : |f(x)| < t \text{ for } \mu - a.e.x\}$$

is called the $\mathbb{L}^\infty$-norm

Example. $(X, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{L}, m)$ and $f(x) = 1_{\{x=0\}}$, them $\sup_{x \in \mathbb{R}} |f(x)| = 1$, but $\text{essup}(|f|) = 0$ so $\|f\|_\infty = 0$

Remark. For any measurable function f, there exists a measurable g such that $f = g$ μ-a.e. and $\sup_{x\in\mathbb{R}} |g(x)| = \|g\|_\infty = \|f\|_\infty$. Indeed, take $g = f1_{\{|f(x)|\leq \|f\|_\infty\}}$

Proposition. (Minkowski inequality) $\forall\ p \in [1,\infty]$, $\forall\ f, g$ measurable on $(X, \mathcal{A}, \mu)$

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. If $p = \infty$ it is obvious from definition. If $p < \infty$, then the inequality is equivalent to

.

$$\begin{aligned}
&\left\| \frac{f}{\|f\|_p + \|g\|_p} + \frac{g}{\|f\|_p + \|g\|_p} \right\|_p \leq 1 \\
\iff &\left\| \frac{\|f\|_p}{\|f\|_p + \|g\|_p}\frac{f}{\|f\|_p} + \frac{\|g\|_p}{\|f\|_p + \|g\|_p}\frac{g}{\|g\|_p} \right\|_p \leq 1 \\
\iff &\|tF + (1-t)G\|_p \leq 1 \\
\iff &\int \|tF + (1-t)G\|^p \mathrm{d}\mu \leq 1.
\end{aligned}$$

where

$$t = \frac{\|f\|_p}{\|f\|_p + \|g\|_p}, F = \frac{f}{\|f\|_p}, G = \frac{g}{\|g\|_p}.$$

The fact $x \mapsto x^p$ is convex if $p \geq 1$ so $\forall\ c \in X$

$$(t|F(x)| + (1-t)|G(x)|)^p \leq t|F(x)|^p + (1-t)|G(x)|^p.$$

So integrating

$$\int |t|F(x)| + (1-t)|G(x)||^p \leq \int t|F(x)|^p + (1-t)|G(x)|^p = 1.$$

□

Proposition. (Holden inequality)
Let f, g be measurable on $(X, \mathcal{A}, \mu)$ any measure space and if $p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} = 1$ then

$$\int_X |fg|\mathrm{d}\mu \leq \|f\|_p \cdot \|g\|_p.$$

(when $p = q = 2$ this is Cauchy-Schwarz) also hold (trivially) if $p = 1$ and $q = \infty$

Notation.

$$\mathcal{L}^p(X, \mathcal{A}, \mu) = \{f : X \to \mathbb{R} \text{ measurable } : \|f\|_p < \infty, \}$$

for $p \in [1, \infty]$. Note that $\mathcal{L}^p(X, \mathcal{A}, \mu)$ is a vector space (because of Minkowski's inequality)

Definition. We sat that 2 measurable functions f, g on $(X, \mathcal{A}, \mu)$ say f and g are *equivalent* (denote $f \equiv g$) if $f(x) = g(x)$ for μ-a.e.x

Lemma. $\equiv$ is an equivalence relation which is stable under addition and multiplication.

Proof.
- $f \equiv g$ and $g \equiv h \implies f \equiv h$
- $f \equiv g, f' \equiv g' \implies f + f' \equiv g + g'$ and $f \cdot f' \equiv g \cdot g'$

□

We denote by $[f]$ the equivalence class of f for $\equiv$

$$[f+g] := [f] + [g], [fg] := [f] \cdot [g], [\lambda f] := \lambda[f].$$

Definition. $\mathbb{L}^p(X, \mathcal{A}, \mu) = \{[f] : f \in \mathcal{L}^p(X, \mathcal{A}, \mu)\} = \mathcal{L}^p \setminus \equiv$. It is a vector space and $\|[f]\|_p$ is well defined as $\|f\|_p$

Proposition. $\mathbb{L}^p$ is a normed vector space when endowed with the $\mathbb{L}^p$-norm $\|[f]\|_p$. Moreover, it is complete, so $\mathbb{L}^p(X, \mathcal{A}, \mu)$ is a Banach space

Proof. $\|f\|_p$ depends only on the class $[f]$, and

- $\|[f] + [g]\|_p \leq \|[f]\|_p + \|[g]\|_p$
- $\|\lambda[f]\|_p = \|[\lambda f]\|_p = |\lambda| \|f\|_p \ \ \forall\ \lambda \in \mathbb{R}$
- $\|[f]\|_p = 0 \implies [f] = 0$ (by positivity of μ-integral)

These properties show that $\mathbb{L}^p$ is a normed vector space. To prove completeness, we need to show a Cauchy sequence must converge. $([f_n])_n$ Cauchy means that

$$\forall\ \varepsilon > 0\ \exists\ n_0\ \forall\ n, m \geq n_0 \|[f_n] - [f_m]\|_p = \|f_n - f_p\| < \varepsilon.$$

Hence $\forall\ k\ \exists\ n_k$ such that $\|f_{n_k} - f_{n_k}\| \leq \frac{1}{2^k}$. Set

$$S_K(x) = \sum_{k=1}^{K} |f_{n_{k+1}}(x) - f_{n_k}(x)|.$$

$S_{K+1} \geq S_K$ so by MCT $\int |S_K|^p \mathrm{d}\mu \underset{K\to\infty}{\to} \int |S_\infty|^p \mathrm{d}\mu$. Moreover,

$$\begin{aligned}\left(\int |S_K|^p \mathrm{d}\mu\right)^{\frac{1}{p}} &= \|S_k\|_p \\ &\leq \sum_{k=1}^{K} \|f_{n_{k+1} - f_{n_k}}\|_p \text{ (Minkowski)} \\ &\leq \sum_{k=1}^{K} \frac{1}{2^k} \leq 1.\end{aligned}$$

So $\|S_\infty\|_p \leq 1$. Hence $S_\infty(x) < \infty$ μ-a.e. Hence for μ-a.e.x. $(f_{n_k}(x))_k$ is a Cauchy sequence in $\mathbb{R}$. $\mathbb{R}$ is complete, so it has a limit say $f(x) := \lim_{k\to\infty} f_{n_k}(x)$. On the complement of this set of x's just set $f(x) = 0$. Then $\forall\ m\ \|f - f_m\|_p \leq \liminf \|f_{n_k} - f_m\|_p$ by Fatou's lemma but this is $\leq \varepsilon(m) \underset{m\to\infty}{\to} 0$ because $(f_n)_n$ was Cauchy. Hence $[f_m] \underset{m\to\infty}{\to} [f]$ in $\mathbb{L}^p(X, \mathcal{A}, \mu)$. □

This proof worked for finite p. When $p = \infty$, do the same argument but replace the MCT and Fatou's lemma by:

Fact. If $f_n \to f$ μ-a.e. then $\|f\|_\infty \leq \liminf_{n\to\infty} \|f_n\|_\infty$

Proof. Pick $t > \liminf \|f_n\|_\infty \implies \exists\ n_k \uparrow \infty : \|f_{n_k}\|_\infty < t\ \forall\ k$. Hence $\forall\ k |f_{n_k}(x)| < t$ for μ-a.e.
Note that $\forall\ k$ μ-a.e. $\iff$ μ-a.e. $\forall\ k$, so μ-a.e. $\forall\ k\ |f_{n_k}(x)| < t$. Hence $|f(x)| = \lim |f_{n_k}(x)| \leq t$ so $\text{essup}|f| = \|f\|_\infty \leq t$ □

Proposition. (Approximating by simple functions) If $p \geq 1$ finite, let V be the linear span of simple functions on $(X, \mathcal{A}, \mu)$ then $V \cap \mathbb{L}^p$ is dense in $\mathbb{L}^p$.

Proof. Wlog assume $f \geq 0$ then we had a lemma $\exists\ g_n$ simple $0 \leq g_n \leq f$ and $g_n \to f$ pointwise. Then $g_n \to f$ in $\mathbb{L}^p$ by DCT □

13 Hilbert spaces

V is a vector space over $\mathbb{C}$

Definition. A *hermitian* inner product on V is a positive definite sesquilinear form on V.

$$\begin{aligned} V \times V &\to \mathbb{C} \\ (x, y) &\mapsto \langle x, y \rangle. \end{aligned}$$

such that (sesquilinear)

- $\langle \alpha x + \beta y \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \ \forall \ x, y, x \in V \ \ \alpha, \beta \in \mathbb{C}$
- $\overline{\langle x, y \rangle} = \langle y, x \rangle$

and (positive definite)

- $\langle x, x \rangle \geq 0$ and it's zero iff $x = 0$

If V is a real vector space, you call $\langle \cdot, \cdot \rangle$ a Euclidean inner product.

Example.
- $V = \mathbb{C}^d$ $\langle x, y \rangle := \sum_{i=1}^d x_i \overline{y}_i$ the standard Hermitian inner product on $\mathbb{C}^d$
- $V = \mathbb{R}^d$ $\langle x, y \rangle := \sum_{i=1}^d x_i y_i$ the standard Euclidean inner product on $\mathbb{R}^d$

Notation. $\|x\| = \sqrt{\langle x, x \rangle}$ is called the *Hermitian norm* associated to $\langle \cdot, \cdot \rangle$.

Proposition.
(i) $\|\alpha x\| = |\alpha| \|X\| \ \forall \ \alpha \in \mathbb{C}, \ \forall \ x \in V$

(ii) $|\langle x, y \rangle| \leq \|x\| \cdot \|y\| \ \forall \ x, y \in V$ (Cauchy-Schwarz)

(iii) $\|x + y\| \leq \|x\| + \|y\| \ \forall \ x, y \in V$ (triangle inequality)

(iv) $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \ \forall \ x, y \in V$ (parallelogram identity)

Proof.
(i) Clear

(ii) The usual C-S proof

(iii) $\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x, y \rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\| \|y\|$

(iv) Just expand the inner products.

□

Corollary. $(V, \| \cdot \|)$ is a normed vector space.

Definition (Hilbert space)**.** A *Hilbert space* is a Hermitian inner product space over $\mathbb{C}$ (or a Euclidean inner product space over $\mathbb{R}$) which is complete as a metric space.

Example. $V = \mathbb{L}^2(X, \mathcal{A}, \mu)$, $(X, \mathcal{A}, \mu)$ is a measure space and

$$\langle [f], [g] \rangle = \int_X f \overline{g} \mathrm{d}\mu$$

is a Hilbert space.

Proposition. (Orthogonal projection to convex sets) Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{C}$ be a closed convex subset of $\mathcal{H}$. Then $\forall\, x \in \mathcal{H}\ \exists!\ y \in \mathcal{C} : \|x - y\| = d(x, \mathcal{C})$ where $d(x, \mathcal{C}) = \inf\{d(x, c), c \in \mathcal{C}\}$.

Proof. Pick $c_n \in \mathcal{C} : \|x - c_n\| \to d(x, \mathcal{C})$. We write

$$\|\frac{x - c_n}{2} + \frac{x - c_m}{2}\|^2 + \|\frac{x - c_n}{2} - \frac{x - c_m}{2}\|^2 = 2(\|\frac{x - c_n}{2}\|^2 + \|\frac{x - c_m}{2}\|^2).$$

so

$$\|x - \frac{c_n + c_m}{2}\|^2 + \frac{1}{4}\|c_n - c_m\|^2 = \frac{1}{2}(\|x - c_n\|^2 + \|x - c_m\|^2) = d(x, \mathcal{C})^2.$$

Convex implies that $\frac{c_n + c_m}{2} \in \mathcal{C} \geq d(x, \mathcal{C})^2$. Hence as $n, m \uparrow \infty, \|c_n - c_m\| \to 0$, so $(c_n)_n$ is a Cauchy sequence. $\mathcal{H}$ is a Hilbert space, hence complete, so $(c_n)_n$ converges to some $c \in \mathcal{H}$. $\mathcal{C}$ is closed, so $c \in \mathcal{C}$. So we've shown $\exists\, c \in \mathcal{C}, d(x, c) = d(x, \mathcal{C})$. For uniqueness, suppose $c_1, c_2 \in \mathcal{C} : \|x - c_i\| = d(x, \mathcal{C})\ i = 1, 2$ write the same parallelogram identity with $c_n = c_1$ and $c_m = c_2$ then we get $c_1 = c_2$ □

Corollary. Let $V \subseteq \mathcal{H}$ be a closed subspace of Hilbert space $\mathcal{H}$ then we have a direct sum decomposition

$$\mathcal{H} = V \oplus V^{\perp},$$

where $V^{\perp} = \{z \in \mathcal{H} : \langle v, z\rangle = 0\ \forall\ v \in V\}$

Remark. Even if V is not closed, $V^{\perp}$ is always closed.

$$x_n \in V^{\perp}, x_n \to x\ \langle x_n, v\rangle = 0\ \forall\ n \implies |\langle x - x_n, v\rangle| \leq \|x - x_n\|\|v\|.$$

Proof. (Of corollary)

- $V \cap V\ \perp = \{0\}$ is clear $\langle v, v\rangle = 0 \implies v = 0$.
- Let $x \in \mathcal{H}$ and let $y \in V$ be the orthogonal projection of x on V (given by the proposition). We need to show that $x - y \in V^{\perp}$ (because then $x = y + x - y$ - a sum of a vector in V and a vector in $V^{\perp}$). By contradiction: suppose not. Then there exists a $z \in V$ such that $\langle x - y, z\rangle \neq 0$ wlog $\exists\, x \in V$ such that $\langle x - y, z\rangle$ is real and > 0. Consider $x - y - tz = x - (y + tz)$. For $t > 0$

  $$\|x - y - tz\|^2 = \|x - y\|^2 - 2t\langle x - y, z\rangle + t^2\|z\|^2 < \|x - y\|^2$$

  if $t > 0$ small enough - this is a contradiction.

□

Definition. A linear form on a Hilbert space is a linear map $\ell : \mathcal{H} \to \mathbb{C}$ i.e.

$$\ell(\alpha x + \beta y) = \alpha\ell(x) + \beta\ell(y)\ \forall\ \alpha, \beta \in \mathbb{C},\ x, y \in \mathcal{H}.$$

A linear form is said to be bounded if $\exists\, C > 0$ such that $|\ell(X)| \leq C \cdot \|x\|\ \forall\, x \in \mathcal{H}$

Exercise. A linear form on a normed vector space is bounded iff it is continuous.

Theorem (Riesz representation theorem)**.** If $\mathcal{H}$ is a Hilbert space and $\ell\mathcal{H} \to \mathbb{C}$ a bounded linear form, then $\exists!\ x_0 \in \mathcal{H}$ such that $\ell(x) = \langle x, x_0\rangle\ \forall\ x \in \mathcal{H}$.

Proof. Uniqueness is clear $\langle x, x_0 \rangle = \langle x, x_1 \rangle \ \forall \ x$ then apply it to $x = x_0 - x_1$ so $\|x_0 - x_1\|^2 = 0$ hence $x_0 = x_1$. Existence follows from the corollary since $\mathcal{H} = \ker(\ell) \oplus (\ker(\ell))^\perp$ $\ker\ell = \{x \in \mathcal{H} : \ell(x) = 0\}$ is a closed subspace. If $\ell \neq 0$ then $\ker\ell \subsetneq \mathcal{H}$ so $\exists \ x_0 \in (\ker\ell)^\perp \setminus \{0\}$. We claim that $(\ker\ell)^\perp = \mathbb{C}x_0$:
Let $x \in (\ker\ell)^\perp$ let α such that $\ell(x) = \alpha\ell(x_0)(\alpha = \frac{\ell(x)}{\ell(x_0)}$ then $\ell(x - \alpha x_0) = 0$ so $x - \alpha x_0 \in \ker\ell \cap (\ker\ell)^\perp = 0$ so $x = \alpha x_0$.
Now finally, write

$$\ell(x) - \langle x, x_0 \rangle \frac{\ell(x_0)}{\|x_0\|^2} = \ell(x) - \langle x, v_0 \rangle,$$

where $v_0 = x_0 \overline{\frac{\ell(x_0)}{\|x_0\|^2}}$ this vanishes on $\ker\ell$ and on $\ker\ell)^\perp$ so it is identically zero $\square$