Probability and Measure

June 2, 2024

Contents

1 Introduction **3**
- 1.1 Some Questions . . . 3
- 1.2 Basic Definitions . . . 4
- 1.3 Probability Measures . . . 15
- 1.4 Borel-Cantelli Lemmas . . . 16

2 Measurable Functions **18**
- 2.1 Random Variables . . . 22
- 2.2 Rademacher functions . . . 24
- 2.3 Convergence of Random Variables . . . 25

3 Integration **29**
- 3.1 Product Measures . . . 38

4 L^p Spaces, Norms, Inequalities **43**
- 4.1 Inequalities . . . 44

5 Fourier Transforms **55**

6 Ergodic Theory **71**

Index **81**

Lectures

Lecture 1
Lecture 2
Lecture 3
Lecture 4
Lecture 5
Lecture 6
Lecture 7
Lecture 8
Lecture 9
Lecture 10
Lecture 11
Lecture 12
Lecture 13
Lecture 14
Lecture 15
Lecture 16
Lecture 17
Lecture 18
Lecture 19
Lecture 20
Lecture 21
Lecture 22
Lecture 23
Lecture 24

Start of

lecture 1

1 Introduction

1.1 Some Questions

Holes in Classical Theory of Analysis

(1) What is the "volume" of a subset of $\mathbb{R}^d$? $d = 2$ we have "area", $d = 1$ we have "length" (we know the length of intervals).

(2) Integration: Riemann Integral has holes. Let $\{f_n\}$ be a sequence of continuous functions on $[0, 1]$ such that $0 \le f_n(x) \le 1$ for all $x \in [0, 1]$, $f_n(x)$ is monotonically decreasing as $n \to \infty$, i.e. $f_n(x) \ge f_{n+1}(x)$ for all x. So, $\lim_{n\to\infty} f_n(x) = f(x)$ exists for all x. So $\lim_{n\to\infty} \int_0^1 f_n(x)\mathrm{d}x$ exists. But f is not *necessarily* Riemann integrable. We want a new theory of integrals such that f is integrable, and such that $\lim_{n\to\infty} \int_0^1 f_n(x)\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x$.

(3) Let $L' = \overline{(\mathcal{C}[0, 1], \|\bullet\|_1)}$. If $f \in L'$, is f Riemann integrable? ($\|f\|_1 = \int_{0^1} |f(x)|\mathrm{d}x$). Will have to change the definition of integral. L^2 = a Hilbert space →Fourier Analysis.

Holes in Classical Theory of Probability

(1) Discrete probability has its limitations.

- Toss an unbiased coin 5 times. What is the probability of getting 3 heads? This is a question we know how to answer.
- Take an infinite sequence of coin tosses and an event A that depends on that infinite sequence. How to define $P(A)$? (For example Strong Law of Large Numbers).
- How to draw a point uniformly at random from $[0, 1]$?

Probability needs *axioms* to be made rigorous.

(2) Define expectation ($\mathbb{E}$) for a random variable. Also would want the following: if $0 \le X_n \le 1$ and $X_n \downarrow X$, then $\mathbb{E}X_n \to \mathbb{E}X$.

1.2 Basic Definitions

Definition (σ-algebra)**.** Let E be a set. A σ-algebra $\mathcal{E}$ is a collection of subsets of E such that

- $\emptyset \in \mathcal{E}$
- if $A \in \mathcal{E}$, then $A^c \in \mathcal{E}$ ($A^c = E \setminus A$)
- if $(A_n : n \in \mathbb{N})$, $A_n \in \mathcal{E} \forall n$, then $\bigcup_n A_n \in \mathcal{E}$ too.

$(E, \mathcal{E})$ is called a *measurable space*.

Example. $\mathcal{E} = (\emptyset, E)$ or $\mathcal{E} = \mathcal{P}(E)$ (power set). Typically, we will deal with things somewhere between these extremes.

Remark. Since $\bigcap_n A_n = \left(\bigcup_n A_n^c\right)^c$, any σ-algebra is stable under countable intersections. Also, if $a, B \in \mathcal{E}$, then $B \setminus A = B \cap A^c \in \mathcal{E}$.

Definition (Measure)**.** A *measure* μ on $(E, \mathcal{E})$ is a non-negative function $\mu : \mathcal{E} \to [0, \infty]$ such that

- $\mu(\emptyset) = 0$
- For all sequences A_n, $n \in \mathbb{N}$ with $A_n \in \mathcal{E}$ and all A_n pairwise disjoint, we have countable additivity:
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

We call $(e, \mathcal{E}, \mu)$ a *measure space*.

Remark. Let E be a countable set, with $\mathcal{E} = \mathcal{P}(E)$. Then $\forall A \subset E$, $\mu(A) = \mu\left(\bigcup_{x \in A}\{x\}\right) = \sum_{x \in A} \mu(\{x\}) = \sum_{x \in A} m(x)$, where we define $m : E \to [0, \infty]$ such that $m(x) = \mu(\{x\}$. We call such an m a "mass function" (or pmf in discrete probability), and measures μ are in one-to-one correspondence with mass function m. Here $\mathcal{E} = \mathcal{P}(E)$ and this is the theory in elementary discrete probability (when $\mu(\{x\}) = 1$ for all $x \in E$, μ is called a counting measure, and here $\mu(A) = |A|$ for all $A \subset E$).

For uncountable E however, the story is not so simple and $\mathcal{E} = \mathcal{P}(E)$ is generally not feasible. Instead measures are defined on σ-algebras "*generated*" by a smaller class $\mathcal{A}$ of simple subsets of E.

Definition (Generated σ-algebra)**.** If $\mathcal{A}$ is any collection of subsets of E, we define

$$\sigma(\mathcal{A}) = \{A \subseteq E : A \in \mathcal{E} \forall \ \sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A}\} = \bigcap_{\substack{\mathcal{E} \supseteq \mathcal{A} \\ \mathcal{E} \text{ a } \sigma\text{-algebra}}} \mathcal{E}$$

We call this *the σ-algebra generated by* $\mathcal{A}$. It is the smallest σ-algebra containing $\mathcal{A}$.

Why is $\sigma(\mathcal{A})$ a σ-algebra? Answer: Example Sheet 1 problem 1.

The class $\mathcal{A}$ will usually satisfy some properties too.

Definition (Ring)**.** Let E be a set and $\mathcal{A}$ a collection of subsets of E. $\mathcal{A}$ is called a *ring* if

- $\emptyset \in \mathcal{A}$
- For all $A, B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$ and $B \setminus A \in \mathcal{A}$.

Remark. If $A, B \in \mathcal{A}$, $\mathcal{A}$ a ring, then

$$A \triangle B = (A \setminus B) \cup (B \setminus A) \in \mathcal{A}$$

$$A \cap B = (A \cup B) \setminus (A \triangle B) \in \mathcal{A}$$

Definition (Algebra)**.** $\mathcal{A}$ is called an *algebra* if

- $\emptyset \in \mathcal{A}$
- If $A, B \in \mathcal{A}$, then $A^c \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.

Remark. If $\mathcal{A}$ an algebra and $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$ and $B \setminus A = B \cap A^c \in \mathcal{A}$. So an algebra is also a ring. $\{\emptyset\}$ is a ring, but not an algebra.

The idea:

- Define a set function on a suitable collection $\mathcal{A}$.

- Extend the set function to a measure on $\sigma(\mathcal{A})$ (Caratheodory's Extension theorem).
- Such an extension is unique (Dynkin's lemma).

Start of lecture 2

Definition (Set-function)**.** Let $\mathcal{A}$ be any collection of subsets of E such that $\emptyset \in \mathcal{A}$. A *set-function* μ is a map $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(\emptyset) = 0$. We say

(1) μ is *increasing* if $\mu(A) \le \mu(B)$ for all $A, B \in \mathcal{A}$ such that $A \subset B$.

(2) μ is *additive* if $A\mu(A \cup B) = \mu(A) + \mu(B)$ for all $A, B \in \mathcal{A}$ with A, B disjoint.

(3) μ is *countably additive* if $\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$ for all A_n disjoint such that $A_n, \bigcup_n A_n \in \mathcal{A}$.

(4) μ is *countably subadditive* if $\mu\left(\bigcup_n A_n\right) \le \sum_n \mu(A_n)$ for all $A_n, \bigcup_n A_n \in \mathcal{A}$.

Remark. If μ is a *countably additive* set function on $\mathcal{A}$, and $\mathcal{A}$ *is a ring*, then μ satisfies (Example Sheet 1) all of (1), (2), (3), (4).

Theorem (Caratheodory)**.** Let $\mathcal{A}$ be a *ring of subsets of* E and $\mu : \mathcal{A} \to [0, \infty]$ be a countably additive set function on $\mathcal{A}$. Then μ extends to a measure on $\sigma(\mathcal{A})$.

Proof. For any $B \subseteq E$, define

$$\mu^*(B) = \inf\left(\left\{\sum \mu(A_i) : B \subseteq \bigcup_i A_i, A_i \in \mathcal{A}\right\} \cup \{\infty\}\right)$$

Clearly $\mu^*(\phi) = 0$ and μ^* is increasing. So μ^* is an increasing set function on $\mathcal{P}(E)$.

Call a set $A \subset E$ μ^* measurable if $\forall B \subseteq E$,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

Define $\mathcal{M} = \{A \subset E : A \text{ is } \mu^* \text{ measurable}\}$. Shall show $\mathcal{M}$ is a σ-algebra that contains $\mathcal{A}$, $\mu^*|_{\mathcal{M}}$ is a measure on $\mathcal{M}$ that extends μ (i.e. $\mu^*|_{\mathcal{A}} = \mu$).

Step 1: μ^* is countably subadditive, i.e. if $B \subseteq \bigcup_n B_n$, will show $\mu^*(B) \le \sum_n \mu^*(B_n)$. Nothing to prove if $\mu^*(B_n) = \infty$ for some n. Sow assume $\mu^*(B_n) < \infty$ for all n. For all

$\varepsilon > 0$, there exists $(A_{n,m}) \in \mathcal{A}$ such that $B_n \subset \bigcup_m A_{n,m}$ and $\mu^*(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu(A_{n,m})$. But then, $B \subseteq \bigcup_n B_n \subseteq \bigcup_{n,m} A_{n,m}$ and $(A_{n,m}) \in \mathcal{A}$, so by definition of μ^*,

$$\mu^*(B) \leq \sum_n \sum_m \mu(A_{n,m}) \leq \sum_n \left(\mu^*(B_n) + \frac{\varepsilon}{2^n}\right) = \sum_n \mu^*(B_n) + \varepsilon$$

As $\varepsilon > 0$ is arbitrary, we get the desired result.

Step 2: μ^* extends μ, i.e. for all $A \in \mathcal{A}$, $\mu^*(A) = \mu(A)$. ($\mu^*(A) \leq \mu(A)$ by definition of μ^* as $A \subseteq A$, $A \in \mathcal{A}$). Now we prove $\mu^*(A) \geq \mu(A)$. As μ is countably additive on $A\mathcal{A}$ and $\mathcal{A}$ is a ring, μ is countably sub-additive on $\mathcal{A}$ and increasing (by earlier Remark). Now, let $A \subset \bigcup_n (A \cap A_n)$, so by countable sub-additivity on $\mathcal{A}$,

$$\mu(A) \leq \sum_n \mu(A \cap A_n) \leq \sum_n \mu(A_n)$$

(the second inequality is because μ is increasing). So by taking inf over all such $\{A_n\}$, $\mu(A) \leq \mu^*(A)$.

Step 3: $M \supseteq A$, i.e. $A \in \mathcal{A}$ and $B \subseteq E$, want to show

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

Since $B \subseteq (B \cap A) \cup (B \cap A^c)$, by Step 1 (countable sub additivity), $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$. To prove $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$, without loss of generality assume $\mu^*(B) < \infty$. So again $\forall \varepsilon > 0$, there exists $(A_n) \in \mathcal{A}$ with $B \subseteq \bigcup_n A_n$ such that $\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n)$. Then $(B \cap A) \subseteq \bigcup_n (A_n \cap A)$ and $(B \cap A^c) \subseteq \bigcup_n (A_n \cap A^c)$. So that

$$\mu^*(B \cap A) \leq \sum \mu(A_n \cap A) \qquad \text{and} \qquad \mu^*(B \cap A^c) \leq \sum_n \mu(A_n \cap A^c),$$

so that

$$\mu^*)B \cap A) + \mu^*(B \cap A^c) \leq \sum_n (\mu(A_n \cap A) + \mu(A_n \cap A^c)) = \sum \mu(A_n) \leq \mu^*(B) + \varepsilon$$

As $\varepsilon > 0$ is arbitrary, we get $\mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \mu^*(B)$.

Step 4: $\mathcal{M}$ is an algebra: $\phi \in M$. Also if $A \in \mathcal{M}$, then $A^c \in \mathcal{M}$. Now let $A_1, A_2 \in \mathcal{M}$ and $B \subset E$. Then

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) && (\text{as } A_1 \in \mathcal{M}) \\
&= \mu^*(B \cap A_1 \cap A_2) = \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) && (\text{as } A_2 \in \mathcal{M}) \\
&= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\
&= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c) && (\text{as } A_1 \in \mathcal{M},\ B \cap (A_1 \cap
\end{aligned}$$

So $A_1 \cap A_2 \in \mathcal{M}$.

Step 5: $\mathcal{M}$ is a σ-algebra and $\mu^*|_{\mathcal{M}}$ is a measure (since $\mathcal{M}$ is an algebra, convince yourself), i.e. for any sequence $(A_n) \in \mathcal{M}$ with A_n pairwise disjoint, we want to prove $A := \bigcup_n A_n \in \mathcal{M}$ and $\mu(A) = \sum_n \mu(A_n)$. So, as before for any $B \subseteq E$,

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) && (\text{as } A_1 \in \mathcal{M}) \\
&= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) && (\text{as } A_2 \in \mathcal{M}) \\
&= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\
&\ \ \vdots \\
&= \sum_{i=1}^{n} \mu^*(B \cap A_i) + \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c) \\
&\geq \sum_{i=1}^{n} \mu^*(B \cap A_1) + \mu^*(B \cap A^c)
\end{aligned}$$

The last inequality comes from the fact that μ^* is increasing and since $A = \bigcup_i A_i$, we have $A^c = \cap A_i^c \subseteq A_1^c \cap \cdots \cap A_n^c$.

So as $n \to \infty$,

$$\begin{aligned}
\mu^*(B) &\geq \sum_{i=1}^{\infty} \mu^*(B \cap A_1) + \mu^*(B \cap A^c) \\
&\geq \mu^*(B \cap A) + \mu^*(B \cap A^c)
\end{aligned}$$

Also,

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

is obvious by sub additivity. So $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$, i.e. $A \in \mathcal{M}$. □

Start of lecture 3

To address the uniqueness of extension, we introduce further subclasses of $\mathcal{P}(E)$. Let $\mathcal{A}$ be a collection of subsets of E.

Definition (π-system)**.** $\mathcal{A}$ is called a *π-system* if $\emptyset \in \mathcal{A}$ and $\forall A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$.

Definition (d-system)**.** $\mathcal{A}$ is called a *d-system* (for Dynkin, alternatively called a *λ-system*) if

- $E \in \mathcal{A}$
- $A, B \in \mathcal{A}$ and $A \subseteq B$, then $B \setminus A \in \mathcal{A}$
- $\{A_n\} \in \mathcal{A}$ and $A_n \uparrow A = \bigcup_n A_n$, then $A \in \mathcal{A}$ ($A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$)

Equivalently, $\mathcal{A}$ is a d-system if

- $\emptyset \in \mathcal{A}$
- $A \in \mathcal{A} \implies A^c \in \mathcal{A}$
- $\{A_n\} \in \mathcal{A}$ and A_n's *disjoint*, then $\bigcup_n A_n \in \mathcal{A}$

Proof is on Example Sheet 1.

Fact (Example Sheet 1): A π-system $\mathcal{A}$ which is also a d-system is actually a σ-algebra.

Lemma (Dynkin's lemma)**.** Let $\mathcal{A}$ be a π-system. Then any d-system that contains $\mathcal{A}$, contains also $\sigma(\mathcal{A})$.

Proof. Define

$$\mathcal{D} = \bigcap_{\substack{\overline{\mathcal{D}} \text{ a } d\text{-system} \\ \overline{\mathcal{D}} \supseteq \mathcal{A}}} \overline{\mathcal{D}}$$

Then $\mathcal{D}$ is itself a d-system (proof same as in $\sigma(\mathcal{A})$). We will show that $\mathcal{D}$ is also a π-system, then we are done.

To achieve this, define

$$\mathcal{D}' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D}\ \forall A \in \mathcal{A}\}$$

Then $\mathcal{A} \subseteq \mathcal{D}'$ (as $\mathcal{A}$ is a π-system). $\mathcal{D}'$ is in fact a d-system ($B \in \mathcal{A}$, then for any $A \in \mathcal{A}$, $B \cap A \in \mathcal{A}$ as $\mathcal{A}$ is a π-system, hence $B \in \mathcal{D}$). Fix $A \in \mathcal{A}$.

- Then $E \cap A = A \in \mathcal{A} \subseteq \mathcal{D}$. So $E \in \mathcal{D}'$.
- If $B_1 \subseteq B_2$, and $B_1, B_2 \in \mathcal{D}'$, then $(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A) \in \mathcal{D}$. But by definition, as $B_1, B_2 \in \mathcal{D}'$, we get $B_1 \cap A, B_2 \cap A \in \mathcal{D}$. Also, $B_1 \cap A \subseteq B_2 \cap A$ and $\mathcal{D}$ is a d-system, so $(B_2 \cap A) \setminus (B_1 \cap A) \in \mathcal{D}$. So $(B_2 \setminus B_1) \cap A \in \mathcal{D}$, i.e. $B_2 \setminus B_1 \in \mathcal{D}'$.

- Finally, if $\{B_n\} \in \mathcal{D}'$ and $B_n \uparrow B = \bigcup_n B_n$, then $B_n \cap A \in \mathcal{D}$ and $(B_n \cap A) \uparrow (B \cap A)$ and $\mathcal{D}$ is a d-system, so $B \cap A \in \mathcal{D}$. So $B \in \mathcal{D}'$.

So $\mathcal{D}'$ is a d-system. Also, $\mathcal{D}' \subseteq \mathcal{D}$. But also, $\mathcal{A} \subseteq \mathcal{D}'$ and $\mathcal{D}'$ is a d-system, so $\mathcal{D} \subseteq \mathcal{D}'$ (by minimality of $\mathcal{D}$). So $\mathcal{D} = \mathcal{D}'$, i.e. for all $B \in \mathcal{D}$ and $A \in \mathcal{A}$, $B \cap A \in \mathcal{D}$ $(*)$. Now we can repeat this argument one level higher. Define

$$\mathcal{D}'' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D}\ \forall A \in \mathcal{D}\}$$

Then $\mathcal{A} \subseteq \mathcal{D}''$ (by $(*)$). Then check, as before, that $\mathcal{D}''$ is a d-system. So $\mathcal{D}'' = \mathcal{D}$. But then (by the definition of $\mathcal{D}''$), $\forall B \in \mathcal{D}$, $A \in \mathcal{D} \implies B \cap A \in \mathcal{D}$, i.e. $\mathcal{D}$ is a π-system. So $\mathcal{D}$ is a σ-algebra containing $\mathcal{A}$, hence $\mathcal{D} \supseteq \sigma(\mathcal{A})$ (check that $\emptyset \in \mathcal{D}$). □

Now we get the *uniqueness theorem*.

Theorem (Uniqueness of Extension)**.** Let μ_1, μ_2 be some measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$. Suppose that $\mu_1 = \mu_2$ on $\mathcal{A}$ for some π-system $\mathcal{A}$ that generates $\mathcal{E}$ (i.e. $\sigma(\mathcal{A}) = \mathcal{E}$). Then $\mu_1 = \mu_2$ on $\mathcal{E}$.

Proof. Consider $\mathcal{D} = \{A \in \mathcal{E} : \mu_1(A) = \mu_2(A)\}$. Then, by hypothesis, $\mathcal{A} \subseteq \mathcal{D}$. Also, $\mathcal{A}$ is a π-system. So enough to show that $\mathcal{D}$ is a d-system (by Dynkin's lemma, $\sigma(\mathcal{A} = \mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{E}$, so $\mathcal{D} = \mathcal{E}$).

- $\emptyset \in \mathcal{D}$ since $\mu_1(\emptyset) = \mu_2(\emptyset) = 0$.
- $A \in \mathcal{D} \implies \mu_1(A) = \mu_2(A)$. So $\mu_1(A^c) = \mu_1(E) - \mu_1(A) = \mu_2(E) - \mu_2(A) = \mu_2(A^c)$. So $A^c \in \mathcal{D}$.
- $\{A_n\} \in \mathcal{D}$ disjoint, then $\mu_1\left(\bigcup_n A_n\right) = \sum \mu_1(A_n) = \sum \mu_2(A_n) = \mu_2\left(\bigcup_n A_n\right)$. So $\bigcup_n A_n \in \mathcal{D}$.

So $\mathcal{D}$ is a d-system. □

Question: How to show all sets of a σ-algebra $\mathcal{E}$ generated by $\mathcal{E}$ has a certain property $\mathcal{P}$?

$$\mathcal{G} = \{A \subseteq E : A \text{ has the property } \mathcal{P}\}$$

and all elements of $\mathcal{A}$ has the property $\mathcal{P}$. Possible methods:

(1) Show that $\mathcal{G}$ is a σ-algebra.

(2) Show that $\mathcal{G}$ is a d-system and $\mathcal{A}$ is a π-system, and use Dynkin's lemma.

(3) MCT (monotone convergence theorem?)

Definition (Borel Sets)**.** Let E be a Hausdorff topological space. The σ-algebra generated by the set of open sets, i.e. $\sigma(\mathcal{A})$ where $\mathcal{A} = \{A \subseteq E : \mathcal{A} \text{ open}\}$ is called the *Borel σ-algebra* $\mathcal{B}(E)$ of E. $\mathcal{B}(\mathbb{R})$ is written as $\mathcal{B}$. A measure μ on $(E, \mathcal{B}(E))$ is called a *Borel measure* on E. If $\mu(K) < \infty$ for all K compact, then μ is called a *Radon* measure on E. If $\mu(E) = 1$, μ is called a *probability measure* on E, and $(E, \mathcal{E}, \mu)$ is called a probability space (usually use $(\Omega, \mathcal{F}, \mathbb{P})$). If $\mu(E) < \infty$, μ is a finite measure on E. If $\exists (E_n)$ in $\mathcal{E}$ such that $\mu(E_n) < \infty$ for all n and $E = \bigcup_n E_n$, then μ is called a *σ-finite measure* (the Uniqueness of Extension holds for σ-finite measures).

Start of lecture 4

Goal: one of the main goals of this course is to construct a Borel measure μ on $\mathcal{B}(\mathbb{R}^d)$ such that such that

$$\mu\left(\prod_{i=1}^{d}(a_i, b_i)\right) = \prod_{i=1}^{d}(b_i - a_i) \qquad a_i < b_i$$

corresponding to the usual notion of volume of rectangles. This measure will be called the *Lebesgue measure* after H. Lebesgue (1902).

We'll first look at $d = 1$.

Theorem. There exists a unique Borel measure μ on $\mathbb{R}$ such that $\forall a, b \in \mathbb{R}$ with $a < b$, $\mu((a, b]) = b - a (\dagger)$. μ is called the Lebesgue measure on $\mathbb{R}$.

Proof. **Existence:** Consider the ring $\mathcal{A}$ of finite unions of disjoint intervals of the form

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n] \qquad a_1 \le b_1 \le a_2 \le b_2 \le \cdots \le a_n \le b_n$$

Note that $\sigma(\mathcal{A}) = \mathcal{B}$ (Example Sheet) (as all open intervals in $\sigma(\mathcal{A})$ and open intervals generate open sets).

Define for such $A \in \mathcal{A}$,

$$\mu(A) = \sum_{i=1}^{n}(b_i - a_i)$$

This agrees with $(\dagger)$ for $(a, b]$. This is additive and well-defined (check). (Note that this is important since representations aren't unique, for example $(0, 2] = (0, 1] \cup (1, 2]$). So, the existence of μ on $\mathcal{B}$ follows from Caratheodory's theorem if we can show that μ is countably additive.

Remark. (Example Sheet) μ a finitely additive set function on a ring $\mathcal{A}$. Then μ is countably additive iff

- $A_n \uparrow A$, $A_n \in \mathcal{A} \implies \mu(A_n) \uparrow \mu(A)$.
- In addition, if μ is finite and $A_n \downarrow A$, $(A_n), A \in \mathcal{A}$, then $\mu(A_n) \downarrow \mu(A)$.

$A_n = [n, \infty)$, $A_n \downarrow \emptyset$.

So, by example sheet, showing μ is countably additive on $\mathcal{A}$ is equivalent to showing the following: If $A_n \in \mathcal{A}$, $A_n \downarrow \emptyset$, then $\mu(A_n) \downarrow 0$, $A_1 \supseteq A_2 \supseteq \cdots$. ($\mu$ finitely additive on a ring $\mathcal{A}$. THen μ is countably additive iff $A_n \downarrow \emptyset$, $\mu(A_1) < \infty$ implies $\mu(A_n) \downarrow 0$).

We shall prove this by contradiction. Suppose $\exists B_n \in \mathcal{A}$, $B_n \downarrow \emptyset$, but for all n, $\mu(B_n) \geq 2\varepsilon > 0$ for some $\varepsilon > 0$. For each n, B_n can be approximated from within by $C_n \in \mathcal{A}$ such that $\overline{C_n} \subseteq B_n$ and $\mu(B_n \setminus C_n) \leq \varepsilon/2^n$. For example,

$$\begin{aligned} B_n &= (a_1, b_1] \cup (a_2, b_2] \\ C_n &= \left(a_1 + \frac{\varepsilon}{2 \cdot 2^n}, b_1\right] \cup \left(a_2 + \frac{\varepsilon}{2 \cdot 2^n}, b_2\right] \end{aligned}$$

Then

$$\begin{aligned} \mu(b_n \setminus C_1 \cap C_2 \cap \cdots \cap C_n) &\leq \mu\left(\bigcup_{i=1}^n (B_i \setminus C_i\right) \\ &\leq \sum_{i=1}^n \mu(B_i \setminus C_i) \\ &\leq \sum_{i=1}^n \frac{\varepsilon}{2^i} \\ &= \varepsilon \end{aligned}$$

Since $\mu(B_n) \geq 2\varepsilon$ and $\mu(B_n \setminus (C_1 \cap \cdots \cap C_n)) \leq \varepsilon$,

$$\implies \mu(C_1 \cap \cdots \cap C_n) \geq \varepsilon \implies C_1 \cap \cdots \cap C_n \neq \emptyset$$

So $K_n = \overline{C_1} \cap \cdots \cap \overline{C_n} \neq \emptyset$. (K_n) is a sequence of decreasing bounded closed sets, each non-empty, so $\bigcap_n K_n \neq \emptyset$ (by completeness of $\mathbb{R}$). But then $\emptyset \neq \bigcap_n K_n \subseteq \bigcap_n B_n = \emptyset$, which is a contradiction.

Uniqueness: For uniqueness, suppose μ, λ are measures on $\mathcal{B}$ such that $(\dagger)$ holds for sets of the form $(a, b]$. Define the truncated measures, for $A \in \mathcal{B}$,

$$\mu_n(A) = \mu(A \cap (n, n+1]) \qquad \text{and} \qquad \lambda_n(A = \lambda(A \cap (n, n+1])$$

Then μ_n and λ_n are probability measures on $\mathcal{B}$ and $\mu_n = \lambda_n$ on the π-system of intervals of the form $(a, b]$, $a < b$, which generates $\mathcal{B}$. So by Uniqueness of Extension, $\lambda_n = \mu_n$ on $\mathcal{B}$. Hence for all $A \in \mathcal{B}$,

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_n (A \cap (n, n+1])\right) \\ &= \sum_n \mu(A \cap (n, n+1]) \\ &= \sum_n \mu_n(A) \\ &= \sum_n \lambda_n(A) \\ &= \lambda(A) \end{aligned}$$

□

Remark.

1. A set $B \in \mathcal{B}$ is called a Lebesgue (λ) null set if $\lambda(B) = 0$. Any singleton set

$$\{x\} = \bigcap_n \left(x - \frac{1}{n}, x\right]$$

is a Lebesgue-null set since $\lambda(\{x\}) = \lim_{n\to\infty} \lambda((x - 1/n, x]) = \lim_{n\to\infty} 1/n = 0$. In particular, $\lambda((a, b)) = \lambda([a, b]) = \lambda([a, b)) = b - a$. Any countable set Q satisfies $\lambda(Q) = 0$. There exist uncountable sets C with $\lambda(C) = 0$ (for example the Cantor set).

2. Lebesgue measure is translation invariant: for $x \in \mathbb{R}$, and $B \in \mathcal{B}$, define $B + x = \{b + x : b \in B\}$, and define $\lambda_x(B) = \lambda(B + x)$. Then $\lambda_x((a, b]) = \lambda((a, b] + x) = \lambda((a + x, b + x]) = b - a = \lambda((a, b])$. So $\lambda_x = \lambda$ on the π-system of intervals, and hence $\lambda_x = \lambda$ on $\mathcal{B}$. Question: Is the Lebesgue measure the only such translation invariant measure on $\mathcal{B}$?

3. Lebesgue-measurable sets: Caratheodory extends λ from $\mathcal{A}$ to not just $\sigma(A) = \mathcal{B}$, but actually to $\mathcal{M}$, the set of outer measurable sets. $\mathcal{M} \supseteq \mathcal{B}$, but how large is $\mathcal{M}$?

$$\mathcal{M} = \{A \cap N : A \in \mathcal{B}, N \subseteq B, B \in \mathcal{B}, \lambda(B) = 0\}$$

(Lebesgue σ-algebra)

Start of

lecture 5

We now show that $\mathcal{B} \subsetneq \mathcal{P}(\mathbb{R})$ (in fact, $\mathcal{M}_{\text{Leb}} \subsetneq \mathcal{P}(\mathbb{R})$). Vitali (1905) first showed such a set exists.

Consider $E = [0, 1)$ with addition modulo 1. Then Lebesgue measure is invariant under this operation, i.e.

$$\lambda(B) = \lambda(B + x)$$

where adding is done modulo 1.

Question: If $B \in \mathcal{B}$, why is $B + x \in \mathcal{B}$? $\mathcal{G} = \{B \in \mathcal{B} : B + x \in \mathcal{B}\}$.

For $x, y \in [0, 1)$, define the equivalence relation $x \sim y$ if $x - y \in \mathbb{Q} \cap [0, 1)$ ($\mathbb{Q} \cap [0, 1)$ a subgroup on $[0, 1)$). Using the Axiom of Choice (uncountable version), we can select a representative from each equivalence class and form the set S. We will show $S \notin \mathcal{B}$.

The sets $(S + q)_{q \in \mathbb{Q} \cap [0,1)}$ are all disjoint and their union is $[0, 1)$, i.e.

$$[0, 1) = \bigcup_{q \in \mathbb{Q} \cap [0,1)} (S + q)$$

Now, if S were a Borel set, so would $S + q$ be, and by translation invariance of λ and countable additivity of λ,

$$1 = \lambda([0, 1]) = \sum_{q \in \mathbb{Q} \cap [0,1)} \lambda(S + q) = \sum_{q \in \mathbb{Q} \cap [0,1)} \lambda(S)$$

giving a contradiction. So $S \notin \mathcal{B}$.

Remark. Extend this proof to show that $S \notin \mathcal{M}_{\text{Leb}}$.

Theorem (Banach-Kuratowski 1929)**.** Assuming the continuum Hypothesis, there does not exist a measure μ on $\mathcal{P}([0, 1])$ such that

$$\mu([0, 1]) = 1 \qquad \text{and} \qquad \mu(\{x\}) = 0 \ \forall x$$

Proof. Dudley (appendix). □

Henceforth, whenever we are on a metric space E, we will work with $\mathcal{B}(E)$, which will be perfectly satisfactory.

1.3 Probability Measures

We usually use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a probability space.

- Ω is the set of possible outcomes of the experiment / sample space.
- $\mathcal{F}$ is the set of events.
- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(A)$ for $A \in \mathcal{F}$ is the probability that the event occurs.

Kolmogorov (1933) set the axioms of probability theory.

Axioms:

(1) $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$.

(2) $\mathbb{P}(E) \geq 0\ \forall E \in \mathcal{F}$.

(3) $\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n)$ for all $A_n \in \mathcal{F}$ disjoint, i.e. $\mathbb{P}$ is a measure.

Remark.

- $\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n)$.
- $A_n \uparrow A \implies \mathbb{P}(A_n) \uparrow \mathbb{P}(A)$.
- $A_n \downarrow A \implies \mathbb{P}(A_n) \downarrow \mathbb{P}(A)$.

Definition (Independence)**.** We say $(A_i, i \in I)$, $A_i \in \mathcal{F}$, are *independent*, if for all finite sets $J \subseteq I$, we have

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

Say that the σ-algebras $(\mathcal{A}_i, i \in I)$, $\mathcal{A}_i \in \mathcal{F}$ for all i, are *independent*, if $(A_i, i \in I)$ is independent for all $A_i \in \mathcal{A}_i$.

Theorem. Let $\mathcal{A}_1$, $\mathcal{A}_2$ be π-systems contained in $\mathcal{F}$ such that

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) \qquad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$$

Proof. Fix $A_1 \in \mathcal{A}_1$, and define for $A \in \sigma(\mathcal{A}_2)$:

$$\mu(A) = \mathbb{P}(A_1 \cap A), \qquad \nu(A) = \mathbb{P}(A_1)\mathbb{P}(A).$$

Then μ and ν are finite measures, and they agree on the π-system $\mathcal{A}_2$. Hence, by Uniqueness of Extension,

$$\mu(A) = \nu(A) \ \forall A \in \sigma(\mathcal{A}_2) \tag{$*$}$$

Repeat the same argument, now by fixing $A_2 \in \sigma(\mathcal{A}_2)$.

$$\mu'(A) = \mathbb{P}(A \cap A_2), \qquad \nu'(A) = \mathbb{P}(A)\mathbb{P}(A_2) \qquad \forall A \in \sigma(\mathcal{A}_1)$$

By $(*)$, $\mu' = \nu'$ on $\mathcal{A}_1$, hence by Uniqueness of Extension, $\mu' = \nu'$ on $\sigma(\mathcal{A}_1)$. □

1.4 Borel-Cantelli Lemmas

Given a sequence of events $(A_n, n \in \mathbb{N})$, we may ask for the probability that infinitely many of them occur.

Definition. For $A_n \in \mathcal{F}\ \forall n$, define

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m \ge n} A_n = \{A_n \text{ infinitely often}\}$$

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m \ge n} A_n = \{A_n \text{ eventually}\}$$

Lemma (Borel-Cantelli Lemma 1)**.** If $\sum \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ infinitely often}) = 0$.

Proof. Fix any $n \in \mathbb{N}$. Then

$$0 \le \mathbb{P}(A_n \text{ infinitely often}) \le \mathbb{P}\left(\bigcup_{m \ge n} A_m\right) \le \sum_{m \ge n} \mathbb{P}(A_m)$$

Take limit as $n \to \infty$. □

Start of

lecture 6

Example.

$$\{A_n \text{ infinite often}\} = \{\text{infinitely many of the } \{A_n\} \text{ occur}\} = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m$$

So if $A_n = \{H \text{ in the } n\text{-th toss}\}$, then

$$\{A_n \text{ infinitely often}\} = \{\text{infinitely many heads}\}$$

Remark. The lemma holds for any measure μ (not just probability measures).

Lemma (Borel Cantelli Lemma 2)**.** Assume the events (A_n) are independent. Then if $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ infinitely often}) = 1$.

Proof. We will use the inequality $1-a \leq e^{-a}$ for all $a \geq 0$. Now, $(A_n)_{n\in\mathbb{N}}$ are independent so $(A_n^c)_{n\in\mathbb{N}}$ are independent. So, for all n and $N \geq n$,

$$0 \leq \mathbb{P}\left(\bigcap_{m=n}^{N} A_m^c\right) = \prod_{m=n}^{N} \mathbb{P}(A_m^c) = \prod_{m=n}^{N} (1 - \mathbb{P}(A_m)) \leq e^{-\sum_{m=n}^{N} \mathbb{P}(A_m)}$$

Taking $N \to \infty$,

$$0 \leq \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) \leq \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{N} A_m^c\right) \leq \lim_{N\to\infty} e^{-\sum_{m=n}^{N} \mathbb{P}(A_m)} = \lim_{n\to\infty} e^{-\sum_{m=n}^{\infty} \mathbb{P}(A_m)} = 0$$

So,

$$\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0$$

i.e.

$$\mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) = 1 \qquad \forall n \tag{$*$}$$

$\bigcup_{m=n}^{\infty} A_m =: B_n$. Then

$$B_n \downarrow \bigcap_n B_n = \bigcap_n \bigcup_{m\geq n} A_m = \{A_n \text{ infinitely often}\}.$$

So, as $\mathbb{P}(B_n) = 1$ for all n (by $(*)$), so $\mathbb{P}(A_n \text{ infinitely often}) = \lim_{n\to\infty} \mathbb{P}(B_n) = 1$. $\square$

Remark. If $(A_n)_{n\in\mathbb{N}}$ independent, then $\{A_n \text{ infinitely often}\}$ is a $0/1$ event. For all "tail events", the probability is $0/1$ (Kolmogorov $0-1$ law, will prove later).

2 Measurable Functions

Definition (measurable function)**.** Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be 2 measurable functions. A map $f : E \to G$ is called *measurable* if $f^{-1}(A) \in \mathcal{E}\ \forall A \in \mathcal{G}$, where $f^{-1}(A)$ is the pre-image of A under f, i.e.

$$f^{-1}(A) = \{x \in E : f(a) \in A\}.$$

When $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we simply say f is measurable. If E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, then f is called Borel.

Remark. Preimages preserve set operations:

$$f^{-1}\left(\bigcup_i A_i\right) = \bigcup_i f^{-1}(A_i) \qquad \text{and} \qquad f^{-1}(G \setminus A) = E \setminus f^{-1}(A).$$

(Checking these is an exercise).

So, $\{f^{-1}(A) : A \in \mathcal{G}\}$ is a σ-algebra on E and $\{A \subset G : f^{-1(A)} \in \mathcal{E}\}$ is a σ-algebra on G. If $\mathcal{G} = \sigma(\mathcal{A})$ and $f^{-1}(A) \in \mathcal{E}\ \forall A \in \mathcal{A}$, then $\{A \subset G : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra containing $\mathcal{A}$, hence it contains $\sigma(\mathcal{A}) = \mathcal{G}$. So f is measurable.

In particular, when $G = \mathbb{R}$, $\mathcal{G} = \mathcal{B}$, then $\mathcal{B} = \sigma(\mathcal{A})$ where $\mathcal{A} = \{-\infty, y] : y \in \mathbb{R}\}$, so f is Borel measurable if and only if $\{x \in E : f(x) \le y\} \in \mathcal{E}\ \forall y \in \mathbb{R}$. If E is a topological space, $f : E \to \mathbb{R}$ continuous, then for $\mathcal{A} = \{U : U \text{ open}\}$, $f^{-1}(E) \in \mathcal{E}$ (as $f^{-1}(U)$ is open). So f is Borel-measurable.

Example. For $A \subseteq E$, the indicator function

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

is measurable if and only if $A \in \mathcal{E}$.

Composition of measurable functions is measurable (easy exercise).

For a family of functions $f_i : E \to G$, $i \in I$, we can make all (f_i) measurable with respect to the σ-algebra

$$\mathcal{E} = \sigma(f_i^{-1}(A) : A \in \mathcal{G}, i \in I).$$

$\mathcal{E}$ is called the σ-algebra generated by $\{f_i\}_{i \in I}$.

Proposition. If $f_1, f_2, \dots$ are measurable $\mathbb{R}$-valued, then

$$f_1 + f_2, \quad f_1 f_2, \quad \inf_n f_n, \quad \sup_n f_n, \quad \liminf_n f_n, \quad \limsup_n f_n$$

are all measurable.

Proof. See Example Sheet 1. $\square$

Theorem (Monotone Class Theorem)**.** Let $(E, \mathcal{E})$ be a measurable space and $\mathcal{A}$ a π-system generating $\mathcal{E}$. Let $\mathcal{V}$ be a vector space of bounded functions $f : E \to \mathbb{R}$ such that

(1) $1 \in \mathcal{V}$ and $\mathbb{1}_A \in \mathcal{V}\ \forall A \in \mathcal{A}$

(2) If $f_n \in \mathcal{V}\ \forall n$ and f bounded with $0 \leq f_n \uparrow f$, then $f \in \mathcal{V}$.

Then $\mathcal{V}$ contains all bounded measurable functions.

Proof. Let $\mathcal{D} = \{A \in \mathcal{E} : \mathbb{1}_A \in \mathcal{V}\}$. Then $\mathcal{D}$ is a d-system. This is because $\mathbb{1} = \mathbb{1}_E \in D$, $\mathbb{1}_{B \setminus A} = \mathbb{1}_B - \mathbb{A} \in \mathcal{V}$, if $A \subseteq B$, as $\mathcal{V}$ is a vector space. If $A_n \in \mathcal{D}$, i.e. $\mathbb{1}_{A_n} \in \mathcal{V}$, $A_n \uparrow A$, then $\mathbb{1}_{A_n} \uparrow \mathbb{1}_A$ so by (2), $\mathbb{1}_A \in \mathcal{V}$, so $A \in \mathcal{D}$.

It contains the π-system $\mathcal{A}$ so by Dynkin's lemma, contains $\sigma(\mathcal{A}) = \mathcal{E}$, so $\mathcal{D} = \mathcal{E}$, i.e. $\mathbb{1}_A \in \mathcal{V}\ \forall A \in \mathcal{E}$. Since $\mathcal{V}$ is a vector space, it contains all finite linear combinations of indicators of measurable sets. So,

$$f_n = 2^{-n} \lfloor 2^n f \rfloor \in \mathcal{V}.$$

Then

$$\begin{aligned} f_n(x) &= 2^{-n} \lfloor 2^n f(x) \rfloor \\ &= 2^{-n} \sum_{j=0}^{n} j \mathbb{1}_{\{2^n f(x) \in [j, j+1)\}} \\ &= 2^{-n} \sum_{j=0}^{K_n} j \mathbb{1}_{\{f^{-1}([j/2^n, (j+1)/2^n))\}} \end{aligned}$$

for some finite K_n since f is bounded. Then $f_n \leq f \leq f_n + 2^{-n}$. So $|f_n - f| \to 0$ as $n \to \infty$ and $f_n \uparrow f$.

So $0 \leq f_n \uparrow f, f_n \in \mathcal{V}$, and f is bounded non-negative. So $f \in \mathcal{V}$ by (2). Finally, for any f bounded measurable, $f = f^+ - f^-$, where $f^+ = \max(f, 0)$, $f^- = \max(-f, 0)$. f^+, f^- are bounded non-negative measurable and $\in \mathcal{V}$. So since $\mathcal{V}$ is a vector space, $f \in \mathcal{V}$. $\square$

Start of

lecture 7

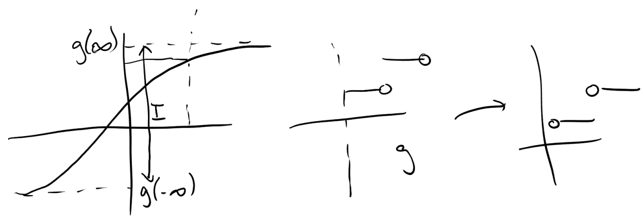
Definition (image / measure)**.** Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be 2 measurable spaces, $f : E \to G$ measurable, and μ a measure on $(E, \mathcal{E})$. Then μ induces an *image / pull-forward measure* $\mathcal{V}$ on $\mathcal{G}$ given by $\mathcal{V} = \mu \circ f^{-1}$, i.e. $\mathcal{V}(A) = \mu(f^{-1}(A))\ \forall A \in \mathcal{G}$. This is well-defined and $\mathcal{V}$ is a measure (Example Sheet 1).

Note. Starting from Lebesgue measure, we can get all probability measures (Radon measures) in this way.

Lemma. Let $g : \mathbb{R} \to \mathbb{R}$ be non-constant, right continuous, and increasing. Set $G(\pm\infty) = \lim_{x \to \pm\infty} g(x)$ and $I = (g(-\infty), g(\infty))$. Define $f : I \to \mathbb{R}$ by $f(x) = \inf\{y \in \mathbb{R} : g(y) \ge x\}$. Then f is left continuous, increasing and $\forall x \in I, y \in \mathbb{R}$,

$$f(x) \le y \iff x \le g(y) \qquad (f(x) > y \iff x > g(y))$$

f is called a generalised inverse of g (if $I = (0, 1)$ then f is the quantile function).



Proof. Fix $x \in I$. Define $J_x = \{y \in \mathbb{R} : g(y) \ge x\}$. Then J_x is non-empty and bounded below and hence $f(x) \in \mathbb{R}$. Since g if increasing, if $y \in J_x$ and $y' \ge y$, then $g(y') \ge g(y) \ge x$, i.e. $y' \in J_x$. Since g is right continuous, if $y_n \in J_x$, $y_n \downarrow y$, then $g(y) = \lim_{n\to\infty} g(y_n) \ge x$, i.e. $y \in J_x$. So $J_x = [f(x), \infty)$. So $x \le g(y) \iff y \in J_x \iff f(x) \le y$. If $x \le x'$, we have $J_x \supseteq J_{x'}$ (as $y \in J_x \Rightarrow y \in J_{x'}$, as $y \in J'_x \iff g(y) \ge x' \implies g(y) \ge x \implies y \in J_x$). So $[f(x), \infty) \supseteq [f(x'), \infty)$, so $f(x) \le f(x')$, i.e. f is increasing.

To show f is left continuous: Let $x_n \uparrow x$. Then $J_x = \bigcap_n J_{x_n}$, i.e. $[f(x), \infty) = \bigcap_n [f(x_n), \infty)$, so $f(x_n) \to f(x)$. $\square$

Theorem. Let $g : \mathbb{R} \to \mathbb{R}$ as in the lemma. Then there exists a unique Radon measure μ_g on $\mathbb{R}$ such that $\forall a, b \in \mathbb{R}$ with $a < b$,

$$\mu_g((a, b]) = g(b) - g(a)$$

Also, every Radon measure can be obtained in this way.

Remark. The measure μ_g is called the Lebesgre-Stieljtes measure associated with g.

Proof. Define I, f as in the lemma, and let λ be the Lebesgue measure on I. f is Borel measurable since

$$f^{-1}((-\infty, z]) = \{x \in I : f(x) \leq z\} = \{x \in I : x \leq g(z)\} = (g(-\infty), g(z)] \in \mathcal{B}$$

and $\{(-\infty, z] : z \in \mathbb{R}\}$ generate $\mathcal{B}$, hence f is measurable.

Thus, the induced measure $\mu_g = \lambda \circ f^{-1}$ exists on $\mathcal{B}$, where $\mu_g(A) = \lambda(f^{-1}(A))$. Then

$$\begin{aligned}
\mu_g((a, b]) &= \lambda(f^{-1}(a, b]) \\
&= \lambda(\{x : f(x) > a, f(x) \leq b\}) \\
&= \lambda(\{x : x > g(a), x \leq g(b)\}) \\
&= \lambda((g(a), g(b)]) \\
&= g(b) - g(a)
\end{aligned}$$

By the Uniqueness of Extension for σ-finite measures, μ_g is uniquely determined.

Conversely, if $\mathcal{V}$ is any Radon measure on $\mathbb{R}$, define

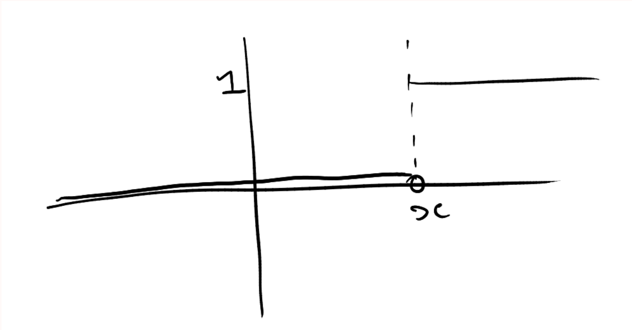
$$g : \mathbb{R} \to \mathbb{R} \quad \text{as} \quad g(y) = \begin{cases} \mathcal{V}((0, y]) & y \geq 0 \\ -\mathcal{V}((y, 0]) & y < 0 \end{cases}$$

$\mathcal{V}$ Radon implies $g : \mathbb{R} \to \mathbb{R}$. Easy to check g is right continuous ($y \geq 0$, $y_n \downarrow y$, then $(0, y_n] \downarrow (0, y]$ and then $\mathcal{V}((0, y_n]) \downarrow \mathcal{V}((0, y])$ by countably additivity, and for $y < 0$, if $y_n \downarrow y$ then use a similar argument). So g is increasing. Lastly,

$$\mathcal{V}((a, b]) = g(b) - g(a)$$

(check cases $0 \in (a, b)$ and $0 \notin (a, b)$. $\square$

Example. Fix $x \in \mathbb{R}$. Take $g = \mathbb{1}_{[x,\infty)}$



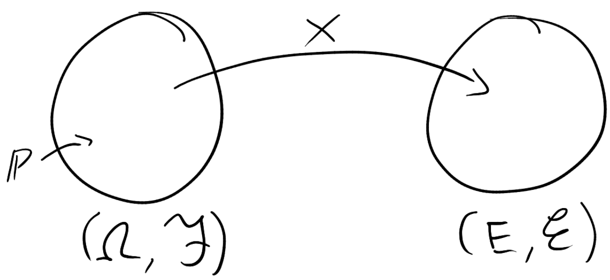
Then $\mu_g = \delta_x$: Dirac measure at x, i.e.

$$\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \forall A \in \mathcal{B}$$

2.1 Random Variables

Definition (random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space. Let $X : \Omega \to E$ a measurable function. Then X is called a *random variable* in E.

X models a "random" outcome of an experiment. For example $\Omega = \{\mathrm{H}, \mathrm{T}\}$, X : # heads : $\Omega \to \{0, 1\}$.



Definition (distribution)**.** The image measure $\mu_X = \mathbb{P} \circ X^{-1}$ is called the *law* or *distribution* of X. It is a measure on $(E, \mathcal{E})$.

If $E = \mathbb{R}$, μ_X is uniquely determined by its values on the π-system $\{(-\infty, x] : x \in \mathbb{R}\}$ given by

$$F_X(x) := \mu_X((-\infty, x]) = \mathbb{P} \circ X^{-1}((-\infty, x]) = \mathbb{P}(\omega \in \Omega : X(\omega) \le x) = \mathbb{P}(X \le x).$$

The function F_X is called the distribution function of X, because it characterises the distribution of X.

By properties of probability measure:

(1) F_X is increasing.

(2) F_X is right continuous ($x_n \downarrow x \implies (-\infty, x_n] \downarrow (-\infty, x]$ hence countability additivity of $\mathbb{P} \circ X^{-1}$)

(3) $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$, $F_X(\infty = \lim_{x \to \infty} F(x) = 1$.

Any $F : \mathbb{R} \to [0, 1]$ satisfying all these properties is called a *distribution function*.

Start of lecture 8

Given any distribution F, there exists a function X such that $F = F_X$, i.e. $F(x) = F_X(x) = \mathbb{P}(X \le x)\ \forall x$ ($\mathbb{P}(\omega \in \Omega : X(\omega) \le x)$.

Proof. Let $\Omega = (0, 1)$ and $\mathbb{P}$ the Lebesgue measure $\lambda|_{(0,1)}$. Let F be any distribution function. Then F is $\uparrow$, right continuous, so we can define

$$X(\omega) = \inf\{x : \omega \le F(x)\} : (0, 1) \to \mathbb{R}$$

Since X is a measurable function, X is a random variable.

$$\forall x, F_X(x) = \mathbb{P}(\le x) = \mathbb{P}(\omega \in \Omega : X(\omega) \le x) = \mathbb{P}(\omega \in \Omega : \omega \le F(x)) = \mathbb{P}((0, F(x)]) = F(x)$$

□

Definition (Independent variables)**.** A (countable) family of random variables $(X_i,\ i \in I)$ is said to be *independent*, if the family of σ-algebras $(\sigma(X_i), i \in I)$ is independent (where recall $\sigma(X) = \sigma\{X^{-1}(A) : A \in \mathcal{E}\}$ for $X : \Omega \to (E, \mathcal{E})$).

Proposition. For a sequence of random variables $(X_n, n \in \mathbb{N})$, this sequence is *independent* if

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \cdots \mathbb{P}(X_n \le x_n) \qquad \forall x_1, \ldots, x_n \in \mathbb{R}, n \in \mathbb{N}.$$

Proof. Example Sheet 1 □

2.2 Rademacher functions

Question: Given a distribution function F, we know there exists a random variable X corresponding to it. But given an infinite sequence of distribution functions $F_1, F_2, \ldots$ does there exist independent random variables $(X_1, X_2, \ldots)$ corresponding to them?

Let $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1), \mathcal{B}(0.1), \lambda|_{(0,1)})$. Any $\omega \in \Omega$ has a binary expansion:

$$\omega = 0 \cdot \underbrace{\omega_1}_{\frac{1}{2}} \underbrace{\omega_2}_{\frac{1}{4}} \underbrace{\omega_3}_{\frac{1}{8}} \cdots, \qquad \omega_i \in \{0,1\}$$

If we exclude representations ending in an infinite sequence of $0s$, then the representation is unique.

Define $R_n : \Omega \to \{0,1\}$ by $R_n(\omega) = \omega_n$, i.e. $R_n = \mathbb{1}_{\{\omega_n = 1\}}$ So

$R_1 = \mathbb{1}_{(1/2,1]}, \quad R_2 = \mathbb{1}_{\{\omega_2=1\}} = \mathbb{1}_{\{\omega_1=0,\omega_2=1\}} + \mathbb{1}_{\{\omega_1=1,\omega_2=1\}} = \mathbb{1}_{(1/4,1/2)} + \mathbb{1}_{(3/4,1]}, \quad R_3 = \cdots$

So R_ns are finite sums of indicators of intervales, hence measurable, i.e. they are random variables. They are called *Rademacher functions*.

Claim: R_i are IID Ber $\left(\frac{1}{2}\right)$. $\mathbb{P}(R_n = 1) = \frac{1}{2} = \mathbb{P}(R_n = 0)\ \forall n$.

$$\mathbb{P}(R_1 = x_1, R_2 = x_2, \ldots, R_n = x_n) = 2^{-n} = \mathbb{P}(R_1 = x_1) \cdots \mathbb{P}(R_n = x_n)$$

and hence $(R_i)_{i \in \mathbb{N}}$ independent.

Now, choose a bijection $m : \mathbb{N}^2 \to \mathbb{N}$ and define $Y_{k,n} = R_{m(k,n)}$ and set $Y_n = \sum_{k=1}^{\infty} 2^{-k} Y_{k,n}$ (converges on $|Y_{k,m}| \le 1$).

Claim: $(Y_n)_n$ are IID $U(0,1)$ (i.e. $\mu_{Y_n} = \lambda|_{(0,1)}$ and (Y_n) are independent). independent is easy (Y_1 is measurable function of $Y_{1,1}, Y_{2,1}, \ldots$, similarly Y_2 is a measurable function of $Y_{1,2}, Y_{2,2}, \ldots$, but note that these two lists are independent). Any measurable functions of independent random variables are independent (check!).

The law of Y_n is identified on the π-systemof intervals $\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]$, $i = 0, 1, \dots, 2^m - 1, m \in \mathbb{N}$. And

$$\begin{aligned}
\mathbb{P}\left(\frac{i}{2^m} < Y_n \le \frac{i+1}{2^m}\right) &= \mathbb{P}\left(\frac{i}{2^m} < \sum_{K=1}^{\infty} 2^{-k} Y_{k,n} \le \frac{i+1}{2^m}\right) \\
&= \mathbb{P}\left(Y_{1,n} = y_1, \dots, Y_{m,n} = y_m\right) \qquad \text{where } \frac{i}{2^m} = 0 \cdot y_1 y_2 \cdots y_m \\
&= \prod_{i=1}^{m} \mathbb{P}(Y_{i,n} = y_i) \\
&= 2^{-m} \\
&= \lambda\left(\frac{i}{2^m}, \frac{i+1}{2^m}\right]
\end{aligned}$$

Hence μ_{Y_n} is $\lambda|_{(0,1)}$, i.e. Y_1 are IID $U(0,1)$. Then, as before, set

$$G_n(x) = F_n^{-}(x) = \inf\{y : x \le F_n(y)\}$$

then G_n's are Borel functions. Set $X_n = G_n(Y_n)$, $n = 1, 2, \dots$. Then as before $F_{X_n} = F_n$ and (X_n) are independent (as $(Y_n)_N$ are independent).

2.3 Convergence of Random Variables

Definition (almost everywhere). $(E, \mathcal{E}, \mu)$ be a measure space. Let $A \in \mathcal{E}$ be defined by some property. We say the property holds almost everywhere (a.e / μ.a.e) if $\mu(A^c) = 0$. If μ is a probability measure, we say almost surely (a.s) if $\mathbb{P}(A^c) = 0$, i.e. $\mathbb{P}(A) = 1$ (w.p.1).

Thus if (f_n), f, $(E, \mathcal{E}, \mu) \to (\mathbb{R}, \mathcal{B})$ measurable, we say

- $f_n \to f$ almost everywhere if

  $$\mu(\{x \in E : f_n(x) \not\to f(x)\}) = 0$$

  for $\mathbb{P}$, almost surely $\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$.

- $f_n \to f$ in (μ-)measure if $\forall \varepsilon > 0$,

  $$\mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) \to 0$$

  as $n \to \infty$, and in ($\mathbb{P}$)-probability if $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$.

Theorem. Let (f_n) be a sequence of measurable functions. Then if $\mu(E) < \infty$, then $f_n \to 0$ almost everywhere $\implies$ $f_n \to 0$ in μ-measure. ($f_n = \mathbb{1}_{(n,\infty)}$ and the Lebesgue measure then $f_n \to 0$ almost everywhere but $\mu(|f_n| > \varepsilon) = \infty$ $\forall n$).

Proof. Fix $\varepsilon > 0$. Suppose $f_n \to 0$ almost everywhere. Then

$$\mu(|f_n| \le \varepsilon) \ge \mu\left(\bigcap_{m=1}^{\infty} \{|f_m| \le \varepsilon\}\right)$$

$$\begin{aligned} &\uparrow \mu(|f_n| \le \text{ eventually}) \\ &\ge \mu(f_n \overset{n\to\infty}{\longrightarrow} 0) \\ &= \mu(E) \\ &< \infty \end{aligned}$$

$$A_n = \bigcap_{m=n}^{\infty} \{|f_m| \le \varepsilon\} \uparrow \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|f_m| \le \varepsilon\}$$

Hence

$$\lim_{n\to\infty} \mu(|f_n| \le \varepsilon) = \mu(E)$$

So, $\lim_{n\to\infty} \mu(|f_n| > \varepsilon) = 0$. □

Start of lecture 9

Theorem. If $f_n \overset{\mu}{\longrightarrow} 0$, then $\exists$ subsequence (n_k) such that $f_{n_k} \to 0$ μ.a.e.

Proof. Suppose $f_n \overset{\mu}{\longrightarrow} 0$. Choosing $\varepsilon = \frac{1}{k}$, then $\mu\left(|f_n| > \frac{1}{l}\right) \to 0$ as $n \to \infty$. We can get n_k such that $\mu\left(|f_{n_k}| > \frac{1}{k}\right) < \frac{1}{k^2}$. We can choose n_{k+1} in the same way (i.e. $\mu\left(|f_{n_{k+1}}| > \frac{1}{k+1}\right) < \frac{1}{(k+1)^2}$, and such that $n_{k+1} > n_k$).

So we get a subsequence (n_k) such that

$$\mu\left(|f_{n_k}| > \frac{1}{k}\right) < \frac{1}{k^2}$$

Also, $\sum_j \frac{1}{k^2} < \infty$.

So $\sum_k \mu\left(|f_{n_k}| > \frac{1}{k}\right) < \infty$. So by Borel-Cantelli Lemma 1,

$$\mu\left(\underbrace{|f_{n_k}| > \frac{1}{k} \text{ i.o}}_{f_{n_k} \not\to 0}\right) = 0$$

$$\implies \mu(f_{n_k} \not\to 0) = 0$$

i.e. $f_{n_k} \to 0$ μ.a.e. □

Remark. Going to a subsequence is necessary, i.e. convergence in μ measure $\not\Longrightarrow$ μ.a.e convergence.

For example, let $(A_n)_{n\in\mathbb{N}}$ be independent events such that $\mathbb{P}(A_n) = \frac{1}{n}$. Let $X_n = \mathbb{1}_{A_n}$. Then $X_n \xrightarrow{\mathbb{P}} 0$ (as $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(A_n) = \frac{1}{n} \to 0\ \forall \varepsilon > 0$). But $\sum_n \mathbb{P}(A_n) = \infty$, i.e. $\sum_n \mathbb{P}(|X_n| > \varepsilon) = \infty$ and $\{|X_n| > \varepsilon\}$ are independent. So by Borel Cantelli Lemma 2 $\implies$ $\mathbb{P}(|X_n| > \varepsilon \text{i.o}) = 1$, hence $X_n \not\to 1$ almost everywhere.

Definition (Convergence in distribution)**.** For $X, (X_n)_n$ a sequence of random variables, we say $X_n \xrightarrow{d} X$ (X_n *converges to* X *in distribution*), if

$$F_{X_n}(t) \overset{n\to\infty}{\longrightarrow} F_X(t)\ \forall t \text{ such that } F_X(t) \text{ is continuous}$$

(this definition does not require (X_n) to be defined on the same probability space).

Remark. If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$ (proof is on Example Sheet 2).

Example. $(X_n)_{n\in\mathbb{N}}$ be IID Exp(1), i.e. $\mathbb{P}(x_n > x) = e^{-x}\ \forall n \in \mathbb{N}, x \geq 0$. Find a deterministic function $g : \mathbb{N} \to \mathbb{R}$ such that almost surely

$$\limsup \frac{X_n}{g(n)} = 1$$

For $\alpha > 1$, $\mathbb{P}(X_n > \alpha \log n) = e^{-\alpha \log n} = n^{-\alpha}$. So $\sum_n \mathbb{P}(X_n > \alpha \log n) < \infty$ if and only if $\alpha > 1$.

For any $\varepsilon > 0$, $\sum_n \mathbb{P}(X_n > (1+\varepsilon)\log n) < \infty$, so by Borel-Cantelli Lemma 1,

$$\mathbb{P}\left(\frac{X_n}{\log n} > 1 + \varepsilon \text{ i.o}\right) = 0$$

Also, $\sum_n \mathbb{P}(X_n > \log n) = \infty$, also $\{X_n > \log n\}$ are independent events (as (X_n) independent), so by Borel Cantelli Lemma 2,

$$\mathbb{P}\left(\frac{X_n}{\log n} > 1 \text{ i.o}\right) = 1$$

So

$$\mathbb{P}\left(\limsup \frac{X_n}{\log n} = 1\right) = 1.$$

Definition (Tail events)**.** Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables. Define

$$\tau_n = \sigma\{X_{n+1}, X_{n+2}, \ldots\}$$

and define $\tau = \bigcap_{n\in\mathbb{N}} \tau_n$.

Then τ is a σ-algebra called the *tail σ-algebra* (contains events that depend only on the "limiting behaviour" of the sequence).

Theorem (Kolmogorov $0-1$ law)**.** Let $(X_n)_n$ be a sequence of independent random variables. Then for the tail σ-algebra τ, if $A \in \tau$, then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. If $Y : (\Omega, \tau) \to \mathbb{R}$ is measurable, then Y is almost surely convergent.

Proof. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n$ is generated by the π-system of sets

$$A = \{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\} \qquad x_1, \ldots, x_n \in \mathbb{R}.$$

and $\tau_n = \sigma(X_{n+1}, \ldots)$ is generated by the π-system of events

$$B = \{X_{n+1} \le x_{n+1}, \ldots, X_{n+k} \le x_{n+k}\}, \qquad x_{n+1}, \ldots, \in \mathbb{R}, k \in \mathbb{N}$$

By by independence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all such A and B. Hence, by an earlier theorem, $\mathcal{F}_n$ and τ_n are independent. But $\tau \subseteq \tau_n$, so $\mathcal{F}_n$ and τ are independent for all n.

Now, consider $\bigcup_n \mathcal{F}_n$ $(\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots)$ is a π-system that generates $\mathcal{F}_\infty := \sigma(X_n, n \in \mathbb{N})$. But $\bigcup_n \mathcal{F}_n$ and τ are independent, so by the theorem again, $\mathcal{F}_\infty$ and τ are independent. But $\tau \subseteq \mathcal{F}_\infty$, so for any $A \in \tau$, $A \in \mathcal{F}_\infty$,

$$\mathbb{P}(A) = \mathbb{P}(\underbrace{A}_{\in\tau} \cap \underbrace{A}_{\in\mathcal{F}_\infty}) = \mathbb{P}(A)\mathbb{P}(A) = \mathbb{P}(A)^2,$$

i.e. $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Finally, if Y is τ measurably, for any $y \in \mathbb{R}$, $\{Y \le y\} \in \tau$, so $\mathbb{P}(Y \le y) = 0$ or 1. Then $c = \inf\{y : \mathbb{P}(Y \le y) = 1\}$, then $\mathbb{P}(Y = c) = 1$. X_i IID, $\mathbb{E}X < \infty$, then

$$\limsup \frac{\sum_{i=1}^n X_i}{n}, \qquad \liminf \frac{\sum_{i=1}^n X_i}{n}$$

are constants almost surely. □

3 Integration

For $(E, \mathcal{E}, \mu)$ a measure space, $f : T \to \mathbb{R}$ measurable and $f \geq 0$, we shall define the integral of f and write it as

$$\mu(f) = \int_E f \mathrm{d}\mu = \int_E f(x) \mathrm{d}\mu(x)$$

When $(E, \mathcal{E}, \mu) = (\mathbb{R}, \mathcal{B}, \lambda)$, we write it as $\int f(x) \mathrm{d}x$.

For $(E, \mathcal{E}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$, and X a random variable, we define its *Expectation*

$$\mathbb{E}(X) = \int_\Omega X \mathrm{d}P = \int_\Omega = X(\omega) \mathrm{d}P(\omega)$$

To start, we say $f : E \to \mathbb{R}$ is *simple* if $f = \sum_{k=1}^m a_K \mathbb{1}_{A_k}$, $0 \leq a_k < \infty$, $A_k \in \mathcal{E}$ $\forall k, m \in \mathbb{N}$. Define for such simple f,

$$\mu(f) = \sum_{k=1}^m a_k \mu(A_k)$$

(where $0 \cdot \infty = 0$). This is well defined (see Example Sheet 2). Check for f, g simple, $\alpha, \beta \geq 0$,

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

(b) $f \leq g \implies \mu(f) \leq \mu(g)$

(c) $f = 0$ almost everywhere $\implies \mu(f) = 0$.

Start of lecture 10

Example. If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$. $X_n \xrightarrow{\mathbb{P}} X$ but $X_n \not\to X$ almost surely $((0,1], \mathcal{B}, \lambda)$.

$$f_1 = \mathbb{1}_{(0,\frac{1}{2}]}, \quad f_2 = \mathbb{1}_{(\frac{1}{2},1]}$$

$$f_3 = \mathbb{1}_{(0,\frac{1}{3}]}, \quad f_4 = \mathbb{1}_{(\frac{1}{3},\frac{2}{3}]}, \quad f_5 = \mathbb{1}_{(\frac{2}{3},1]}$$

$$f_6 = \mathbb{1}_{(0,\frac{1}{4}]}, \dots$$

Then $f_n \to 0$ in λ-measure, but $f_n \not\to 0$ λ almost everywhere. For any $x \in (0,1]$, $(f_n(x))$ has an infinite sequence of 1s, hence $f_n(x) \not\to 0$.

Recall, for f simple, i.e. $f = \sum_{k=1}^m a_k \mathbb{1}_{A_k}$, $a_k \geq 0$, $A_k \in \mathcal{E}$, then

$$\mu(f) = \int f \mathrm{d}\mu = \sum_{k=1}^m a_k \mu(A_k)$$

(recall the properties given last lecture).

Definition (Measure of function)**.** For $f : E \to \mathbb{R}$ measurable, $f \geq 0$, define

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

Clearly, if $0 \leq f_1 \leq f_2$, then $\mu(f_1) \leq \mu(f_2)$.

For general $f : E \to \mathbb{R}$ measurable, $f = f^+ - f^-$, where $f^+ = \max(f, 0)$, $f^- = \max(-f, 0)$ and $|f| = f^+ + f^-$.

Definition (Integrable function)**.** We say f is *integrable* if $\mu(|f|) < \infty$ and then define

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

$(\mu(|f|) = \mu(f^+) + \mu(f^-)$, hence $|\mu(f)| \leq \mu(|f|))$.

If one of $\mu(f^+)$ or $\mu(f^-)$ is ∞ and the other is finite, we define $\mu(f)$ to be ∞ or $-\infty$ respectively (though f is not integrable).

- $x_n \uparrow x$ to mean $x_n \leq x_{n+1}$ $\forall n$, $x_n \to x$
- $f_n \uparrow f$ to mean $f_n(x) \leq f_{n+1}(x)$ $\forall x \in E$ and $f_n(x) \to f(x)$.

Theorem (Monotone Convergence Theorem)**.** Let $(f_n)_n, f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ measurable and non-negative, and suppose $f_n \uparrow f$. Then $\mu(f_n) \uparrow \mu(f)$.

Proof. Recall $\mu(f) = \sup\{\mu(g) : g \leq f, g \text{ simple}\}$. Let $M = \sup_n(\mu(f_n))$, then $\mu(f_n) \uparrow M$. Need to show that $M = \mu(f)$.

Since $f_n \leq f$, $\mu(f_n) \leq \mu(f)$, so by taking sup, $M \leq \mu(f)$.

Now need to show $M \geq \mu(f)$. So it is enough to show $M \geq \mu(g)$ for all g simple, $g \leq f$. Let $g = \sum_{k=1}^m a_k \mathbb{1}_{A_k} \leq f$. Assume without loss of generality, that (A_k)s are disjoint. Define the approximation g_n as

$$g_n(x) = (2^{-n} \lfloor 2^n f_n(x) \rfloor) \underbrace{\wedge}_{\text{min}} g(x)$$

So g_n is simple, $g_n \leq \overline{f_n} \leq f_n \uparrow f$, so $g_n = \overline{f_n} \wedge g \uparrow f \wedge g = g$, i.e. $g_n \uparrow g$, g_n simple, $g_n \leq f_n$.

Fix $\varepsilon \in (0, 1)$ and consider the sets

$$A_k(n) = \{x \in A_k : g_n(x) \geq (1 - \varepsilon)a_k\}$$

Now, $g = a_k$ on the set A_k, and $g_n \uparrow g$, so $A_k(n) \uparrow A_k$, hence $\mu(A_k(n)) \uparrow \mu(A_k)$. Also,

$$\mathbb{1}_{A_k(n)}(1-\varepsilon)a_k \leq \mathbb{1}_{A_k(n)} g_n \leq \mathbb{1}_{A_k} g_n$$

So as $\mu(f)$ is increasing, we have,

$$\begin{aligned} &\mu(\mathbb{1}_{A_k(n)}(1-\varepsilon)a_k) \leq \mu(\mathbb{1}_{A_k} g_n) \\ \implies &(1-\varepsilon)a_k\mu(A_k(n)) \leq \mu(\mathbb{1}_{A_k} g_n) \end{aligned} \tag{$*$}$$

Finally, $g_n = \sum_{k=1}^n \mathbb{1}_{A_k} g_n$ ($g_n \leq g$ and g support on $\bigcup_{k=1}^n A_k$ and (A_k) are disjoint), so

$$\begin{aligned} \mu(g_n) &= \mu\left(\sum_{k=1}^n \mathbb{1}_{A_k} g_n\right) \\ &= \sum_{k=1}^n \mu(\mathbb{1}_{A_k} g_n) \\ &\geq \sum_{k=1}^n (1-\varepsilon)a_k\mu(A_k(n)) && \text{by } (*) \\ &\uparrow \sum_{k=1}^n (1-\varepsilon)a_k\mu(A_k) \\ &= (1-\varepsilon)\mu(g) \end{aligned}$$

Then

$$(1-\varepsilon)\mu(g) \leq \lim_{n\to\infty} \mu(g_n) \underbrace{\leq}_{g_n \leq f_n} \lim_{n\to\infty} \mu(f_n) \leq M$$

i.e. $\mu(g) \leq \frac{M}{1-\varepsilon}$ for all $\varepsilon > 0$, hence $\mu(g) \leq M$. □

Theorem. Let $(f, g) : (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable, non-negative. Then $\forall \alpha, \beta \geq 0$,

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

(b) $f \leq g \implies \mu(f) \leq \mu(g)$

(c) $f = 0$ almost everywhere $\iff \mu(f) = 0$.

Proof. (a) Let $f_n = (2^{-n}\lfloor 2^n f \rfloor) \wedge n$, $g_n = (2^{-n}\lfloor 2^n g \rfloor) \wedge n$. Then, f_n, g_n are simple and $f_n \uparrow f$, $g_n \uparrow g$. Then $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$. So by Monotone Convergence Theorem, $\mu(f_n) \uparrow \mu(f)$, $\mu(g_n) \uparrow \mu(g)$, $\mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g)$.

(b) Obvious from the definition.

(c) If $f = 0$ almost everywhere, then $0 \leq f_n \leq f$, so $f_n = 0$ almost everywhere, but f_n simple $\implies \mu(f_n) = 0$ and $\mu(f_n) \uparrow \mu(f)$ so $\mu(f) = 0$. □

Theorem. Now, let $f, g : (E, \mathcal{E}, \mu) \to \mathbb{R}$ be integrable. Then $\forall \alpha, \beta \in \mathbb{R}$,

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

(b) $f \leq g \implies \mu(f) \leq \mu(g)$

(c) $f = 0$ almost everywhere $\implies \mu(f) = 0$.

Proof. Exercise. Set $f = f^+ - f^-$, $g = g^+ - g^-$, and use definition, $\mu(f) = \mu(f^+) - \mu(f^-)$. If $\mu(f^+) = \mu(f^-)$, then $\mu(f) = 0$ but f need not be 0 almost everywhere. □

Remark.

(1) If $0 \leq f_n \uparrow f$ almost everywhere, then $\mu(f_n) \uparrow \mu(f)$.

(2) Monotone Convergence Theorem $\implies \lim_n \int f_n \mathrm{d}\mu = \int \lim_n f_n \mathrm{d}\mu$ for $0 \leq f_n \uparrow \lim f_n = f$. If $g_n \geq 0$, then (writing $f_n = \sum_{k=1}^n g_k$, $f_n \uparrow f = \sum_{k=1}^\infty g_k$),

$$\begin{aligned} &\implies \lim_n \int \sum_{k=1}^n g_k \mathrm{d}\mu = \int \left(\sum_{k=1}^\infty g_k \right) \mathrm{d}\mu \\ &\implies \sum_{k=1}^\infty \int g_k \mathrm{d}\mu = \int \sum_{k=1}^\infty g_k \mathrm{d}\mu \end{aligned}$$

i.e.

$$\boxed{\sum_{k=1}^\infty \mu(g_k) = \mu\left(\sum_{k=1}^\infty g_k\right)}$$

This generalises the countable additivity of μ to integrals of non-negative functions. In part, if $g_k = \mathbb{1}_{A_k}$ where (A_k) disjoint, implies countable additivity of μ.

Start of lecture 11

- $f \geq 0$, $\mu^*(f) = \sup\{\mu(g), g \text{ simple}, g \leq f\}$. For f simple, ≥ 0, is $\mu^*(f) = \mu(f)$? $\mu^*(f) \geq \mu(f)$ is easy. For any g simple, $g \leq f$, $\mu(g) \leq \mu(f)$, taking supremum of LHS, we get $\mu^*(f) \leq \mu(f)$.

- f measurable, $f \geq 0$, f bounded, then $2^{-n} \lfloor 2^n f \rfloor$ is simple. For unbounded f, we truncate to $f_n = 2^{-n} \lfloor 2^n f \rfloor \wedge n$ (which is simple and has $f_n \uparrow f$).

A general question we begun to explore last time: when can we say that

$$\lim \int f_n \mathrm{d}\mu = \int \lim f_n \mathrm{d}\mu?$$

Example. $f_n = \mathbb{1}_{(n,n+1)}$, $f_n \geq 0$, $f_n \to 0$ as $n \to \infty$. But $\lim_{n\to\infty} \lambda(f_n) = 1 > \lambda(0) = 0$.

Lemma (Fatou's Lemma)**.** Let $f_n : (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable, non-negative. Then

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

Recall that

$$\boxed{\liminf x_n = \sup_m \inf_{k\geq m} x_k}$$

Proof. For $k \geq n$, $\inf_{m\geq n} f_m \leq f_k$. So

$$\mu(\inf_{m\geq n} f_m) \leq \mu(f_k) \quad \forall k \geq n$$

i.e.

$$\mu(\inf_{m\geq n} f_m) \leq \inf_{k\geq n} \mu(f_k) \leq \liminf \mu(f_k) \tag{$*$}$$

Let $g_n = \inf_{m\geq n} f_m$. Then $g_n \geq 0$, and $g_n \uparrow \sup_n g_n = \sup_n \inf_{m\geq n} f_m = \liminf f_n$. So by Monotone Convergence Theorem, $\mu(g_n) \uparrow \mu(\liminf f_n)$. Taking limit in $(*)$, we get

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

Theorem (Dominated Convergence Theorem)**.** Let $(f_n) : (E, \mathcal{E}, \mu) \to \mathbb{R}$ be measurable. Suppose $|f_n| \leq g$ for all n, for some integrable function g (i.e. $\mu(g) < \infty$). Also suppose $f_n \to f$ as $n \to \infty$ on E. Then f, f_n are integrable and

$$\mu(f_n) \to \mu(f)$$

Proof. f is measurable since it is a limit of measurable functions. Also, taking limits of $|f_n| \leq g$, we get $|f| \leq g$. So, $|f_n| \leq g \implies \mu(|f_n|) \leq \mu(g) < \infty$, $\mu(|f|) \leq \mu(g) < \infty$. So f_n, f are integrable.

We have $0 \le g \pm g_n \to g \pm f$, so since $g \pm f_n$ are non-negative, by Fatou's Lemma, we have

$$\mu(g \pm f) \le \liminf \mu(g \pm f_n)$$

i.e.

$$\mu(g) + \mu(f) \le \liminf(\mu(g) + \mu(f_n)) \le \mu(g) + \liminf \mu(f_n)$$

$$\mu(g) - \mu(f) \le \liminf(\mu(g) - \mu(f_n)) \le \mu(g) - \limsup \mu(f_n)$$

As $\mu(g) < \infty$, we get

$$\mu(f) \le \liminf \mu(f_n) \le \limsup \mu(f_n) \le \mu(f)$$

i.e. $\mu(f_n) \to \mu(f)$. □

Remark.

(1) The theorem is still true if we change all the conditions to hold almost everywhere (instead of everywhere).

(2) In fact, $\mu(|f_n - f|) \to 0$ (recall that $\mu(|g|) \ge |\mu(g)|$, which in this case shows $\mu(|f_n - f|) \ge |\mu(f_n) - \mu(f)|$).

We prove this by noting $|f_n - f| \le |f_n| + |f| \le g + g = 2g$ and $2g$ is integrable, so applying Dominated Convergence Theorem we get that $\mu(|f_n - f|) \to 0$.

(3) If $(E, \mathcal{E}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$, $X_n \to X$, $\mathbb{P}$ almost surely, and $|X_n| \le Y$ with $\mathbb{E}Y < \infty$. Then $\mathbb{E}X_n \to \mathbb{E}X$ and $\mathbb{E}|X_n - X| \to 0$. In particular, if $|X_n| \le M$ $\forall n$, for some constant $M > 0$, $M \in \mathbb{R}$, then $\mathbb{E}|X_n - X| \to 0$. (Bounded Convergence Theorem).

(4) f_n measurable on $[0, 1]$ and $|f_n| \le 1$, and $f_n \to f$ pointwise, then $\int f_n \mathrm{d}x \to \int f \mathrm{d}x$ ($\int f_n \mathrm{d}\lambda(x)$). So stronger than Riemann integral as it requires uniform convergence.

Comparisons with Riemann integral

(a) FTC:

(1) Let $f : [a, b] \to \mathbb{R}$ be continuous and set $F(t) = \int_a^t f(x)\mathrm{d}x$. Then F is differentiable on $[a, b]$ with $F' = f$.

(2) Let $F : [a, b] \to \mathbb{R}$ be differentiable and F' is continuous, then $\int_a^b F'(x)\mathrm{d}x = F(b) - F(a)$.

Proofs are the same as before (just use $\int_t^{t+h} \mathrm{d}x = h$).

(a′) If $f : [a,b] \to \mathbb{R}$ is Lebesgue integrable and $F(t) = \int_a^t f(x)\mathrm{d}x$. Then $\lim_{h\to 0} \frac{F(t+h)-F(t)}{h} = \lim_{h\to 0} \frac{\int_t^{t+h} f(x)\mathrm{d}x}{h} = f(t)$ almost everywhere (Lebesgue differentiation Theorem in Analysis of Functions).

(b) Substitution formula: Let $\varphi : [a,b] \to \mathbb{R}$, φ strictly increasing and continuously differentiable. Then for all g Borel measurable, $g \ge 0$ on $[\varphi(q), \varphi(b)]$,

$$\int_{\varphi(a)}^{\varphi(b)} g(y)\mathrm{d}y = \int_a^b g(\varphi(x))\varphi'(x)\mathrm{d}x \qquad (*)$$

Proof. Let $\mathcal{V}$ be the set of all measurable functions g for which $(*)$ this holds. Then by lineariy of integrals, $\mathcal{V}$ is a vector space.

- $\mathbb{1} \in \mathcal{V}$ by FTC(2). holds. Also, $\mathbb{1}_{(c,d]} \in \mathcal{V}$ by FTC(2).
- If $f_n \in \mathcal{V}$, $f_n \uparrow f$, $f_n \ge 0$, then by Monotone Convergence Theorem, $f \in \mathcal{V}$.

Hence by Monotone Class Theorem, $(*)$ holds $\forall g \ge 0$ measurable. □

(c) A (bounded) Riemann integrable (RI) function $f : [a,b] \to \mathbb{R}$ is Lebesgue integrable in the following sense. If f is bounded on $[a,b]$, f is RI if and only if $\mathcal{D} = \{x \in [a,b] : f \text{ is not continuous at } x\}$ has $\lambda(\mathcal{D}) = 0$ (Lebesgue 1904), i.e. f is continuous almost everywhere. Such an f need not be Borel (but is Lebesgue measurable), and can be modified on a Lebesgue measure 0 set to make it Borel, i.e. $\exists \tilde{f}$ Borel such that $\tilde{f} = f$ on A and $\lambda(A^c) = 0$, and $\int \tilde{f}\mathrm{d}x = \int f\mathrm{d}x$ (where we use Lebesgue integral on the left, and Riemann integral on the right).

(d) $\mathbb{1}_{\mathbb{Q}}$ on $[0,1]$. $\mathcal{D} = [0,1]$, $\lambda(\mathcal{D}) \neq 0$, so $\mathbb{1}_{\mathbb{Q}}$ is not Riemann integrable. But $\mathbb{1}_{\mathbb{Q}}$ is Lebesgue integrable and $\mathbb{1}_{\mathbb{Q}} = 0$ λ almost surely, so $\lambda(\mathbb{1}_{\mathbb{Q}}) = 0$.

Start of lecture 12

Theorem (DCT for almost surely convergence)**.** $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and (X_n), X are random variables. Suppose $X_n \to X$ $\mathbb{P}$ almost surely and $|X_n| \le Y$ $\forall n$ for some integrable random variable Y (i.e. $\mathbb{E}Y < \infty$), then

$$\mathbb{E}|X_n - X| \to 0$$

as $n \to \infty$.

Theorem (DCT for in $\mathbb{P}$ convergence)**.** $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and (X_n), X are random variables. SUppose $X_n \to X$ in $\mathbb{P}$ probability, and $|X_n| \leq Y$ $\forall n$ for some integrable random variable Y. Then

$$\mathbb{E}|X_n - X| \to 0$$

as $n \to \infty$.

Proof. Suppose $\mathbb{E}|X_n - X| \not\to 0$. Then there exists a subsequence (n_k) such that $\mathbb{E}|X_{n_k} - X| > \varepsilon$ for all k, for some $\varepsilon > 0$. Now, $X_n \xrightarrow{\mathbb{P}} X$ implies $X_{n_k} \xrightarrow{\mathbb{P}} X$. Hence $\exists (n_{k_l})$ such that $X_{n_{k_l}} \xrightarrow{a.s} X$ and $|X_{n_{k_l}}| \leq Y$. But then by Dominated Convergence Theorem, $\mathbb{E}|X_{n_{k_l}}| \to 0$, but that contradicts the constructed property of (n_k). $\square$

Theorem (BCT for in $\mathbb{P}$ convergence)**.** $X_n \xrightarrow{\mathbb{P}} X$ and $|X_n| \leq M$ for some constant $M > 0$, $\forall n \geq 0$. Then $\mathbb{E}|X_n - X| \to 0$.

Theorem (Differentiation under the integral sign)**.** Let $U \subseteq \mathbb{R}$ be open and $f : U \times E \to \mathbb{R}$ such that

(i) $x \mapsto f(t, x)$ is integrable for all $t \in U$.

(ii) $t \mapsto f(t, x)$ is differentiable $\forall x \in E$.

(iii) $\exists g : E \to \mathbb{R}$ integrable such that $\forall x \in E$, $\forall t \in U$,

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

THen $x \mapsto \frac{\partial f}{\partial t}(t, x)$ is integrable $\forall t$ and $F : U \to \mathbb{R}$ defined by

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

is differentiable and

$$\frac{\mathrm{d}}{\mathrm{d}t} F(t) = \int \frac{\partial f}{\partial t}(t, x) \mu(\mathrm{d}x).$$

Proof. For $h_n \to 0$, set

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

For any fixed t, $g_n(x) \to 0$ for all $x \in E$ (by (ii)), and

$$|g_n(x)| = \left| \underbrace{\frac{\partial f}{\partial t}(\tilde{t},x)}_{\text{MVT}} - \frac{\partial f}{\partial t}(t,x) \right| \overset{\text{(iii)}}{\leq} \leq 2g(x)$$

and $2g$ is integrable. Hence, $\mu(g_n) \to 0$ by Dominated Convergence Theorem, i.e.

$$\int \frac{f(t+h_n)-f(t,x)}{h_n}\mu(\mathrm{d}x) - \int \frac{\partial f}{\partial t}(t,x)\mu(\mathrm{d}x) \to 0$$

i.e.

$$\underbrace{\lim_{n\to\infty} \frac{F(t+h_n)-F(t)}{h_n}}_{=F'(t)} = \int \frac{\partial f}{\partial t}(t,x)\mu(\mathrm{d}x).$$ □

Integrals and image measures

Let $f : (E, \mathcal{E}, \mu) \to (G, \mathcal{G})$ be measurable with the image measure $\nu = \mu \circ f^{-1}$ on $(G, \mathcal{G})$. If $g : (G, \mathcal{G}) \to \mathbb{R}$ measurable, ≥ 0, then

$$\nu(g) = \int_G g(x)\mathrm{d}\nu(x) = \int_G g\mathrm{d}\mu \circ f^{-1} \overset{?}{=} \int_E g(f(x))\mathrm{d}\mu(x)$$

(the ? equality is an exercise on Example Sheet 2). Then

$$\boxed{\mu \circ f^{-1}(g) = \mu(g \circ f)}$$

In part, for $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ measurable, $X \geq 0$, we have

$$\mathbb{E}(g(X)) = \int_\Omega g(X(\omega))\mathrm{d}\mathbb{P}(\omega) = \int g(x)\mathrm{d}\mu_X(x)$$

where $\mu_X = \mathbb{P} \circ X^{-1}$ is the *law* of X.

Densities of measures

For $f : (E, \mathcal{E}, \mu) \to \mathbb{R}$ measurable, ≥ 0, let

$$\nu(A) = \mu(f \mathbb{1}_A) \ \forall A \in \mathcal{E} \qquad \left(= \int_A f \mathrm{d}\mu \right)$$

Then ν is a measure on $(E, \mathcal{E})$ (check).

For any g measurable, $g \geq 0$ on E, $\nu(g) = \mu(fg)$, i.e. $\int g \mathrm{d}\nu = \int gf \mathrm{d}\mu$.

Proof. Holds for indicators by definition, then holds for simple functions by additivity, and for non-negative measurable functions by MCT. ☐

We say that ν has the density f with respect to μ. ($\mu(f\mathbb{1}_A) = \mu(g\mathbb{1}_A)$ $\forall A \in \mathcal{E} \implies f = g$ μ almost everywhere (Example Sheet 2)).

IN part, for $\mu = \lambda$ (Lebesgue measure), $\forall f$ Borel, if $\mu(f\mathbb{1}_A) = 0$ for all A in a π-system generating $\mathcal{E}$, then $f = 0$ almost everywhere.

There exists a Borel measure ν on $\mathbb{R}$ given by $\nu(A) = \int_A f(x)\mathrm{d}x$ and then $\forall g$ Borel, $g \geq 0$, $\nu(g) = \int f(x)g(x)\mathrm{d}x$. We say that ν has density f. This ν is a probability measure on $(\mathbb{R}, \mathcal{B})$ if and only if $\int f(x)\mathrm{d}x = 1$. For $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, if the law μ_X (i.e. $\mathbb{P} \circ X^{-1}$) has the density f_X (with respect to λ), we call f_X the probability density of X. Then

$$\mathbb{P}(X \in A) = \mathbb{P} \circ X^{-1}(A) = \mu_X(A) = \int_A f_X(x)\mathrm{d}x \qquad \forall A \in \mathcal{B} \forall g \text{ Borel}, g \geq 0$$

(taking $A = (-\infty, x]$, $F_X(x) = \int_{-\infty}^{\infty} f_X(x)\mathrm{d}x$. If $F_X' = f_X$, then this holds).

$$= \int g(x)\mathrm{d}\mu_X(x) = \int g(x)f_X(x)\mathrm{d}x$$

3.1 Product Measures

Definition (Product measure)**.** Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be 2 *finite* measure spaces. On the Cartesian product $E := E_1 \times E_2$, we consider the set of 'rectangles'

$$\mathcal{A} = \{A_1 \times A_2 : A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$$

Then $\mathcal{A}$ is a π-system. Define the product σ-algebra $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 := \sigma(\mathcal{A})$.

One can show that if E_i are topological spaces with a countable basis, then

$$\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$$

(where $E_1 \times E_2$ is the product topology on $E_1 \times E_2$; see Dudley for a proof).

Goal: To construct a product measure on $(E_1 \times E_2, \mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2)$.

Start of

lecture 13

Lemma (Lemma 1)**.** Let $f : E \to \mathbb{R}$ be $\mathcal{E}$-measurable. Then $\forall x_1 \in E_1$, the function

$$x_2 \mapsto f(x_1, x_2) : E_2 \to \mathbb{R} \tag{$*$}$$

is $\mathcal{E}_2$-measurable.

Proof. $f = \mathbb{1}_A$ where $A = A_1 \times A_2$.

$$\mathbb{1}_A(x_1, x_2) = \mathbb{1}_{A_1 \times A_2}(x_1, x_2) = \mathbb{1}_{A_1}(x_1) \cdot \mathbb{1}_{A_2}(x_2) = \begin{cases} \mathbb{1}_{A_2}(x_2) & \text{if } x_1 \in A \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathcal{V}$ be the set of all functions for which $(*)$ holds. If $f_n \in \mathcal{V}$ for all n, $f_n \uparrow f$, then $f \in \mathcal{V}$. By Monotone Class Theorem, $\mathcal{V}$ contains all bounded measurable functions f. $f_n = (-n) \vee f \wedge n$, $f_n \in \mathcal{V}$ and $f_n \to f$. So $f \in \mathcal{V}$. □

Lemma (Lemma 2)**.** Let $f : (E, \mathcal{E}) \to \mathbb{R}$ be measurable and

(i) f is bounded or

(ii) f is non-negative.

Then

$$f_1(x_1) = \int_{E_2} f(x_1, x_2) \mu_2(\mathrm{d}x_2), \quad x_1 \in E_1$$

is $\mathcal{E}_1$-measurable and (i) bounded or (ii) non-negative, taking values in $[0, \infty]$.

Remark. A function f taking values in $[0, \infty]$ is measurable means $f^{-1}(\infty) \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{B}$.

Proof. $f = \mathbb{1}_{A_1 \times A_2}$ then $f_1(x_1) = \int \mathbb{1}_{A_1 \times A_2}(x_1, x_2) \mu_2(\mathrm{d}x_2) = \mathbb{1}_{A_1}(x_1) \times \mu_2(A_2)$ is $\mathcal{E}_1$-measurable. $f_n \uparrow f$. Use Monotone Convergence Theorem and limit of measurable functions if measurable. Conclude using the Monotone Class Theorem. □

Theorem. There exists a unique measure $\mu := \mu_1 \otimes \mu_2$ on $\mathcal{E}$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \qquad \forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$$

Proof. Uniqueness obvious as $\mathcal{A}$ is a π-system generating $\mathcal{E}$ and μ is a finite measure. For existence, define the iterated integral

$$\mu(A) = \int_{E_1} \left(\int_{E_2} \mathbb{1}_A(x_1, x_2) \mu_2(\mathrm{d}x_2) \right) \mu_1(\mathrm{d}x_1) \qquad A \in \mathcal{E}$$

This definition makes sense by the previous two lemmas. Clearly, $\mu(\emptyset) = 0$, and $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. μ is countably additive: if (A_i) disjoint, $A = \bigcup i = 1^\infty A_i$, then $\mathbb{1}_A = \sum_{i=1}^\infty \mathbb{1}_{A_i}$, so apply Monotone Convergence Theorem twice. $\square$

Remark. Note that

$$\mu(A) = \int_{E_2} \left(\int_{E_1} \mathbb{1}_A(x_1, x_2) \mu_1(\mathrm{d}x_1) \right) \mu_2(\mathrm{d}x_2)$$

by Dynkin's lemma.

Theorem (Fubini-Tonelli)**.** Consider $(E, \mathcal{E}, \mu) = (E_1 \times E_2, \mathcal{E} \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$, $\mu_i(E_i) < \infty$.

(1) Let $f : E \to \mathbb{R}$ be measurable, $f \geq 0$. Then

$$\mu(f) \overset{(\dagger)}{=} \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(\mathrm{d}x_2) \right) \mu_1(\mathrm{d}x_1) \overset{(*)}{=} \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mu_1(\mathrm{d}x_1) \right) \mu_2(\mathrm{d}x_2)$$

(2) Let $f : E \to \mathbb{R}$ be μ-integrable (i.e. $\int |f| \mathrm{d}\mu < \infty$). If we set

$$A_1 = \{x_1 \in E_1 : \int_{E_2} |f(x_1, x_2)| \mathrm{d}\mu_2(x_2) < \infty\}$$

and define $f_1 : E_1 \to \mathbb{R}$ by $f_1(x_1) = \int_{E_2} f(x_1, x_2) \mathrm{d}\mu_2(x_2)$ for all $x_1 \in A_1$ and 0 otherwise. Then $\mu_1(A_1^c) = 0$ and f_1 is μ_1-integrable and $\mu_1(f_1) = \mu(f)$.

Proof.

(1) The identities $(\dagger)$ and $(*)$ hold for $f = \mathbb{1}_A$ for $A \in \mathcal{E}$, by definition of product measure μ. Hence they extend to simple functions by linearity and for general functions $f \geq 0$ by Monotone Convergence Theorem and approximation by simple functions $f_n = 2^{-n} \lfloor 2^n f \rfloor \wedge n$.

(2) Define $h : E_1 \to [0, \infty]$ as $h(x_1) = \int_{E_2} |f(x_1, x_2)| \mu_2(\mathrm{d}x_2)$. By the Lemma 2, h is measurable (as $|f| \geq 0$), is non-negative, so $A_1 \in \mathcal{E}_1$ (as $h^{-1}(\{\infty\}) \in \mathcal{E}_1$, $\{\infty\} = A_1^c$).

So by (1),

$$\int_{E_1}\left(\int_{E_2}|f(x_1,x_2)|\mu_2(\mathrm{d}x_2)\right)\mu_1(\mathrm{d}x_1)=\mu(|f|)<\infty.$$

Hence, $\mu_1(A_1^c)=0$ (hence f_1 integrable) (as $\mu(h)\geq\mu(h\mathbb{1}_{A_1^c})=\infty$ if $\mu(A_1^c)>\infty$). Setting

$$f_1^+(x_1)=\int_{E_2}f^+(x_1,x_2)\mu_2(\mathrm{d}x_2),\qquad f_1^-(x)=\int_{E_2}f^-(x_1,x_2)\mu_2(\mathrm{d}x_2),$$

we see

$$f_1=(f_1^+-f_1^-)\mathbb{1}_{A_1}=f_1^+\mathbb{1}_{A_1}-f_1^-\mathbb{1}_{A_1}$$

Also, $\mu_1(f_1^+)\overset{(1)}{=}\mu(f^+)<\infty$ and $\mu_1(f_1^-)\overset{(1)}{=}\mu(f^-)<\infty$, so,

$$\mu(f)=\mu(f^+)-\mu(f^-)=\mu_1(f_1^+)-\mu_1(f_1^-)=\mu_1(f_1)$$ □

Remark.

(1) The proof of (2) is symmetric in μ_1, f_1, so $\mu_1(f)=\mu(f)=\mu_2(f_2)$. So we can interchange the order of integrals whenever $f\geq 0$ or f integrable.

(2) The theorems extend to σ-finite measures μ.

(3) Associativity is easy to check, i.e. $(\mathcal{E}_1\otimes\mathcal{E}_2)\otimes\mathcal{E}_3=\mathcal{E}_1\otimes(\mathcal{E}_2\otimes\mathcal{E}_3)$ and $\mu_1\otimes(\mu_2\otimes\mu_3)=(\mu_1\otimes\mu_2)\otimes\mu_3$. So we can define the n-fold products $\bigotimes_{i=1}^n\mu_i$ on $(E_1\times\cdots\times E_n,\mathcal{E}\otimes\cdots\otimes\mathcal{E}_n)$ and n-fold integrals. In particular, when $E_i=\mathbb{R}$ and $\mu_i=\lambda$, we get $\bigotimes_{i=1}^n\lambda_i$ on $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$.

Start of lecture 14

Proposition. Let $X_1,\ldots,X_n$ be random variables $X_i:(\Omega,\mathcal{F},\mathbb{P})\to(E_i,\mathcal{E}_i)$. Set $E=E_1\times\cdots\times E_n$, $\mathcal{E}=\mathcal{E}_1\otimes\cdots\otimes\mathcal{E}_n$. Consider $X:(\Omega,\mathcal{F},\mathbb{P})\to(E,\mathcal{E})$ given by $X(\omega)=(X_1(\omega),\ldots,X_n(\omega))$. Then X is $\mathcal{E}$-measurable and the following are equivalent:

(i) $X_1,\ldots,X_n$ are independent (i.e. $\sigma\{X_i^{-1}(A):A\in\mathcal{E}_i\}$ are independent).

(ii) $\mu_X=\mu_{X_1}\otimes\mu_{X_2}\otimes\cdots\otimes\mu_{X_n}$.

(iii) For all bounded measurable $f_i:E_i\to\mathbb{R}$,

$$\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right)=\prod_{i=1}^n\mathbb{E}(f_i(X_i)).$$

Proof. X measurable: $X^{-1}(A) \in \mathcal{F}\ \forall A \in \mathcal{E}$. Enough to check

$$\begin{aligned} X^{-1}(A_1 \times \cdots \times A_n) &= \{\omega : X_i(\omega) \in A_1, \ldots, X_n(\omega) \in A_n\} \\ &= \bigcap_{i=1}^n X_i^{-1}(A_i) \end{aligned}$$

is in $\mathcal{F}$. But this is true since $X_i : (\Omega, \mathcal{F}) \to (E_i, \mathcal{E}_i)$ are all measurable, so $X_i^{-1}(A_i) \in \mathcal{F}$ for all $A_i \in \mathcal{E}_i$. Now we will show (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (i).

(i) $\implies$ (ii) Let $\nu = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$. Enough to show that

$$\begin{aligned} \mu_X(A_1 \times \cdots \times A_n) &\stackrel{?}{=} \nu(A_1 \times \cdots \times A_n) \\ &= \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) \\ &\stackrel{\text{(a)}}{=} \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n) \\ &= \mu_{X_1}(A_1) \cdots \mu_{X_n}(A_n) \\ &= \nu(A) \end{aligned}$$

Now finish since $\mathcal{A}$ is a π-system generating $\mathcal{E}$...

(ii) $\implies$ (iii)

$$\begin{aligned} \mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) &= \int \prod_{i=1}^n f_i(x_i) \mathrm{d}\mu_X(x) \\ &= \prod_{i=1}^n \int f_i(x_i) \mathrm{d}\mu_{X_i}(x_i) \\ &= \prod_{i=1}^n \mathbb{E}(f_i(X_i)) \end{aligned}$$

(iii) $\implies$ (i) Use $f_i = \mathbb{1}_{A_i}$, $A_i \in \mathcal{E}_i$. Then

$$\begin{aligned} \mathbb{E}\left(\prod_{i=1}^n \mathbb{1}_{A_i}(X_i)\right) &= \mathbb{E}(\mathbb{1}_{A_1 \times \cdots \times A_n}(X)) \\ &= \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) \\ &\stackrel{\text{(a)}}{=} \prod_{i=1}^n \mathbb{E}(\mathbb{1}_{A_i}(X_i)) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \end{aligned}$$

□

4 L^p Spaces, Norms, Inequalities

Definition (Norm)**.** Recall that a norm on a real vector space V is $\|\bullet\| : V \to [0, \infty)$ such that

(1) $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda \in \mathbb{R}, v \in V$.

(2) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in W$.

(3) $\|v\| = 0 \iff v = 0$.

Definition (L^p norm)**.** For $(E, \mathcal{E}, \mu)$ a measure space and $1 \leq p \leq \infty$, define

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

where

$$\|f\|_p = \left(\int |f(x)|^p \mathrm{d}\mu(x)\right)^{1/p}$$

for $1 \leq p < \infty$, and

$$\|f\|_\infty = \operatorname{ess\,sup} |f| = \inf\{\lambda \geq 0 : |f| \leq \lambda \ \mu\text{.almost everywhere}\}$$

By linearity of integral, this satisfies the properties of a norm:

(1) This holds for $1 \leq p < \infty$ and for $p = \infty$ it's obvious.

(2) This holds for $p = 1, \infty$ easily. For other p, we shall show by Minkowski inequality.

(3) $f = 0$ implies $\|f\|_p = 0$. But $\|f\|_p = 0$ implies $f = 0$ almost everywhere. So we fix this by defining equivalence classes

$$[f] = \{g : g = f \text{ almost everywhere}\}$$

and

$$\mathcal{L}^p = \{[f] : f \in L^p\}$$

Then $\mathcal{L}^p$, $1 \leq p \leq \infty$ are normed vector spaces.

4.1 Inequalities

Markov / Chebyshev's Inequality

Let $f \geq 0$ measurable, then $\forall \lambda > 0$,

$$\underbrace{\mu(\{x \in E : f(x) \geq \lambda\})}_{\mu(f \geq \lambda)} \leq \frac{\mu(f)}{\lambda}$$

Proof. $\lambda \mathbb{1}_{\{f \geq \lambda\}} \leq f$. So integrating with respect to μ, we get

$$\lambda \mu(f \geq \lambda) \leq \mu(f).$$

□

In particular, if $g \in L^p$, $p < \infty$, then $\mu(|g| \geq \lambda) \leq \frac{\mu(|g|^p)}{\lambda^p} < \infty$. This gives the tail estimates as $\lambda \to \infty$.

Definition (Convex function)**.** For $I \subseteq \mathbb{R}$ an interval, say a function $c : I \to \mathbb{R}$ is *convex* if $\forall x, y \in I$, $t \in [0, 1]$,

$$c(tx + (1-t)x) \leq tc(x) + (1-t)c(y)$$

equivalently:

$$\frac{c(\tilde{t}) - c(t)}{\tilde{t} - x} \leq \frac{f(y) - c(\tilde{t})}{y - \tilde{t}} \qquad (*)$$

for $x < \tilde{t} < y$ in I. (In particular, this second definition shows that c is continuous on I, hence Borel measurable).

Lemma. Let $c : I \to \mathbb{R}$ be convex, $m \in I$ (an interval). Then there exists $a, b \in \mathbb{R}$ such that $c(x) \geq ax + b$ for all $x \in I$, and equality at $x = m$.

Proof. Let

$$a = \sup \left\{ \frac{c(m) - c(x)}{m - x} : x \in I, x < m \right\} < \infty.$$

Then by $(*)$, $\forall y > m, y \in I$,

$$\frac{c(m) - c(x)}{m - x} \leq a \leq \frac{c(y) - c(m)}{y - m}$$

These inequalities imply

$$c(y) \geq ay - am + c(m) \qquad \forall y \geq m$$

and

$$c(x) \geq ax - am + c(m) \qquad \forall x \leq m. \qquad \square$$

Theorem (Jensen's Inequality)**.** Let X be an integrable random variable (i.e. $\mathbb{E}|X| < \infty$), taking values in an interval $I \subset \mathbb{R}$, and let $c : I \to \mathbb{R}$ be convex. Then $\mathbb{E}(c(x))$ is well-defined, and

$$\boxed{\mathbb{E}(c(X)) \geq c(\mathbb{E}(X))}$$

Proof. If X is a constant almost surely, then nothing to prove. Assume otherwise. Then $\mathbb{E}(x) = m \in I$. Using the previous lemma, $\exists a, b \in \mathbb{R}$ such that

$$c(X) \geq aX + b \tag{$*$}$$

In particular,

$$c((X))^- \leq |a||X| + |b|$$

so $\mathbb{E}(c(X)^-) < \infty$. Hence $\mathbb{E}(c(X)) = \mathbb{E}(c(X)^+) - \mathbb{E}(c(X)^-)$ is well-defined on $(-\infty, \infty]$.

Claim: $(*) \implies \mathbb{E}(c(X)) \geq a\mathbb{E}(X) + b$. If $\mathbb{E}(c(X)) = \infty$, nothing to prove. Otherwise, $c(X)$ and $aX + b$ are integrable random variables satisfying $(*)$, so taking expectatation,

$$\mathbb{E}(c(X)) \geq a\mathbb{E}(X) + b = am + b = c(m) = c(\mathbb{E}(X)). \qquad \square$$

As an application: $(\Omega, \mathcal{F}, \mathbb{P})$ and $1 \leq p \leq \infty$. If $X \in L^{\infty(\mathbb{P})}$, then $X \in L^{pp(\mathbb{P})}$ for all $1 \leq p < \infty$ as $\|X\|_p \leq \|X\|_\infty$.

Claim: If $X \in L^q$ and $q > p \geq 1$, then $X \in L^p$.

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p} = (c(\mathbb{E}(|X|^p)))^{1/q} \leq (\mathbb{E}(c(|X|^p)))^{1/q} = (\mathbb{E}(|X|^q))^{1/q} = \|X\|_q$$

Hence, $X \in L^q$ implies $X \in L^p$ for all $1 \leq p < q$, i.e.

$$L^\infty(\mathbb{P}) \subseteq L^q(\mathbb{P}) \subseteq L^p(\mathbb{P}) \subseteq L^1(\mathbb{P})$$

for all $1 \leq p \leq q \leq \infty$.

Start of

lecture 15

Theorem (Holder Inequality)**.** Let f, g measurable and $1 \leq p \leq g \leq \infty$ be conjugate, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\mu(|fg|) \leq \|f\|_p \|g\|_q.$$

Proof. If p or $q = 1$ or ∞ then clear. So are the cases when $\|f\|_p$ or $\|g\|_q = 0$. Exclude these.

Dividing both sides by $\|f\|_p$, we may assume $\|f\|_p = 1$, i.e. $\int |f|^p \mathrm{d}\mu = 1$. So, define a probability measure $\mathbb{P}$ on $\mathcal{E}$ by

$$\mathbb{P}(A) = \int_A |f|^p \mathrm{d}\mu$$

($\mathbb{P}$ has probability density $|f|^p$ with respect to μ). Also, for $h \geq 0$ measurable,

$$\int h \mathrm{d}\mathbb{P} = \int h|f|^p \mathrm{d}\mu \qquad (*)$$

Then,

$$\begin{aligned}
\mu(|fg|) &= \mu(|fg|\mathbb{1}_{\{|f|>0\}}) \\
&= \int \frac{|f|^p|g|}{|f|^{p-1}} \mathbb{1}_{\{|f|>0\}} \mathrm{d}\mu \\
&= \int \frac{|g|}{|f|^{p-1}} \mathbb{1}_{\{|f|>0\}} |f|^p \mathrm{d}\mu \\
&= \int \frac{|g|}{|f|^{p-1}} \mathbb{1}_{\{|f|>0\}} \mathrm{d}\mathbb{P} \\
&= \mathbb{E}\left(\frac{|g|}{|f|^{p-1}} \mathbb{1}_{\{|f|>0\}}\right) \\
&\leq \left(\mathbb{E}\left(\frac{|g|^q}{|f|^{q(p-1)}} \mathbb{1}_{|f|>0}\right)\right)^{1/q} && \text{(Jensen's Inequality)} \\
&= \left(\int \left(\frac{|g|^q}{|g|^p} \mathbb{1}_{\{|f|>0\}}\right) \mathrm{d}\mathbb{P}\right)^{1/q} \\
&= \left(\int |g|^q \mathbb{1}_{\{|f|>0\}} \mathrm{d}\mu\right)^{1/q} && \text{(by } (*)) \\
&\leq \left(\int |g|^q \mathrm{d}\mu\right)^{1/q} \\
&= \|g\|_q
\end{aligned}$$

□

Remark. $p = q = 2$ is the Cauchy-Schwarz inequality (see Example Sheet 3).

Theorem (Minkowski inequality)**.** For $p \in [1, \infty]$ and f, g measurable, $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Proof. $p = 1, \infty$ obvious. Also obvious when $\|f\|_p$ or $\|g\|_p = 0$ or $\|f+g\|_p$. Assume otherwise. Since

$$|f+g|^p \leq 2^p(|f|^p + |g|^p)$$

we have

$$\mu(|f+g|^p) \leq 2^p(\mu(|f|^p) + \mu(|g|^p)) < \infty$$

if $f, g \in L^p$. So $f + g \in L^p$. With q the conjugate of p,

$$\begin{aligned}
\|f+g\|_p^p &= \int |f+g|^p \mathrm{d}\mu \\
&= \int |f+g||f+g|^{p-1} \mathrm{d}\mu \\
&\leq \int |f+g|^{p-1}|f| \mathrm{d}\mu + \int |f+g|^{p-1}|g| \mathrm{d}\mu \\
&\leq \|f\|_p \|(f+g)^{p-1}\|_q + \|g\|_p \|(f+g)^{p-1}\|_q && \text{(Holder Inequality)} \\
&= \left(\int |f+g|^{\overbrace{(p-1)q}^{=p}} \mathrm{d}\mu \right)^{1/q} (\|f\|_p + \|g\|_p) \\
&= \underbrace{\left(\int |f+g|^p \mathrm{d}\mu \right)^{1/q}}_{=\|f+g\|_p^{p/q}} (\|f\|_p + \|g\|_p)
\end{aligned}$$

Hence, dividing by $\|f+g\|_p^{p/q}$, we get

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

□

Theorem ($\mathcal{L}^p$ is a complete normed vector space (Banach space))**.** Let $p \in [1, \infty]$. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions in L^p such that

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall m, n \geq N \quad \|f_m - f_n\|_p < \varepsilon$$

(i.e. (f_n) is Cauchy in L^p). Then $\exists \; f \in L^p$ such that $\|f_n - f\|_p \to 0$ as $n \to \infty$.

Proof. Assume $1 \leq p < \infty$ ($p = \infty$ is an exercise).

Choose a subsequence (n_k) such that $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$. Then $S = \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty$. By Minkowski inequality, for any $K \in \mathbb{N}$,

$$\left\| \sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k} \right\|_p \leq \sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p = S < \infty.$$

So,

$$\int \left(\sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}| \right)^p \mathrm{d}\mu \le S^p < \infty.$$

But

$$\left(\sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k} \right)^p \uparrow \left(\sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k} \right)^p.$$

So by Monotone Convergence Theorem,

$$\left\| \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \right\|_p < \infty.$$

So, in particular,

$$\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty \mu.\text{almost everywhere.}$$

Let A be the set where this is $< \infty$. Then $\mu(A^c) = 0$. For any $x \in A$, $(f_{n_k}(x))$ is Cauchy, and since $\mathbb{R}$ is complete, it converges to $f(x)$ say. Define $f(x) = 0$ for all $x \in A^c$. Then f is measurable and $f_{n_k} \to f$ as $k \to \infty$ μ almost everywhere. Then,

$$\|f_n - f\|_p^p = \mu(|f_n - f|^p) = \mu(\liminf_k |f_n - f_{n_k}|_p^p) \overset{\text{Fatou's Lemma}}{\le} \liminf_k \mu(|f_n - f_{n_k}|^p) \le \varepsilon^p$$

and

$$\|f\|_p \overset{\text{Minkowski inequality}}{\le} \|f_N - f\|_p + \|f_n\|_p \le \varepsilon^p + \|f_N\|_p < \infty.$$

Hence $f \in L^p$ and $f_n \overset{L^p}{\to} f$. □

Remark. One can show that any choice of vector spaces

$V = \mathcal{C}[0,1], \{\text{simple functions}\}, \{\text{finite linear combination of indicators of intervals}\}$

are dense in $L^p((0,1), \mathcal{B}, \lambda)$ and so $(\mathcal{C}[0,1], \|\bullet\|_1)$ is L^1 space of Lebesgue integrable functions (exercise in Example Sheet 3).

$\mathcal{L}^2$ as Hilbert space: On a vector space V, a symmetric bilinear form $V \times V \to \mathbb{R}$, $(u, v) \mapsto \langle u, v \rangle$ is called an inner product if $\langle v, v \rangle \ge 0$ $\forall v \in V$ and $\langle v, v \rangle = 0$ if and only if $v = 0$.

Then $\sqrt{\langle v, v \rangle} = \|v\|$ is a norm (Cauchy-Schwarz inequality gives the triangle inequality for $\|\bullet\|$). If $(V, \|\bullet\|)$ is complete, it is called a Hilbert space.

Corollary. $\mathcal{L}^2$ with the inner product $\langle f, g \rangle = \int fg \mathrm{d}\mu$ is a Hilbert space.

Basic Geometry

(1) Pythagoras theorem $\|f+g\|_2^2 = \|f\|_2^2 + \|g\|_2^2 + 2\langle f, g\rangle$.

(2) Parallelogram law: $\|f+g\|_2^2 + \|f-g\|_2^2 = 2(\|f\|_2^2 + \|g\|_2^2)$

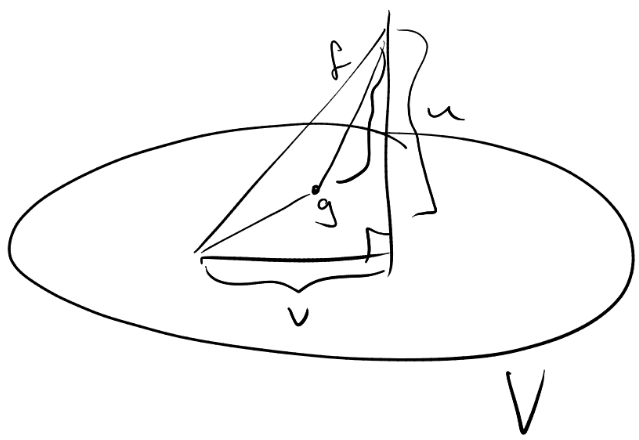
We say f is *orthogonal* to f (written $f \perp g$) if $\langle f, g\rangle = 0$. Then $\|f+g\|_2^2 = \|f\|_2^2 + \|g\|_2^2$. For a subset $V \subseteq L^2$, we define its orthogonal complement

$$V^\perp = \{f \in L^2 : \langle f, v\rangle = 0 \; \forall v \in V\}$$

A subset V is closed if $(f_n) \in V$ and $f_n \overset{L^2}{\to} f$ implies $f \in V$.

Start of lecture 16

Theorem (Orthogonal projection)**.** If V is a *closed subspace* of L^2, then $\forall f \in L^2$, $f = v + u$ where $v \in V$, $u \in V^\perp$. Moreover, $\|f - v\|_2 \le \|f - g\|_2$ for all $g \in V$ with equality if and only if $g = v$ almost everywhere. In particular, v is unique (almost everywhere) and is called the orthogonal projection of f on V.



Proof. Define $d(f, V) = \inf_{g\in V} \|f - g\|_2$. Let $(g_n) \in V$ be a sequence such that $\|f - g_n\|_2 \to d(f, V)$. Now by parallelogram law,

$$2(\|f - g_n\|_2^2 + \|f - g_m\|_2^2) = \|g_n - g_m\|_2^2 + \underbrace{4\left\|f - \frac{g_n + g_m}{2}\right\|^2}_{\ge 4d(f,V)^2}$$

so by taking $\lim_{n,m\to\infty}$, we deduce that $\|g_n - g_m\|_2^2 \to 0$, i.e. (g_n) is Cauchy. As L^2 is complete and V is closed, $g_n \xrightarrow{L^2} v \in V$. Then $\|f - g_n\|_2^2 \to \|f - v\|_2^2$ hence $\|f - v\|_2^2 = d(f, V)^2$, i.e. $d(f, V) = \|f - v\|_2$.

Then for any $h \in V$, $t \in \mathbb{R}$,

$$d(f, v)^2 \le \|f - (v + th)\|_2^2 = d(f, v)^2 - 2t\langle f - v, h\rangle + t^2\|h\|_2^2 \qquad (*)$$

Letting $t \downarrow 0$ and $t \uparrow 0$, $\langle f - v, h\rangle = 0$, hence $f - v \in V^\perp$. Now

$$f = \underbrace{v}_{\in V} + \underbrace{f - v}_{\in V^\perp}$$

as desired. For any $g \in V$,

$$f - g = \underbrace{f - v}_{\in V^\perp} + \underbrace{v - g}_{\in V}$$

and

$$\|f - g\|_2^2 = \|f - v\|_2^2 + \|v - g\|_2^2$$

Hence $\|f - g\|_2 \ge \|f - v\|_2$ with equality if and only if $\|v - g\|_2 = 0$, i.e. $v = g$ almost everywhere. □

$(\Omega, \mathcal{F}, \mathbb{P})$ and $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}X = \mathbb{E}Y = 0$. Then $\mathrm{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY = \langle X, Y\rangle$. $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$. If X and Y independent, $\langle X, Y\rangle = 0$, converse not true.

If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$ (i.e. $\mathcal{G} \subseteq \mathcal{F}$), then $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. For $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, (a variant of) the conditional expectation of X given $\mathcal{G}$, $\mathbb{E}(X \mid \mathcal{G})$ is defined as the orthogonal projection of X on $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (X should be measurable with respect to $\mathcal{G}$ and $\|X - Y\|_2 \ge \|X - \mathbb{E}(X \mid \mathcal{G})\|_2$ for all Y $\mathcal{G}$-measurable).

Question: How to define $\mathbb{E}(X \mid \mathcal{G})$ is $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$? (advanced probability).

Exercise: Let $(G_i)_{i\in I}$ be a countable family of disjoint events whose union is Ω and set $\mathcal{G} = \sigma(G_i : i \in I)$. Let X be integrable. Then the conditional expression of X given $\mathcal{G}$ is given by

$$Y = \sum_i \mathbb{E}(X \mid G_i)\mathbb{1}_{G_i}, \qquad \mathbb{E}(X \mid G_i) = \frac{\mathbb{E}(x\mathbb{1}_{G_i})}{\mathbb{P}(G_i)} \ \forall i \in I$$

Check:

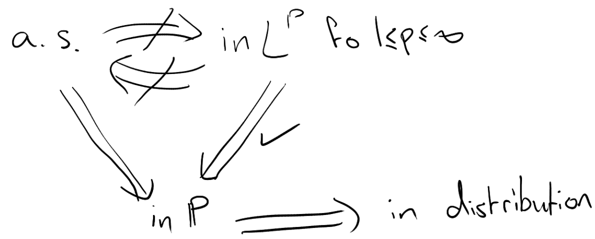
(1) Y is $\mathcal{G}$-measurable

(2) $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$

(3) Y is "the" orthogonal projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$ if $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

L^p Convergence and Uniform Integrability

$(\Omega, \mathcal{F}, \mathbb{P})$.



Explanations:

$f_n = n\mathbb{1}_{(0,1/n)}$ on $((0,1), \mathcal{B}, \lambda)$. Then $f_n \to 0$ almost surely. But $\mathbb{E}|f_n| = \mathbb{E}f_n = 1\ \forall n$, i.e. almost surely $\not\Longrightarrow$ L^p convergence.

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p} \to 0$$

(Markov inequality).

Theorem (DCT)**.** Let (X_n) be such that $X_n \xrightarrow{\mathbb{P}} X$ and $|X_n| \le Y$ for all n, for some integrable random variable. Then $X_n \xrightarrow{L^1} X$, i.e. $\mathbb{E}|X_n - X| \to 0$ as $n \to \infty$.

Question: What is the "minimum condition" on (X_n) under which $X_n \xrightarrow{\mathbb{P}} X$ implies $X_n \xrightarrow{L^1} X$? $\implies$ $\mathbb{E}|X_n| \to \mathbb{E}|X|$, "Uniformly Integrable".

For $X \in L^1(\mathbb{P})$ define

$$I_X(\delta) = \sup\{\mathbb{E}(|X|\mathbb{1}_A) : A \in \mathcal{F}, \mathbb{P}(A) \le \delta\}$$

Then $I_X(\delta) \to 0$ as $\delta \to 0$. (If not then $\exists \varepsilon > 0$ and $(A_n) \in \mathcal{F}$ such that $\mathbb{P}(A_n) \le 2^{-n}$ and $\mathbb{E}|X|\mathbb{1}_{A_n} \ge \varepsilon$. Then $\sum \mathbb{P}(A_n) < \infty$ so $\mathbb{P}(A_n \text{ i.o.}) = 0$ by Borel-Cantelli Lemma 1. Then

$$\varepsilon \le \mathbb{E}|X|\mathbb{1}_{A_n} \le \mathbb{E}|X|\mathbb{1}_{\bigcup_{m=n}^\infty A_m} \xrightarrow{\text{DCT}} 0$$

since

$$\mathbb{1}_{\bigcup_{m=n}^\infty A_m} \to \mathbb{1}_{\bigcap_m \bigcup_{m=n}^\infty A_m} = \mathbb{1}_{\{A_n \text{ i.o.}\}} = 0 \text{ almost surely}$$

contradiction).

Definition. Let χ be a collection of random variables in $L^1(\mathbb{P})$. Define

$$I_\chi(\delta) = \sup\{\mathbb{E}(|X|\mathbb{1}_A) : X \in \chi, A \in \mathcal{F}, \mathbb{P}(A) \le \delta\}$$

We say χ is *uniformly integrabl* (UI) if

(1) χ is bounded in L^1 (i.e. $\sup_{X\in\chi} \|X\|_1 = \sup_{X\in\chi} \mathbb{E}|X| = I_\chi < \infty$)

(2) $I_\chi(\delta) \to 0$ as $\delta \to 0$.

Remark. Note to reader: I didn't follow what the lecturer was writing in these remarks so they are probably nonsense, but if you rearrange things appropriately it should hopefully make sense (I've tried thinking about it for a bit but haven't figured it out).

(1) Any single integrable random variable is UI (so does any finite collection of integrable random variables, also if

$$\chi = \{X : X \text{ a random variable such that } |X| \le Y \text{ for some } Y \in L^1\}$$

) Then X is UI,

$$I_\chi(\delta) \le I_Y(\delta) \to 0$$

as $\delta \to 0$.

(2) If χ is bounded in L^p for some $p > 1$ then

$$\sup \mathbb{E}(X\mathbb{1}_A) \overset{\text{Holder Inequality}}{\le} \|X\|_p(\mathbb{P}(A))^{1/q} \le C\delta^{1/q}$$

$X \in \chi$ such that $\mathbb{P}(A) \le \delta$ i.e. $I_\chi(\delta) \le C\delta^{1/q} \to 0$ as $\delta \to 0$ for all A such that $\mathbb{P}(A) \le \delta$ as $\sup_{X\in\chi} \mathbb{E}|X|\mathbb{1}_A \le \mathbb{E}Y\mathbb{1}_A$.

(3) L^1 bounded $\not\Rightarrow$ UI.

Start of lecture 17

Recall for a collection χ of random variables, χ is UI if

(1) χ is L^1 bounded

(2) $I_\chi(\delta) = \sup_{X\in\chi}\{\mathbb{E}(|X|\mathbb{1}_A) : A \in \mathcal{F}, \mathbb{P}(A) \le \delta\} \downarrow 0$ as $\delta \downarrow 0$.

Lemma (Alternative definition of UI)**.** χ is UI if and only if

$$\sup_{X\in\chi} \mathbb{E}(|X|\mathbb{1}_{(|X|\geq K)}) \to 0$$

as $K \to \infty$.

Proof.

$\Rightarrow$ Fix any $\varepsilon > 0$. Then $\exists\delta > 0$ such that $I_\chi(\delta) < \varepsilon$ (as χ is UI). Choose $K < \infty$ such that

$$\frac{I_\chi(1)}{\delta} = \frac{\sup_{X\in\chi} \mathbb{E}|X|}{\delta} \leq K.$$

Then for any $X \in \chi$,

$$\mathbb{P}(|X| \geq K) \overset{\text{Markov}}{\leq} \frac{\mathbb{E}|X|}{K} \leq \frac{I_\chi(1)}{K} \leq \delta.$$

So, with $A = \{|X| \geq K\}$,

$$\mathbb{E}(|X|\mathbb{1}_{|X|\geq K}) \leq I_\chi(\delta) < \varepsilon.$$

$\Leftarrow$

$$\mathbb{E}|X| = \mathbb{E}|X|\mathbb{1}_{(|X|\geq K)} + \mathbb{E}\underbrace{|X|\mathbb{1}_{(|X|\leq K)}}_{\leq K}$$

So,

$$\sup_{X\in\chi} \mathbb{E}|X| \leq K + \sup_{X\in\chi} \mathbb{E}|X|\mathbb{1}_{(|X|\leq K)}$$

Choose K large so that the second term on the RHS is ≤ 1. Then χ is L^1 bounded. Fix any $\varepsilon > 0$. Choose K so that $\mathbb{E}(|X|\mathbb{1}_{|X|\geq K}) < \varepsilon/2$ for all $X \in \chi$. Then choose $\delta > 0$ such that $\delta \leq \frac{\varepsilon}{2K}$. Then $\forall X \in \mathcal{X}$ and $A \in \mathcal{F}$ with $\mathbb{P}(A) \leq \delta$,

$$\begin{aligned}\mathbb{E}(|X|\mathbb{1}_A) &= \mathbb{E}(|X|\mathbb{1}_A\mathbb{1}_{|X|\geq K}) + \mathbb{E}(|X|\mathbb{1}_A\mathbb{1}_{|X|\leq K}) \\ &\leq \underbrace{\mathbb{E}(|X|\mathbb{1}_{|X|\geq K})}_{<\varepsilon} + \underbrace{K\mathbb{P}(A)}_{\leq K\delta\leq\varepsilon/2} \\ &\leq \varepsilon\end{aligned}$$

□

Theorem. Let (X_n), X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent:

(a) $X, X_n \in L^1$ for all n and $X_n \xrightarrow{L^1} X$.

(b) $\chi = (X_n : n \in \mathbb{N})$ is UI and $X_n \xrightarrow{\mathbb{P}} X$.

Proof.

(a) $\implies$ (b) By Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \frac{\mathbb{E}|X_n - X|)}{\varepsilon} \to 0$$

as $n \to \infty$, so $X_n \xrightarrow{\mathbb{P}} X$. Choose N such that $\mathbb{E}|X_n - X| < \frac{\varepsilon}{2}$ for all $n \ge N$. Choose δ so that $\mathbb{E}|X|\mathbb{1}_A \le \frac{\varepsilon}{2}$ whenever $\mathbb{P}(A) < \delta$.

$$\mathbb{E}|X_n|\mathbb{1}_A \le \underbrace{\mathbb{E}|X_n - X|}_{\le \delta/2} + \underbrace{\mathbb{E}|X|\mathbb{1}_A}_{\le \varepsilon/2} \le \varepsilon \qquad \forall n \ge N.$$

For all $n = 1, 2, \dots, N-1$,

$$\mathbb{E}|X_n|\mathbb{1}_A \le \varepsilon$$

by the choice of δ. Hence χ is UI.

(b) $\implies$ (a) $X_n \xrightarrow{\mathbb{P}} X$. So take a subsequence (n_k) such that $X_{n_k} \xrightarrow{\text{almost surely}} X$. Then

$$\mathbb{E}|X| = \mathbb{E} \liminf |X_{n_k}| \overset{\text{Fatou's Lemma}}{\le} \liminf \mathbb{E}|X_{n_K}| \le \sup_{X_n \in \chi} \mathbb{E}|X_n| < \infty.$$

(as χ is UI, hence L^1 bounded). So $X \in L^1$. Define the truncated random variables

$$X_n^K = (-K) \vee X_n \wedge K$$
$$X^k(-k) \vee X \wedge K$$

Then $X_n^K \xrightarrow{\mathbb{P}} X^K$ (as $\mathbb{P}(|X_n^K - X^K| > \varepsilon) \le \mathbb{P}(|X_n - X| > \varepsilon)$).

Aside: If $X_n \xrightarrow{\mathbb{P}} X$ and f is a continuous function, then $f(x_n) \xrightarrow{\mathbb{P}} f(x)$. Also $|X_n^K| \le K$ for all n. Hence by BCT, $X_n^K \xrightarrow{L^1} X^k$. Now,

$$\mathbb{E}|X_n - X| \le \mathbb{E} \overbrace{|X_n - X_n^K|}^{\le |X_n|\mathbb{1}_{(|X_n| \ge K)} \ (1)} + \mathbb{E} \overbrace{|X - X^K|}^{\le \mathbb{E}|X|\mathbb{1}_{(|X| \le K)} \ (2)} + \mathbb{E}|X_n^K - X^K| \le \varepsilon$$

for all $n \ge N$. By UI choose K large so that (1) and (2) are $\le \frac{\varepsilon}{3}$. Then choose N large so that the last term is $\le \frac{\varepsilon}{3}$ for all $n \ge N$. $\square$

5 Fourier Transforms

For g measurable such that $\int |G| \mathrm{d}x < \infty$, define

$$\int g(x)\mu(\mathrm{d}x) = \int \mathrm{Re}(g(x))\mu(\mathrm{d}x) + i \int \mathrm{Im}(g(x))\mu(\mathrm{d}x)$$

Here $L^p = L^p(\mathbb{R}^d)$ is the space of *complex valued* Borel measurable functions on $\mathbb{R}^d$, i.e. $f : \mathbb{R}^d \to \mathbb{C}$ for which

$$\underbrace{\left(\int_{\mathbb{R}^d} |f(x)|^p \mu(\mathrm{d}x)\right)^{1/p}}_{=\|f\|_p} < \infty$$

for all $1 \le p < \infty$. and

$$\left|\int g(x)\mu(\mathrm{d}x)\right| \le \int |g(x)|\mu(\mathrm{d}x)$$

(Example Sheet 3). We also define for $f, g \in L^2$,

$$\langle f, g \rangle = \int f(x)\overline{g(x)}\mathrm{d}\mu(x)$$

which is an inner product on $L^2(\mu)$. For any $y \in \mathbb{R}^d$,

$$\begin{aligned}\int f(x-y)\mathrm{d}x &= \int f(y-x)\mathrm{d}x - \int f(x)\mathrm{d}x \\ &= \int f(-x)\mathrm{d}x\end{aligned}$$

(translation invariance and $x \mapsto -x$ symmetry of λ, see Example Sheet 3). Also for $a \in \mathbb{R}$, $a \neq 0$,

$$\int f(ax)\mathrm{d}x = \frac{1}{a^d}\int f(x)\mathrm{d}x$$

Definition (Fourier Transform)**.** The *Fourier transform* $\hat{f}$ of $f \in L^1(\mathbb{R}^d)$ is defined as

$$\hat{f}(u) = \int_{\mathbb{R}^d} f(x)e^{i\langle u,x\rangle}\mathrm{d}x$$

for all $u \in \mathbb{R}^d$ and $\langle u, x \rangle = \sum_{i=1}^d u_i x_i$.

For all $u \in \mathbb{R}^d$,

$$\sup_u |\hat{f}(u)| \le \int |f(x)|\mathrm{d}x = \|f\|_1 < \infty$$

i.e. $\hat{f} \in L^\infty$. Also, for $u_n \to u$,

$$f(x)e^{i\langle u_n, x\rangle} \to f(x)e^{i\langle u,x\rangle}$$

and $|f(x)e^{i\langle u_n,x\rangle}| \le |f(x)|$ and $f \in L^1$, so by DCT, $\hat{f}(u_n) \to \hat{f}(u)$. Moreover,

$$\lim_{\|u\|\to\infty} \hat{f}(u) = 0$$

(Riemann-Lebesgue Lemma, Example Sheet 3). Thus

$$\hat{f} \in C_0(\mathbb{R}^d) = \{f \text{ bounded continuous vanishing at } \infty\}$$

The map is $1-1$ (but not onto).

For a finite / probability measure μ on $\mathbb{R}^d$, define similarly,

$$\hat{\mu}(u) = \int e^{i\langle ux\rangle} \mathrm{d}\mu(x) \qquad u \in \mathbb{R}^d$$

Then $\hat{\mu}$ is a bounded continuous function on $\mathbb{R}^d$ and $|\hat{\mu}(u)| \le \mu(\mathbb{R}^d) < \infty$. If μ has density f (with respect to λ), then

$$\hat{\mu}(u) = \int e^{i\langle u,x\rangle} f(x)\mathrm{d}x = \hat{f}(u).$$

Definition (Characteristic function)**.** The *characteristic function* (c.f.) ϕ_X of a random variable X on $\mathbb{R}^d$ is the Fourier transform of its law $\mu_X = \mathbb{P} \circ X^{-1}$. So

$$\phi_X(u) = \hat{\mu}_X(u) = \int e^{i\langle u,x\rangle} \mathrm{d}\mu_X(x) = \int e^{i\langle u,x\rangle} \mathrm{d}\mathbb{P} = \mathbb{E}e^{i\langle u,X\rangle}$$

$(\nu \circ f^{-1}(g) = \nu(f \circ g))$. In particular if X has pdf f, then $\phi_X(u) = \hat{f}(u)$.

Start of lecture 18

Definition. For $f \in L^1(\mathbb{R}^d)$, with $\widehat{f} \in L^1(\mathbb{R}^d)$, we say that the *Fourier Inversion* holds for f if

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(u) e^{-i\langle u,x\rangle} \mathrm{d}u$$

almost everywhere in $\mathbb{R}^d$.

Remark.

(1) The RHS is continuous by DCT, so for f continuous, the equality is everywhere.

(2) $f \mapsto \widehat{f}$, $L^1 \to C_0$ is $1-1$. (for $f, g \in L^1$ with $\widehat{f} = \widehat{g}$, then $f - g \in L^1$ and $\widehat{f-g} = \widehat{f} - \widehat{g} = 0$. So by Fourier Inversion, $f - g = 0$ almost everywhere).

A key concept in Fourier analysis is *convolution*.

Definition (Convolution)**.** For $f \in L^p(\mathbb{R}^d)$, $1 \leq p < \infty$, and ν a probability measure,

$$f * \nu(x) = \begin{cases} \int_{\mathbb{R}^d} f(x-y)\nu(\mathrm{d}y) & \text{if the integral exists} \\ 0 & \text{otherwise} \end{cases} \qquad x \in \mathbb{R}^d$$

$$\begin{aligned} \int |f * \nu(x)|^p \mathrm{d}x &\leq \int \left(\int |f(x-y)| \nu(\mathrm{d}y) \right)^p \mathrm{d}x \\ &\leq \iint |f(x-y)|^p \nu(\mathrm{d}y) \mathrm{d}x && \text{(as } p \geq 1\text{, use Jensen's Inequality)} \\ &= \int \left(\int |f(x-y)^p| \mathrm{d}x \right) \nu(\mathrm{d}y) && \text{(Fubini-Tonelli)} \\ &= \int \left(\int |f(x)|^p \mathrm{d}x \right) \nu(\mathrm{d}y) && (\lambda \text{ is translation invariant}) \\ &= \|f\|_p^p \\ &< \infty \end{aligned}$$

Hence $f * \nu$ is defined almost everywhere, and $\|f * \nu\|_p \leq \|f\|_p < \infty$. When ν has pdf $g \in L^1$,

$$f * \nu(x) = \int f(x-y) g(y) \mathrm{d}y = f * g(x).$$

For 2 probability measures μ, ν on $\mathbb{R}^d$, the convolution $\mu * \nu$ is a new probability measure defined as

$$\mu * \nu(A) = \iint \mathbb{1}_A(x+y) \mu(\mathrm{d}x) \nu(\mathrm{d}y) = \mu \otimes \nu(x + y \in A) = \mathbb{P}(X + Y \in A).$$

where X, Y are independent, $X \sim \mu$, $Y \sim \nu$. In other words, $(X+Y) \sim \mu * \nu$.

If μ has pdf $f \in L^1$, then

$$\begin{aligned}
\mu * \nu(A) &= \int \left(\int \mathbb{1}_A(x+y) f(x) \mathrm{d}x \right) \nu(\mathrm{d}y) \\
&= \int \left(\int \mathbb{1}_A(x) f(x-y) \mathrm{d}x \right) \nu(\mathrm{d}y) && \text{(translation invariance)} \\
&= \int \mathbb{1}_A(x) \left(\int \overbrace{f(x-y)\nu(\mathrm{d}y)}^{f*\nu(x)} \right) \mathrm{d}x && \text{(Fubini-Tonelli)} \\
&= \int \mathbb{1}_A f * \nu(x) \mathrm{d}x
\end{aligned}$$

So, $\mu * \nu$ has the pdf $f * \nu$.

Easy to check:

(1) $\widehat{f * \nu}(u) = \widehat{f}(u)\widehat{\nu}(u)$ for all $f \in L^1$, ν a probability measure.

(2) $\widehat{\mu * \nu}(u) = \widehat{u}(u) \cdot \widehat{\nu}(u)$ for all μ, ν probability measures. X, Y independent, $X \sim \mu$, $Y \sim \nu$, then $X + Y \sim \nu * \mu$, then

$$\widehat{\mu * \nu} = \mathbb{E}e^{i\langle u, X+Y\rangle} \overset{\text{ind}}{=} \mathbb{E}e^{i\langle u, X\rangle} \cdot \mathbb{E}e^{i\langle u, Y\rangle} = \widehat{\mu}(u)\widehat{\nu}(u).$$

Fourier transform of Gaussians

If ϕ_z is the characteristic function of $Z \sim \mathrm{N}(0,1)$, i.e.

$$\phi_Z(u) = \mathbb{E}e^{iuZ} = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} e^{iuz} \mathrm{d}z$$

then by a previous theorem, ϕ_Z is differentiable and can be differentiated under the integral sign, i.e.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}u}\phi_Z(u) &= \frac{1}{\sqrt{2\pi}} \int \frac{\mathrm{d}}{\mathrm{d}u}(e^{-\frac{z^2}{2}} e^{iuz}) \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int ize^{iuz} e^{-\frac{z^2}{2}} \mathrm{d}z \\
&= \frac{i}{\sqrt{2\pi}} \int e^{iuz} (ze^{-\frac{z^2}{2}}) \mathrm{d}z \\
&= \frac{i}{\sqrt{2\pi}} \int iue^{iuz} e^{-\frac{z^2}{2}} \mathrm{d}z && \text{(integration by parts)} \\
&= -u \int e^{iuz} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \mathrm{d}z \\
&= -u\phi_Z(u) && (*)
\end{aligned}$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}u}(e^{\frac{u^2}{2}}\phi_Z(u)) = e^{\frac{u^2}{2}}\phi_Z'(u) + \phi_Z(u)ue^{\frac{u^2}{2}} \overset{(*)}{=} 0.$$

i.e. $e^{\frac{u^2}{2}}\phi_Z(u) = \phi_Z(0) = 1$, so $\phi_Z(u) = e^{-\frac{u^2}{2}}$.

Consider for $t \in (0,\infty)$, the centered Gaussian random variable on $\mathbb{R}^d$ which has pdf g_t (with respect to λ^d),

$$g_t(x) = \frac{1}{(2\pi t)^{d/2}}e^{-\frac{\|x\|^2}{2t}}, \qquad \|x\|^2 = \sum_{i=1}^d x_i^2.$$

So, if $(Z_1,\ldots,Z_d)$ are IID $\mathrm{N}(0,1)$, then $\sqrt{t}Z$ has density g. So,

$$\begin{aligned}
\widehat{g_t}(y) &= \mathbb{E}(e^{i\langle u,\sqrt{t}z\rangle}) \\
&= \mathbb{E}\left(e^{i\sum_{i=1}^d u_i\sqrt{t}z_i}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^d e^{iu_i\sqrt{t}z_i}\right) \\
&= \prod_{i=1}^d \mathbb{E}(e^{iu_i\sqrt{t}z_i}) && (z_i \text{ independent}) \\
&= \prod_{i=1}^d \phi_Z(\sqrt{t}u_i) \\
&= \prod_{i=1}^d e^{-t\frac{u_i^2}{2}} \\
&= e^{-\frac{t\|u\|^2}{2}}
\end{aligned}$$

Hence,

$$\widehat{g_t}(y) = e^{-t\|u\|^2/2} = \frac{(2\pi)^{d/2}}{t^{d/2}} = \left(\frac{t}{2\pi}\right)^{d/2} e^{-t\|u\|^2/2} = \frac{(2\pi)^{d/2}}{t^{d/2}}g_{1/t}(u).$$

So,

$$\widehat{\widehat{g}}_t(u) = \frac{(2\pi)^{d/2}}{t^{d/2}}\widehat{g_{1/t}}(u) = (2\pi)^d g_t(u).$$

Then

$$g_t(x) = g_t(-x) = (2\pi)^{-d}\widehat{\widehat{g}}_t(-x) = (2\pi)^{-d}\int \widehat{g_t}(u)e^{-i\langle u,x\rangle}\mathrm{d}u.$$

Thus the Fourier Inversion holds for g_t.

Definition (Gaussian convolution)**.** For $f \in L^1(\mathbb{R}^d)$, the *Gaussian convolution* of f is $f * g_t$ where

$$g_t(u) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{|u|^2}{2t}}.$$

Start of

lecture 19

Facts

(1) $\|f * g_t\|_1 \le \|f\|_1$.

(2) $f * g_t$ is continuous.

(3) $f * g_t$ is bounded.

(4) $\widehat{f * g_t}(w) = \widehat{f}(u)\widehat{g_t}(u) = \widehat{f}(u) e^{-\frac{t|u|^2}{2}}$.

(5) $\widehat{f * g_t}$ is bounded continuous.

(6) $\|\widehat{f * g_t}\|_1 \le c_t \|\widehat{f}\|_\infty \le c_t \|f\|_1$.

(7) For μ a probability measure, and any $t > 0$, $\mu * g_t$ is a Gaussian convolution. Note that

$$\mu * g_t = \mu * (g_{\frac{t}{2}} * g_{\frac{t}{2}}) = \underbrace{(\mu * g_{\frac{t}{2}})}_{\in L^1} * g_{\frac{t}{2}}.$$

Lemma. Fourier Inversion holds for Gaussian convolutions.

Proof. Let $f \in L^1$, $t > 0$. Then

$$\begin{aligned}
(2\pi)^d f * g_t(x) &= (2\pi)^d \int f(x-y) g_t(y) \mathrm{d}y \\
&= \int f9x-y) \cdot (2\pi)^d g_t(y) \mathrm{d}y \\
&= \int f(x-y) \int \widehat{g_t}(u) e^{-i\langle u,y\rangle} \mathrm{d}u \mathrm{d}y && \text{(Fourier Inversion)} \\
&= \iint f(x-y) \widehat{g_t}(u) e^{-i\langle u,y\rangle} \mathrm{d}u \mathrm{d}y \\
&= \int \widehat{g_t}(u) \left(\int f(x-y) e^{-\langle u,y\rangle} \mathrm{d}y \right) \mathrm{d}u && \text{(Fubini-Tonelli)} \\
&= \int \widehat{g_t}(u) \left(\int f(y) e^{-i\langle u,x-y\rangle} \mathrm{d}y \right) \mathrm{d}u \\
&= \int \widehat{g_t}(u) e^{-i\langle u,x\rangle} \left(\int f(y) e^{i\langle u,y\rangle} \mathrm{d}y \right) \mathrm{d}u \\
&= \int \widehat{g_t}(u) e^{-i\langle u,x\rangle} \widehat{f}(u) \mathrm{d}u \\
&= \int \widehat{g_t}(u) \widehat{f}(u) e^{-\langle u,x\rangle} \mathrm{d}u \\
&= \widehat{f * g_t}(u) e^{-i\langle u,x\rangle} \mathrm{d}u
\end{aligned}$$

□

$f * g_t$ as $t \to 0$, $f * g_t$ "$\to$" $f * \delta_0 = f$.

Lemma. Let $f \in L^p(\mathbb{R}^d)$, $1 \le p < \infty$. Then $\|f * g_t - f\|_p \to 0$ as $t \to 0$.

Proof. Given $\varepsilon > 0$, there exists $h \in C_c(\mathbb{R}^d)$ (continuous functions with compact support), such that $\|f - h\|_p \le \frac{\varepsilon}{3}$. Then by linearity of $*$,

$$\|f * g_t - h * g_t\|_p = \|\underbrace{(f-h)}_{\in L^p} * g_t\|_p \le \|f - h\|_p \le \frac{\varepsilon}{3}.$$

So by Minkowski inequality,

$$\begin{aligned}
\|f * g_t - f\|_p &\le \underbrace{\|f * g_t - h * g_t\|_p}_{\le \frac{\varepsilon}{3}} + \underbrace{\|f - h\|_p}_{\le \frac{\varepsilon}{3}} + \|h * g_t - h\|_p \\
&\le 2\frac{\varepsilon}{3} + \|h * g_t - h\|_p
\end{aligned}$$

So it is enough to prove that $\|h * g_t - h\|_p \to 0$. So h is bounded and h is supported on $[-M, M]^d$ say, for some $M > 0$. Define

$$e(y) = \int |h(x-y) - h(x)|^p \mathrm{d}x.$$

Then as $y \to 0$, $|h(x-y)-h(x)|^p \to 0$ as h is continuous. Also,

$$|h(x-y)-h(x)|^p \le 2^p \|h\|_\infty^p \mathbb{1}_{|x|\le M+1}$$

for $|y|<1$. Hence, by DCT, $e(y) \to 0$ as $y \to 0$. Also,

$$\begin{aligned}
\|h * g_t - h\|_p^p &= \int \left| \int h(x-y) g_t(y) \mathrm{d}y - \int h(x) g_t(y) \mathrm{d}y \right|^p \mathrm{d}x \\
&= \int \left| \int (h(x-y)-h(x)) g_t(y) \mathrm{d}y \right|^p \mathrm{d}x \\
&\le \iint |h(x-y)-h(x)|^p g_t(y) \mathrm{d}y \mathrm{d}x && (p \ge 1\text{, Jensen's Inequality}) \\
&= \int \left(\int |h(x-y)-h(x)|^p \mathrm{d}x \right) g_t(y) \mathrm{d}y && \text{(Fubini-Tonelli)} \\
&= \int e(y) g_t(y) \mathrm{d}y \\
&= \int e(y) \frac{1}{t^{d/2}} g_1\left(\frac{y}{\sqrt{t}}\right) \mathrm{d}y \\
&= \int e(\sqrt{t}y) g_1(y) \mathrm{d}y \\
&\to 0 && \text{(DCT)}
\end{aligned}$$

(e is bounded so $e(\sqrt{t}y)g_1 \le C g_1$). □

Theorem (Fourier Inversion)**.** Let $f \in L^1(\mathbb{R}^d)$ and $\widehat{f} \in L^1(\mathbb{R}^d)$. Then

$$f(x) = \frac{1}{(2\pi)^d} \int e^{-i\langle u,x\rangle} \widehat{f}(u) \mathrm{d}u$$

almost everywhere in $\mathbb{R}^d$.

Proof. Consider $f * g_t$ and

$$f_t(x) = \frac{1}{(2\pi)^d} \int e^{-i\langle x,u\rangle} \underbrace{\widehat{f}(u) \underbrace{e^{-\frac{t|u|^2}{2}}}_{\widehat{g_t}(u)}}_{\widehat{f * g_t}(u)} \mathrm{d}u. \qquad (*)$$

As Fourier Inversion holds for $f * g_t$ ($f * g_t$ is a Gaussian convolution), we have $f * g_t = f_t$. So, $\|f_t - f\|_1 \xrightarrow{t \to 0} 0$, i.e. there exists a subsequence $t_n \downarrow 0$ such that $f_{t_n} \to f$ almost everywhere (so $f_t \xrightarrow{\mathbb{P}} f$). But from $(*)$, as $t \to 0$, $e^{-i\langle x,u\rangle}\widehat{f}(u)e^{-t\frac{|u|^2}{2}} \to e^{i\langle x,u\rangle}\widehat{f}(u)$ and bounded by $|\widehat{f}(u)|$ which is integrable. So by DCT,

$$f_t(x) \overset{t\to 0}{\longrightarrow} \frac{1}{(2\pi)^d}\int e^{-i\langle x,u\rangle}\widehat{f}(u)\mathrm{d}u$$

almost everywhere. Hence

$$f = \frac{1}{(2\pi)^d}\int e^{-i\langle x,u\rangle}\widehat{f}(u)\mathrm{d}u$$

almost everywhere. □

Theorem (Plancherel)**.** For $f, g \in L^1 \cap L^2(\mathbb{R}^d)$,

$$\|f\|_2 = \frac{1}{(2\pi)^{d/2}}\|\widehat{f}\|_2 \qquad \text{and} \qquad \langle f, g\rangle_2 = \frac{1}{(2\pi)^{\frac{d}{2}}}\langle \widehat{f}, \widehat{g}\rangle.$$

$f \mapsto \widehat{f}$.

Remark. Since $L^1 \cap L^2$ is dense in L^2, the linear operator

$$F_0 : L^1 \cap L^2 \to L^2$$
$$F_0(f) = (2\pi)^{-\frac{d}{2}}\widehat{f}$$

extends to an isometry $F : L^2 \to L^2$, which is an isometry by Fourier inversion formula.

Start of lecture 20

Proof. First assume $f \in L^1$ with $\widehat{f} \in L^1$. Then $(x, u) \mapsto f(x)\widehat{f}(u)$ is $\mathrm{d}x\mathrm{d}u$-integrable. So

$$\begin{aligned}
(2\pi)^d\|f\|_2^2 &= (2\pi)^d\int f(x)\overline{f}(x)\mathrm{d}x \\
&= \iint \widehat{f}(u)e^{-i\langle x,u\rangle}\overline{f}(x)\mathrm{d}u\mathrm{d}x < \infty && \text{(Fourier Inversion, and } f \in L^2) \\
&= \int \widehat{F}(u)\overline{\left(\int \underbrace{f(x)e^{i\langle x,u\rangle}}_{\widehat{f}(u)}\right)}\mathrm{d}u && \text{(Fubini-Tonelli)} \\
&= \int \widehat{f}(u)\overline{\widehat{f}(u)}\mathrm{d}u && \text{(Fourier transform)} \\
&= \|\widehat{f}\|_2^2 && (*)
\end{aligned}$$

(Note: $f \in L^1$ and $\widehat{f} \in L^1$ implies $f \in L^2 \cap L^\infty$). Now, let $f \in L^1 \cap L^2$. For $t > 0$, take $f_t = f * g_t \overset{t\to\infty}{\longrightarrow} f$ in L^2 and so

$$\|f_t\|_2 \overset{t\to 0}{\longrightarrow} \|f\|_2. \tag{$**$}$$

Also,

$$|\widehat{f_t}(u)| = |\widehat{f}(u)\widehat{g_t}(u)| = |\widehat{f}(u)|e^{-t\frac{|u|^2}{2}} \uparrow |\widehat{f}(u)|$$

as $t \to 0$.

$$\|\widehat{f_t}\|_2^2 = \int |\widehat{f_t}(u)|^2 \mathrm{d}u \overset{t\to 0}{\longrightarrow} \int |\widehat{f}(u)|^2 \mathrm{d}u = \|\widehat{f}\|_2^2 \tag{$\dagger$}$$

by Monotone Convergence Theorem But, $f_t = f * g_t \in L^1$, and $\widehat{f_t} \in L^1$. So by $(*)$, $(2\pi)^d \|f_t\|_2^2 = \|\widehat{f_t}\|_2^2$. Let $t \to 0$, then LHS $\to (2\pi)^d\|f\|_2^2$ by $(**)$, and RHS $\to \|\widehat{f}\|_2^2$ by (**??**). Hence $(2\pi)^d\|f\|_2^2 = \|\widehat{f}\|_2^2$. Similar proof for $\langle f, g\rangle$. □

Characteristic functions, weak convergence and the CLT

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \widehat{\mu_X} = \int e^{i\langle t,x\rangle}\mathrm{d}\mu_X(x)$$

For dirac measure δ_0, $\widehat{\delta_0} = \int e^{itx}\mathrm{d}\delta_0(x) = 1$ not integrable on $\mathbb{R}$ so Fourier Inversion does not make sense. To circumvent this, we 'test' μ on nice test functions f.

Remark.

(0) 2 probability measures μ and ν on $\mathbb{R}^d$ coincide if and only if

$$\int f \mathrm{d}\mu = \int f \mathrm{d}\nu \tag{$*$}$$

for all $f; : \mathbb{R}^d \to \mathbb{R}$ bounded continuous (Example Sheet 2). In fat, enough to have $(*)$ holds for all $f \in C_c^\infty$ (space of infinitely differentiable functions with compact support). ($\mu : C_c^\infty \to \mathbb{R}, f \mapsto \mu(f)$ linear, continuous (Lf top), hence μ is "Schwarz distribution" $\mathcal{A} \circ f$, $\mu \in (C_c^\infty)^*$).

Definition (Converges weakly)**.** Let (μ_n), μ be Borel probability measures on $\mathbb{R}^d$. Then μ_n *converges to* μ *weakly* if

$$\int f \mathrm{d}\mu_n \overset{n\to\infty}{\longrightarrow} \int f \mathrm{d}\mu \tag{$*$}$$

for all $f : \mathbb{R}^d \to \mathbb{R}$ bounded and continuous.

Remark.

(1) For a sequence of random variables (X_n) and X another random variable, $X_n \to X$ weakly if $\mu_{X_n} \to \mu_X$ weakly.

(2) A sequence (μ_n) can have at most one weak limit (by Remark (0)).

(3) If $X_n \to X$ weakly, and $h : \mathbb{R}^d \to \mathbb{R}^k$ is continuous, then $h(X_n) \to h(X)$ weakly (as random variables in $\mathbb{R}^k$) (continuous mapping theorem) (from definition as $f \circ h$ bounded continuous if f bounded continuous).

(4) Sufficient to check $(*)$ for all $f \in C_c^\infty$ ("tightness" argument, i.e. there exists K compact such that $\mu_n(K^c) < \varepsilon$ for all n if $\mu_n \to \mu$ weakly, see Example Sheet 4).

(5) When $d = 1$, this is equivalent to $X_n \xrightarrow{d} X$ (i.e. $F_{X_n}(x) \to F_X(x)$ at all points where $x \mapsto F_X(x)$ is continuous). (Example Sheet 4) $\mathbb{R}^d$, $F(x_1, \dots, x_d) = \mathbb{P}(x \le (x_1, \dots, x_d))$.

Theorem. Let X be a random variable on $\mathbb{R}^d$. Then μ_X is uniquely determined by $\widehat{\mu_X} = \phi_X$. Further, if $\phi_X \in L^1$, then μ_X has a bounded continuous pdf given by

$$f(x) := \frac{1}{(2\pi)^d} \int \phi_X(u) e^{-i\langle x,u\rangle} \mathrm{d}u.$$

Proof. Take $Z \sim \mathrm{N}(0, I_d)$ independent of X. Thus $\sqrt{t}Z$ has pdf g_t and $X + \sqrt{t}Z$ has pdf $\mu_X * g_t =: f_t$. Then

$$\widehat{f_t}(u) = \widehat{\mu_X}(u)\widehat{g_t}(u) = \phi_X(u) e^{-t\frac{|u|^2}{2}}.$$

So by Fourier Inversion of Gaussian convolution,

$$f_t(x) = \frac{1}{(2\pi)^d} \int \phi_X(u) e^{-t\frac{|u|^2}{2}} e^{-i\langle u,x\rangle} \mathrm{d}u \qquad \forall x$$

i.e. f_t is uniquely determined by ϕ_X. Now for any g bounded continuous, $g : \mathbb{R}^d \to \mathbb{R}$, as $t \to 0$,

$$\int g(x) f_t(x) \mathrm{d}x = \mathbb{E}(g(x + \sqrt{t}Z)) \xrightarrow{\text{BCT}} \mathbb{E}(g(X)) = \int g(x) \mu_X(\mathrm{d}x) \tag{$*$}$$

i.e. $\int g(x) \mathrm{d}\mu_X$ is uniquely determined by ϕ_X. Hence μ_X is uniquely determined by ϕ_X (Remark (0)). If $\phi_X \in L^1$, then

$$\phi_X(u) e^{-t\frac{|u|}{2}} e^{-i\langle u,x\rangle} \xrightarrow{t \to 0} \phi_X(u) e^{-i\langle u,x\rangle}$$

By DCT, $f_t(x) \xrightarrow{t \to 0} f_X(t)$ for all x. In particular, $f_X(x) \geq 0$ for all x and $|f_t(x)| \leq \frac{1}{(2\pi)^d} \|\phi_X\|_1$. Then for any g bounded continuous with compact support,

$$\int \underbrace{g(x) f_t(x)}_{\to g(x) f_X(x)} \, \mathrm{d}x \overset{\text{DCT}}{\longrightarrow} \int g(x) f_X(x) \mathrm{d}x$$

Also, LHS $\to \int f(x) \mu_X(\mathrm{d}x)$ from $(*)$. So

$$\int g(x) \mu_X(\mathrm{d}x) = \int g(x) f_X(x) \mathrm{d}x$$

for all g bounded continuous with compat support. Hence μ_X has density f_X (Remark (0)). $\square$

Theorem (Levy)**.** Let (X_n), X be random variables on $\mathbb{R}^d$ with $\phi_{X_n}(u) \to \phi_X(u)$ for all u as $n \to \infty$. Then $X_n \to X$ weakly.

Remark.

(1) Levy's continuous theorem states that if $\phi_{X_n}(u) \to \phi(u)$ for all u for some function ϕ that is continuous in a neighbourhood of 0, then ϕ is the characteristic function of some random variable X and $X_n \to X$ weakly.

(2) Cramér Wold device: Let (X_n), X be random variables on $\mathbb{R}^d$, then $X_n \to X$ weakly if and only if

$$\langle u, X_n \rangle \overset{\text{weakly}}{\longrightarrow} \langle u, X \rangle \qquad \forall u \in \mathbb{R}^d$$

(hence $\phi_{X_n}(u) \to \phi_X(u)$ by BCT and Levy).

Start of lecture 21

Remark. For probability measures, if X_n to X weakly, then $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ for all f bounded continuous.

Proof of Levy. Let $g : \mathbb{R}^d \to \mathbb{R}$ be compactly supported and Lipschitz continuous, i.e.

$$|g(x) - g(y)| \leq Cg|x - y| \qquad \forall x, y \in \mathbb{R}^d$$

(any $g \in C_c^\infty$ will do). Enough to show $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ (by Remark (4) last lecture). Let $Z \sim \mathrm{N}(0, I_d)$ independent of (X_n), X. Then for fixed $\varepsilon > 0$, choose t small enough so that

$$Cg\sqrt{t}\mathbb{E}|Z| \leq \frac{\varepsilon}{3}.$$

Then,

$$\begin{aligned}
|\mathbb{E}g(X_n) - \mathbb{E}g(X)| &\leq |\mathbb{E}g(X_n) - \mathbb{E}g(X_n + \sqrt{t}Z)| + |\mathbb{E}g(X) - \mathbb{E}g(X + \sqrt{t}Z)| \\
&\quad + |\mathbb{E}g(X_n + \sqrt{t}Z) - \mathbb{E}g(X + \sqrt{t}Z)| \\
&\leq \mathbb{E}|g(X_n) - g(X_n + \sqrt{t}Z)| + \mathbb{E}|g(X) - g(X + \sqrt{t}Z)| \\
&\quad + |\mathbb{E}g(X_n + \sqrt{t}Z) - \mathbb{E}g(X + \sqrt{t}Z)| \\
&\leq \mathbb{E}|Cg\sqrt{t}Z| + \mathbb{E}|Cg\sqrt{t}\mathbb{E}Z| + |\mathbb{E}g(X_n + |sqrttZ) - \mathbb{E}g(X + \sqrt{Z})| \\
&\leq \underbrace{Cg\sqrt{t}\mathbb{E}|Z| + Cg\sqrt{t}\mathbb{E}|Z|}_{\leq 2\frac{\varepsilon}{3}} + |\mathbb{E}g(X_n + \sqrt{t}Z) - \mathbb{E}g(X + \sqrt{t}Z)|
\end{aligned}$$

$X_n + \sqrt{t}Z$ has density $\mu_{X_n} * g_t =: f_{t,n}$. Then by Fourier Inversion,

$$f_{t,n}(x) = \frac{1}{(2\pi)^d} \int \phi_{X_n}(u) e^{-\frac{t|u|^2}{2}} e^{-i\langle u, x\rangle} \mathrm{d}u$$

So

$$\begin{aligned}
\mathbb{E}g(X_n + \sqrt{t}Z) &= \frac{1}{(2\pi)^d} \iint g(x)\phi_{X_n}(u) e^{-\frac{t|u|^2}{2}} e^{-i\langle u, x\rangle} \mathrm{d}u\mathrm{d}x \\
&\underset{n\to\infty}{\overset{\text{DCT}}{\longrightarrow}} \frac{1}{(2\pi)^d} \iint g(x)\phi_{X}(u) e^{-\frac{t|u|^2}{2}} e^{-i\langle u, x\rangle} \mathrm{d}u\mathrm{d}x \\
&= \mathbb{E}(X + \sqrt{t}Z)
\end{aligned}$$

(using $g(x)e^{-\frac{t|u|^2}{2}}$ as the bounding function to apply DCT). □

Theorem (Central Limit Theorem)**.** Let (X_n) be IID random variables on $\mathbb{R}$ with $\mathbb{E}(X_i) = 0$, $\mathrm{Var}(X_i) = 1$ for all i. Then for $S_n = X_1 + \cdots + X_n$, we have, $\frac{S_n}{\sqrt{n}} \to Z \sim \mathrm{N}(0, 1)$ weakly or *in distribution*, i.e.

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \leq x\right) \overset{n\to\infty}{\longrightarrow} \mathbb{P}(Z \leq x) \qquad \forall x$$

Proof. Set $\phi(u) = \phi_{X_1}(u) = \mathbb{E}e^{iuX_1}$. Then $\phi(0) = 1$ and since $\mathbb{E}X_1^2 < \infty$, we can differentiate under the integral sign and get

$$\begin{aligned}
\phi'(u) &= i\mathbb{E}X_1 e^{iuX_1} \\
\phi''(u) &= i^2\mathbb{E}X_1^2 e^{iuX_1}
\end{aligned}$$

(see Example Sheet 3, Question 7.4), i.e. $\phi'(0) = 0$, $\phi''(0) = -1$. By Taylor's Theorem, as $u \to 0$, $\phi(u) = 1 - \frac{u^2}{2} + o(u^2)$. Let ϕ_n be the characteristic function of $\frac{S_n}{\sqrt{n}}$. Then

$$\begin{aligned}
\phi_n(u) &= \mathbb{E}e^{iu\frac{S_n}{\sqrt{n}}} \\
&= \mathbb{E}e^{i\frac{u}{\sqrt{n}}(X_1+\cdots+X_n)} \\
&\overset{\text{IID}}{=} (\mathbb{E}e^{i\frac{u}{\sqrt{n}}X_1})^n \\
&= \left(\phi\left(\frac{u}{\sqrt{n}}\right)\right)^n \\
&= \left(1 - \frac{u^2}{2n} + o\left(\frac{u^2}{n}\right)\right)^n \\
&= \left(1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right)^n
\end{aligned}$$

(as u fixed, $n \to \infty$). The complex logarithm satisfies, as $z \to 0$, $\log(1 + z) = z + o(z)$. So,

$$\begin{aligned}
\log \phi_n(u) &= n \log \left(1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right) \\
&= n\left(-\frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right) \\
&= -\frac{u^2}{n} + o(1)
\end{aligned}$$

So, $\phi_n(u) \to e^{-\frac{u^2}{2}} = \phi_Z(u)$. So by Levy, $\frac{S_n}{\sqrt{n}} \to Z$ weakly. □

Remark. The Central Limit Theorem in $\mathbb{R}^d$ can be proved similarly using the Cramér Wold device and properties of multivariate Gaussians (Exercise).

A random variable on $\mathbb{R}$ is Gaussian ($\mathrm{N}(\mu, \sigma^2)$) if it has density $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for $\mu \in \mathbb{R}$, $\sigma > 0$.

Definition. X in $\mathbb{R}^d$ is Gaussian if

$$\langle u, X\rangle = \sum_{i=1}^d u_i X_i$$

is Gaussian for all $u \in \mathbb{R}^d$.

Example. If $X_1, \dots, X_n \overset{\text{IID}}{\sim} \mathrm{N}(0,1)$, then $X = (X_1, \dots, X_n)$ is Gaussian in $\mathbb{R}^n$.

Proposition. Let X be Gaussian in $\mathbb{R}^n$ and A be an $m \times n$ matrix, $b \in \mathbb{R}^m$. Then

(a) $AX + b$ is Gaussian in $\mathbb{R}^m$.

(b) $X \in L^2$ and μ_X is determined by $\mathbb{E}(X) = \mu$ and $(\mathrm{Cov}(X_i, X_j))_{i,j} = V$.

(c) $\phi_X(u) = e^{i\langle u, \mu\rangle - \frac{\langle u, Vu\rangle}{2}}$ for all $u \in \mathbb{R}^n$.

(d) If V is invertile, then X has a pdf on $\mathbb{R}^n$ given by

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|V|^{1/2}} e^{-\langle x-\mu, V^{-1}(x-\mu)\rangle/2}$$

(e) If $X = (X_1, X_2)$ with $X_1 \in \mathbb{R}^{n_1}$, $X_2 \in \mathbb{R}^{n_2}$, $(n_1 + n_2 = n)$, then X_1, X_2 independent if and only if $\mathrm{Cov}(X_1, X_2) = 0$.

Proof. Easy. Also see Example Sheet 4 and lecturer's online notes. □

Law of Large Numbers

Weak Law of Large Numbers: For (X_i) IID with $\mathbb{E}X_i = \mu$, $\mathrm{Var}(X_i) < \infty$. Then $\forall \varepsilon > 0$,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) &\le \frac{1}{n^2\varepsilon^2} \mathrm{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{\cancel{n}\mathrm{Var}(X_1)}{n^{\cancel{2}}\varepsilon^2} \\ &\to 0 \end{aligned}$$

as $n \to \infty$. So,

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu$$

Strong Law of Large Numbers: if (X_i) are IID, $\mathbb{E}(X_1) = \mu < \infty$, then

$$\frac{1}{n}\sum_{i=1}^n X_i \to \mu$$

almost surely as $n \to \infty$.

Start of lecture 22

Proposition. Let (X_n) be independent with $\mathbb{E}X_n = \mu$ and $\mathbb{E}X_n^4 \le M$ for all n. Then

$$\frac{S_n}{n} \to \mu$$

almost surely as $n \to \infty$ (where $S_n = X_1 + \cdots + X_n$).

Proof. Let $X_n' = X_n - \mu$. Then

$$\mathbb{E}X_n'^4 \le 2^4(\mathbb{E}X_n^4 + \mu^4) \le \underbrace{2^4(M + \mu^4)}_{=M'} \qquad \forall n.$$

So we assume that $\mu = 0$ (without loss of generality). Then using independence, for distinct indices i, j, k, l,

$$0 = \mathbb{E}(X_i X_j^3) = \mathbb{E}(X_i X_j X_k^2) = \mathbb{E}(X_i X_j X_k X_l)$$

Hence,

$$\begin{aligned}
\mathbb{E}(S_n^4) &= \mathbb{E}(X_1 + \cdots + X_n)^4 \\
&= \mathbb{E}\left(\sum_{1 \le i \le n} X_i^4 + 6 \sum_{1 \le i < j \le n} X_i^2 X_j^2 \right) \\
&\le nM + 6\frac{n(n-1)}{2} \\
&\le nM + 3n(n-1) \\
&\le 3n^2 M
\end{aligned}$$

But

$$\mathbb{E}(X_i^2 X_j^2) \overset{\text{C-S}}{\le} \sqrt{\mathbb{E}X_i^4 X_j^4} \le M,$$

i.e.

$$\mathbb{E}\left(\left(\frac{S_n}{n} \right)^4 \right) \le \frac{3M}{n^2}$$

Now,

$$\mathbb{E}\left(\sum_{n=1}^\infty \left(\frac{S_n}{n} \right)^4 \right) \overset{\text{Monotone Convergence Theorem}}{=} \sum_{n=1}^\infty \mathbb{E}\left(\frac{S_n}{n} \right)^4 \le \sum_n \frac{3M}{n^2} < \infty.$$

So,

$$\sum_{n=1}^\infty \left(\frac{S_n}{n} \right)^4 < \infty$$

almost surely. Hence $\left(\frac{S_n}{n}\right)^4 \to 0$ almost surely, i.e. $\frac{S_n}{n} \to 0$ almost surely. $\square$

6 Ergodic Theory

Definition (Measure preserving map)**.** Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space. A measurable map $\theta : E \to E$ is called (μ-)measure-preserving (m.p.) if

$$\mu \circ \theta^{-1} = \mu,$$

i.e. $\mu(\theta^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{E}$. In this case, for all $f \in L^1$,

$$\int_E f \mathrm{d}\mu = \int_E f \circ \theta \mathrm{d}\mu = \int f \mathrm{d}\mu \circ \theta^{-1}$$

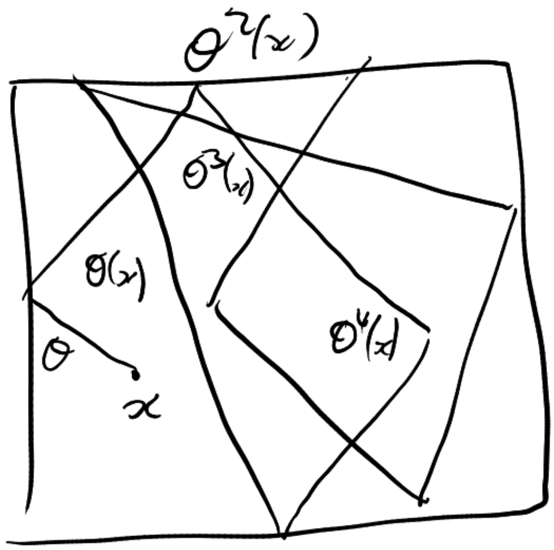
Definition (θ-invariant)**.** A set $A \in \mathcal{E}$ is *θ-invariant* if $\theta^{-1}(A) = A$.

Definition (θ-invariant function)**.** A measurable function f is *θ-invariant* if $f = f \circ \theta$.

The space of all θ-invariant sets $\mathcal{E}_\theta$ is a σ-algebra and f is θ-invariant if and only if f is $\mathcal{E}_\theta$-measurable (exercise on Example Sheet 4).

Definition (Ergodic map)**.** The map θ is called *ergodic* if $\mathcal{E}_\theta$ is μ-trivial, i.e. $\forall A \in \mathcal{E}_\theta$, $\mu(A) = 0$ or $\mu(A^c) = 0$. “well-mixed”.

Boltzman (1880) Ergodic hypothesis for dynamical systems



"space filling". For Markov chains, ergodicity $\iff$ irreducibility.

Fact: If $f : E \to \mathbb{R}$ is θ-invariant, θ Is ergodic if and only if $f = c$, a constant, almost surely (exercise on Example Sheet 4). $(\mu(f^{-1}(-\infty, x)) = 0$ or $\mu(f^{-1}[x, \infty)) = 0)$.

Example. On $((0,1], \mathcal{B}, \lambda_{(0,1]})$ the maps

(1) $\theta_a(x) = x + a \pmod 1$ (rotation of a circle).

(2) $\theta(x) = 2x \pmod 1$

are measure preserving and ergodic unless $a \in \mathbb{Q}$ (see Example Sheet 4).

Theorem (Birkhoff's ergodic theorem (1931))**.** SUppose $(E, \mathcal{E}, \mu)$ is σ-finite and $f \in L^1(E, \mathcal{E}, \mu)$ and $\theta : E \to E$ measure preserving. Define $S_0 = 0$ and

$$S_n = S_n(f) = f + f \circ \theta + f \circ \theta^2 + \cdots + f \circ \theta^{n-1}.$$

Then there exists a θ-invariant function $\overline{f}$ with $\mu(|\overline{f}|) \le \mu(|f|)$ such that $\frac{S_n(f)}{n} \to \overline{f}$ almost everywhere as $n \to \infty$.

Remark. If θ is ergodic, $\overline{f} = c$ almost everywhere.

Lemma (Maximal ergodic theorem)**.** For $f \in L^1(\mu)$, set $S^* = S^*(f) = \sup_{n \geq 0} S_n(f)$. Then
$$\int_{\{S^*>0\}} f \mathrm{d}\mu \geq 0.$$

Proof. Set $S_n^* = \max_{0 \leq m \leq n} S_m$. Then for $m = 1, 2, \dots, n+1$,
$$S_m = f + S_{m-1} \circ \theta \leq f + S_n^* \circ \theta$$
(as for $m = 1, \dots, n+1$, $S_{m-1} \leq S_n^*$, hence $S_{m-1} \circ \theta \leq S_n^* \circ \theta$). On $A_n := \{S_n^* > 0\}$, we have
$$S_n^* = \max_{1 \leq m \leq n} S_m \leq \max_{1 \leq m \leq n+1} S_m \leq f + S_n^* \circ \theta$$
So integrating,
$$\int_{A_n} S_n^* \mathrm{d}\mu \leq \int_{A_n} f \mathrm{d}\mu + \int_{A_n} S_n^* \circ \theta \mathrm{d}\mu \tag{1}$$
On A_n^c, we have $S_n^* = 0 \leq S_n^* \circ \theta$ (as $S_n^* \geq 0$ since $S_0 = 0$). Thus,
$$\int_{A_n^c} S_n^* \mathrm{d}\mu \leq \int_{A_n^c} S_n^* \circ \theta \mathrm{d}\mu \tag{2}$$
Then (1) + (2) gives
$$\int_E S_n^* \mathrm{d}\mu \leq \int_{A_n} f \mathrm{d}\mu + \int_E S_n^* \circ \theta \mathrm{d}\mu$$
i.e.
$$\int S_n^* \mathrm{d}\mu \leq \int_{A_n} f \mathrm{d}\mu + \int S_n^* \mathrm{d}\mu$$
Hence, (as $S_n^* \in L^1$), $\int_{A_n} f \mathrm{d}\mu \geq 0$ $(*)$ for all n.
$$A_n = \{S_n^* > 0\} = \{\max_{0 \leq m \leq n} S_m > 0\} = \bigcup_{m=0}^{n} \{S_m > 0\} \uparrow \bigcup_{m=0}^{\infty} \{S_m > 0\} = \{\underbrace{\sup S_m}_{=S^*} > 0\}.$$
Hence $f \mathbb{1}_{A_n} \to f \mathbb{1}_{(S^*>0)}$. Hence (as $|f \mathbb{1}_{A_n}| \leq |f|$ and $f \in L^1$) by DCT (and using $(*)$),
$$0 \leq \int_{A_n} f \mathrm{d}\mu \to \int_{(S^*>0)} f \mathrm{d}\mu$$
Hence $\int f \mathrm{d}\mu \geq 0$. □

Remark. Let μ be a finite measure. Then for $f \in L^1$ and any $\alpha > 0$, define $\overline{S_k} = \frac{S_k(f)}{k}$ and $\overline{S}^* = \sup_{k\geq 0} \overline{S_k}$, then

$$\underbrace{\mu(\overline{S}^* > \alpha)}_{\mu(\sup\limits_{\underbrace{k\geq 0}_{\geq S_1 = f}} \overline{S_k} > \alpha)} \leq \frac{1}{\alpha}\int_{\{\overline{S}^* > \alpha\}} f \mathrm{d}\mu \leq \frac{1}{\alpha}\int |f| \mathrm{d}\mu$$

Proof. Exercise. □

$\mu(f > \alpha) \leq \frac{\int |f| \mathrm{d}\mu}{\alpha}$ is Markov.

Remark. For μ a probability maesure and $f \in L^1(\mu)$, show that $\left\{\frac{S_n(f)}{n} : n \in \mathbb{N}\right\}$ is UI (exercise). Hence $\frac{S_n(f)}{n} \xrightarrow{L^1} \overline{f}$. If θ ergodic, $\overline{f} = \int f \mathrm{d}\mu$ almost surely.

Start of lecture 23

Theorem (Birkhoff's Ergodic Theorem). $(E, \mathcal{E}, \mu)$ σ-finite and $f \in L^1(\mu)$ and $\theta : E \to E$ is measure preserving. $S_0 = 0$ and

$$S_n = S_n(f) = f + f \circ \theta + f \circ \theta^2 + \dots + f \circ \theta^{n-1} = f + S_{n-1} \circ \theta.$$

Then there exists a θ-invariant $\overline{f}$ with $\mu(|\overline{f}|) \leq \mu(|f|)$ such that $\frac{S_n}{n} \to \overline{f}$ almost everywhere as $n \to \infty$.

Lemma (Maximal Ergodic lemma). For $f \in L^1(\mu)$ and $S^* = \sup_{n\geq 0} S_n(f)$, we must have

$$\int_{\{S^* > 0\}} f \mathrm{d}\mu \geq 0.$$

Proof.

$$\mu(|S_n|) \leq \sum_{i=0}^{n-1} \mu(|f \circ \theta^i|) \overset{\theta \text{ measure preserving}}{=} \sum_{i=1}^{n-1} \mu(|f|) = n|f|$$

So $\mu\left(\left|\frac{S_n}{n}\right|\right) \le \mu(|f|)$. So

$$\mu(|\overline{f}|) = \mu\left(\liminf\left|\frac{S_n}{n}\right|\right) \overset{\text{Fatou's Lemma}}{\le} \liminf \mu\left(\left|\frac{S_n}{n}\right|\right) \le \mu(|f|).$$

Note that $\frac{S_n \circ \theta}{n} = \frac{S_{n+1} - f}{n+1} \times \frac{n+1}{n}$, so $\limsup \frac{S_n}{n} \circ \theta = \limsup \frac{S_{n+1}}{n+1} = \limsup \frac{S_n}{n}$. Similarly, $\liminf \frac{S_n}{n} \circ \theta = \liminf \frac{S_n}{n}$. For $a < b$,

$$D = D(a,b) = \left\{\liminf_n \left(\frac{S_n}{n}\right) < a < b < \limsup_n \left(\frac{S_n}{n}\right)\right\}$$

is θ-invariant. Shall show $\mu(D) = 0$.

$$\Delta = \left\{\liminf\left(\frac{S_n}{n}\right) < \limsup\left(\frac{S_n}{n}\right)\right\} = \bigcup_{\substack{a,b\in\mathbb{Q} \\ a<b}} D(a,b)$$

Hence if $\mu(D) = 0$ for all $a < b$, then

$$\mu\left(\bigcup_{\substack{a,b\in\mathbb{Q} \\ a<b}} D(a,b)\right) = 0 \implies \mu(\Delta) = 0.$$

Define

$$\overline{f} = \begin{cases} \liminf\left(\frac{S_n}{n}\right) = \limsup\left(\frac{S_n}{n}\right) & \text{on } \Delta^c \\ 0 & \text{on } \Delta \end{cases}$$

Then $\frac{S_n}{n} \to \overline{f}$ μ almost everywhere and $\overline{f}$ is θ-invariant (as $\liminf \frac{S_n}{n}$ is θ-invariant and Δ is θ-invariant).

Fix $a < b$. Note that $\theta : D \to D$ by invariance and θ is $\mu|_D$-measure preserving. Also, either $b > 0$ or $a < 0$ (if $a < 0$, change f to $-f$, b to $-a$, a to $-b$, then $b = -a > 0$). So assume $b > 0$ without loss of generality. Shall apply Maximal Ergodic lemma on D with $\mu|_D$. For any $B \subseteq D$ measurable and $\mu(B) < \infty$, let $g = f - b\mathbb{1}_B$. Then $g \in L^1$, and on D,

$$S_n(g) = S_n(f) - bS_n(\mathbb{1}_B) \ge S_n(f) - nb > 0$$

for some n. Hence, $S^*(g) = \sup_{n\ge 0} S_n(g) > 0$ on D. Hence

$$\{S^*(g).0\} \cap D = D.$$

Thus by Maximal Ergodic lemma on D, $\int_{\{S^*(g)>0\}\cap D} g \mathrm{d}\mu \ge 0$, i.e.

$$0 \le \int_D (f - b\mathbb{1}_B)\mathrm{d}\mu = \int_D f\mathrm{d}\mu - b\mu(B)$$

hence $b\mu(B) \leq \int_D f \mathrm{d}\mu$. Since μ is σ-finite, there exists (B_n) measurable sets, $B_n \uparrow D$ and $\mu(B_n) < \infty$ for all n. Hence

$$b\mu(D) = b\mu(B_n) \leq \int_D f \mathrm{d}\mu < \infty$$

(as $f \in L^1$). Hence $\mu(D) < \infty$. A similar argument applied to $(-f)$ and $(-a)$ will give

$$(-a)\mu(D) \leq \int_D (-f) \mathrm{d}\mu$$

(just take D instead of B now, $(-f) - (-a)\mathbb{1}_D$). Hence

$$b\mu(D) \leq \int_D f \mathrm{d}\mu \leq a\mu(D)$$

But $a < b$. As $\mu(D) < \infty$, hence $\mu(D) = 0$. □

Theorem (von Neumann's L^p Ergodic Theorem)**.** Let $\mu(E) < \infty$ and $1 \leq p < \infty$. Then for $f \in L^p(\mu)$,

$$\frac{S_n(f)}{n} \to \overline{f}$$

in L^p as $n \to \infty$.

Proof. For any $f \in L^p(\mu)$, $\|f \circ \theta^n\|_p = \|f\|_p$ as θ is μ measure preserving. So by Minkowski inequality,

$$\left\| \frac{S_n(f)}{n} \right\|_p \leq \frac{1}{n} \sum_{i=0}^{n-1} \underbrace{\|f \circ \theta^1\|_p}_{= \|f\|_p} = \|f\|_p$$

Now, let $\varepsilon > 0$ be given, then choose $K < \infty$ such that $\|f - f_K\|_p < \frac{\varepsilon}{3}$ where $f_K = (-K) \vee f \wedge (K)$.

$$\left(\|f - f_K\|_p^p = \int |f - f_K|^p \mathrm{d}\mu \leq \int |f|^p \mathbb{1}_{|f| > K} \mathrm{d}\mu \overset{\text{DCT}}{\longrightarrow} 0 \right)$$

But f_K is bounded and μ a finite measure, so $f_K \in L^1(\mu)$, hence by Birkhoff's Ergodic Theorem, there exists $\overline{f_K} \in L^1$ such that $\frac{S_n(f_K)}{n} \to \overline{f_K}$ almost everywhere. Again, $f \in L^p(\mu) \subset L^1(\mu)$ (as μ is a finite measure), so by Birkhoff, there exists $\overline{f} \in L^1$ such that $\frac{S_n(f)}{n} \to \overline{f}$ almost everywhere as $n \to \infty$. Then

$$\begin{aligned}
\|\overline{f} - \overline{f_K}\|_p^p &= \int |\overline{f} - \overline{f_K}|^p \mathrm{d}\mu \\
&= \int \liminf_n \left| \frac{S_n(f)}{n} - \frac{S_n(f_K)}{n} \right|^p \mathrm{d}\mu \\
&= \int \liminf_n \left| \frac{S_n(f - f_K)}{n} \right|^p \mathrm{d}\mu \\
&\le \liminf \int \left| \frac{S_n(f - f_K)}{n} \right|^p \mathrm{d}\mu \qquad \text{Fatou's Lemma} \\
&= \liminf_n \left\| \frac{S_n(f - f_K)}{n} \right\|_p^p
\end{aligned}$$

Then

$$\begin{aligned}
\left\| \frac{S_n}{n} - \overline{f} \right\|_p &\overset{\text{Fatou's Lemma}}{\le} \left\| \frac{S_n(f)}{n} - \frac{S_n(f_K)}{n} \right\|_p + \|\overline{f} - \overline{f_K}\|_p + \left\| \frac{S_n(f_K)}{n} - \overline{f_K} \right\|_p \\
&= \underbrace{\left\| \frac{S_n(f - f_K)}{n} \right\|_p}_{= \|f - f_K\|_p \le \varepsilon/3} + \underbrace{\|\overline{f} - \overline{f_K}\|_p}_{\le \varepsilon/3} + \underbrace{\left\| \frac{S_n(f_K)}{n} - \overline{f_K} \right\|_p}_{\to 0}
\end{aligned}$$

so $\frac{S_n}{n} \to \overline{f}$ in L^p as $n \to \infty$ as desired. □

Start of lecture 24

Birkhoff: $(E, \mathcal{E}, \mu)$ σ-finite, $f \in L^1(\mu)$ and θ measure preserving, and $S_n(f) = f + f \circ \theta + \cdots + f \circ \theta^{n-1}$. Then there exists a θ-invariant $\overline{f}$ such that $\frac{S_n(f)}{n} \to \overline{f}$ as $n \to \infty$ μ almost everywhere.

Von Neumann: μ is finite, then $\frac{S_n(f)}{n} \to \overline{f}$ in L^1 as $n \to \infty$.

Remark.

(1) If μ a probability measure and θ **??**, then $\overline{f}$ is a constant almost surely, so $\overline{f} = \int \overline{f} \mathrm{d}\mu = \int f \mathrm{d}\mu$. Also,

$$\int f \mathrm{d}\mu = \int \frac{S_n(f)}{n} \mathrm{d}\mu \to \int \overline{f} \mathrm{d}\mu.$$

Hence $\inf \overline{f} \mathrm{d}\mu = \int f \mathrm{d}\mu$. Then, $\frac{S_n(f)}{n} \to \mathbb{E}(f)$ as $n \to \infty$ μ almost surely and in L^1.

(2) For θ measure preserving and $f \in L^1$, $\frac{S_n(f)}{n} \to \mathbb{E}(f \mid \mathcal{E}_0)$ μ almost surely in L^1 as $n \to \infty$. For $f \in L^2$, just (a version of) the projection of f on $L^2(\mathcal{E}_0)$.

Bernoulli Shifts and Strong Law of Large Numbers

On the infinite product space $E = \mathbb{R}^{\mathbb{N}} = \{x = (x_1, x_2, \dots) : x_i \in \mathbb{R} \forall i\}$ consider the cylinder sets

$$\mathcal{A} = \left\{ \prod_{n=1}^{\infty} A_n : A_n \in \mathcal{B} = \mathcal{B}(\mathbb{R}) \forall n, \& A_n = \mathbb{R} \forall n \geq N \text{ for some } N \in \mathbb{N} \right\}.$$

For example,

$$(0,1) \cap \mathbb{R} \cap \mathbb{R} \cap \dots \in \mathcal{A}$$

whereas

$$(0,1) \cap (0,1) \cap (0,1) \cap \dots \notin \mathcal{A}.$$

Then $\mathcal{A}$ is a π-system and $\mathcal{E} = \sigma(\mathcal{A})$. Check:

- $\mathcal{E} = \sigma(\mathcal{A}) = \sigma(f_n : n \in \mathbb{N})$ where $f_n : E \to \mathbb{R}$, $f_n(x) = x_n$ are the coordinate maps.
- $\mathcal{E}$ is the Borel σ-algebra generated by the toplogy of poinwise convergence.

Now consider a sequence of IID random variables $(X_n)_{n \in \mathbb{N}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, (such a sequence always exists), with the common distribution or law $\mu_{X_n} = \mathbb{P} \circ X_n^{-1} = m \forall n$. The map $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$, $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$ is measurable.

The image measure $\mathbb{P} \circ X^{-1} =: \mu$ is a probability measure on $(E, \mathcal{E})$, that satisfies for any $A = A_1 \times \cdots \times A_n \times \mathbb{R} \times \cdots \in \mathcal{A}$,

$$\begin{aligned} \mu(A) &= \mathbb{P} \circ X^{-1}(A) \\ &= \mathbb{P}(X_1 \in A, X_2 \in A_2, \ldots, X_N \in A_N) \\ &= \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2) \cdots \mathbb{P}(X_N \in A_N) && \text{as } X_i \text{ IID} \\ &= m(A_1)m(A_2) \cdots m(A_n) \\ &= \prod_{n=1}^{\infty} m(A_n) \end{aligned}$$

and μ is the unique probability measure on $\mathcal{E}$ such that

$$\mu(A) = \prod_{n=1}^{\infty} m(A_n).$$

Under μ, the coordinate maps f_n are IID with law m.

The probability space $(E, \mathcal{E}, \mu)$ is called the canonical model for an IID sequence of random variables with law m. Define the shift map $\theta : E \to E$ by $\theta(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$ (similar to $x \to 2x \pmod 1$ on $((0,1), \lambda)$).

Theorem. On $(E, \mathcal{E}, \mu)$, the shift map θ is measurable, measure preserving and ergodic.

Proof. Measurable is obvious. measure preserving? Enough to check on $\mathcal{A}$, i.e. for $A = A_1 \times \cdots \times A_n \times \mathbb{R} \times \cdots$. Indeed:

$$\begin{aligned} &\mu \circ \theta^{-1}(A) \\ &\qquad = \mu(\mathbb{R} \times A_1 \times A_2 \times \cdots) \\ &\qquad = \prod_{i=1}^{\infty} m(A_i) \\ &\qquad = \mu(A) \end{aligned}$$

Ergodicity: Recall the tail σ-algebra

$$\tau = \bigcap_n \tau_n$$

where $\tau_n = \sigma(x_{n+1}, x_{n+2}, \ldots) = \sigma(f_{n+1}, f_{n+2})$. For $A = \prod_n A_n \in \mathcal{A}$,

$$\theta^{-n}(A) = \mathbb{R} \times \cdots \times \mathbb{R} \times A_1 \times \cdots = \{x_{n+1} \in A_1, x_{n+2} \in A_2, \ldots\} \in \tau_n \qquad \forall n, \forall A \in \mathcal{A}$$

If $A \in \mathcal{E}_0$, then $\theta^{-1}(A) = A$, so $\theta^{-n}(A) = A$ for all n. So $A =\in \tau_n$ for all n. So $A \in \bigcap_n \tau_n = \tau$, i.e. $\mathcal{E} \subseteq \tau$. But the (x_i) IID and hence τ is μ-trivial (Kolmogorov $0-1$ law), so $\mathcal{E}_0$ is μ-trivial. □

Theorem. Let m be a probability measure on $\mathbb{R}$ such that $\int_{\mathbb{R}} |x| dm(x) < \infty$ and $\int_{\mathbb{R}} x \mathrm{d}m(x) = \nu$. Let $(E, \mathcal{E}, \mu)$ be the canonical model, where the coordinate maps $f_n(x) = x_n$ are IID with law m. Then

$$\mu\left(\left\{x : \frac{x_1 + \cdots + x_n}{n} \to \nu \text{ as } n \to \infty\right\}\right) = 1.$$

Proof. Let $\theta : E \to E$ be the shift map $\theta(x_1, x_2, \dots) = (x_2, x_3, \dots)$. It is measure preserving and ergodic. Consider $f : E \to \mathbb{R}$ as $f(x) = x_1$. Then $f \in L^1(\mu)$ as $\int |f| \mathrm{d}\mu = \int |x_1| \mathrm{d}m(x_1) < \infty$. Also,

$$S_n(f) = f + f \circ \theta + \cdots + f \circ \theta^{n-1} = x_1 + x_2 + \cdots + x_n.$$

Hence by Birkhoff and von Neumann, as θ ergodic, by Remark (1) earlier this lecture,

$$\frac{S_n(f)}{n} = \frac{x_1 + \cdots + x_n}{n} \to \overline{f} = \int f \mathrm{d}\mu \int x_1 \mathrm{d}m(x_1) = \nu$$

μ almost surely. □

Theorem (Kolmogorov Strong Law of Large Numbers (1930))**.** Let (X_n) be a sequence of IID integrable random variables, with $\mathbb{E}X_1 = \nu$. Set $S_n = \sum_{i=1}^n X_i$. Then $\frac{S_n}{n} \to \nu$ almost surely as $n \to \infty$.

Proof. Let m be the law of X_n, $\mu = \mathbb{P} \circ X^{-1}$ where $X : \Omega \to E$ is $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$. Then apply the previous theorem. □

This is the end of the course.

Remark.

(1) If (μ_n) is a sequence of probability measures that converges weakly to μ, then (μ_n) is "tight", i.e. $\forall \varepsilon > 0$, $\exists$ a compact set K such that $\mu_n(K^c) < \varepsilon$ for all n.

(2) If (μ_n) is a sequence of probability measures that are tight, then there exists a subsequence (n_k) and a probability measure μ such that $(\mu_{n_k}) \to \mu$ weakly (Banach-Alaoglu Theorem) (Prokhorov Theorem).

Index

Fourier Inversion 56, 59, 60, 62, 63, 64, 65, 67

Fourier transform 55, 56, 63

$\|f\|_p$ 43, 45, 46, 47, 48, 49, 50, 51, 52, 55, 57, 60, 61, 62, 63, 64, 65, 76, 77

L^p 43, 45, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 68, 71, 72, 73, 74, 75, 76, 77, 80

L^p 43, 47, 48

almost everywhere 25, 26, 29, 31, 32, 34, 35, 38, 43, 48, 49, 50, 56, 57, 62, 63, 72, 74, 75, 76, 77

almost surely 25, 27, 28, 29, 34, 35, 36, 45, 51, 54, 69, 70, 72, 74, 77, 80

ϕ 56, 64, 65, 66, 68

characteristic function 56, 58, 66, 67

converges in distribution 27

weakly 64, 80

convex 44

C_c^∞ 64

?? : ergeodic 77

ergodic 71, 72, 74, 79, 80

eventually 16

convolution 57

FT of f 55, 56, 58, 59, 60, 62, 63, 64, 65

$*$ 57, 58, 59, 60, 61, 62, 63, 64, 65, 67

Gaussian convolution 59, 60, 62, 65

g_t 59, 60, 61, 62, 63, 64, 65, 67

independent 15, 26, 27

independent 23, 24, 25, 26, 27, 28, 41

independent 15, 28

integrable 30, 31

i.o 16, 26, 27

measure preserving 71, 72, 74, 75, 76, 77, 79

$\otimes$ 39, 40, 41, 42

$\otimes$ 38, 40, 41, 57

Rademacher function 24

simple 29, 30, 31, 32

tail σ-algebra 27, 28, 79

τ 27, 28, 79

θ-invariant 71, 75

θ-invariant 71, 72, 74, 75, 77

uniformly integrable 51

UI 51, 52, 53, 54, 74

weakly 64, 66, 67, 68