Measure Theory

Felix Chen

Contents

§1 Introduction

Teacher: Zhang Fuxi
Email: zhangfxi@math.pku.edu.cn
Homepage: http://www.math.pku.edu.cn/teachers/zhangfxi

§1.1 Starting from probablistics

Definition 1.1.1 (σ-algebra)**.** Let $\mathscr{F}$ be a family of subsets of a set Ω, if

- $\Omega \in \mathscr{F}$;
- If $A \in \mathscr{F}$, $A^c \in \mathscr{F}$;
- If $A_1, A_2, \cdots \in \mathscr{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.(Countable union)

then we call $\mathscr{F}$ a σ-algebra.

Some intros about probablistics (left out because I haven't studied probablistics yet;)

§1.2 What is measure theory?

It's an abstract theory, different from probablistics and real analysis. In this course we study a general set X, focus on mathematical thinking and skills, from the simple to construct the complex.

Measure theory studies the intrinsic structure of mathematical objects, and the map between different measure spaces.

§2 Measure spaces and measurable maps

§2.1 Sets and set operations

Definition 2.1.1. A non-empty set X is our space(universal set), its elements (points) are denoted by lower case letters $x, y, \ldots$.

Some notations:

$$x \in A, x \notin A, x \in A^c, A \subset B, A \cup B, AB = A \cap B,$$

$$B \backslash A (B - A \text{ when } A \subset B), A \Delta B.$$

A family of sets $\{A_t, t \in T\}$.

$$\bigcup_{t \in T} A_t := \{x : \exists t \in T, s.t. x \in A_t\}, \quad \bigcap_{t \in T} A_t := \{x : x \in A_t, \forall t \in T\}.$$

Sometimes we write the union of disjoint sets as sums to emphasize the disjoint property.

Monotone sequence of sets:

$$A_n \uparrow : A_n \subset A_{n+1}, \forall n; \quad A_n \downarrow : A_n \supseteq A_{n+1}, \forall n.$$

Next we define the limits of sets:

Definition 2.1.2. For monotone sequences:

$$\lim_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} A_n \text{ or } \bigcap_{n=1}^{\infty} A_n.$$

For general sequence of sets:

$$\limsup A_n := \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_n; \quad \liminf A_n := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_n.$$

A clearer intepretation of limsup and liminf:

limsup is the set of elements which occurs infinitely many times in A_n, and liminf is the elements which doesn't occur in only finitely many A_n's.

§2.2 Families of sets

Definition 2.2.1. A family of sets is denoted by script letters $\mathscr{A}, \mathscr{B}, \ldots$.

- A family is a **π-system** if $\mathscr{P} \neq \emptyset$ and it's closed under intersections, e.g. $\{(-\infty, a] : a \in \mathbb{R}\}$.
- **Semi-rings**: $\mathscr{Q}$ is a π-system, and for all $A \subset B$, then there exists finitely many pairwise disjoint sets $C_1, \ldots, C_n \in \mathscr{Q}$ s.t.

  $$B \backslash A = \bigcup_{k=1}^{n} C_k = \sum_{k=1}^{n} C_k.$$

  e.g. $\mathscr{Q} = \{(a, b] : a, b \in \mathbb{R}\}$.

  Remark 2.2.2 — The condition $A \subset B$ can be removed.

- **Rings**: $\mathscr{R}$ is nonempty, and it's closed under union and substraction.
  e.g. $\mathscr{R} = \{\bigcup_{k=1}^{n}(a_k, b_k] : a_k, b_k \in \mathbb{R}\}$.
- **Algebras (fields)**: $\mathscr{A}$ is a π-system that contains X, and is closed under completion.

Proposition 2.2.3

Semi-rings are π-systems, rings are semi-rings, algebras are rings.

Proof. By definition we only need to check rings are π-systems: $A \cap B = A \backslash (A \backslash B)$.

For algebras, $A \cup B = (A^c \cap B^c)^c, A \backslash B = A \cap B^c$, so they are rings. □

Remark 2.2.4 — Rings are semi-rings with unions, Algebras are rings with universal set X.

Definition 2.2.5. Some other families that start from taking limits:

- **Monotone class**: If $A_1, \cdots \in \mathscr{U}$ and A_n monotone, then $\lim_{n\to\infty} A_n \in \mathscr{U}$.

- **λ-system**:
$$X \in \mathscr{L}; \quad A_1, A_2, \cdots \in \mathscr{L}, A_n \uparrow A \implies A \in \mathscr{L};$$
$$A, B \in \mathscr{L}, A \supseteq B \implies A \backslash B \in \mathscr{L}.$$
notes: $A_n \in \mathscr{L} \iff B_n = A_n^c \in \mathscr{L}$.
- **σ-algebra**:
$$X \in \mathscr{F}; \quad A \in \mathscr{F} \implies A^c \in \mathscr{F};$$
$$A_1, A_2, \cdots \in \mathscr{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{F}.$$

Proposition 2.2.6

σ-algebra = algebra & monotone class = λ-system & π-system.

Definition 2.2.7. **σ-rings**: $\mathscr{R}$ nonempty, $A, B \in \mathscr{R} \implies A \backslash B \in \mathscr{R}$;
$$A_1, A_2, \cdots \in \mathscr{R} \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{R}.$$
Note: We only need to verify the case when A_n's are disjoint.

Definition 2.2.8 (Measurable space)**.** Let $\mathscr{F}$ be a σ-algebra on a set X, we say $(X, \mathscr{F})$ is a **measurable space**.

Proposition 2.2.9

Let $(X, \mathscr{F})$ be a measurable space, A is a subset of X. Then $(A, A \cap \mathscr{F})$ is also a measurable space.

The smallest σ-algebra is $\{\emptyset, X\}$, the largest σ-algebra is the power set $\mathscr{T} = \mathcal{P}(X)$.
In some cases, $\mathscr{T}$ is too large, for example, when $X = \mathbb{R}$, we can't assign a "measure" to every subset that fits our common sense.

§2.3 Generation of σ-algebras

Let $\mathscr{E}$ be a nonempty collection of sets.

Definition 2.3.1 (Generate rings)**.** We say $\mathscr{G}$ is the ring (algebra, etc.) generated by $\mathscr{E}$, if

- $\mathscr{G} \supseteq \mathscr{E}$;
- For any ring $\mathscr{G}'$, $\mathscr{G}' \supseteq \mathscr{E} \implies \mathscr{G}' \supseteq \mathscr{G}$

Proposition 2.3.2

The ring (or whatever) generated by $\mathscr{E}$ always exists.

Proof. Let $\mathbf{A}$ be the set consisting of the rings containing $\mathscr{E}$, then $\bigcap_{\mathscr{G} \in \mathbf{A}} \mathscr{G}$ is the desired ring. $\square$

Denote $r(\mathscr{E}), m(\mathscr{E}), p(\mathscr{E}), l(\mathscr{E}), \sigma(\mathscr{E})$ the ring/monotone class/π-system/λ-system/σ-algebra generated by $\mathscr{E}$.

Theorem 2.3.3

Let $\mathscr{A}$ be an algebra, then $\sigma(\mathscr{A}) = m(\mathscr{A})$.

Proof. Clearly $\sigma(\mathscr{A}) \supseteq m(\mathscr{A})$.

On the other hand, we only need to prove $m(\mathscr{A})$ is a σ-algebra.

Since $\mathscr{A}$ is an algebra, so $X \in \mathscr{A} \subset m(\mathscr{A})$.

For the completion:

Let $\mathscr{G} := \{A : A^c \in m(\mathscr{A})\}$, we want to prove $\mathscr{G} \supseteq m(\mathscr{A})$.

Clearly $\mathscr{A} \subset \mathscr{G}$; If $A_1, A_2, \cdots \in \mathscr{G}$, $A_n \uparrow A$, then

$$A_n^c \in m(\mathscr{A}) \implies A^c =\downarrow \lim_n A_n^c \in m(\mathscr{A}).$$

Similarly if $A_n \downarrow A$, we can also deduce $A^c \in m(\mathscr{A})$.

So $\mathscr{G}$ is a monotone class containing $\mathscr{A}$, hence it must contain $m(\mathscr{A}) \implies \forall A \in m(\mathscr{A})$, $A^c \in m(\mathscr{A})$.

For the intersection:

- $\forall A \in \mathscr{A}, B \in m(\mathscr{A}), AB \in m(\mathscr{A})$: If $B \in \mathscr{A}$, this clearly holds;

  Moreover, such B's constitude a monotone class:

  > **Claim 2.3.4.** Let $\mathscr{M}$ be a monotone class, then $\forall C \in \mathscr{M}$, $\mathscr{G}_C = \{D : CD \in \mathscr{M}\}$ is a monotone class.

  If $D_1, D_2, \cdots \to D$ satisfy $C \cap D_i \in m(\mathscr{A})$, then $D \cap C = \lim_n D_i \cap C \in \mathscr{M}$.

  Therefore such B's constitude a monotone class $\mathscr{G}_A$ containing $\mathscr{A} \implies \mathscr{G}_A \supseteq m(\mathscr{A})$.

- All the A's which satisfies the first condition constitude a monotone class:

  Let $\mathscr{G}_B = \{A : AB \in m(\mathscr{A})\}$, then $\mathscr{G} = \bigcup_{B \in m(\mathscr{A})} \mathscr{G}_B$ is a monotone class containing $\mathscr{A}$.

  Hence $\mathscr{G} \supseteq m(\mathscr{A}) \implies \forall A \in m(\mathscr{A})$, $\forall B \in m(\mathscr{A})$, we have $AB \in m(\mathscr{A})$.

□

Theorem 2.3.5 (λ-π theorem)

Let $\mathscr{P}$ be a π-system, then $\sigma(\mathscr{P}) = l(\mathscr{P})$.

Proof. Obviously $\sigma(\mathscr{P}) \supseteq l(\mathscr{P})$.

We only need to check that $l(\mathscr{P})$ is a π-system, i.e. closed under intersection.

> **Claim 2.3.6.** If $\mathscr{L}$ is a λ-system, then $\forall C \in \mathscr{L}$, $\mathscr{G}_C$ is a λ-system, where
>
> $$\mathscr{G}_C := \{D : CD \in \mathscr{L}\}.$$

Proof of the claim. First of all, $X \in \mathscr{G}_C$ as $CX = C \in \mathscr{G}_C$.

Second, if $D_1, D_2 \in \mathscr{G}_C$,

$$CD_1, CD_2 \in \mathscr{L} \implies C(D_1 - D_2) = CD_1 - CD_2 \in \mathscr{L} \implies D_1 - D_2 \in \mathscr{G}_C.$$

Lastly, if $D_n \in \mathscr{G}_C$, $D_n \to D$,

$$CD_n \in \mathscr{L} \implies CD = \lim_n CD_n \in \mathscr{L} \implies D \in \mathscr{G}_C$$

□

The rest is similar to the previous theorem:

- $\forall A \in \mathscr{P}, B \in l(\mathscr{P}), AB \in l(\mathscr{P})$: If $B \in \mathscr{P}$ this clearly holds;
  By the claim, $\mathscr{G}_A = \{B : AB \in l(\mathscr{P})\}$ is a λ-system, so $\mathscr{G}_A \supseteq l(\mathscr{P})$.
- For $B \in l(\mathscr{P})$, let
  $$\mathscr{G}_B = \{A : AB \in l(\mathscr{P})\}.$$
  By our claim, $\mathscr{G}_B$'s are λ-systems. So $\mathscr{G} = \bigcap_{B \in l(\mathscr{P})} \mathscr{G}_B$ is a λ-system.
  Moreover $\mathscr{G} \supseteq \mathscr{P}$ (This is proved above), so $\mathscr{G} \supseteq l(\mathscr{P})$.
  This means $\forall A, B \in l(\mathscr{P}), AB \in l(\mathscr{P})$.

□

Remark 2.3.7 — These two proofs are very similar. Note how we make use of the conditions.

Let X be a topological space, $\mathscr{O}$ is the collection of all the open sets.

Let $\mathscr{B}_X := \sigma(\mathscr{O})$ be the **Borel σ-algebra** on the space X, $B \in \mathscr{B}_X$ are called **Borel sets**, and $(X, \mathscr{B}_X)$ is called the **topological measurable space**.

Theorem 2.3.8

Let $\mathscr{Q}$ be a semi-ring, then

$$r(\mathscr{Q}) = \mathscr{G} := \bigcup_{n=1}^{\infty} \left\{ \sum_{k=1}^{n} A_k : A_1, \dots, A_n \in \mathscr{Q} \text{ and pairwise disjoint} \right\}.$$

Remark 2.3.9 — If $\mathscr{R}$ is a ring, then $\mathscr{A} = a(\mathscr{R}) = \mathscr{R} \cup \{A^c : A \in \mathscr{R}\}$ can also be written out explicitly, while $\sigma(\mathscr{A})$ usually cannot be expressed explicitly.

Proof. Since $r(\mathscr{Q})$ is closed under finite unions, so $r(\mathscr{Q}) \supseteq \mathscr{G}$.

Reversely, $\mathscr{G}$ is nonempty. In fact we only need to prove it's closed under subtraction, as

$$A \cap B = A \backslash (A \backslash B) \in \mathscr{G}, A \cup B = (A \backslash B) \cup (B \backslash A) \cup (A \cap B) \in \mathscr{G}.$$

Suppose $A = \sum_{i=1}^n A_i, B = \sum_{j=1}^m B_j$.

Then $A_i \backslash B_1$ can be split to several disjoint sets C_k in $\mathscr{Q}$. Continue this process, each C_k can be split again into smaller set. When all of the B_j's are removed, we end up with many tiny sets which are in $\mathscr{Q}$ and pairwise disjoint. (This process can be formalized using induction)

Therefore $A \backslash B \in \mathscr{G}$, the conclusion follows. □

§2.4 Measurable maps and measurable functions

For a map $f : X \to Y$, we say the **preimage** of $B \subset Y$ is $f^{-1}(B) := \{x : f(x) \in B\}$.

Some properties of preimage:

$$f^{-1}(\emptyset) = \emptyset, \quad f^{-1}(Y) = X;$$
$$B_1 \subset B_2 \implies f^{-1}(B_1) \subset f^{-1}(B_2), \quad (f^{-1}(B))^c = f^{-1}(B^c);$$
$$f^{-1}\left(\bigcup_{t \in T} A_t\right) = \bigcup_{t \in T} f^{-1}(A_t), \quad f^{-1}\left(\bigcap_{t \in T} A_t\right) = \bigcap_{t \in T} f^{-1}(A_t).$$

Proposition 2.4.1

Let $\mathscr{T}$ be a σ-algebra on Y, then $f^{-1}(\mathscr{T})$ is also a σ-algebra on X.

Furthermore, for $\mathscr{E}$ on Y,

$$\sigma(f^{-1}(\mathscr{E})) = f^{-1}(\sigma(\mathscr{E})).$$

Proof. $f^{-1}(\mathscr{E}) \subset f^{-1}(\sigma(\mathscr{E})) \implies f^{-1}(\sigma(E)) \supseteq \sigma(f^{-1}(\mathscr{E}))$.

Again, let

$$\mathscr{G} := \{B \subset Y : f^{-1}(B) \in \sigma(f^{-1}(\mathscr{E}))\}.$$

We need to prove $\mathscr{G}$ is a σ-algebra. This can be checked easily by previous properties, so I leave them out. Hence $\mathscr{G} \supseteq \mathscr{E} \implies \mathscr{G} \supseteq \sigma(\mathscr{E}) \implies f^{-1}(\sigma(\mathscr{E})) \subset \sigma(f^{-1}(\mathscr{E}))$. $\square$

Definition 2.4.2 (Measurable maps)**.** Let $(X, \mathscr{F})$ and $(Y, \mathscr{S})$, and $f : X \to Y$ a map. We say f is **measurable** if $f^{-1}(\mathscr{S}) \subset \mathscr{F}$, i.e. the preimage of measurable sets are also measurable, denoted by

$$f : (X, \mathscr{F}) \to (Y, \mathscr{S}) \quad \text{or} \quad (X, \mathscr{F}) \xrightarrow{f} (Y, \mathscr{S}) \quad \text{or} \quad f \in \mathscr{F}.$$

Clearly the composition of measurable maps is measurable as well.

A map f is measurable is equivalent to $\sigma(f) \subset \mathscr{F}$, where

$$\sigma(f) := f^{-1}(\mathscr{S})$$

is the smallest σ-algebra which makes f measurable, called the generate σ-algebra of f.

Theorem 2.4.3

Let $\mathscr{E}$ be a nonempty collection on Y, then

$$f : (X, \mathscr{F}) \to (Y, \sigma(\mathscr{E})) \iff f^{-1}(\mathscr{E}) \subset \mathscr{F}.$$

Proof. Trivial. $\square$

Definition 2.4.4 (Generalize real numbers)**.** Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. Similarly we can assign an order to $\overline{\mathbb{R}}$.

For the calculations, we assign 0 to $0 \cdot \pm\infty$, and $\infty - \infty$, $\frac{\infty}{\infty}$ is undefined.

For all $a \in \overline{\mathbb{R}}$, define $a^+ = \max\{a, 0\}, a^- = \max\{-a, 0\}$, so $a = a^+ - a^-$.

Define the Borel σ-algebra on $\overline{\mathbb{R}}$:

$$\mathscr{B}_{\overline{\mathbb{R}}} := \sigma(\mathscr{B}_{\mathbb{R}} \cup \{+\infty, -\infty\}).$$

For any set A, $A \in \mathscr{B}_{\overline{\mathbb{R}}} \iff A = B \cup C$, where $B \in \mathscr{B}_{\mathbb{R}}, C \subset \{+\infty, -\infty\}$.

Definition 2.4.5 (Measurable functions)**.** We say a function f is **measurable** if

$$f : (X, \mathscr{F}) \to (\overline{\mathbb{R}}, \mathscr{B}_{\overline{\mathbb{R}}}).$$

A **random variable (r.v.)** is a measurable map to $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.

Measurable functions are in fact random variables that can take $\pm\infty$ as its value.

Theorem 2.4.6
Let $(X, \mathscr{F})$ be a measurable space, $f : X \to \overline{\mathbb{R}}$ if and only if

$$\{f \le a\} \in \mathscr{F}, \quad \forall a \in \mathbb{R}.$$

Proof. Just note that these sets can generate $\mathscr{B}_{\overline{\mathbb{R}}}$.
Let $\mathscr{E} = \{[-\infty, a] : \forall a \in \mathbb{R}\}$. Then

$$f \text{ measurable} \iff \sigma(f) = f^{-1}\mathscr{B}_{\overline{\mathbb{R}}} = f^{-1}\sigma(\mathscr{E}) \subset \mathscr{F} \iff \sigma(f^{-1}\mathscr{E}) \subset \mathscr{F}.$$

□

Example 2.4.7
The contant functions are measurable; the indicator functions of a measurable set are measurable $\implies$ *step functions* are measurable.

We say a function f is **Borel function** if it's a measurable function from Borel measurable space to itself.

Corollary 2.4.8
If f, g are measurable functions, then $\{f = a\}, \{f > g\}, \dots$ are measurable sets.

Theorem 2.4.9
The arithmetic of measurable funtions are also measurable functions (if they are well-defined).

Proof. Here we only proof $f + g$ is measurable for f, g measurable. For all $a \in \mathbb{R}$, decompose $\{f + g < a\}$ to $A_1 \cup A_2 \cup A_3$:

$$A_1 := \{f = -\infty, g < +\infty\} \cup \{g = -\infty, f < +\infty\} \in \mathscr{F};$$

$$A_2 := \{f = +\infty, g > -\infty\} \cup \{g = +\infty, f > -\infty\} \in \mathscr{F};$$

$$A_3 := \{f < a - g\} \cap \{f, g \in \mathbb{R}\} = \left(\bigcup_{r \in \mathbb{Q}} (\{f < r\} \cap \{g < a - r\})\right) \cap \{f, g \in \mathbb{R}\} \in \mathscr{F}.$$

□

Remark 2.4.10 — All the measurable functions (or random variables) constitude a vector space.

Theorem 2.4.11
The limit inferior and limit superior of measurable functions are measurable.

Proof. If $f_1, f_2, \dots$ are measurable, then $\inf f_n$ is measurable:

$$\left\{\inf_{n\geq 1} f_n \geq a\right\} = \bigcap_{n=1}^{\infty}\{f_n \geq a\}.$$

Remark 2.4.12 — In particular, f measurable $\implies f^+, f^-$ measurable.

Hence

$$\liminf_{n\to\infty} f_n = \lim_{N\to\infty}\inf_{n\geq N} f_n = \sup_{N\geq 1}\inf_{n\geq N} f_n,$$

which is clearly measurable. □

Remark 2.4.13 — The inferior or superior of **countable** many measurable functions are measurable as well.

Definition 2.4.14 (Simple functions)**.** Let $(X, \mathscr{F})$ be a measurable space. A **measurable partition** of X is a collection of subsets $\{A_1, \dots, A_n\}$ with $\sum_{i=1}^n A_i = X$, and $A_i \in \mathscr{F}$.

A **simple function** is defined as

$$f = \sum_{i=1}^n a_i \mathbf{I}_{A_i}.$$

where $\{A_1, \dots, A_n\}$ is a measurable partition of X, and $a_i \in \mathbb{R}$.

It's clear that simple functions are measurable.

Theorem 2.4.15

Let f be a measurable function, there exists simple functions $f_1, \dots$ s.t. $f_n \to f$.

- If $f \geq 0$, we have $0 \leq f_n \leq f$;
- If f is bounded, we have $f_n \rightrightarrows f$.

Proof. This is a standard truncation. For $f \geq 0$, let

$$f_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n}\mathbf{I}_{\{k\leq 2^n f\leq k+1\}} + n\mathbf{I}_{f\geq n}.$$

It's clear that $f_n \geq 0$, $f_n \uparrow$, and $f_n(x) \to f(x)$:

$$0 \leq f(x) - f_n(x) \leq \frac{1}{2^n}, \quad f(x) < n;$$

$$n = f_n(x) \leq f(x), \quad f(x) \geq n.$$

Therefore if f is bounded, when $n > \max f(x)$ we have $|f_n(x) - f(x)| < \frac{1}{2^n}$ for all $x \in X$.

For general measurable functions, just decompose f to $f^+ - f^-$. □

Theorem 2.4.16

Let $g : (X, \mathscr{F}) \to (Y, \mathscr{S})$. Let h be a map $X \to \mathbb{R}$.

Then $h : (X, g^{-1}\mathscr{S})$ iff $h = f \circ g$, where $f : (Y, \mathscr{S}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.

Remark 2.4.17 — For $\overline{\mathbb{R}}$ or $[a, b]$, this theorem also holds.

Proof. There's a typical method for proving something related to measurable functions:

We'll prove the statement for $h \in \mathcal{H}_i$ in order:

- $\mathcal{H}_1$: indicator functions $h = \mathbf{I}_A$, $\forall A \in g^{-1}\mathscr{S}$;
- $\mathcal{H}_2$: non-negative simple functions;
- $\mathcal{H}_3$: non-negative measurable functions;
- $\mathcal{H}_4$: measurable functions.

When $h \in \mathcal{H}_1$, suppose $h = \mathbf{I}_A$, then

$$A = g^{-1}B, B \in \mathscr{S} \implies f = \mathbf{I}_B \text{ suffices.}$$

When $h = \sum_{i=1}^{n} a_i \mathbf{I}_{A_i} \in \mathcal{H}_2$, since $A_i \in g^{-1}\mathscr{S}$,

$$\exists B_i \in \mathscr{S} \quad s.t. \quad A_i = \{h = a_i\} = g^{-1}B_i.$$

Thus $f = \sum_{i=1}^{n} a_i \mathbf{I}_{B_i}$ is the desired function.

When $h \in \mathcal{H}_3$, $\exists h_1, h_2, \ldots \uparrow h$.

Assume $h_n = f_n \circ g$, let

$$f(y) := \begin{cases} \lim\limits_{n\to\infty} f_n(y), & \text{if it exists;} \\ 0, & \text{otherwise.} \end{cases}$$

Remark 2.4.18 — Here we still need to prove f is measurable.

Hence for any $x \in X$,

$$h(x) = \lim_{n\to\infty} h_n(x) = \lim_{n\to\infty} f_n(g(x)) = f(g(x)),$$

as f_n's limit must exist at $y = g(x)$.

So for general h, let $h = h^+ - h^-$ and we're done. NOTE: We need to assert that $\infty - \infty$ doesn't occur. □

Remark 2.4.19 — This is the typical method we'll use frequently in measure theory: to start from simple functions and extend to general functions.

§3 Measure spaces

§3.1 The definition of measure and its properties

The concept of "measure" is frequently used in our everyday life: length, area, weight and even propbability. They all share a similarly: the measure of a whole is equal to the sum of the measure of each part.

In the language of mathematics, let $\mathscr{E}$ be a collection of sets, and there's a function $\mu : \mathscr{E} \to [0, \infty]$ which stands for the measure.

countable additivity: Let $A_1, A_2, \cdots \in \mathscr{E}$ be pairwise disjoint sets, and $\sum_{i=1}^{\infty} A_i \in \mathscr{E}$, then

$$\mu\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Definition 3.1.1 (Measure)**.** Suppose $\emptyset \in \mathscr{E}$, if a non-negative function

$$\mu : \mathscr{E} \to [0, \infty]$$

satisfies countable additivity, and $\mu(\emptyset) = 0$, then we say μ is a **measure** on $\mathscr{E}$.

If $\mu(A) < \infty$ for all $A \in \mathscr{E}$, we say μ is finite. (In practice we'll just simplify this to $\mu(X) < \infty$)
If $\exists A_1, A_2, \cdots \in \mathscr{E}$ are pairwise disjoint sets, s.t.

$$X = \sum_{n=1}^{\infty} A_n, \quad \mu(A_n) < \infty, \forall n.$$

Then we say μ is σ-finite.

There's a weaker version of countable additivity, that is **finite additivity**: If $A_1, \ldots, A_n \in \mathscr{E}$, pairwise disjoint, and $\sum A_i \in \mathscr{E}$,

$$\mu\left(\sum_{i=1}^{n} A_i\right) = \sum_{1=i}^{n} \mu(A_i),$$

then we say μ is finite additive.

Subtractivity: $\mu(B - A) = \mu(B) - \mu(A)$, where $A, B, B - A \in \mathscr{E}$, and $\mu(A) < \infty$.

Proposition 3.1.2
Measure satisfies finite additivity and subtractivity.

Example 3.1.3 (Counting measure)
Let $\mu(A) = \#A$, $\forall A \in \mathscr{T}_X$. Then μ is a measure.

Example 3.1.4 (Point measure)
Let $(X, \mathscr{F})$ be a measurable space, define $\delta_x(A) = \mathbf{I}_A(x)$. Then we can define a measure

$$\mu(A) = \sum_{i=1}^{n} p_i \delta_{x_i}(A)$$

Example 3.1.5 (Length)

Let $\mathscr{E} = \mathscr{Q}_{\mathbb{R}} = \{(a,b] : a,b \in \mathbb{R}\}$, $a \le b$, then $\mu((a,b]) = b-a$ gives a measure.

Another classical example is the so-called "coin space":
Let $X = \{x = (x_1, x_2, \dots) : x_i \in 0, 1, \forall n\}$.

$$C_{i_1,\dots,i_n} := \{x : x_1 = i_1, \dots, x_n = i_n\},$$

Let

$$\mathscr{Q} = \{\emptyset, X\} \cup \{C_{i_1,\dots,i_n} : n \in \mathbb{N}, i_1, \dots, i_n \in \{0,1\}\}$$

be a semi-ring. Then $\mu(C_{i_1,\dots,i_n}) = \frac{1}{2^n}$ gives a measure.

We need to check the countable additivity, but actually this can be realized as a compact space and the C's are open sets, so in fact we only need to check finite additivity. (Or we can prove this explicitly)

Another more complex example: finite markov chain.

Proposition 3.1.6

Let $X = \mathbb{R}$, $\mathscr{E} = \mathscr{R}_{\mathbb{R}}$. $F : \mathbb{R} \to \mathbb{R}$ is non-decreasing, right continuous, then $\mu((a,b]) = F(b) - F(a)$ gives a measure on $\mathscr{E}$.

Proof. First $\mu(\emptyset) = 0$, suppose

$$\sum_{i=1}^{\infty} (a_i, b_i] = (a,b].$$

Since every partial sum has measure at most $F(b_{n+1}) - F(a_1) < F(b) - F(a)$,
$\implies \sum_{i=1}^{n} \mu((a_i, b_i]) \le \mu((a,b])$.

For the reversed inequality, first we prove that for intervals

$$\bigcup_{i=1}^{n} (c_i, d_i] \supseteq (a,b] \implies \sum_{i=1}^{n} \mu((c_i, d_i]) \ge \mu((a,b]).$$

This can be easily proved by induction, WLOG $b_{n+1} = \max_i b_i$.

Our idea is to extend each $(a_i, b_i]$ a little bit to apply above inequality.

For all $\varepsilon > 0$, take $\delta_i > 0$ s.t.

$$\tilde{b}_i := b_i + \delta_i, \quad F(\tilde{b}_i) - F(b_i) \le \frac{\varepsilon}{2}.$$

Hence for all $\delta > 0$, $\bigcup_{i=1}^{\infty} (a_i, \tilde{b}_i) \supseteq [a+\delta, b]$, by compactness exists a finite open cover.

$$F(b) - F(a+\delta) \le \sum_{i=1}^{n} \left(F(\tilde{b}_i) - F(a_i) \right) \le \varepsilon + \sum_{i=1}^{\infty} (F(b) - F(a)).$$

Let $\varepsilon, \delta \to 0$ to conclude. $\square$

Definition 3.1.7 (Measure space)**.** A triple $(X, \mathscr{F}, \mu)$ is called a **measure space**, if $(X, \mathscr{F})$ is a measurable space and μ is a measure on $\mathscr{F}$.

If $N \in \mathscr{F}$ s.t. $\mu(N) = 0$, we say N is a **null set**.

A probability space is a measure space $(X, \mathscr{F}, P)$ with $P(X) = 1$.

Example 3.1.8 (Discrete measure)
If X is countable, $p : X \to [0,\infty]$, $\mu(A) := \sum_{x\in A} p(x)$ is a measure.

There are other important properties which we think a sensible measure would have:

- Monotonicity: If $A, B \in \mathscr{E}$, $A \subset B$, then $\mu(A) \le \mu(B)$.
- Countable subadditivity: $A_1, A_2, \cdots \in \mathscr{E}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu(A_i).$$

- Lower continuity: $A_1, A_2, \cdots \in \mathscr{E}$ and $A_n \uparrow A \in \mathscr{E}$.

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

- Similarly there's upper continuity (which requires $\mu(A_1) < \infty$).

Theorem 3.1.9
The measure on a semi-ring has all the above properties.

Proof. In fact,

- Finite additivity $\implies$ monotonicity, subtractivity;
- Countable additivity $\implies$ subadditivity, upper and lower continuity.

Here we only prove the subadditivity, since others are trivial.
Let $A_1, A_2, \cdots \in \mathscr{Q}$, and $\bigcup_{i=1}^{\infty} A_i \in \mathscr{Q}$.

$$B_n := A_n \backslash \bigcup_{i=1}^{n-1} A_i \in r(\mathscr{Q}) \implies B_n = \sum_{k=1}^{k_n} C_{n,k}, \quad C_{n,k} \in \mathscr{Q}.$$

$$A_n \backslash B_n \in r(\mathscr{Q}) \implies A_n \backslash B_n = \sum_{l=1}^{l_n} D_{n,l}, \quad D_{n,l} \in \mathscr{Q}.$$

Thus by countable additivity,

$$\begin{aligned}\mu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{n=1}^{\infty} \mu(B_n) = \sum_{n=1}^{\infty}\left(\sum_{k=1}^{k_n} \mu(C_{n,k})\right) \\ &\le \sum_{n=1}^{\infty}\left(\sum_{k=1}^{k_n} \mu(C_{n,k}) + \sum_{l=1}^{l_n} \mu(D_{n,l})\right) = \sum_{n=1}^{\infty} \mu(A_n).\end{aligned}$$

Using similar technique we can deduce the upper and lower continuity. $\square$

Theorem 3.1.10

Let μ be a set function on a ring with finite additivity, then $1 \iff 2 \iff 3 \implies 4 \implies 5$.

- μ is countablely additive;
- μ is countablely subadditive;
- μ is lower continuous;
- μ is upper continuous;
- μ is continuous at $\emptyset$.

§3.2 Outer measure

Once we construct a measure on a semi-ring, we want to extend it to a σ-algebra. Since we can't directly do this, we shall relax some of our restrictions, say reduce countable additivity to subadditivity.

Definition 3.2.1 (Outer measure)**.** Let $\tau : \mathscr{T} \to [0, \infty]$ satisfying:

- $\tau(\emptyset) = 0$;
- If $A \subset B \subset X$, then $\tau(A) \le \tau(B)$;
- (Countable subadditivity) $\forall A_1, A_2, \cdots \in \mathscr{T}$, we have

$$\tau\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \tau(A_n).$$

We call τ an **outer measure** on X.

It's easier to extend a measure on semi-ring to an outer measure:

Theorem 3.2.2

Let μ be a non-negative set function on a collection $\mathscr{E}$, where $\emptyset \in \mathscr{E}$ and $\mu(\emptyset) = 0$. Let

$$\tau(A) := \inf\left\{\sum_{n=1}^{\infty} \mu(B_n) : B_n \in \mathscr{E}, \bigcup_{n=1}^{\infty} B_n \supseteq A\right\}, \quad \forall A \in \mathscr{T}.$$

By convention, $\inf \emptyset = \infty$. (*μ need not be a measure!*)

Then τ is called the outer measure generated by μ.

Proof. Clearly $\tau(\emptyset) = 0$, and $\tau(A) \le \tau(B)$ for $A \subset B$.

$$\bigcup_{n=1}^{\infty} B_n \supseteq B \implies \bigcup_{n=1}^{\infty} B_n \supseteq A.$$

For all $A_1, A_2, \cdots \in \mathscr{T}$, WLOG $\tau(A_n) < \infty$. Take $B_{n,k}$ s.t. $\bigcup_{k=1}^{\infty} B_{n,k} \supseteq A_n$, such that

$$\sum_{k=1}^{\infty} \mu(B_{n,k}) < \tau(A_n) + \frac{\varepsilon}{2^n}.$$

Therefore

$$\bigcup_{n=1}^{\infty}\bigcup_{k=1}^{\infty} B_{n,k} \supseteq A_n,$$

$$\tau\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\mu(B_{n,k}) + \varepsilon \le \sum_{n=1}^{\infty}\tau(A_n) + \varepsilon.$$

□

Example 3.2.3

Let $\mathscr{E} = \{X, \emptyset\}$, $\mu(X) = 1$, $\mu(\emptyset) = 0$. Then $\tau(A) = 1$, $\forall A \ne \emptyset$.

Example 3.2.4

Let $X = \{a, b, c\}$, $\mathscr{E} = \{\emptyset, \{a\}, \{a, b\}, \{c\}\}$. $\mu(A) = \#A$ for $A \in \mathscr{E}$.

Here something strange happens: $\tau(\{b\}) = 2$ instead of 1, and $\tau(\{b, c\}) = 3$ instead of 2.

In the above example, we found the set $\{b\}$ somehow behaves badly: if we divide $\{a, b\}$ to $\{a\} + \{b\}$, the outer measure is not the sum of two smaller measure.

Hence we want to get rid of this kind of inconsistency to get a proper measure:

Definition 3.2.5 (Measurable sets)**.** Let τ be an outer measure, if a set A satisfies *Caratheodory condition*:

$$\tau(D) = \tau(D \cap A) + \tau(D \cap A^c), \quad \forall D \in \mathscr{T},$$

we say A is **measurable**.

Remark 3.2.6 — Inorder to prove A measurable, we only need to check

$$\tau(D) \ge \tau(D \cap A) + \tau(D \cap A^c), \quad \forall D \in \mathscr{T}.$$

Let $\mathscr{F}_\tau$ be the collection of all the τ measurable sets,

Definition 3.2.7 (Complete measure space)**.** Let $(X, \mathscr{F}, \mu)$ be a measure space, if for all null set A, and $\forall B \subset A, B \in \mathscr{F} \implies \mu(B) = 0$, we say $(X, \mathscr{F}, \mu)$ is **complete**.

Theorem 3.2.8 (Caratheodory's theorem)

Let τ be an outer measure, then $\mathscr{F} := \mathscr{F}_\tau$ is a σ-algebra, and $(X, \mathscr{F}, \tau)$ is a complete measure space.

Proof. First we prove $\mathscr{F}$ is an algebra:

Note $\emptyset \in \mathscr{F}$, and $\mathscr{F}$ is closed under completements.

For measurable sets A_1, A_2,

$$\begin{aligned}\tau(D) &= \tau(D \cap A_1) + \tau(D \cap A_1^c)\\ &= \tau(D \cap A_1 \cap A_2) + \tau(D \cap (A_1) \cap A_2^c) + \tau(D \cap A_1^c)\\ &= \tau(D \cap (A_1 \cap A_2)) + \tau(D \cap (A_1 \cap A_2)^c).\end{aligned}$$

So $A_1 \cap A_2$ is measurable.

Secondly, we prove $\mathscr{F}$ is a σ-algebra.

Let $A_1, A_2, \cdots \in \mathscr{F}$,

$$B_n := A_n \backslash \bigcup_{i=1}^{n-1} A_i \in \mathscr{F},$$

Then B_i pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$. Let $B_f = \bigcup_{i=1}^{\infty} B_i$.

It's sufficient to prove

$$\tau(D) \geq \tau(D \cap B_f) + \tau(D \cap B_f^c).$$

Let $D_n = \sum_{i=1}^{n} B_i \cap D$, $D_f = D \cap B_f$, $D_\infty = D \backslash D_f$.

Since B_i are measurable,

$$\tau(D) = \tau(D_n) + \tau(D \backslash D_n) \geq \tau(D_n) + \tau(D_\infty) = \sum_{i=1}^{n} \tau(D \cap B_i) + \tau(D_\infty).$$

Now we take $n \to \infty$,

$$\tau(D) \geq \sum_{i=1}^{\infty} \tau(D \cap B_i) + \tau(D_\infty) \geq \tau\left(D \cap \sum_{i=1}^{\infty} B_i\right) + \tau(D_\infty).$$

Where the last step follows from countable subadditivity.

This implies B_f measurable $\implies \mathscr{F}$ is a σ-algebra.

Next we prove $\tau|\mathscr{F}$ is a measure: Just let $D = \sum_{i=1}^{\infty} B_i$ in the previous equation.

Last we prove $(X, \mathscr{F}, \tau)$ is complete:

If $\tau(A) = 0$, $\tau(D) \geq \tau(D \cap A^c) = \tau(D \cap A) + \tau(D \cap A^c)$. Thus $A \in \mathscr{F}$. □

§3.3 Measure extension

Definition 3.3.1 (Measure extension). Let μ, ν be measures on $\mathscr{E}$ and $\overline{\mathscr{E}}$, and $\mathscr{E} \subset \overline{\mathscr{E}}$. If

$$\nu(A) = \mu(A), \quad \forall A \in \mathscr{E},$$

we say ν is a extension of μ on $\overline{\mathscr{E}}$.

If we start from a measure μ on $\mathscr{E}$, ideally, μ can generate an outer measure τ, and we can take $\mathscr{F}_\tau$ to construct a mesaure space.

However, things could go wrong:

Example 3.3.2

Let $X = \{a, b, c\}, \mathscr{E} = \{\emptyset, \{a, b\}, \{b, c\}, X\}$ with

$$\mu(\emptyset) = 0, \mu(\{a, b\}) = 1, \mu(\{b, c\}) = 1, \mu(X) = 2.$$

Then μ is a measure on $\mathscr{E}$, and the outer measure

$$\tau(\emptyset) = 0.$$

Observe that $\mathscr{F}_\tau = \{\emptyset, X\}$, so in this case $\tau|_{\mathscr{F}}$ is the trivial measure.

Example 3.3.3

Let $X = \mathbb{R}$, $\mathscr{E} = \{(a, b] : a < b, a, b \in \mathbb{R}\}$. Let $\mu(\emptyset) = 0$, and $\mu(A) = \infty$ for $A \neq \emptyset$.

Then μ can be extend to the Borel σ-algebra on $\mathbb{R}$ with $\mu_\alpha = \sum_{q \in \mathbb{Q}} \alpha \delta_q$, $\forall \alpha \geq 0$. So the extension is not unique.

Therefore in order to get a "proper" extension, we must put some requirements on both the starting collection and the set function μ.

Proposition 3.3.4

Let $\mathscr{P}$ be a π system. If two measures μ, ν on $\sigma(\mathscr{P})$ satisfying

$$\mu|_{\mathscr{P}} = \nu|_{\mathscr{P}}, \quad \mu|_{\mathscr{P}} \text{ is } \sigma\text{-finite},$$

Then $\mu = \nu$.

Proof. Let $A_1, A_2, \cdots \in \mathscr{P}$ s.t. $X = \sum_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty$.

Fix n, let $B = A_n$, we want to prove that

$$\mu(B \cap A) = \nu(B \cap A), \quad \forall A \in \sigma(\mathscr{P}).$$

Let $B \in \mathscr{P}$ with $\mu(B) < \infty$,

$$\mathscr{L} := \{A \in \sigma(\mathscr{P}) : \mu(A \cap B) = \nu(A \cap B)\}.$$

We'll prove $\mathscr{L}$ is a λ system, so that $\mathscr{L} \supseteq \sigma(\mathscr{P})$.

Suppose $A_1, A_2 \in \mathscr{L}$ and $A_1 \supseteq A_2$, by $\mu(B) < \infty$,

$$\mu((A_1 - A_2)B) = \mu(A_1 B) - \mu(A_2 B) = \nu(A_1 B - A_2 B) = \nu((A_1 - A_2)B).$$

So $A_1 - A_2 \in \mathscr{L}$.

Let $A_1, A_2, \cdots \in \mathscr{L}$ and $A_n \uparrow A$, then

$$\mu(AB) = \lim_{n \to \infty} \mu(A_n B) = \lim_{n \to \infty} \nu(A_n B) = \nu(AB).$$

Which implies $A \in \mathscr{L}$.

Hence $\sigma(\mathscr{P}) \subset \mathscr{L}$, i.e.

$$\mu(A \cap A_n) = \nu(A \cap A_n), \quad \forall A \in \sigma(\mathscr{P}).$$

Therefore

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap A_n) = \sum_{n=1}^{\infty} \nu(A \cap A_n) = \nu(A), \quad \forall A \in \sigma(\mathscr{P}).$$

□

Example 3.3.5

In probability, let $\mathscr{E}_1, \mathscr{E}_2$ be collections of sets. We say they're independent if

$$P(AB) = P(A)P(B), \quad \forall A \in \mathscr{E}_1, B \in \mathscr{E}_2.$$

By the previous theorem we can derive $\lambda(\mathscr{E}_1), \lambda(\mathscr{E}_2)$ are independent.

If $A_1, A_2, \ldots$ satisfy

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}),$$

we say they are independent.

Let $\{1, 2, \ldots\} = I + J$, then the σ-algebra generated by

$$\mathscr{E}_1 = \{A_\alpha \mid \alpha \in I\}, \quad \mathscr{E}_2 = \{A_\alpha \mid \alpha \in J\}$$

are independent.

Theorem 3.3.6 (Measure extension theorem)

Let μ be a measure on a semi-ring $\mathscr{Q}$, τ is the outer meaasure generated by μ. We have

$$\sigma(\mathscr{Q}) \in \mathscr{F}_\tau, \quad \tau|_{\mathscr{Q}} = \mu.$$

Remark 3.3.7 — Any measure on a semi-ring $\mathscr{Q}$ can extend to the $\sigma(\mathscr{Q})$, and if μ is σ-finite, the extension is unique.

Proof. For any $A \in \mathscr{Q}$, let $B_1 = A$, $B_n = \emptyset, n \geq 2$. Then $\tau(A) \leq \sum \mu(B_n) = \mu(A)$.

On the other hand, if $A_1, A_2, \dots \in \mathscr{Q}$ s.t. $\bigcup_{n=1}^\infty A_n \supseteq A$, then

$$\mu(A) = \mu\left(\bigcup_{n=1}^\infty \mu(AA_n)\right) \leq \sum_{n=1}^\infty \mu(AA_n) \leq \sum_{n=1}^\infty \mu(A_n).$$

Thus $\tau(A) = \mu(A)$, where we used the fact that μ is countable subadditive.

Next we prove $A \in \mathscr{F}_\tau$. We need to show that

$$\tau(D) \geq \tau(D \cap A) + \tau(D \cap A^c).$$

WLOG $\tau(D) < \infty$. Take $B_1, B_2, \dots \in \mathscr{Q}$ s.t.

$$\bigcup_{n=1}^\infty B_n \supseteq D, \quad \sum_{n=1}^\infty \mu(B_n) < \tau(D) + \varepsilon.$$

Denote $\hat{D} := B_n \in \mathscr{Q}$ for a fixed n. Suppose $\hat{D} \cap A^c = \hat{D} \backslash A = \sum_{i=1}^n C_i$.

$$\mu(\hat{D}) = \mu(\hat{D} \cap A) + \sum_{i=1}^n \mu(C_i) \geq \tau(\hat{D} \cap A) + \tau(\hat{D} \cap A^c).$$

Apply this inequality to each B_n,

$$\tau(D) + \varepsilon > \sum_{n=1}^\infty (\tau(B_n \cap A) + \tau(B_n \cap A^c)) \geq \tau(D \cap A) + \tau(D \cap A^c).$$

this implies $\tau(D) \geq \tau(D \cap A) + \tau(D \cap A^c) \implies A \in \mathscr{F}_\tau$.

At last by Caratheodory's theorem, τ is a measure on $\mathscr{F}_\tau \supseteq \sigma(\mathscr{Q})$. $\square$

Theorem 3.3.8 (Equi-measure hull)
Let τ be the outer measure generated by μ,

- $\forall A \in \mathscr{F}_\tau$, $\exists B \in \sigma(\mathscr{Q})$ s.t. $B \supseteq A$ and $\tau(A) = \tau(B)$;
- If μ is σ-finite, then $\tau(B \backslash A) = 0$.

Remark 3.3.9 — This theroem states that $\mathscr{F}_\tau$ is just $\sigma(\mathscr{Q})$ appended with null sets.

Proof. If $\tau(A) = \infty$, $B = X$ suffices.

By definition, there exists $B_n = \bigcup_{k=1}^{k_n} B_{n,k} \supseteq A$ s.t. $\tau(B_n) < \tau(A) + \frac{1}{n}$. Let $B = \bigcap_{n=1}^{\infty} B_n$, we must have $\tau(B) = \tau(A)$.

Now for the second part, let $X = \sum_{n=1}^{\infty} A_n$, $A_n \in \mathscr{Q}$, $\mu(A_n) < \infty$.

Since $A = \sum_{n=1}^{\infty} AA_n$, we have

$$AA_n \in \mathscr{F}_\tau, \quad \tau(AA_n) \leq \tau(A_n) = \mu(A_n) < \infty.$$

Let $B_n \in \sigma(\mathscr{Q})$ s.t. $B_n \supseteq AA_n$ and $\tau(B_n) = \tau(AA_n) < \infty$. Let $B := \bigcup_{n=1}^{\infty} B_n$ we have

$$\tau(B - A) = \tau\left(\bigcup_{n=1}^{\infty}(B_n - AA_n)\right) \leq \sum_{n=1}^{\infty} \tau(B_n - AA_n) = 0.$$

$\square$

Let $\mathscr{R}, \mathscr{A}, \mathscr{F}$ be the ring, algebra, σ-algebra generated by $\mathscr{Q}$, respectively. The outer measure τ restricts to a measure on each of these collections, denoted by μ_1, μ_2, μ_3. Each μ_i can generate an outer measure τ_i, but actually they're all the same as our original τ, since τ_i are "build up" from τ, intuitively τ_i cannot be any better than τ. (The proof says exactly the same thing, so I'll omit it)

Proposition 3.3.10
Let μ be a measure on an algebra $\mathscr{A}$. τ is the outer measure generated by μ, for all $A \in \sigma(\mathscr{A})$, if $\tau(A) < \infty$, then $\forall \varepsilon > 0$, $\exists B \in \mathscr{A}$ s.t. $\tau(A \Delta B) < \varepsilon$.

Remark 3.3.11 — In practice we often replace τ with a σ-finite measure μ on $\sigma(\mathscr{A})$. (Here σ-finite is on $\mathscr{A}$)

Proof. Choose $B_1, B_2, \dots \in \mathscr{A}$ s.t.

$$\hat{B} := \bigcup_{n=1}^{\infty} B_n \supseteq A, \quad \sum_{n=1}^{\infty} \mu(B_n) < \tau(A) + \frac{\varepsilon}{2}.$$

Let N be a sufficiently large number, $B := \bigcup_{n=1}^{N} B_n \in \mathscr{A}$,

$$\tau(A \backslash B) \leq \tau\left(\bigcup_{n=N+1}^{\infty} B_n\right) \leq \sum_{n=N+1}^{\infty} \tau(B_n) \leq \frac{\varepsilon}{2}.$$

As $\tau(B \backslash A) \leq \tau(\hat{B} \backslash A) < \frac{\varepsilon}{2}$, $\tau(A \Delta B) < \varepsilon$. □

Example 3.3.12

Consider the Bernoulli test, recall $C_{i_1,\ldots,i_n}$ we defined earlier. A measure(probability) μ is defined on the semi-ring $\{C_{i_1,\ldots,i_n}\} \cup \{\emptyset, X\}$, then it can extend uniquely to the σ-algebra generated by it. This is how the probability of Bernoulli test comes from.

Let $(X, \mathscr{F}, P)$ be a probability space, $A_1, A_2, \cdots \in \mathscr{F}$. We define the **tail σ-algebra** $\mathscr{T}$:

$$\mathscr{G}_n := \sigma(\{A_{n+1}, A_{n+2}, \ldots\}), \quad \mathscr{T} := \bigcap_{n=1}^{\infty} \mathscr{G}_n.$$

Let $f_1, f_2, \ldots$ be random variable, the tail σ-algebra generated by them is defined similarly:

$$\mathscr{G}_n := \sigma(\{f_{n+1}, f_{n+2}, \ldots\}), \quad \mathscr{T} := \bigcap_{n=1}^{\infty} \mathscr{G}_n.$$

Theorem 3.3.13 (Kolmogorov's 0-1 law)

If $A_1, A_2, \cdots \in \mathscr{F}$ are independent, then $P(A) \in \{0, 1\}$, $\forall A \in \mathscr{T}$.

Proof. Let $\mathscr{F}_n := \sigma(\{A_1, \ldots, A_n\})$ and $\mathscr{G}_n$. They are clearly independent.

Note that $\mathscr{A} := \bigcup_{n=1}^{\infty} \mathscr{F}_n$ is an algebra.

Let $\mathscr{H} := \sigma(\mathscr{A}) \supseteq \mathscr{G}_n \supseteq \mathscr{T}$.

Hence $\forall A \in \mathscr{T} \subset \mathscr{H}$, $\forall \varepsilon > 0$, exists $B \in \mathscr{A}$ s.t. $P(A \Delta B) < \varepsilon$, so

$$P(A) - P(AB) \leq \varepsilon, \quad |P(A) - P(B)| \leq \varepsilon.$$

Since $B \in \mathscr{F}_n$ for some n, thus it is independent to A.

$$|P(A) - P(A)^2| \leq |P(A) - P(AB)| + |P(AB) - P(A)^2| \leq 2\varepsilon.$$

Let $\varepsilon \to 0$, we'll get $P(A) \in \{0, 1\}$. □

Remark 3.3.14 — When A_i's are replace by random variables, this theorem also holds.

Example 3.3.15

finite Markov chain

§3.4 The completion of measure spaces

Let $(X, \mathscr{F}, \mu)$ be a measure space, and

$$\widetilde{\mathscr{F}} := \{A \cup N : A \in \mathscr{F}, \exists B \in \mathscr{F} \ s.t. \ \mu(B) = 0, N \subset B\}.$$

Another way to define it is: $\widetilde{\mathscr{F}} := \{A \backslash N\}$, since

$$A \cup N = A + NA^c = (A \cup B) \backslash (BA^c \backslash N);$$

$$A \backslash N = A - NA = (A \backslash B) + (BA \backslash N).$$

In fact, we can do even more: $\widetilde{\mathscr{F}} := \{A \Delta N\}$.

Next we define the measure

$$\widetilde{\mu}(A \cup N) := \mu(A), \quad \forall A \cup N \in \widetilde{\mathscr{F}}$$

We need to check several things:

- $\widetilde{\mathscr{F}}$ is a σ-algebra.
- $\widetilde{\mu}$ is well-defined.
- $(X, \widetilde{\mathscr{F}}, \widetilde{\mu})$ is a complete measure space.

Remark 3.4.1 — The mesure $\widetilde{\mu}$ is the *minimal complete extension* of μ, i.e. if $(X, \mathscr{G}, \nu)$ is another complete extension, then

$$\nu(B) = \mu(B) = 0 \implies \forall N \subset B, N \in \mathscr{G}, \nu(N) = 0.$$

$$\mu(A) = \nu(A) \leq \nu(A \cup N) \leq \nu(A) + \nu(N) = \nu(A).$$

Thus $\mathscr{G} \supseteq \widetilde{\mathscr{F}}$ and $\nu(A) = \widetilde{\mu}(A)$ for $A \in \widetilde{\mathscr{F}}$.

Therefore we call $(X, \widetilde{\mathscr{F}}, \widetilde{\mu})$ the **completion** of $(X, \mathscr{F}, \mu)$.

Obviously $\emptyset \in \widetilde{\mathscr{F}}$; For $A \cup N \in \widetilde{\mathscr{F}}$, $(A \cup N)^c = A^c - A^c N \in \widetilde{\mathscr{F}}$.

$$\bigcup_{n=1}^{\infty}(A_n \cup N_n) = \bigcup_{n=1}^{\infty} A_n \cup \bigcup_{n=1}^{\infty} N_n, \quad N = \bigcup_{n=1}^{\infty} N_n \subset \bigcup_{n=1}^{\infty} B_n.$$

Thus $\widetilde{\mathscr{F}}$ is a σ-algebra.

For $\widetilde{\mu}$, if $A_1 \cup N_1 = A_2 \cup N_2$,

$$\mu(A_1) = \mu(A_1 \cup B_2) \geq \mu(A_2).$$

Last we prove the countable additivity of $\widetilde{\mu}$. It's easy to check, so left out.

For the completeness, if $C \subset A \cup N, \mu(A) = 0$, then $C \subset A \cup B$ which is null.

Combining with the previous results we have

Theorem 3.4.2

Let τ be the outer measure generated by μ, a σ-finite measure on a semi-ring $\mathscr{Q}$. We have $(X \mathscr{F}_\tau, \tau)$ is the completion of $(X, \sigma(\mathscr{Q}), \tau)$.

Proof. Let $\mathscr{F} = \sigma(\mathscr{Q})$, we'll prove that $\widetilde{\mathscr{F}} = \mathscr{F}_\tau$.

Since $(X, \mathscr{F}_\tau, \tau)$ is complete, we have $\mathscr{F}_\tau \supseteq \widetilde{\mathscr{F}}$.

For all $C \in \mathscr{F}_\tau$, it suffices to prove $C = A + N$ for some $A \in \mathscr{F}, N \subset B$ with B null.

Since $C^c \in \mathscr{F}_\tau$, $\exists B \in \mathscr{F}$ s.t.

$$B \supseteq C^c, \quad \tau(B \backslash C^c) = 0.$$

□

§3.5 Distributions

Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing, right continuous function (called a **quasi-distribution function**). Let ν be the measure on $\mathscr{Q}_\mathbb{R}$,

$$\nu : (a, b] \mapsto \max\{F(b) - F(a), 0\}.$$

Let τ be the outer measure generated by ν. We call the sets in $\mathscr{F}_\tau$ to be the Lebesgue-Stieljes measurable sets (L-S measurable), a measurable function

$$f : (\mathbb{R}, \mathscr{F}_\tau) \to (\mathbb{R}, \mathscr{B}_\mathbb{R})$$

is L-S measurable, and $\tau\big|_{\mathscr{F}_\tau}$ is the L-S measure.

In fact finite L-S measures and the quasi-distribution functions are 1-1 correspondent (ignoring the difference of a constant), since $\mathscr{B}_\mathbb{R} = \sigma(\mathscr{Q}_\mathbb{R})$, $(\mathbb{R}, \mathscr{F}_\tau, \tau)$ is the completion of $(\mathbb{R}, \mathscr{B}_\mathbb{R}, \tau)$, and $\mu_F = \tau\big|_{\mathscr{B}_\mathbb{R}}$ is the unique extension of ν.

Conversely, given a measure μ on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$, if $\mu((a, b]) < \infty$ for all $a < b$, then $\mu = \mu_F$, where

$$F = F_\mu : x \mapsto \mu((-\infty, x]), \quad x \in \mathbb{R}.$$

We say a probability measure on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ is a **distribution**. Let $F : \mathbb{R} \to \mathbb{R}$ be a quasi-distribution function, if F satisfies:

$$F(-\infty) := \lim_{x \to -\infty} F(x) = 0, \quad F(+\infty) := \lim_{x \to +\infty} F(x) = 1,$$

then we say F is a distribution function (d.f.).

From the previous example we know distribution and d.f. are one-to-one correspondent.

Theorem 3.5.1

Let $g : (X, \mathscr{F}) \to (Y, \mathscr{S})$, μ is a measure on $\mathscr{F}$. Let

$$\nu(B) := \mu(g^{-1}(B)) = \mu \circ g^{-1}(B), \quad \forall B \in \mathscr{S}.$$

Then ν is a measure on $\mathscr{S}$.

Proof. Trivial. Just check the definition one by one. □

Let $(\Omega, \mathscr{F}, P)$ be a probability space, $f : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}_\mathscr{R})$. We say

$$P \circ f^{-1} : B \mapsto P(f \in B)$$

is the **distribution** of f, denoted by μ_f, i.e. $\mu_f(B) = P(f \in B)$ for Borel sets B.

If $\mu_f = \mu$, we say f obeys the distribution μ, denoted by $f \sim \mu$.

Let $F_f = F_{\mu_f}$ be the distribution function of f.

$$F_f := \mu_f((-\infty, x]) = P(f \leq x), \quad x \in \mathbb{R}.$$

We can also say f obeys F_f, denoted by $f \sim F_f$.

If $F_f = F_g$, then we say f and g is **equal in distribution**, denoted by $f \overset{d}{=} g$.

Theorem 3.5.2

Any d.f. is the distribution function of some random variable.

Proof. Let $\Omega = \mathbb{R}, \mathscr{F} = \mathscr{B}_{\mathbb{R}}, P = \mu_F$, and $f = \mathrm{id}$. It's clear that the distribution function of f is precisely F. $\square$

§3.6 The convergence of measurable functions

Let $(X, \mathscr{F}, \mu)$ be a measure space.

For any statement, if there exists null set N s.t. it holds for all $x \in N^c$, then we say this statement holds *almost everywhere*. (Often abbreviated as *a.e.*)

Definition 3.6.1. If a sequence of functions f_n satisfies

$$\mu\left(\lim_{n\to\infty} f_n \neq f\right) = 0,$$

(here f is finite a.e.) we say $\{f_n\}$ converges to f **almost everywhere**, denoted by $f_n \to f, a.e.$.

Definition 3.6.2. If $\forall \delta > 0$, $\exists A \in \mathscr{F}$ s.t. $\mu(A) < \delta$ and

$$\lim_{n\to\infty} \sup_{x \notin A} |f_n(x) - f(x)| = 0,$$

then we say $\{f_n\}$ converges to f **almost uniformly**, denoted by $f_n \to f, a.u.$.

If $f_n \to f, a.u.$, $\forall \varepsilon > 0$, $\exists m = m_k(\varepsilon)$ s.t. when $n \geq m$, $|f_n(x) - f(x)| < \varepsilon, \forall x \in C_k$, but we could have $\sup_k m_k(\varepsilon) = \infty$, thus $f_n \rightrightarrows f$ doesn't hold. e.g. $f_n(x) = x^n, f(x) = 0, x \in [0, 1), f(1) = 1$.

Proposition 3.6.3

$f_n \to f, a.u. \implies f_n \to f, a.e.$.

Proof. For all n, $\exists A_n$ s.t. $\mu(A_n) < \frac{1}{n}$, and $f_n \to f$ in A_n^c. Let $A := \bigcap_n A_n$.

Then $\{f_n \not\to f\} \cup \{|f| = \infty\} \subset A$, $\mu(A) = 0$, hence $f_n \to f, a.e.$. $\square$

Proposition 3.6.4

$f_n \to f, a.e.$ iff $\forall \varepsilon > 0$,

$$\mu\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|f_m - f| \geq \varepsilon\}\right) = 0.$$

Note: If $f(x) - g(x)$ is not defined, we regard it as $+\infty$.

Proof. Let $A_\varepsilon := \bigcap \bigcup \{|f_m - f| > \varepsilon\}$.

$$\left\{\lim_{n\to\infty} f_n \neq f\right\} \cup \{|f| = \infty\} = \bigcup_{k=1}^{\infty} A_{\frac{1}{k}} =\uparrow \lim_{k\to\infty} A_{\frac{1}{k}}.$$

$\square$

Proposition 3.6.5

$f_n \to f, a.u.$ iff $\forall \varepsilon > 0$, we have

$$\downarrow \lim_{m\to\infty} \mu\left(\bigcup_{n=m}^{\infty} \{|f_n - f| \geq \varepsilon\}\right) = 0.$$

Proof. If $f_n \to f, a.u.$, $\forall \delta, \exists A \in \mathscr{F}$ s.t. $\mu(A) < \delta$ and $f_n \rightrightarrows f, x \in A^c$.

This means for any fixed ε, $\exists m$ s.t. when $n \geq m$, $x \notin A \implies |f_n(x) - f(x)| < \varepsilon$. Thus $A \supseteq \bigcup_{n=m}^{\infty} |f_n - f| \geq \varepsilon$.

Conversely, $\forall \delta > 0$, $\exists m_k$ s.t.

$$\mu\left(\bigcup_{n=m_k}^{\infty} \{|f_n - f| \geq \frac{1}{k}\}\right) < \frac{\delta}{2^k}.$$

Denote the above set by A_k, and $A = \bigcup_{k=1}^{\infty} A_k$, then $\mu(A) < \delta$, and $f_n(x) \rightrightarrows f(x)$ for $x \in A^c$. □

Definition 3.6.6. If $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mu(|f_n - f| \geq \varepsilon) = 0,$$

then we say $\{f_n\}$ converges to f **in measure**, denoted by $f_n \xrightarrow{\mu} f$.

Theorem 3.6.7

$$f_n \to f, a.u. \implies f_n \to f, a.e., \quad f_n \xrightarrow{\mu} f.$$

If $\mu(X) < \infty$, then

$$f_n \to f, a.u. \iff f_n \to f, a.e. \implies f_n \xrightarrow{\mu} f.$$

Theorem 3.6.8

$f_n \to f$ in measure iff for any subsequence of $\{f_n\}$, exists its subsequence $\{f_{n'}\}$ s.t.

$$f_{n'} \to f, a.u.$$

Proof. When $f_n \to f$ in measure, let $n_0 = 0$. Take $n_k > n_{k-1}$ inductively such that

$$\mu\left(\left\{|f_n - f| \geq \frac{1}{k}\right\}\right) \leq \frac{1}{2^k}, \quad \forall n \geq n_k.$$

Then $\forall \varepsilon > 0$, $\exists \frac{1}{m} < \varepsilon$, $\{|f_{n_k} - f| \geq \varepsilon\} \subset \{|f_{n_k} - f| \geq \frac{1}{k}\}$,

$$\mu\left(\bigcup_{k=m}^{\infty} \{|f_{n_k} - f| \geq \varepsilon\}\right) \leq \mu\left(\bigcup_{k=m}^{\infty} \left\{|f_{n_k} - f| \geq \frac{1}{k}\right\}\right) \leq \frac{1}{2^{m-1}} \to 0.$$

Conversely, we assume for contradiction that $\exists \varepsilon > 0$ s.t. $\mu(\{|f_n - f| \geq \varepsilon\}) \not\to 0$.

So $\exists \delta > 0$ and subsequence $\{n_k\}$ s.t. $\mu(\{|f_{n_k} - f| \geq \varepsilon\}) > \delta$.

Hence there doesn't exist a subsequence $\{f_{n'}\}$ of $\{f_{n_k}\}$ s.t. $f_{n'} \to f, a.u.$. □

Example 3.6.9

Consider measure space $(\mathbb{R}, \mathscr{B}_{\mathbb{R}}, \lambda)$, the Lebesgue measure, $f_n = \mathbf{I}_{|x|>n}$, then

$$f_n \to 0, \forall x \implies f_n \to 0, a.e..$$

let $\varepsilon = 1$, it's clear that f_n doesn't converge to f in measure, hence not almost uniformly.

Example 3.6.10

Let $f_{k,i} = \mathbf{I}_{\frac{i-1}{k} < x \le \frac{i}{k}}$, $i = 1, \dots, k$. It's clear that $f_{k,i} \to 0$ in measure, but not almost everywhere, and hence not almost uniformly.

§3.7 Probability space

Let $(\Omega, \mathscr{F}, P)$ be a probability space. Here almost everywhere is renamed to almost surely.

Let F be a real function, let $C(F)$ be the continuous points of F.

Let $F, F_1, F_2, \dots$ be non-decreasing functions, if

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in C(F),$$

then we say $\{F_n\}$ converge to F weakly, $F_n \xrightarrow{w} F$.

Let $F, F_1, F_2, \dots$ be distribution functions, $f_n \sim F_n$,

Definition 3.7.1. If $F_n \xrightarrow{w} F$, then we say $\{f_n\}$ converge to F in distribution, denoted by $f_n \xrightarrow{d} F$. For $f \sim F$, we can also write $f_n \xrightarrow{d} f$.

Theorem 3.7.2

$f_n \xrightarrow{P} f \implies f_n \xrightarrow{d} f$.

Proof.

$$\begin{aligned} P(h \le y) &\le P(h \le y, |h-g| < \varepsilon) + P(h \le y, |h-g| \ge \varepsilon) \\ &\le P(g \le y + \varepsilon) + P(|h-g| \ge \varepsilon). \end{aligned}$$

Let $F_n(x) = P_n(f \le x)$ Let $h = f_n$, $g = f$, $y = x$.

$$\limsup_{n\to\infty} F_n(x) \le F(x+\varepsilon), \quad \forall \varepsilon > 0.$$

Thus $\limsup_{n\to\infty} F_n(x) \le F(x)$. TODO □

Theorem 3.7.3 (Skorokhod)

If $f_n \xrightarrow{d} f$, then exists a probability space $(\widetilde{X}, \widetilde{\mathscr{F}}, \widetilde{P})$, with random variables $\{\tilde{f}_n\}$ and $\tilde{f}$, such that

$$\tilde{f}_n \stackrel{d}{=} f_n, \tilde{f} \stackrel{d}{=} f, \quad \tilde{f}_n \to \tilde{f}, a.s.$$

Proof. If $F_n \to F$ weakly, then $F_n^{\leftarrow} \to F^{\leftarrow}$ weakly. (Prove this yourself!)
Since $\mathbb{R} \backslash C(F_n^{\leftarrow})$ is countable, TODO □

If f is defined almost everywhere, we can extend it to $\tilde{f} = f \cdot \mathbf{I}_{N^c}$. So from now on when we talk about $f = g$, we mean $f = g, a.e.$.

§3.8 Review of first two sections

Here we list some concepts so that you can recall their definition and properties.
Collections of sets:

- π-system
- Semi-ring
- Ring, algebra
- σ-algebra
- Monotone class, λ-system

Measure:

- σ-finite
- Outer measure
- Caratheodory condition, measurable sets
- Measure extension, semi-ring $\to$ σ-algebra
- Complete measure space, completion
- For $\mathscr{F} = \sigma(\mathscr{A})$, $\forall F \in \mathscr{F}$, $\varepsilon > 0, \exists A \in \mathscr{A}$ s.t. $F = A \Delta N_\varepsilon$, $\mu(N_\varepsilon) \le \varepsilon$.

Functions:

- Measurable map
- $h \in \sigma(g) \implies h = f \circ g$ for some f.
- Typical method, simple non-negative functions $\to$ measurable functions
- Almost uniformly, almost everywhere, converge in measure

§4 Integrals

§4.1 Definition of Integrals

The idea of integration of f over μ is to compute the weighted sum of the values of f.
The definition of integrals is another example of typical method.

- For an indicator function $\mathbf{I}_A$, define $\int \mathbf{I}_A \, \mathrm{d}\mu = \mu(A)$.
- For simple function $f = \sum_{i=1}^n a_i \mathbf{I}_{A_i}$, just let $\int f \, \mathrm{d}\mu = \sum_{i=1}^n a_i \mu(A_i)$.
- For non-negative measurable function f, let $\int f \, \mathrm{d}\mu = \sup_{g \le f} \int g \, \mathrm{d}\mu$, where g is non-negative simple functions.

- For generic function f, write $f = f_+ - f_-$, define $\int f = \int f_+ - \int f_-$.

Definition 4.1.1 (Measurable partitions)**.** If a collection of sets $\{A_i\}$ satisfies

$$\mu(A_i \cap A_j) = 0, \quad \mu((\bigcup A_i)^c) = 0,$$

then we say $\{A_i\}$ is a **measurable partition** of X.

Definition 4.1.2 (Integrals for simple functions)**.** Let $\{A_i\}$ be a partition of X, $a_i \geq 0$ are reals. Let

$$f = \sum_{i=1}^{n} a_i \mathbf{I}_{A_i},$$

define

$$\int_X f \, \mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(A_i).$$

Check it's well-defined: if $f = \sum_{j=1}^{m} b_j \mathbf{I}_{B_j}$, then

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_i m(A_i \cap B_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(A_i \cap B_j).$$

Proposition 4.1.3

Let f, g be non-negative simple functions.

(1) $\int_X \mathbf{I}_A \, \mathrm{d}\mu = \mu(A), \quad \forall A \in \mathscr{F}$;

(2) $\int_X f \, \mathrm{d}\mu \geq 0$;

(3) $\int_X (af) \, \mathrm{d}\mu = a \int_X f \, \mathrm{d}\mu$;

(4) $\int_X (f+g) \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu$;

(5) If $f \geq g$, then $\int_X f \, \mathrm{d}\mu \geq \int_X g \, \mathrm{d}\mu$.

(6) If $f_n \uparrow$ and $\lim_{n\to\infty} f_n \geq g$, then $\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu \geq \int_X g \, \mathrm{d}\mu$.

Remark 4.1.4 — $f :=\uparrow \lim_{n\to\infty} f_n$ need not be simple function. Even if f is simple, we don't know $\lim \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$ yet.

Proof of (4), (5). Since $\{A_i \cap B_j\}$ is a partition of X, on $A_i \cap B_j$,

$$f + g = a_i + b_j, \quad f = a_i, g = b_j.$$

□

Proof of (6). For all $\alpha \in (0,1)$, let $A_n(\alpha) := \{f_n \geq \alpha g\} \uparrow X$. Then

$$f_n \mathbf{I}_{A_n(\alpha)} \geq \alpha g \mathbf{I}_{A_n(\alpha)}.$$

Thus if $g = \sum_{j=1}^m b_j \mathbf{I}_{B_j}$,

$$\int_X f_n \,\mathrm{d}\mu \geq \int_X f_n \mathbf{I}_{A_n(\alpha)} \,\mathrm{d}\mu \geq \alpha \int_X g \mathbf{I}_{A_n(\alpha)} \,\mathrm{d}\mu.$$

$$RHS = \alpha \sum_{j=1}^m b_j \mu(B_j \cap A_n(\alpha)) \uparrow \alpha \int_X g \,\mathrm{d}\mu.$$

Hence

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{d}\mu \geq \alpha \int_X g \,\mathrm{d}\mu, \quad \forall \alpha < 1,$$

which completes the proof. □

Definition 4.1.5 (Integrals for non-negative measurable functions)**.** Let f be a non-negative measurable function. We know that $\exists f_1, f_2, \ldots$ s.t. $f_n \uparrow f$. If we define the integral of f to be the limit of $\int f_n \,\mathrm{d}\mu$, we still need to prove this is well-defined. Therefore we use another definition:

$$\int_X f \,\mathrm{d}\mu := \sup \left\{ \int_X g \,\mathrm{d}\mu : g \leq f \text{ is simple and non-negative} \right\}.$$

Proposition 4.1.6

Let f be a non-negative measurable function.

(1) If f is simple, then the two definition is the same.

(2) If $\{f_n\}$ is a series of simple non-negative functions, and $f_n \uparrow f$, then

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu.$$

(3)

$$\int_X f \,\mathrm{d}\mu = \lim_{n\to\infty} \left[\sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mu \left(\left\{ \frac{k}{2^n} \leq f < \frac{k+1}{2^n} \right\} \right) + n\mu(\{f \geq n\}) \right].$$

Proof of (2). By definition, $\int_X f_n \,\mathrm{d}\mu \leq \int_X f \,\mathrm{d}\mu$. Since for all simple function g, if $f_n \uparrow f \geq g$,

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{d}\mu \geq \int_X g \,\mathrm{d}\mu.$$

Hence the desired equality holds. □

Remark 4.1.7 — The integral of f relies only on $\mu\big|_{\sigma(f)}$: if $f \in \mathscr{G} \subset \mathscr{F}$, then the integral of f is the same on $(X, \mathscr{G}, \mu\big|_{\mathscr{G}})$ and $(X, \mathscr{F}, \mu\big|_{\mathscr{F}})$.

Proposition 4.1.8

Continuing on the properties of integrals:

(1) $\int_X f \, \mathrm{d}\mu \geq 0$;

(2) $\int_X (af+g) \, \mathrm{d}\mu = a \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu$;

(3) If $f \geq g$, then $\int_X f \, \mathrm{d}\mu \geq \int_X g \, \mathrm{d}\mu$.

Proof. Use the previous proposition. ☐

Definition 4.1.9 (Integrals for generic functions)**.** Let f be a measurable function, and $f = f^+ - f^-$. If

$$\min\left\{\int_X f^+ \, \mathrm{d}\mu, \int_X f^- \, \mathrm{d}\mu\right\} < \infty,$$

we say the integral of f exists and define it to be

$$\int_X f \, \mathrm{d}\mu := \int_X f^+ \, \mathrm{d}\mu - \int_X f^- \, \mathrm{d}\mu.$$

If $\int_X f \, \mathrm{d}\mu \neq \pm\infty$, we say f is **integrable**.

For any $A \in \mathscr{F}$, $(A, \mathscr{F}_A, \mu_A)$ is a measure space. Define the integral of f on A to be

$$\int_A f \, \mathrm{d}\mu := \int_A f\big|_A \, \mathrm{d}\mu_A = \int_X f \mathbf{I}_A \, \mathrm{d}\mu.$$

where the latter equality holds since it holds for indicator functions.

Example 4.1.10 (The Lebesgue-Stieljes integral)

Let $(\mathbb{R}, \mathscr{B}_\mathbb{R}, \mu_F)$ be a measure space, where F is a quasi-distribution function. For a Borel function g,

$$\int_\mathbb{R} g \, \mathrm{d}F = \int_\mathbb{R} g(x) \, \mathrm{d}F(x) = \int_\mathbb{R} g(x) F(\mathrm{d}x) := \int_\mathbb{R} g \, \mathrm{d}\mu_F.$$

In particular, when $F(x) = x$, the integral is Lebesgue integral. Let λ be Lebesgue measure,

$$\int_\mathbb{R} g(x) \, \mathrm{d}x := \int_\mathbb{R} g \, \mathrm{d}\lambda.$$

If μ is a distribution, $F = F_\mu$, $g = \mathrm{id}$, we say

$$\int_\mathbb{R} x \, \mathrm{d}F(x) = \int_\mathbb{R} x\mu(\mathrm{d}x) = \int_\mathbb{R} \mathrm{id} \, \mathrm{d}\mu.$$

is the **expectation** of the distribution μ.

Example 4.1.11 (The integral on discrete measure)

Let $X = \{x_1, x_2, \dots\} = \{1, 2, \dots\}$, $\mu(\{x_i\}) = a_i$.

Let $I^+ = \{i : f(x_i) \geq 0\}, I^- = \{i : f(x_i) < 0\}$.

Let $I_n^+ = I^+ \cap \{1, \dots, n\}$, $f\mathbf{I}_{I_n^+}$ is a non-negative simple function and converges to f^+. Hence

$$\int_X f^+ \, \mathrm{d}\mu = \sum_{i \in I^+} f(x_i)a_i, \quad \int_X f^- \, \mathrm{d}\mu = -\sum_{i \in I^-} f(x_i)a_i.$$

$$\int_X f \, \mathrm{d}\mu = \sum_{i \in I} \sum_{i=1}^{\infty} f(x_i)a_i.$$

So f is integrable iff the series absolutely converges.

Theorem 4.1.12

Let f be a measurable function.

(1) If $\int_X f \, \mathrm{d}\mu$ exists, then $|\int_X f \, \mathrm{d}\mu| \leq \int_X |f| \, \mathrm{d}\mu$.

(2) f integrable $\iff$ $|f|$ integrable.

(3) If f is integrable, then $|f| < \infty, a.e.$.

Proof of (3). WLOG $f \geq 0$, then $f \geq f\mathbf{I}_{\{f=\infty\}}$.

$$\int_X f \, \mathrm{d}\mu \geq \int_X f\mathbf{I}_{\{f=\infty\}} \geq n\mu(\{f = \infty\}), \quad \forall n.$$

Thus $\mu(\{f = \infty\})$ must be 0. □

Theorem 4.1.13

Let f, g be measurable functions whose integral exists.

- $\int_A f \, \mathrm{d}\mu = 0$ for all null set A;
- If $f \geq g, a.e.$ then $\int_X f \, \mathrm{d}\mu \geq \int_X g \, \mathrm{d}\mu$.
- If $f = g, a.e.$, then their integrals exist simultaneously, $\int_X f \, \mathrm{d}\mu = \int_X g \, \mathrm{d}\mu$.

Proof. By definition, just check them one by one. □

Corollary 4.1.14

If $f = 0, a.e.$, then $\int_X f \, \mathrm{d}\mu = 0$; If $f \geq 0, a.e.$ and $\int_X f \, \mathrm{d}\mu = 0$, then $f = 0, a.e.$.

§4.2 Properties of integrals

Theorem 4.2.1 (Linearity of integrals)

Let f, g be functions whose integral exists.

- $\forall a \in \mathbb{R}$, the integral of af exists, and $\int_X (af)\,\mathrm{d}\mu = a\int_X f\,\mathrm{d}\mu$;
- If $\int_X f\,\mathrm{d}\mu + \int_X g\,\mathrm{d}\mu$ exists, then $f+g$ *a.e.* exists, its integral exists and

$$\int_X (f+g)\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu + \int_X g\,\mathrm{d}\mu.$$

Proof. The first one is trivial by definition.

As for the second,

1. First we prove $f+g$ *a.e.* exists. If $|f| < \infty, a.e.$, we're done.

   If $\mu(f=\infty) > 0$, then $\int_X f\,\mathrm{d}\mu = \infty$. This means $\int_X g\,\mathrm{d}\mu \neq -\infty$, so $\mu(g=-\infty)=0$. Thus $f+g$ *a.e.* exists. Similarly we can deal with the case $\mu(f=-\infty)>0$.

2. Next we prove the equality. $f+g = (f^+ + g^+) - (f^- + g^-)$. Let $\varphi = f^+ + g^+, \psi = f^- + g^-$. Our goal is

   $$\int_X (\varphi - \psi)\,\mathrm{d}\mu = \int_X \varphi\,\mathrm{d}\mu - \int_X \psi\,\mathrm{d}\mu.$$

   Since $f+g$ *a.e.* exists, so $\varphi - \psi$ exists almost everywhere. If $\int_X \varphi\,\mathrm{d}\mu = \int_X \psi\,\mathrm{d}\mu = \infty$, then the integral of f, g must be $+\infty$ and $-\infty$, which contradicts with our condition. So both sides of above equation exist.

   Since $\max\{\varphi, \psi\} = \psi + (\varphi-\psi)^+ = \varphi + (\varphi - \psi)^-$, by the linearity of non-negative integrals,

   $$\int_X \psi\,\mathrm{d}\mu + \int_X (\varphi-\psi)^+\,\mathrm{d}\mu = \int_X \varphi\,\mathrm{d}\mu + \int_X (\varphi-\psi)^-\,\mathrm{d}\mu.$$

   which rearranges to the desired equality.

Note: we need to verify that we didn't add infinity to the equation in the last step. □

Proposition 4.2.2

Let f, g be integrable functions, If $\int_A f\,\mathrm{d}\mu \geq \int_A g\,\mathrm{d}\mu, \forall A \in \mathscr{F}$, then $f \geq g, a.e.$.

Proof. Let $B = \{f < g\}$, then $(g-f)\mathbf{I}_B \geq 0$,

$$\int_B (g-f)\,\mathrm{d}\mu = \int_B (g-f)\mathbf{I}_B\,\mathrm{d}\mu \geq 0.$$

By the linearity of integrals we get $(g-f)\mathbf{I}_B = 0, a.e.$, i.e. $\mu(B) = 0$. □

Proposition 4.2.3

If μ is σ-finite, the integral of f, g exists, the conclusion of previous proposition also holds.

Proof. Let $X = \sum_n X_n$, $\mu(X_n) < \infty$. By looking at X_n, we may assume $\mu(X) < \infty$.

Since $\{f < g\} = \{-\infty \neq f < g\} + \{f = -\infty < g\}$.

Let $B_{M,n} = \{|f| \leq M, f + \frac{1}{n} < g\}$. By condition,

$$\int_{B_{M,n}} f \,\mathrm{d}\mu \geq \int_{B_{M,n}} g \,\mathrm{d}\mu \geq \int_{B_{M,n}} f \,\mathrm{d}\mu + \frac{1}{n}\mu(B_{M,n}).$$

Since $\int_{B_{M,n}} f \,\mathrm{d}\mu \leq M\mu(X)$ is finite, we get $\mu(B_{M,n}) = 0$. This implies $\{-\infty \neq f < g\} = \bigcup B_{M,n}$ is null.

Let $C_M = \{g > -M\}$, similarly,

$$-\infty \cdot \mu(C_M) = \int_{C_M} f \,\mathrm{d}\mu \geq \int_{C_M} g \,\mathrm{d}\mu = -M\mu(C_M).$$

Hence $\mu(C_M) = 0$, $\{-\infty = f < g\} = \bigcup C_M$ is null. □

Remark 4.2.4 — When $\geq$ is replaced by $=$, the conclusion holds as well. This proposition tells us that the integrals of f totally determines f. (In calculus, taking the derivative of integrals gives original functions)

Theorem 4.2.5 (Absolute continuity of integrals)

Let f be an integrable function, $\forall \varepsilon > 0$, $\exists \delta > 0$, such that $\forall A \in \mathscr{F}$,

$$\mu(A) < \delta \implies \int_A |f| \,\mathrm{d}\mu < \varepsilon.$$

Proof. Take non-negative simple functions $g_n \uparrow |f|$. Since $\int |f| \,\mathrm{d}\mu < \infty$, $\exists N$ s.t.

$$\int_X (|f| - g_N) \,\mathrm{d}\mu = \int_X |f| \,\mathrm{d}\mu - \int_X g_N \,\mathrm{d}\mu < \frac{\varepsilon}{2}.$$

Let $M = \max_{x \in X} g_N(x)$, $\delta = \frac{\varepsilon}{2M}$, so

$$\int_A |f| \,\mathrm{d}\mu < \frac{\varepsilon}{2} + \int_A g_N \,\mathrm{d}\mu = \frac{\varepsilon}{2} + M\mu(A) < \varepsilon.$$

□

Example 4.2.6

Fundamental theorem of Calculus, Lebesgue version: Let g be a measurable function, then g is absolutely continuous iff $\exists f : [a, b] \to \mathbb{R}$ Lebesgue integrable, s.t.

$$g(x) - g(a) = \int_a^x f(z) \,\mathrm{d}z.$$

The absolute continuity can be implied by the absolute continuity of integrals.

§4.3 Convergence theorems

Levi, Fatou, Lebesgue.

In this section we mainly discuss the commutativity of integrals and limits, i.e. if $f_n \to f$, we care when does the following holds:

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu.$$

Theorem 4.3.1 (Monotone convergence theorem, Levi's theorem)

Let $f_n \uparrow f, a.e.$ be non-negative functions, then

$$\int_X f_n \, \mathrm{d}\mu \uparrow \int_X f \, \mathrm{d}\mu.$$

Proof. By removing countable null sets, we may assume $0 \le f_n(x) \uparrow f$.

Take non-negative simple functions $f_{n,k} \uparrow f_n$. Let $g_k = \max_{1\le n\le k} f_{n,k}$ be simple functions.

$$g_k = \max_{1\le n\le k} f_{n,k} \le \max_{1\le n\le k+1} f_{n,k+1} = g_{k+1}.$$

So $g_k \uparrow$, say $g_k \to g$ for some function g. Clearly $g \le f$ as $g_k \le f_k$, $\forall k$.

Note as $k \to \infty$, $g_k \ge f_{n,k} \implies g \ge f_n, \forall n$. so $g = f$.

By definition of integrals,

$$\int_X f \, \mathrm{d}\mu = \lim_{k\to\infty} \int_X g_n \, \mathrm{d}\mu,$$

and

$$\int_X g_n \, \mathrm{d}\mu \le \int_X f_n \, \mathrm{d}\mu \le \int_X f \, \mathrm{d}\mu.$$

So the conclusion follows. □

Corollary 4.3.2

Let f_n be functions whose integrals exist, if

$$f_n \uparrow f, a.e. \quad \int_X f_1^- \, \mathrm{d}\mu < \infty, \quad \text{or} \quad f_n \downarrow f, a.e. \quad \int_X f_1^+ \, \mathrm{d}\mu < \infty,$$

then the integral of f exists, and $\int_X f_n \, \mathrm{d}\mu \to \int_X f \, \mathrm{d}\mu$.

Remark 4.3.3 — Counter example when $\int_X f_1^+ \, \mathrm{d}\mu = \infty$: let $X = \mathbb{R}$,

$$f_n = \mathbf{I}_{[n,\infty)} \downarrow f = 0, \quad \int_X f_n \, \mathrm{d}\mu = \infty, \quad \int_X f \, \mathrm{d}\mu = 0.$$

Corollary 4.3.4
If the integral of f exists, then for any measure partition $\{A_n\}$,

$$\int_X f \, \mathrm{d}\mu = \sum_{n=1}^{\infty} \int_{A_n} f \, \mathrm{d}\mu.$$

If $f \geq 0$, then $\nu : A \mapsto \int_A f \, \mathrm{d}\mu$ is a measure on $\mathscr{F}$. If we don't require $f \geq 0$, ν will become a signed measure which we'll cover later.

Theorem 4.3.5 (Fauto's Lemma)
Let $\{f_n\}$ be non-negative measurable functions almost everywhere, then

$$\int_X \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu.$$

Proof. Let $g_k := \inf_{n\geq k} f_n \uparrow g := \liminf_{n\to\infty} f_n$. By monotone convergence theorem,

$$\int_X g \, \mathrm{d}\mu = \lim_{k\to\infty} \int_X g_k \, \mathrm{d}\mu \leq \lim_{k\to\infty} \inf_{n\geq k} \int_X f_n \, \mathrm{d}\mu = \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu.$$

□

Corollary 4.3.6
If there exists integrable g s.t. $f_n \geq g$, then $\int_X \liminf_{n\to\infty} f_n$ exists and

$$\int_X \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu.$$

Theorem 4.3.7 (Lebesgue)
Let $f_n \to f, a.e.$ or $f_n \xrightarrow{\mu} f$, if there exists non-negative integrable function g s.t. $|f_n| \leq g, \forall n$, then

$$\lim_{n\to\infty} \int_X f_n \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu.$$

Proof. When $f_n \to f, a.e.$, by Fatou's lemma,

$$\int_X f \, \mathrm{d}\mu = \int_X \lim_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_X f_n \, \mathrm{d}\mu.$$

Since $|f_n| \leq g$,

$$\int_X f \, \mathrm{d}\mu = \int_X \lim_{n\to\infty} f_n \, \mathrm{d}\mu \geq \limsup_{n\to\infty} \int_X f_n \, \mathrm{d}\mu,$$

which gives the desired.

When $f_n \xrightarrow{\mu} f$, for all subsequence $\{n_k\}$, exists a subsequence $\{n'\}$ s.t. $f_{n'} \to f, a.e.$.

Thus $\int_X f_{n'} \, \mathrm{d}\mu \to \int_X f \, \mathrm{d}\mu$, hence $\int_X f_n \, \mathrm{d}\mu \to \int_X f \, \mathrm{d}\mu$.(Why?) □

Corollary 4.3.8

Let f_n be random variable on $(\Omega_n, \mathscr{F}_n, P_n)$, $f_n \xrightarrow{d} f$, then we have

$$\lim_{n\to\infty} \int_{X_n} f_n \,\mathrm{d}P_n = \int_X f \,\mathrm{d}P.$$

Proposition 4.3.9 (Transformation formula of integrals)

Let $g : (X, \mathscr{F}, \mu) \to (Y, \mathscr{S})$ be a measurable map. For all measurable f on $(Y, \mathscr{S})$, then

$$\int_Y f \,\mathrm{d}\mu \circ g^{-1} = \int_X f \circ g \,\mathrm{d}\mu$$

if one of them exists.

Proof. By the typical method, we only need to prove for indicator function f. □

Remark 4.3.10 — μ and $\mu \circ g^{-1}$ are the same measure in different spaces.

§4.4 Expectations

Let ξ be a r.v. on $(\Omega, \mathscr{F}, P)$,

Definition 4.4.1 (Expectations)**.** If $\int_\Omega \xi \,\mathrm{d}P$ exists, then we call it the **expectation** of ξ, denoted by $E(\xi)$ or $E\xi$.

Consider the distribution $\mu_\xi = P \circ \xi^{-1}$, $F_\xi(x) = P(\xi \le x)$.
Let $f = \mathrm{id} : \mathbb{R} \to \mathbb{R}$, then $E(\xi) = E(\mu_\xi)$:

$$\int_{\mathbb{R}} x \,\mathrm{d}F_\xi(x) = \int_{\mathbb{R}} f \,\mathrm{d}\mu_\xi = \int_{\mathbb{R}} f \,\mathrm{d}P \circ \xi^{-1} = \int_{\mathbb{R}} f \circ \xi \,\mathrm{d}P = \int_\Omega \xi \,\mathrm{d}P = E(\xi).$$

Let f be a measurable function on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$, then $f(\xi)$ is a measurable function on $(\Omega, \mathscr{F})$, and

$$Ef(\xi) = \int_{\mathbb{R}} f \,\mathrm{d}F_\xi.$$

Let $\eta = f \circ \xi$, by the transformation formula,

$$\begin{aligned} Ef(\xi) &= \int_\Omega \eta(\omega) \,\mathrm{d}P(\omega) \\ &= \int_{\overline{\mathbb{R}}} y \,\mathrm{d}P \circ \eta^{-1}(y) = \int_{\overline{\mathbb{R}}} y \,\mathrm{d}\mu_\eta(y) = \int_{\overline{\mathbb{R}}} y \,\mathrm{d}\mu_\xi \circ f^{-1}(y) \\ &= \int_{\mathbb{R}} f(x) \,\mathrm{d}\mu_\xi(x) = \int_{\mathbb{R}} f \,\mathrm{d}F_\xi. \end{aligned}$$

Example 4.4.2

Possion distribution: $P(\xi = i) = \frac{\lambda^i}{i!}e^{-\lambda} =: p_i$. Its expectation is

$$\int_{\mathbb{R}} x \,\mathrm{d}\mu = \int_{\mathbb{N}} x \,\mathrm{d}\mu = \sum_{i=0}^{\infty} i p_i.$$

For continuous distribution, the density function p is actually a non-negative, integrable function, and $\int_{\mathbb{R}} p(x)\,\mathrm{d}x = 1$. So $\mu(B) = \int_B p(x)\,\mathrm{d}x$ is a probability measure.

Since $\mu_\xi\big|_{\mathscr{P}_{\mathbb{R}}} = \mu\big|_{\mathscr{P}_{\mathbb{R}}}$, $\mu_\xi = \mu$. By typical method, we can prove

$$Ef(\xi) = \int_{\mathbb{R}} f \,\mathrm{d}\mu_\xi = \int_{\mathbb{R}} f(x)p(x)\,\mathrm{d}x.$$

§4.5 L_p spaces

Definition 4.5.1 (L_p spaces)**.** Let $1 \le p < \infty$. Define

$$\|f\|_p := \left(\int_X |f|^p\right)^{\frac{1}{p}}, \quad L_p(X, \mathscr{F}, \mu) := \{f : \|f\|_p < \infty\}.$$

Sometimes we'll simplify the notation as $L_p(\mu), L_p(\mathscr{F})$ or just L_p.

- $f \in L_1$ iff f integrable, let $\|f\| := \|f\|_1$.
- $f \in L_p \iff f^p \in L_1 \implies f$ is finite a.e..

In fact, L_p is a normed vector space under the norm $\|\cdot\|_p$:

Lemma 4.5.2

Let $1 \le p < \infty$, let $C_p = 2^{p-1}$, then

$$|a + b|^p \le C_p(|a|^p + |b|^p), \quad a, b \in \mathbb{R}.$$

Proof. It's a single-variable inequality, it's obvious by taking the derivative. □

Thus by taking integral on both sides,

$$\int_X |f + g|^p \,\mathrm{d}\mu \le C_p \left(\int_X |f|^p \,\mathrm{d}\mu + \int_X |g|^p \,\mathrm{d}\mu\right).$$

So L_p space is a vector space.

Lemma 4.5.3 (Holder's inequality)

Let $1 < p, q < \infty$, and $\frac{1}{p} + \frac{1}{q} = 1$.

$$\|fg\| \le \|f\|_p \|g\|_q, \quad \forall f \in L_p, g \text{ measurable}.$$

Proof. WLOG $\|f\|_p > 0$, $0 < \|g\|_q < \infty$. Let

$$a = \left(\frac{|f|}{\|f\|_p}\right)^p = \frac{|f|^p}{\int_X |f|^p \,\mathrm{d}\mu}, \quad b = \left(\frac{|g|}{\|g\|_q}\right)^q = \frac{|g|^q}{\int_X |g|^q \,\mathrm{d}\mu}.$$

By weighted AM-GM,

$$\int_X \frac{|fg|}{\|f\|_p \|g\|_q} \,\mathrm{d}\mu \le \int_X \left(\frac{a}{p} + \frac{b}{q}\right) \mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

The equality holds iff $a = b$, i.e. $\exists \alpha, \beta \ge 0$ not all zero s.t. $\alpha |f|^p = \beta |g|^q, a.e.$. ☐

Theorem 4.5.4 (Minkowski's inequality)
Let $1 \le p < \infty$, then

$$\|f + g\|_p \le \|f\|_p + \|g\|_p, \quad \forall f, g \in L_p.$$

The equality holds iff $(1) : p = 1, fg \ge 0$; $(2) p > 1, \exists \alpha, \beta \ge 0, s.t. \alpha f = \beta g, a.e.$.

Proof. When $p = 1$, it follows by $|f + g| \le |f| + |g|$.

When $p \ge 1$, let $q = \frac{p}{p-1}$, by Holder's inequality,

$$|f + g|^p \le |f||f + g|^{p-1} + |g||f + g|^{p-1},$$

$$\implies \|f + g\|_p^p \le (\|f\|_p + \|g\|_p) \cdot \||f + g|^{p-1}\|_q.$$

Note that

$$\||f + g|^{p-1}\|_q = \left(\int_X |f + g|^p \,\mathrm{d}\mu\right)^{\frac{1}{q}} = \|f + g\|_p^{\frac{p}{q}}.$$

Since $f + g \in L_p$, we can divide both sides by $\|f + g\|_p^{\frac{p}{q}}$ to get the result. ☐

In L_p space, we view two functions $f = g, a.e.$ as the same function, i.e. the original function space modding the equivalence relation out.

Hence $(L_p/\sim, \|\cdot\|_p)$ is a normed vector space.

When $p = \infty$, define

$$\|f\|_\infty := \inf\{a \in \mathbb{R} : \mu(|f| > a) = 0\}, \quad L_\infty := \{f : \|f\|_\infty < \infty\}.$$

We call the functions in L_∞ **essentially bounded**.

Let $\mu(X) < \infty$, then $f \in L_\infty \implies f \in L_p$, and $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$: For all $0 < a < \|f\|_\infty$,

$$a^p \mu(|f| > a) \le \int_X |f|^p \mathbf{I}_{|f|>a} \,\mathrm{d}\mu \le \int_X |f|^p \,\mathrm{d}\mu \le \|f\|_\infty^p \mu(X),$$

So taking the exponent $\frac{1}{p}$,

$$a \leftarrow a\mu(|f| > a)^{\frac{1}{p}} \le \|f\|_p \le \|f\|_\infty$$

But when $\mu(X) = \infty$, let $f \equiv 1$, then $f \in L_\infty$ but $f \notin L_p$.

Theorem 4.5.5
Let $f, g \in L_\infty$,
$$\|fg\| \le \|f\|\|g\|_\infty,$$
$$\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty.$$

Proof.
$$\int_X |fg| \, \mathrm{d}\mu \le \int_X |f|\|g\|_\infty \, \mathrm{d}\mu = \|f\|\|g\|_\infty.$$
Since $|f(x)+g(x)| \le |f(x)| + |g(x)| \le \|f\|_\infty + \|g\|_\infty, a.e.$, we get the second inequality. □

Similarly we get $(L_\infty, \|\cdot\|_\infty)$ is a normed vector space.

The norm can deduce a *distance*:
$$\rho(f,g) := \|f-g\|.$$

Theorem 4.5.6 (L_p space is complete)
Let $1 \le p \le \infty$. If $\{f_n\} \subset L_p$ satisfying $\lim_{n,m\to\infty}\|f_n - f_m\|_p = 0$, then there exist $f \in L_p$ s.t. $\lim_{n\to\infty}\|f - f_n\|_p = 0$.

Proof. Take $n_1 < n_2 < \cdots$ such that
$$\|f_m - f_n\|_p \le \frac{1}{2^k}, \quad \forall n, m \ge n_k.$$
Let $g =\uparrow \lim_{k\to\infty} g_k$, where
$$g_k := |f_{n_1}| + \sum_{i=1}^{k} |f_{n_{i+1}} - f_{n_i}| \in L_p, \quad g_k \ge 0.$$
Since
$$\|g_k\|_p \le \|f_{n_1}\|_p + \sum_{i=1}^{k} \|f_{n_{i+1}} - f_{n_i}\|_p \le \|f_{n_1}\|_p + 1.$$
$$\implies \|g\|_p =\uparrow \lim_{k\to\infty} \|g_k\|_p \le \|f_{n_1}\|_p + 1.$$
Here we use the monotone convergence theorem. We can check the above also holds for $p = \infty$.

Therefore $g \in L_p \implies g < \infty, a.e.$. We have
$$f := f_{n_1} + \sum_{i=1}^{\infty} (f_{n_{i+1}} - f_{n_i}) = \lim_{k\to\infty} f_k, a.e.$$
the series is absolutely convergent, so f exists a.e. and $|f| \le g, a.e.$.

Lastly we can check: when $p = \infty$,
$$\|f_n - f\|_\infty \le \|f_n - f_{n_k}\|_\infty + \|f_{n_k} - f\|_\infty,$$
where the both term approach to 0 as $n \to \infty$.

When $p < \infty$, by Fatou's lemma,
$$\|f_n - f\|_p^p = \int_X |f_n - f|^p \, \mathrm{d}\mu = \int_X \lim_{k\to\infty} |f_n - f_{n_k}|^p \, \mathrm{d}\mu \le \liminf_{k\to\infty} \int_X |f_n - f_{n_k}|^p \, \mathrm{d}\mu \le \varepsilon.$$
□

Remark 4.5.7 — Using the same technique we can prove that if f_n is Cauchy in measure, then f_n converge to some f in measure:

Let $A_i = \{|f_{n_{i+1}} - f_{n_i}| > 2^{-i}\}$ s.t. $\mu(A_i) < 2^{-i}$.

Define $f = f_{n_1} + \sum_{i\geq 1}(f_{n_{i+1}} - f_{n_i})$ on the set $\bigcup_{k\geq 1}\bigcap_{i\geq k} A_i^c$.

This theorem implies that $(L_p, \|\cdot\|_p)$ is a Banach space. So we can try to define an *inner product* on L_p space:

$$\langle x, y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2).$$

We can check $\langle\cdot,\cdot\rangle$ is bilinear only if $p = 2$, so L_2 is actually a Hilbert space.

When $0 < p < 1$, let

$$\|f\|_p := \int_X |f|^p \,\mathrm{d}\mu, \quad L_p = \{f : \|f\|_p < \infty\}.$$

Lemma 4.5.8

Let $0 < p < 1$, $C_p = 1$, then

$$|a+b|^p \leq C_p(|a|^p + |b|^p), \quad \forall a, b \in \mathbb{R}.$$

So L_p is a vector space.

Theorem 4.5.9 (Minkowski)

Let $0 < p < 1$ then

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

Remark 4.5.10 — When $\|f\|_p = (\int_X |f|^p \,\mathrm{d}\mu)^{\frac{1}{p}}$, $0 < p < 1$. then it won't satisfy Minkowski's inequality.

Thus L_p is only a metric space but not a normed vector space. Using the same method we can prove L_p is a complete metric space.

§4.6 Convergence in L_p space

Definition 4.6.1. Let $0 < p \leq \infty$, $f, f_1, f_2, \cdots \in L_p$. When $\|f_n - f\|_p \to 0$, then we write $f_n \xrightarrow{L_p} f$, called **average converge of order p**.

Theorem 4.6.2

Let $0 < p < \infty$, $f, f_1, \cdots \in L_p$,

- If $f_n \xrightarrow{L_p} f$, then $f_n \xrightarrow{\mu} f$, and $\|f_n\|_p \to \|f\|_p$.
- If $f_n \to f$, *a.e.* or in measure, then $\|f_n\|_p \to \|f\|_p \iff f_n \xrightarrow{L_p} f$.

Proof. When $f_n \xrightarrow{L_p} f$, let $A := \{|f_n - f| > \varepsilon\}$,

$$\mu(A) \leq \frac{1}{\varepsilon^p}\int_X |f_n - f|^p \mathbf{I}_A \,\mathrm{d}\mu \leq \frac{1}{\varepsilon^p}\|f_n - f\|_p^p \to 0.$$

and obviously $\|f_n\|_p \to \|f\|_p$

On the other hand, when $f_n \to f, a.e.$ and $\|f_n\|_p \to \|f\|_p$, From $|a+b|^p \leq C_p(|a|^p + |b|^p)$,

$$g_n := C_p(|f_n|^p + |f|^p) - |f_n - f|^p \geq 0.$$

$g_n \to 2C_p|f|^p, a.e.$, so

$$\int_X 2C_p|f|^p \,\mathrm{d}\mu \leq \liminf_{n\to\infty}\int_X g_n \,\mathrm{d}\mu = 2C_p\int_X |f|^p \,\mathrm{d}\mu - \limsup_{n\to\infty}\int_X |f_n - f|^p \,\mathrm{d}\mu.$$

When $f_n \to f$ in measure, for any subsequence there exist its subsequence $f_{n'} \to f, a.e.$, so $\|f_{n'} - f\|_p \to 0$, hence $\|f_n - f\|_p \to 0$. □

Remark 4.6.3 — This theorem implies for any L_p function f, we can take simple functions $f_1, f_2, \cdots \to f$ and $|f_n| \uparrow |f|$, so $f_n \xrightarrow{L_p} f$.

Definition 4.6.4 (Weak convergence)**.** Let $1 < p < \infty$, and $f_1, f_2 \cdots \in L_p$. If

$$\lim_{n\to\infty}\int_X f_n g \,\mathrm{d}\mu = \int_X fg \,\mathrm{d}\mu, \quad \forall g \in L_q.$$

Then we say f_n **weak convergent** to f, denoted by $f_n \xrightarrow{(w)L_p} f$.

When $p = 1$ and $(X, \mathscr{F}, \mu)$ is a σ-finite measure space, and the condition also holds, we say $\{f_n\}$ weak convergent to f in L_1.

Corollary 4.6.5

Let $1 \leq p < \infty$, then

$$f_n \xrightarrow{L_p} f \implies f_n \xrightarrow{(w)L_p} f.$$

Proof. By Holder's inequality,

$$\left|\int_X (f_n - f)g \,\mathrm{d}\mu\right| \leq \|f_n - f\|_p\|g\|_q \to 0.$$

□

If $\sup_{t\in T}\|f_t\|_p =: M < \infty$, then we say $\{f_t, t \in T\}$ is **bounded in L_p**.

Theorem 4.6.6

Let $1 < p < \infty$, $\{f_n\} \subset L_p$, there exists M s.t. $\|f_n\|_p \leq M, \forall n$. If $f_n \to f, a.e.$ or in measure, then $f \in L_p$ and $f_n \to f$ weakly.

Proof. First $\|f\|_p \le M$:

$$\int_X |f|^p \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int_X |f_n|^p \,\mathrm{d}\mu \le M^p.$$

Next we prove the weak convergence: For all $g \in L_q$, recall the bounded convergence theorem in probability, we can view M as a bound of f_n, and $\|g\|_q$ as P.

Let $B = \{|f_n - f| \le \hat{\varepsilon}\}$, consider

$$a := \int_B (f_n - f)g \,\mathrm{d}\mu, \quad b := \int_{B^c} (f_n - f)g \,\mathrm{d}\mu.$$

Note that

$$|a| \le \hat{\varepsilon} \int_X |g| \,\mathrm{d}\mu.$$

But $\int_X |g| \,\mathrm{d}\mu$ might be infinity, so let $A_k := \{\frac{1}{k} \le |g|^q \le k\}$, we have

$$\int_{A_k} |g| \,\mathrm{d}\mu \le k^{\frac{1}{q}} \mu(A_k) < \infty.$$

($\frac{1}{k}\mu(A_k) < \int_{A_k} |g|^q \,\mathrm{d}\mu < \infty$ since $g \in L_q$).

Now we can proceed:

$$a := \int_{A_k B} (f_n - f)g \,\mathrm{d}\mu, \quad b := \int A_k^c \cup B^c (f_n - f)g \,\mathrm{d}\mu.$$

Now $|a| \le \hat{\varepsilon} k^{\frac{1}{q}} \mu(A_k) < \varepsilon$.

$$\left| \int_X (f_n - f) g \mathbf{I}_{A_k^c \cup B^c} \,\mathrm{d}\mu \right| \le \|f_n - f\|_p \|g \mathbf{I}_{A_k^c \cup B^c}\|_q \le 2M \left(\int_{A_k^c} |g|^q \,\mathrm{d}\mu + \int_{A_k \backslash B} |g|^q \,\mathrm{d}\mu \right).$$

By LDC(Dominated convergence), $A_k^c \to \{g = 0, \infty\}$, so $\int_{A_k^c} |g|^q \,\mathrm{d}\mu < \varepsilon$.

Since $\mu(A_k) < \infty$, $f_n \to f, a.e. \implies f_n \xrightarrow{\mu} f$. By the continuity of integrals, $\mu(A_k \backslash B) \le \mu(B^c) < \delta \implies \int_{A_k \backslash B} |g|^q \,\mathrm{d}\mu < \varepsilon$.

Now we can conclude: $\forall \varepsilon > 0$, first choose k large, then $\hat{\varepsilon}$ small, we get

$$\int_X (f_n - f)g \,\mathrm{d}\mu \le \varepsilon + 4M\varepsilon \implies f_n \xrightarrow{(w)L_p} f.$$

□

Remark 4.6.7 — The proof is a little complicated, we divide the entire integral to three part, and estimate them respectively.

When $p = 1$, f_n bounded in L_p *cannot* imply weak convergence.

Example 4.6.8

Let $X = \mathbb{N}$, $\mu(\{k\}) = 1, \forall k$, clearly it's σ-finite.

Let $f_n(k) = \mathbf{I}_{k=n}$, then $\|f_n\| = \sum_k \mu(k)|f_n(k)| = 1$, and $f_n \to 0, a.e.$.

But let $g = 1 \in L_\infty$, $\int_X (f_n - f)g \,\mathrm{d}\mu = 1 \not\to 0$.

Proposition 4.6.9

Let $f_1, f_2, \cdots \in L_1$, then:

$$\|f_n\| \to \|f\| \& f_n \to f, a.e. \implies f_n \xrightarrow{L_1} f \implies f_n \xrightarrow{(w)L_1} f \implies \int_A f_n \,\mathrm{d}\mu \to \int_A f \,\mathrm{d}\mu, \forall A.$$

Proof. For the last part let $g = \mathbf{I}_A$, the rest is trivial. □

§4.7 Integrals in probability space

We can also consider L_p space in probability space $(\Omega, \mathscr{F}, P)$.

Theorem 4.7.1

Let $0 < s < t < \infty$. Then $L_t \subset L_s$. If $s \geq 1$, we have $\|f\|_s \leq \|f\|_t$, with equality f constant.

Proof. When $f \in L_t$, let $p = \frac{t}{s}, q = \frac{t}{t-s}$.

$$\int_\Omega |f|^s \cdot 1 \,\mathrm{d}P \leq \||f|^s\|_p \|1\|_q = (E|f|^{sp})^{\frac{1}{p}} = (E|f|^t)^{\frac{1}{p}}.$$

So $f \in L_s \implies L_t \subset L_s$. When $s \geq 1$,

$$\|f\|_s^s \leq (\|f\|_t)^{\frac{t}{p}} = \|f\|_t^s \implies \|f\|_s \leq \|f\|_t.$$

□

From this we know $L_\infty \subset L_p$, and $\|f\|_p \uparrow \|f\|_\infty$.

Remark 4.7.2 — This theorem does not hold for general space. Let $X = \mathbb{N}$, $\mu(\{n\}) = 1$, $f(n) = \frac{1}{n}$, then $f \in L_2 \backslash L_1$.

The expectation Ef^k is called **k-order moment** of random variable f.

Definition 4.7.3 (Uniformly integrable)**.** Let $\{f_t, t \in T\}$ be r.v.'s, if $\forall \varepsilon > 0, \exists \lambda > 0$, such that

$$E|f_t|\mathbf{I}_{\{|f_t| > \lambda\}} < \varepsilon, \quad \forall t \in T,$$

then we say $\{f_t, t \in T\}$ **uniformly integrable**.

If $\forall \varepsilon > 0, \exists \delta > 0$ s.t $\forall A \in \mathscr{F}$,

$$P(A) < \delta \implies E|f_t|\mathbf{I}_A < \varepsilon, \forall t \in T,$$

we say $\{f_t\}$ is uniformly absolutely continuous, which is abbreviated as absolutely continuous.

Theorem 4.7.4

Uniformly integrable $\iff$ absolute continuity and bounded in L_1.

Proof. Firstly when $\{f_t\}$ uniformly integrable, $\forall A \in \mathscr{F}, \lambda > 0$,

$$\begin{aligned} E|f_t|\mathbf{I}_A &= E|f_t|\mathbf{I}_{A\cap\{|f_t|\le\lambda\}} + E|f_t|\mathbf{I}_{A\cap\{|f_t|>\lambda\}} \\ &\le \lambda P(A) + E|f_t|\mathbf{I}_{\{|f_t|>\lambda\}} \end{aligned}$$

Let $A = X$ we know $E|f_t| \le \lambda + \frac{\varepsilon}{2}, \forall t \in T$. Now let $\delta = \frac{\varepsilon}{2\lambda}$ we get AC property.

On the other hand,

$$\lambda P(|f_t| > \lambda) \le E|f_t|\mathbf{I}_{\{|f_t|>\lambda\}} \le E|f_t| \le M, \forall t \in T.$$

So when $\lambda > \frac{M}{\delta}$, $P(|f_t| > \lambda) < \delta$, hence $E|f_t|\mathbf{I}_{\{|f_t|>\lambda\}} \le \varepsilon$, $\forall t \in T$. □

Theorem 4.7.5

Let $0 < p < \infty$, and $f_n \to f$ in probability. TFAE:

(1) $\{|f_n|^p\}$ uniformly integrable;

(2) $f_n \xrightarrow{L_p} f$;

(3) $f \in L_p$ and $\|f_n\|_p \to \|f\|_p$.

Proof. (1) $\implies$ (2): Take subsequence $f_{n'} \to f, a.s.$,

$$E|f|^p \le \liminf_{n\to\infty} E|f_n|^p < \infty,$$

since $\{|f_n|^p\}$ is bounded in L_1. This means $f \in L_p$.

Let $A_n = \{|f_n - f| > \varepsilon\}$, now we compute

$$E|f_n - f|^p \le \varepsilon^p + E|f_n - f|^p\mathbf{I}_{A_n} \le \varepsilon^p + C_pE|f_n|^p\mathbf{I}_{A_n} + C_pE|f|^p\mathbf{I}_{A_n}$$

Since $P(A_n) \to 0$ and $\{|f_n|^p\}$ absolutely continuous (also note $E|f|^p\mathbf{I}_{A_n} \to 0$), RHS converges to 0. Therefore $f_n \xrightarrow{L_p} f$.

As for (3) $\implies$ (1), we'll prove a lemma:

Lemma 4.7.6

If $f_n \xrightarrow{P} f$, then $\forall 0 < p < \infty$,

$$|f_n|^p\mathbf{I}_{\{|f_n|\le\lambda\}} \xrightarrow{P} |f|^p\mathbf{I}_{\{|f|\le\lambda\}}, \quad \forall \lambda \in C(F_{|f|}).$$

By lemma and bounded convergence theorem, their expectation also converges. Note that $\|f_n\|_p \to \|f\|_p$, so

$$E|f_n|^p\mathbf{I}_{\{|f_n|>\lambda\}} \to E|f|^p\mathbf{I}_{\{|f|>\lambda\}},$$

thus $\forall \varepsilon > 0, \exists \lambda_0 \in C(F_{|f|})$, s.t. $E|f|^p\mathbf{I}_{\{|f|>\lambda_0\}} < \frac{\varepsilon}{2}$, thus

$$\exists N, \quad E|f_n|^p\mathbf{I}_{\{|f_n|>\lambda_0\}} < \varepsilon, \quad \forall n > N.$$

Now we can take $\lambda > \lambda_0$ such that $\max_{n\le N} E|f_n|^p\mathbf{I}_{\{|f_n|^p>\lambda\}} < \varepsilon$, and we're done.

Proof of the lemma. Since $|f_n| \to |f|$ in probability, WLOG $f_n, f \geq 0$. Define

$$A_n := (\{f_n \leq \lambda\}\Delta\{f \leq \lambda\}) \cap \{|f_n^p - f^p| > \varepsilon\}$$

$$B_n := \{f_n, f \leq \lambda, |f_n^p - f^p| > \varepsilon\}.$$

Since x^p is uniformly continuous in $[0, \lambda]$, $B_n \subset \{|f_n - f| > \kappa_{\varepsilon,\lambda}\}$, $P(B_n) \to 0$.

Also $P(A_n) \to 0$ as

$$A_n \subset \{\lambda - \delta < f \leq \lambda + \delta\} \cup \{|f_n - f| > \delta\},$$

and $F_{|f|}$ continuous at λ. □

□

§5 Signed measure

§5.1 Definitions

Let $(X, \mathscr{F}, \mu)$ be a measure space, consider

$$\varphi(A) := \int_A f \, \mathrm{d}\mu, \quad \forall A \in \mathscr{F}.$$

If the integral of f exists, then φ has countable additivity. Also note $\varphi(\emptyset) = 0$, so φ looks like a measure, except it can take negative values.

In fact, denote $X^+ = \{f \geq 0\}, X^- = \{f < 0\}$, then $\varphi(A) = \varphi(AX^+) + \varphi(AX^-)$.

Definition 5.1.1 (Signed measure)**.** If a set function $\varphi : \mathscr{F} \to \overline{\mathbb{R}}$ which satisfies countable additivity and $\varphi(\emptyset) = 0$, then we call φ a **signed measure**.

If $|\varphi(A)| < \infty, \forall A \in \mathscr{F}$, then φ is **finite**; Similarly we define **σ-finite**.

Since $\int_A f \, \mathrm{d}\mu$ can't reach both $\pm\infty$(otherwise the integral doesn't exist), so

Proposition 5.1.2

Let φ be a signed measure, then:

$$\varphi(A) < \infty, \quad \forall A \in \mathscr{F}, \quad \textit{or} \quad \varphi(A) > -\infty, \quad \forall A \in \mathscr{F}.$$

Proof. Assume that $\varphi(A) = \infty, \varphi(B) = -\infty$, then:

$$\varphi(A \cup B) = \varphi(A) + \varphi(A \backslash B) = +\infty,$$

and similarly $\varphi(A \cup B) = -\infty$, contradiction! □

Remark 5.1.3 — From now on we may assmue $\varphi(A) > -\infty$.

Proposition 5.1.4

If $A \supseteq B$, and $|\varphi(A)| < \infty$, then $|\varphi(B)| < \infty$.

Proof. Trivial, same as above proposition. □

Proposition 5.1.5

Let $A_1, A_2, \dots$ be pairwise disjoint sets, and $|\varphi(\sum_{n=1}^{\infty} A_n)| < \infty$, then

$$\sum_{n=1}^{\infty} |\varphi(A_n)| < \infty.$$

Proof. Let $I = \{n : \varphi(A_n) > 0\}$, $J = \{n : \varphi(A_n) < 0\}$,

$$B = \sum_{n \in I} A_n, \quad C = \sum_{n \in J} A_n,$$

since $B, C \subset \sum_{n=1}^{\infty} A_n$, thus $\varphi(B), \varphi(C) \in \mathbb{R}$.

Note that $\sum_{n \in I} |\varphi(A_n)| = |\varphi(B)|$, $\sum_{n \in J} \varphi(A_n) = |\varphi(C)|$, and we're done. □

§5.2 Hahn decomposition and Jordan decomposition

Let's look at the indefinite integral again, notice that

$$\varphi(A) = \int_{A \cap \{f > 0\}} f \,\mathrm{d}\mu + \int_{A \cap \{f < 0\}} f \,\mathrm{d}\mu = \int_A f^+ \,\mathrm{d}\mu - \int_A f^- \,\mathrm{d}\mu.$$

It turns out that this property holds for any signed measure.

Definition 5.2.1 (Hahn decomposition)**.** If a patition $\{X^+, X^-\}$ of X satisfies:

$$\varphi(A) \geq 0, \forall A \subset X^+, \quad \varphi(A) \leq 0, \forall A \subset X^-,$$

then $\{X^+, X^-\}$ is called a **Hahn decomposition** of φ.

Definition 5.2.2 (Jordan decomposition)**.** Let $\varphi^\pm = \int_A f^\pm \,\mathrm{d}\mu$ be measures, if

$$\varphi = \varphi^+ - \varphi^-,$$

then it's called a **Jordan decomposition** of φ.

We're going to find X^+, or equivalently, find φ^+. Let $\varphi^*(A) := \sup\{\varphi(B) : B \subseteq A\}$.
It's clear that φ^* is non-negative, monotone, and $\varphi^*(\emptyset) = 0$.
Consider $\mathscr{F}^- = \{A : \varphi^*(A) = 0\}$. Intuitively, this is all the subsets of X^-, unioned with "null sets" in X^+.

Theorem 5.2.3 (Hahn decomposition)

Let X^- be a set with maximum $|\varphi|$ in $\mathscr{F}^-$, (since $\varphi > -\infty$, X^- must exist) and $X^+ = X \backslash X^-$ doesn't contain any set A with $\varphi(A) < 0$.

Furthermore, the Hahn decomposition is unique:

$$\varphi(A) = 0, \quad \forall A \in X_1^+ \Delta X_2^+ = X_1^- \Delta X_2^-.$$

The critical part of this theorem is:

Lemma 5.2.4

If $\varphi(A) < 0$, then we can find $A_0 \subset A$ s.t. $\varphi^*(A_0) = 0$, $\varphi(A_0) < 0$.

To prove this lemma, we need another lemma:

Lemma 5.2.5

If $\varphi(A) < \infty$, then $\forall \varepsilon > 0$, $\exists A_\varepsilon \subset A$ s.t.

$$\varphi(A_\varepsilon) \geq 0, \quad \varphi^*(A \backslash A_\varepsilon) \leq \varepsilon.$$

Proof. Assume by contradiction that $\exists \varepsilon_0 \geq 0$ s.t. $\forall A_0 \subset A$, $\varphi(A_0) < 0$ or $\varphi^*(A \backslash A_0) > \varepsilon_0$, this means,

$$\varphi(A_0) \geq 0 \implies \varphi^*(A \backslash A_0) > \varepsilon_0.$$

This will clearly yield a contradiction:

Take any $\varphi(A_0) \geq 0$(say $A_0 = \emptyset$), then exists $A_1 \subset A \backslash A_0$ s.t. $\varphi(A_1) > \varepsilon_0$, and $\varphi(A_0 \cup A_1) \geq 0$, continuing this process we can get infinitely many pairwise disjoint sets $A_1, A_2, \ldots$, with $\varphi(A_n) > \varepsilon_0$, so $\varphi(\sum_{i=1}^\infty A_n) = \infty \implies \varphi(A) = \infty$, contradiction! □

Proof of Lemma 5.2.4. Applying above lemma repeatedly and take a limit:

Take $C_1 \subset A$ s.t. $\varphi(C_1) \geq 0$ and $\varphi^*(A \backslash C_1) \leq 1$. Let $A_1 = A \backslash C_1$, $\varphi(A_1) < 0$.

Again take

$$C_{k+1} \subset A_k, A_{k+1} = A_k \backslash C_{k+1} \implies \varphi^*(A_{k+1}) \leq \frac{1}{k+1}, \varphi(A_{k+1}) < 0.$$

Since $A_k \downarrow$, let $A_0 = \lim_{k\to\infty} A_k$, note $\varphi^*(A_k) \downarrow 0$, we must have $\varphi^*(A_0) = 0$.

Also $\varphi(\sum C_k) = \sum \varphi(C_k) \geq 0$, so $\varphi(A_0) < 0$. □

Proof of Theorem 5.2.3. First we prove that $\mathscr{F}^-$ is a σ-ring: $\emptyset \in \mathscr{F}^-$, if $A_1, A_2 \in \mathscr{F}^-$,

$$0 \leq \varphi^*(A_1 \backslash A_2) \leq \varphi(A_1) = 0.$$

Thus $A_1 \backslash A_2 \in \mathscr{F}^-$.

If $A_1, A_2, \cdots \in \mathscr{F}^-$ pairwise disjoint,

$$\varphi(B) = \sum_{n=1}^\infty \varphi(B \cap A_n) \leq 0, \quad \forall B \subset \sum_{n=1}^\infty A_n.$$

Hence $\sum_{n=1}^\infty A_n \in \mathscr{F}^-$.

Next we'll prove Hahn decomposition exists:

Let $\alpha := \inf\{\varphi(A) : A \in \mathscr{F}^-\}$, $\alpha \leq 0$.

Let $\{A_n\} \in \mathscr{F}^-$ s.t. $\varphi(A_n) \to \alpha$, then $X^- := \bigcup_{n=1}^\infty A_n \in \mathscr{F}^-$.

$$\varphi(X^-) = \varphi(A_n) + \varphi(X^- \backslash A_n) \leq \varphi(A_n) + \varphi^*(X^- \backslash A_n) = \varphi(A_n) \to \alpha.$$

Therefore $-\infty < \varphi(X^-) = \alpha$.

Hence $\forall A, \varphi(AX^-) \leq \varphi^*(X^-) = 0$. By Lemma 5.2.4 we get $\forall A, \varphi(AX^+) \geq 0$, otherwise $\exists A_0 \subset A$ s.t. $\varphi^*(A_0) = 0, \varphi(A_0) < 0$. Then $\varphi(X^- \cup A_0) = \alpha + \varphi(A_0) < \alpha$, contradiction!

At last we'll prove the uniqueness:

If $X_1^\pm, X_2^\pm$ are both Hahn decompositions, then $A \in X_1^+ \cap X_2^- + X_1^- \cap X_2^+$, it's clear $\varphi(A) = 0$. □

Theorem 5.2.6 (Jordan decomposition)
The Jordan decomposition exists and is unique:

$$\varphi = \varphi^+ - \varphi^-, \quad \varphi^+ = \varphi^*, \varphi^- = (-\varphi)^*.$$

Proof. Let $\varphi^\pm$ be measures with $\varphi^\pm = \pm\varphi(A \cap X^\pm)$. It's clear that this is a Jordan decomposition.

Now given any Jordan decomposition $\varphi^\pm$.

Since

$$\forall B \subset A, \varphi(B) \le \varphi^+(B) \le \varphi^+(A),$$

so $\varphi^* \le \varphi^+$. But $A \cap X^+ \subset A$, so $\varphi^* \ge \varphi^+$, which proves the result.

Similarly $\varphi^- = (-\varphi)^*$, so it is unique. □

Remark 5.2.7 — The support of $\varphi^\pm$ are disjoint, but if $\phi \ne 0$, then the support of $\varphi^\pm + \phi$ intersects. $\varphi^\pm$ are called the **upper variation** and **lower variation**, respectively, and $|\varphi| = \varphi^+ + \varphi^-$ is called the **total variation**.

Lemma 5.2.8
$|\varphi|(A) = 0 \iff \varphi(B) = 0, \forall B \subset A.$

Proof. Just write $|\varphi| = \varphi^+ + \varphi^-$, we know $\varphi(B) = 0$.

Conversely, $\varphi(X^\pm \cap A) = 0 \implies |\varphi|(A) = 0.$ □

§5.3 Radon-Nikodym theorem

We assume the functions and sets below are all measurable. Let $(X, \mathscr{F})$ be a measurable space, φ a signed measure.

Definition 5.3.1 (R-N derivative)**.** If there exists a a.e. unique function f s.t.

$$\varphi(A) = \int_A f \, \mathrm{d}\mu, \quad \forall A \in \mathscr{F},$$

we say f is the **Radon-Nikodym derivative** of φ with respect to μ, abbreviated by R-N derivative or derivative, denoted by $\frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$.

Remark 5.3.2 — When μ is σ-finite, then f must be unique a.e..

Definition 5.3.3 (Absolute continuity)**.** If $\forall A \in \mathscr{F}$,

$$\mu(A) = 0 \implies \varphi(A) = 0,$$

then we say φ is **absolutely continuous** with respect to μ, denoted by $\varphi \ll \mu$.

Observe that

$$\mu(A) = 0 \implies \mu(A \cap X^\pm) = 0 \implies \varphi^\pm(A) = 0,$$

so $\varphi \ll \mu \iff \varphi^\pm \ll \mu \iff |\varphi| \ll \mu$.

It's obvious that $\frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$ exists only if $\varphi \ll \mu$, but it turns out that this is also the sufficient condition when μ is a σ-finite measure.

We can't prove this directly, so we'll prove some easy cases first.

Lemma 5.3.4

Let φ, μ be finite measures. Then

$$\exists f \in \mathscr{L} := \left\{ g \in L_1 : g \geq 0, \int_A g \,\mathrm{d}\mu \leq \varphi(A), \forall A \right\},$$

such that $\int_X f \,\mathrm{d}\mu = \sup \int_X g \,\mathrm{d}\mu$.

Proof. This is somehow similar to find simple functions approaching non-negative measurable functions.

First let $\beta = \sup \int_X g \,\mathrm{d}\mu$, and choose g_k s.t. $\int_X g_k \,\mathrm{d}\mu \to \beta$.

Let $f_n := \max_{k \leq n} g_k$, and $f_n \uparrow f$. By Levi's theorem, $\int_A f \,\mathrm{d}\mu = \lim_{n\to\infty} f_n \,\mathrm{d}\mu$, so if $f_n \in \mathscr{L}$, $f \in \mathscr{L}$ as well. Let $A_k = A \cap \{f_n = g_k, f_n \neq g_j, j < k\}$ be a partition of A,

$$\int_A f_n \,\mathrm{d}\mu = \sum_{k=1}^{n} \int_{A_k} g_k \,\mathrm{d}\mu \leq \sum_{k=1}^{n} \varphi(A_k) = \varphi(A).$$

Thus $f_n \in \mathscr{L}$, we have $\int_X f \,\mathrm{d}\mu = \beta \geq \int_X g \,\mathrm{d}\mu$, for all $g \in \mathscr{L}$. □

Proposition 5.3.5

Suppose φ, μ are both finite, then $\varphi \ll \mu \implies \frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$ exists.

Proof. Decompose φ to $\varphi^+ - \varphi^-$, we may assume $\varphi \geq 0$.

Starting from previous lemma, we'll prove that $\int_A f \,\mathrm{d}\mu = \varphi(A)$. Let $\nu(A) = \varphi(A) - \int_A f \,\mathrm{d}\mu$ be a measure.

Let ν_n be increasing signed measures.

$$\nu_n(A) := \nu(A) - \frac{1}{n}\mu(A), \quad \forall A \in \mathscr{F}.$$

Let $X_n^{\pm}$ be the Hahn decomposition of ν_n, and

$$X^+ = \bigcup_{n=1}^{\infty} X_n^+, \quad X^- = \bigcap_{n=1}^{\infty} X_n^-.$$

First since $X^- \subset X_n^-$,

$$\nu(X^-) = \nu_n(X^-) + \frac{1}{n}\mu(X^-) \leq \frac{1}{n}\mu(X^-) \to 0.$$

We have $f + \frac{1}{n}\mathbf{I}_{X_n^+} \in \mathscr{L}$ since

$$\begin{aligned}\int_A \left(f + \frac{1}{n}\mathbf{I}_{X_n^+} \right) \mathrm{d}\mu &= \varphi(A) - \nu(A) + \frac{1}{n}\mu(X_n^+ \cap A) \\ &\leq \varphi(A) - \nu(X_n^+ \cap A) + \frac{1}{n}\mu(X_n^+ \cap A) \\ &= \varphi(A) - \nu_n(X_n^+ \cap A) \leq \varphi(A).\end{aligned}$$

So we have $\int_X f \,\mathrm{d}\mu \geq \int_X (f + \frac{1}{n}\mathbf{I}_{X_n^+}) \,\mathrm{d}\mu$, $\mu(X_n^+) = 0 \implies \mu(X^+) = 0$.

Since $\varphi \ll \mu$, $\varphi(X^+) = 0 \implies \nu(X^+) = 0$. □

Proposition 5.3.6

Let φ be a σ-fintie signed measure, μ be a finite measure, if $\varphi \ll \mu$, then $\frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$ exists and its integral exists.

Proof. Let $X = \sum_{n=1}^{\infty} A_n$, $|\varphi(A_n)| < \infty$, then the R-N derivative f_n exists on A_n,

Let $f = \sum_{n=1}^{\infty} f_n \mathbf{I}_{A_n}$, then f finite a.e.,

$$\varphi(A \cap A_n) = \int_{A \cap A_n} f_n \, \mathrm{d}\mu = \int_{A \cap A_n} f \, \mathrm{d}\mu.$$

WLOG φ^- finite, then

$$\varphi(\{f < 0\} \cap A_n) = \int_{A_n} f^- \, \mathrm{d}\mu = \int_{A_n} f_n^- \, \mathrm{d}\mu \geq -\varphi^-(A_n)$$

So the integral of f exists.

Since φ is countably additive and the integral of f exists, we can add the above equality to get the desired. $\square$

Proposition 5.3.7

Let φ be an arbitary signed measure, the above conclusion also holds.

Proof. Let

$$\mathscr{G} := \left\{ \sum_{n=1}^{\infty} A_n : |\varphi(A_n)| < \infty, n = 1, 2, \ldots \right\}.$$

Since $\emptyset \in \mathscr{G}$, and it's closed under set difference:

$$\sum_{n=1}^{\infty} A_n \backslash \sum_{n=1}^{\infty} B_n = \sum_{n=1}^{\infty} (A_n \backslash B)$$

by $A_n \backslash B \subset A_n$, we have $|\varphi(A_n \backslash B)| < \infty$.

Clearly it's closed under countable disjoint union, combined with difference sets we deduce it's closed under countable union, thus $\mathscr{G}$ is a σ-ring.

Note that there exists B s.t. $\mu(B) = \gamma := \sup\{\mu(A) : A \in \mathscr{G}\}$. (Since we can take $\mu(B_n) \to \gamma, B = \bigcup_{n=1}^{\infty} B_n$.)

So φ is σ-finite on $(B, B \cap \mathscr{F})$, the R-N derivative exists.

For all $C \subset B^c$, we must have $\varphi(C) = 0$ or ∞. TODO!! $\square$

At last we come to the full statement:

Theorem 5.3.8

Let φ be a signed measure, μ a σ-finite measure, if $\varphi \ll \mu$, then $\frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$ exists.

Example 5.3.9

Let $X = \mathbb{R}$, $\mu(A) = \#A$, μ is not σ-finite. Let $\varphi(A) = 0$ when A countable, 1 otherwise.

In this case the R-N derivative doesn't exist, otherwise

$$0 = \varphi(\{x\}) = \int_{\{x\}} f \, \mathrm{d}\mu = f(x)\mu(x) = f(x),$$

contradiction!

Remark 5.3.10 — If μ, ν are σ-finite measures, $\nu \ll \mu$, then

$$\int_X \mathbf{I}_A \, \mathrm{d}\nu = \int_X \mathbf{I}_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \implies \int_X f \, \mathrm{d}\nu = \int_X f \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

§5.4 The dual space of L_p

Let $(X, \mathscr{F}, \mu)$ be a measure space, $1 < p < \infty$.

Recall that $f_n \xrightarrow{(w)L_p} f$ is defined as

$$\lim_{n\to\infty} \int_X f_n g \, \mathrm{d}\mu = \int_X fg \, \mathrm{d}\mu, \quad \forall g \in L_q.$$

By Holder's inequality,

$$\left| \int_X fg \, \mathrm{d}\mu \right| \le \|g\|_q \|f\|_p, \quad \forall f \in L_p, g \in L_q.$$

Thus given any $g \in L_q$, we can induce a **funtional** on L_p, moreover it's linear and bounded.

Definition 5.4.1. We say a funtional $\Phi : L_p \to \mathbb{R}$ is bounded linear if:

$$|\Phi(f)| \le C\|f\|_p, \quad \Phi(f_1 + cf_2) = \Phi(f_1) + c\Phi(f_2).$$

We can easily see that Φ is continuous:

$$\|f_n - f\|_p \to 0 \implies |\Phi(f_n) - \Phi(f)| \to 0.$$

Let $\|\Phi\| := \inf C = \sup_{\|f\|_p=1} |\Phi(f)|$.

For all $A \in \mathscr{F}$, $\Phi_A := \Phi(f\mathbf{I}_A)$ is also a linear and bounded functional. It's clear that $\|\Phi_A\| \le \|\Phi\|$.

Let Φ_g denote the functional induced by $g \in L_q$:

$$\Phi_g : f \mapsto \int_X fg \, \mathrm{d}\mu, \quad |\Phi_g(f)| \le \|g\|_q \|f\|_p.$$

Moreover, take $f = |g|^{q-1}\mathrm{sgn}(g)$, we found that $\|\Phi_g\| = \|g\|_q$. We check it here:

$$\int_X |f|^p \, \mathrm{d}\mu = \int_X |g|^{p(q-1)} \, \mathrm{d}\mu = \int_X |g|^q \, \mathrm{d}\mu,$$

so $f \in L_p$, $\|f\|_p = \|g\|_q^{\frac{q}{p}} = \|g\|_q^{q-1}$. Thus the equality of Holder's inequality holds.

In fact L_q contains all the bounded linear functionals of L_p:

Theorem 5.4.2

The dual space of L_p is L_q, i.e. $L_p^* = L_q$.

The critical part is to use a signed measure φ to determine g:

$$\varphi(A) = \int_A g \, \mathrm{d}\mu = \int_X \mathbf{I}_A g \, \mathrm{d}\mu = \Phi(\mathbf{I}_A), \quad A \in \mathscr{F}.$$

We're faced with two main problems:

- $\mathbf{I}_A$ may not be in L_p.
- μ may not be σ-finite, so the derivative may not be unique.

To solve these problem, we'll start from finite measure, and proceed by finite $\to$ σ-finite $\to$ arbitary.

Proposition 5.4.3

If μ is a finite measure, then $L_p^* = L_q$.

Proof. For any bounded linear functional Φ, let $\varphi(A) = \Phi(\mathbf{I}_A)$,

$$|\varphi(A)| \le C\|\mathbf{I}_A\|_p = C\mu(A)^{\frac{1}{p}},$$

so φ is finite and $\varphi \ll \mu$.

Clearly $\varphi(\emptyset) = 0$, and $\varphi(A + B) = \varphi(A) + \varphi(B)$.

For countable additivity, let $A = \sum_{n=1}^{\infty} A_n$, $B_N = \sum_{n=N+1}^{\infty} A_n$, since $\mu(A)$ finite,

$$\left|\varphi(A) - \sum_{n=1}^{N} \varphi(A_n)\right| = |\varphi(B_N)| \le C\mu(B_N)^{\frac{1}{p}} \to 0.$$

By $\varphi \ll \mu$, let $g = \frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$. We have $|g| < \infty, a.e.$ and $g \in L^1$, so

$$\Phi(\mathbf{I}_A) = \varphi(A) = \int_A g \, \mathrm{d}\mu = \int_X \mathbf{I}_A g \, \mathrm{d}\mu, \quad \forall A \in \mathscr{F}.$$

By the linearity of Φ, we know for simple functions the above equation holds.

For $f \in L_p$ non-negative, we can take simple $f_n \uparrow f$, so $\int f_n^p \, \mathrm{d}\mu \uparrow \int f^p \, \mathrm{d}\mu \implies f_n \xrightarrow{L_p} f$.

By the continuity of Φ, $\Phi(f_n) \to \Phi(f)$.

For the integral part, let $X^+ = \{g \ge 0\}, X^- = \{g < 0\}$. Then $f_n^{\pm} := f_n \mathbf{I}_{X^{\pm}}$ non-negative simple, and $f_n^{\pm} \xrightarrow{L_p} f^{\pm} := f\mathbf{I}_{X^{\pm}}$.

Now we can use Levi's theorem to get

$$\int_X f_n^{\pm} g \, \mathrm{d}\mu \to \int_X f^{\pm} g \, \mathrm{d}\mu.$$

Note since LHS is $\Phi(f_n^{\pm})$, RHS must be $\Phi(f^{\pm}) \in \mathbb{R}$, so we can safely apply $f = f^+ + f^-$. At last f non-negative $\implies$ f measurable is easy, so we've proven

$$\Phi(f) = \int_X fg \, \mathrm{d}\mu, \quad \forall f \in L_p.$$

Next we'll prove $g \in L_q$. Let $A_n = \{|g| \le n\}$, let $g_n := g\mathbf{I}_{A_n}$, clearly $g_n \in L_q$ as the base measure is finite.

Since $\Phi_{g_n} = \Phi_{A_n}$, so

$$\|g_n\|_q = \|\Phi_{A_n}\| \le \|\Phi\|.$$

Now $|g_n| \uparrow |g|, a.e.$, by Levi $\|g_n\|_q \to \|g\|_q$, so $\|g\|_q < \infty$. □

Proposition 5.4.4

When μ is σ-finite, $L_p^* = L_q$.

Proof. Let $X = \sum_{n=1}^{\infty} X_n$, $\mu(X_n) < \infty$.

There exists g_n on X_n s.t. $\Phi_{X_n} = \Phi_{g_n}$. Let $g = \sum_{n=1}^{\infty} g_n \mathbf{I}_{X_n}$.

For $f \in L_p$, $\sum_{n=1}^{N} f\mathbf{I}_{X_n} \xrightarrow{L_p} f$, we have

$$\Phi(f) \leftarrow \Phi\left(\sum_{n=1}^{N} f\mathbf{I}_{X_n}\right) = \sum_{n=1}^{N} \Phi_{X_n}(f) = \sum_{n=1}^{N} \int_{X_n} fg \,\mathrm{d}\mu.$$

Similarly, let $A^+ = \{fg \ge 0\}, A^- = \{fg < 0\}$, $f^{\pm} = f\mathbf{I}_{A^{\pm}}$, we know the integral converges.

$g \in L_q$ is also the same as before.TODO

$$\|g\|_q = \lim_{N\to\infty} \left\|\sum_{n=1}^{N} g_n \mathbf{I}_{X_n}\right\| \le \|\Phi_g\| = \|\Phi\|.$$

□

Proposition 5.4.5

μ is an arbitary measure.

Proof. If $\mu(A) < \infty$, consider $\Phi_A : f \mapsto \Phi(f\mathbf{I}_A)$, we can get g_A.

If $A \subset B$, $\mu(B) < \infty$, then $g_B \mathbf{I}_A = g_A, a.e.$, $\|\Phi_A\| \le \|\Phi_B\|$.

We can take $A_n \uparrow, \mu(A_n) < \infty$ s.t.

$$\sup_n \|\Phi_{A_n}\| = \sup\{\|\Phi_A\| : \mu(A) < \infty\}.$$

Remark 5.4.6 — Here we're using A_n to replace $X_1 + \dots X_n$ in the previous proof.

Let $g_n := g_{A_n} \uparrow g$, then $g \in L_q$:

$$\|g\|_q^q = \int_X \lim_{n\to\infty} |g_n|^q \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int_X |g_n|^q \,\mathrm{d}\mu \le \|\Phi\|^q.$$

Let $A = \bigcup_{n=1}^{\infty} A_n$, since $g \in L_q$, by Holder and LDC,

$$\int_X fg \,\mathrm{d}\mu \leftarrow \int_X fg_n \,\mathrm{d}\mu = \Phi_{A_n}(f) = \Phi(f\mathbf{I}_{A_n}) \to \Phi(f\mathbf{I}_A).$$

The last part is to prove $\Phi(f\mathbf{I}_{A^c}) = 0$. Otherwise let $D_n = \{|f| > \frac{1}{n}\} \cap A^c$, then $\mu(D_n) < \infty$ since

$$\mu(D_n) \le \mu\left(|f| > \frac{1}{n}\right) \le \int_X (n|f|\mathbf{I}_{D_n})^p \, \mathrm{d}\mu < \infty.$$

By LDC, $f\mathbf{I}_{D_n} \xrightarrow{L_p} f\mathbf{I}_{A^c}$, so $\Phi(f\mathbf{I}_{D_n}) \neq 0$ for some n. But $\mu(D) < \infty$, let $B_n = A_n + D$ we'll find a contradiction on $\sup_n \|\Phi_{B_n}\| > \sup_n \|\Phi_{A_n}\|$. □

When $p = 1$, we can prove for σ-fintie measure μ that $L_1^* = L_\infty$. The method is the same as above.

§5.5 Lebesgue decomposition

Let φ, ϕ be two signed measures.

If $\varphi \ll |\phi|$, then we say φ is absolute continuous with respect to ϕ, denoted by $\varphi \ll \phi$. We can see that $\varphi \ll \phi \iff |\varphi| \ll |\phi|$.

Definition 5.5.1. If $\exists N \in \mathscr{F}$ such that

$$|\varphi|(N^c) = |\phi|(N) = 0,$$

then we say φ and ϕ are **mutually singular**, denoted by $\varphi \perp \phi$.

Lemma 5.5.2

$\varphi \perp \phi$ iff there exists $N \in \mathscr{F}$ such that

$$\varphi(A \cap N^c) = \phi(A \cap N) = 0, \quad \forall A.$$

Proof. This is trivial by $|\varphi|(A) = 0 \iff \varphi(B) = 0, \forall B \subset A$. □

Two measures are mutually singular is to say their supports are disjoint.

Lemma 5.5.3

If $\varphi \ll \phi$ and $\varphi \perp \phi$, then $\varphi \equiv 0$.

Proof. Take N s.t. $|\varphi|(N^c) = |\phi|(N) = 0$, since $\varphi \ll \phi$, $|\varphi|(N) = 0$ as well, thus $|\varphi|(X) = 0$. □

Theorem 5.5.4 (Lebesgue decomposition)

Let φ, ϕ be σ-finite signed measures, there exists unique σ-finite signed measures φ_c, φ_s s.t.

$$\varphi = \varphi_c + \varphi_s, \quad \varphi_c \ll \phi, \varphi_s \perp \phi.$$

Again, we'll start from finite measures, and reach σ-finite signed measures step by step.

Proposition 5.5.5

Let φ, μ be finite measures, then the Lebesgue decomposition holds.

Proof. Since $\varphi \ll \varphi + \mu$, let $f = \frac{d\varphi}{d(\varphi+\mu)}$, note that $0 \le f \le 1$, $(\varphi + \mu)$-a.e. (here we use the finite condition) and $1 - f = \frac{d\mu}{d(\varphi+\mu)}$.

Let $N = \{f = 1\}$,

$$\varphi_c(A) = \varphi(A \cap N^c), \quad \varphi_s(A) = \varphi(A \cap N).$$

Clearly $\varphi_s(N^c) = 0$,

$$\varphi(N) = \int_N f \, d(\varphi + \mu) = \int_N 1 \, d(\varphi + \mu) = \varphi(N) + \mu(N)$$

so $\mu(N) = 0, \varphi_s \perp \mu$.

On the other hand, if $\mu(A) = 0$, since $1 - f > 0$,

$$0 = \mu(AN^c) = \int_{AN^c} (1 - f) \, d(\varphi + \mu) \implies \varphi_c(A) \le (\varphi + \mu)(AN^c) = 0.$$

Thus $\varphi_c \ll \mu$, we're done. $\square$

From this proof, we can see that the critical point is to find a set N, s.t. $\mu(N) = 0$ and $\varphi_c = \varphi(\cdot \cap N^c) \ll \mu$, i.e. in some sense the "largest" null set of μ.

So this can give another proof:

Proof. Let $\gamma := \sup\{\varphi(A) : A \in \mathscr{F}, \mu(A) = 0\}$.

Let $A_n \in \mathscr{F}, \mu(A_n) = 0$ and $\varphi(A_n) \to \gamma$. Let $N = \bigcup A_n$, then $\varphi(N) = \gamma, \mu(N) = 0$.

If $\mu(A) = 0, \varphi_c(A) > 0$ for some A, then $\mu(N \cup A) = 0$,

$$\varphi(N \cup A) = \varphi(N) + \varphi(A \cap N^c) > \varphi(N) = \gamma,$$

contradiction!

Hence $\varphi_c \ll \mu$. $\square$

Proposition 5.5.6

Let φ, μ be σ-finite measures, the Lebesgue decomposition holds.

Proof. Let $\{A_n\}$ be a partition of X, $\varphi(A_n) < \infty, \mu(A_n) < \infty$.

On $(A_n, A_n \cap \mathscr{F})$, there exists Lebesgue decomposition $\varphi_{n,c}, \varphi_{n,s}$, let $\varphi_c(A) = \sum_{n=1}^\infty \varphi_{n,c}(A \cap A_n)$, φ_s similarly defined, we can easily check that $\varphi_c \ll \mu$ and $\varphi_s \perp \mu$. $\square$

At last we prove the Lebesgue decomposition: Let X^+, X^- be the Hahn decomposition of φ, WLOG φ^- finite.

By previous propositions, we have $\varphi_c^\pm, \varphi_s^\pm$, since φ_s^-, φ_c^- finite, so φ_c, φ_s is well-defined. The rest is some trivial work to check they satisfy the condition.

Now it remains to check the uniqueness. Suppose $\varphi_{c,i}, \varphi_{s,i}$ are two decompositions, $i = 1, 2$.

Let N_i be sets s.t. $\mu(N_i) = |\varphi_{s,i}|(N_i^c) = 0$, let $N = N_1 \cup N_2$, we have

$$\mu(N) = 0 \implies \varphi_{c,i}(N) = 0; \quad |\varphi_{s,i}|(N^c) = 0, i = 1, 2.$$

Thus $\varphi_{c,i}(A) = \varphi_{c,i}(AN^c) = \varphi(AN^c)$, and $\varphi_{s,i}(A) = \varphi_{s,i}(AN) = \varphi(AN)$.

At last we take $\mu = |\phi|$ to finally conclude.

Example 5.5.7

Let μ be a probability on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, λ is Lebesgue measure.

If $\mu \ll \lambda$, we say μ is continuous, and $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is the density function of μ.

If $\mu(\{x\}) > 0$, then we say x is an atom of μ,

$$D = D_\mu := \{x \in \mathbb{R} : \mu(\{x\}) > 0\},$$

then μ finite $\implies$ D countable.

If $\mu(D) = 1$, then we say μ is discrete.

If $\mu \perp \lambda$ and $D_\mu = \emptyset$, then we say μ is singular.

Then for any finite measure μ, let $\mu = \mu_c + \mu_s$ be the Lebesgue decomposition with respect to λ. Let $\mu_1 = \mu_c, \mu_2 = \mu(\cdot \cap D_\mu), \mu_3 = \mu_s - \mu_2$.

Then μ_1, μ_2, μ_3 are pairwise singular.

§5.6 Conditional expectations

Let $(X, \mathscr{F}, P)$ be a probability space. Let $\mathscr{G}$ be a sub σ-algebra of $\mathscr{F}$. Then we have another probability space $(X, \mathscr{G}, P)$.

Recall that $L_2(\mathscr{G}) \subset L_2(\mathscr{F})$ are Hilbert spaces.

Let $g \in \mathscr{G}$ be a function, $g \geq 0$, then $\int_X g \, \mathrm{d}P$ is the same in two spaces.(By Levi's theorem)

By linear algebra, for any $f \in \mathscr{F}$, there's a unique optimal approximation (or orthogonal projection) $f^* \in \mathscr{G}$ s.t.

$$\|f - f^*\|_2 = \inf_{g \in L_2(\mathscr{G})} \|f - g\|_2.$$

Therefore by orthogonality,

$$Efg = Ef^*g, \forall g \in L_2(\mathscr{G}) \iff Ef\mathbf{I}_A = Ef^*\mathbf{I}_A, \forall A \in \mathscr{G}.$$

Let $\varphi(A) = Ef\mathbf{I}_A$, $\varphi \ll P$, in fact we have $f^* = \frac{\mathrm{d}\varphi}{\mathrm{d}P}$ in $\mathscr{G}$.

Remark 5.6.1 — $\int_X f \, \mathrm{d}\mu$ only depends on $\sigma(f)$, so when $f \in \mathscr{G} \subset \mathscr{F}$, the integral is the same under both σ-algebra.

We can see that the condition L_2 is a little strong, so we can reduce it to existence of integrals.

Definition 5.6.2 (Conditional expectation)**.** Let $f \in \mathscr{F}$ whose integral exists, we say the **conditional expectation** of f under $\mathscr{G}$ is the function f^* with integral which satisfies:

$$f^* \in \mathscr{G}, \quad Ef^*\mathbf{I}_A = \int_A f \, \mathrm{d}P, \forall A \in \mathscr{G}.$$

This function is denoted by $E(f|\mathscr{G})$.

By the notation $E(f|\mathscr{G})$ we mean a family of *almost surely* equal functions which are measurable in $(X, \mathscr{G}, P)$.

The **conditional probability** of A under $\mathscr{G}$ is

$$P(A|\mathscr{G}) := E(\mathbf{I}_A|\mathscr{G}).$$

As we've said, let $\phi(A) = Ef\mathbf{I}_A$ be a signed measure, we have

$$\frac{\mathrm{d}\phi}{\mathrm{d}P} = f \in (X, \mathscr{F}), \quad \frac{\mathrm{d}\phi|_{\mathscr{G}}}{\mathrm{d}P} = f^* \in (X, \mathscr{G}).$$

All we've done is to find a approximation of f which isn't necessarily in $\mathscr{G}$

Let $g : (X, \mathscr{F}) \to (Y, \mathscr{S})$. We say the conditional expectation of f with respect to g is defined as

$$E(f|g) = E(f|\sigma(g))$$

In probability courses we learned:

$$Ef = EE(f|g),$$

now it's almost trivial since $\int_X f \, \mathrm{d}P = \int_X f^* \, \mathrm{d}P$.

Example 5.6.3

Let $\mathscr{G} = \{\emptyset, B, B^c, X\}$, where $B \in \mathscr{F}$. Then $E(f|\mathscr{G}) = \int_B f \, \mathrm{d}P P(B)^{-1}\mathbf{I}_B + \int_{B^c} f \, \mathrm{d}P P(B^c)^{-1}\mathbf{I}_{B^c}$.

We can see that the conditional expectation is indeed an "expectation".

Also, $P(A|\mathscr{G}) = P(A \cap B)P(B)^{-1}\mathbf{I}_B + P(A \cap B^c)P(B^c)^{-1}\mathbf{I}_{B^c}$, thus $P(A|B) = \frac{P(A\cap B)}{P(B)}$, which coincides with elementary probability.

Definition 5.6.4. Let $\{A_t, t \in T\}$ be a family of sets in $\mathscr{F}$, if $\forall n \geq 2, \{t_1, \ldots, t_n\} \subset T$,

$$P\left(\bigcap_{k=1}^n A_{t_k}\right) = \prod_{k=1}^n P(A_{t_k}),$$

we say $\{A_t, t \in T\}$ are independent.

Likely, we can define the independence of collections of sets, and therefore random variables.

Lemma 5.6.5

Let f be a random variable whose integral exists, if f and $\mathscr{E}$ are independent, we have

$$E(f\mathbf{I}_A) = (Ef) \cdot P(A), \quad \forall A \in \mathscr{E}$$

Next we'll study the properties of conditional expectations: Let f, g be functions whose integrals exist, $\mathscr{G}, \mathscr{G}_0$ are sub σ-algebras of $\mathscr{F}$,

(1) If $f \in \mathscr{G}$, then $E(f|\mathscr{G}) = f, a.s.$. (Trivial)

(2) If f and $\mathscr{G}$ are independent, then $E(f|\mathscr{G}) = Ef, a.s.$.

Let $f^* = Ef$, we can see

$$Ef\mathbf{I}_A = (Ef)P(A) = Ef^*\mathbf{I}_A,$$

(3) Let $\mathscr{G} \subset \mathscr{G}_0$,

$$E(E(f|\mathscr{G})|\mathscr{G}_0) = E(f|\mathscr{G}) = E(E(f|\mathscr{G}_0)|\mathscr{G}), a.s.$$

The left hand side is immediate by (1). The right hand side can be checked directly using definition.

(4) If $f \leq g, a.s.$ then $E(f|\mathscr{G}) \leq E(g|\mathscr{G}), a.s.$.

$$Ef^*\mathbf{I}_A = Ef\mathbf{I}_A \leq Eg\mathbf{I}_A = Eg^*\mathbf{I}_A, \quad \forall A \in \mathscr{G}.$$

(5) For all $a, b \in \mathbb{R}$, if $aEf + bEg$ exists, then

$$E(af + bg|\mathscr{G}) = aE(f|\mathscr{G}) + bE(g|\mathscr{G}).$$

This also can be checked using definition (let $h = af + bg$).

Theorem 5.6.6

Let $f_1, f_2, \ldots$ be r.v. whose integrals exist, $\mathscr{G} \subset \mathscr{F}$, then the limit theorems also holds:

- If $0 \le f_n \uparrow f, a.s.$, then
$$0 \le E(f_n|\mathscr{G}) \uparrow E(f|\mathscr{G}), a.s.;$$
- If $f_n \ge 0, a.s.$, then
$$E\left(\liminf_{n\to\infty} f_n|\mathscr{G}\right) \le \liminf_{n\to\infty} E(f_n|\mathscr{G}), a.s.;$$
- If $|f_n| \le g, a.s.$ and $g \in L_1$, $f_n \to f, a.s.$ or in measure.
$$E(f|\mathscr{G}) = \lim_{n\to\infty} E(f_n|\mathscr{G}), a.s.$$

Proof.

- Let $f_n^* = E(f_n|\mathscr{G})$, then they are a.s. increasing, let $\hat{f} = \lim_{n\to\infty} f_n^*$, then $\hat{f} \in \mathscr{G}$, and
$$E\hat{f}\mathbf{I}_A = \lim_{n\to\infty} Ef_n^*\mathbf{I}_A = Ef\mathbf{I}_A.$$
- Similarly, let
$$g_n := \inf_{m\ge n} f_m \uparrow \liminf_{n\to\infty} f_n =: f.$$
We have $g_n^* \uparrow f^*$, so
$$g_n \le f_n \implies g_n^* \le f^* \implies f^* \le \liminf_{n\to\infty} f_n^*, a.s.$$
- Lebesgue dominated theorem can be proved similarly.

□

Theorem 5.6.7

Let f, g are r.v. whose integrals exist, $g \in \mathscr{G} \subset \mathscr{F}$.
$$E(fg|\mathscr{G}) = gE(f|\mathscr{G}), a.s.$$

Proof. Fix f, we use typical method on g. When $g = \mathbf{I}_A$, $A \in \mathscr{G}$, then the conclusion holds:
$$E(f^*\mathbf{I}_A\mathbf{I}_B) = E(f^*\mathbf{I}_{AB}) = Ef\mathbf{I}_{AB} = E(f\mathbf{I}_A\mathbf{I}_B).$$
Since $AB \in \mathscr{G}$.

Now using the linearity and limit theorems we're done. Note that we need to prove on $\{f, g \ge 0\}$ and other 3 sets respectively. □

§5.7 Regular conditional distribution

Let $\{A_n\}$ be a partition of X, $\mathscr{G} = \sigma(\{A_n\})$, $P(A_n) > 0$. Thus if $B \in \mathscr{G}$ and $P(B) = 0 \implies B = \emptyset$. So the conditional expectations are uniquely determined (the only null set is the empty set).

We'll compute the conditional expectation of f under $\mathscr{G}$.
$$f^*(x) = \sum_n a_n\mathbf{I}_{A_n}(x), \quad \forall x. \quad Ef^*\mathbf{I}_{A_n} = Ef\mathbf{I}_{A_n} \implies a_n = \frac{Ef\mathbf{I}_{A_n}}{P(A_n)}.$$

Hence $\forall x \in X, A \in \mathscr{F}$,

$$p(x, A) = P(A|\mathscr{G})(x) = (\mathbf{I}_A)^*(x) = \sum_n \frac{P(A \cap A_n)}{P(A_n)} \mathbf{I}_A(x).$$

We get a function $p(x, \cdot)$, which is a probability on $\mathscr{F}$, and $p(x, \cdot) = P(\cdot|A_n)$ when $x \in A_n$.

For a fixed x,

$$(\mathbf{I}_A)^*(x) = \int_X \mathbf{I}_A(y) \, \mathrm{d}p(x, \cdot), \quad \forall A \in \mathscr{F}.$$

Now using typical method we can generalize $\mathbf{I}_A$ to any measurable function f. Since here $a.s.$ means equal at every point, so all the functions are determined.

Now what we've done is to realize conditional expectation as an integral over **conditional probabilities** $p(x, \cdot)$:

$$f^*(x) = \int_X f(y) \, \mathrm{d}p(x, \cdot) = \int_X f(y) p(x, \mathrm{d}y).$$

Next we'll generalize this observation to generic $\mathscr{G}$.

Since $(\mathbf{I}_A)^*$ is not a implicit function, we'll specify a function $p(x, A)$ for each $(\mathbf{I}_A)^*$. We want $p(x, A)$ is a probability, so we need to check countable additivity: let $A = \sum_n A_n$, we only have

$$p(x, A) = \sum_n p(x, A_n), a.s.$$

but there's uncountably many such $A_1, A_2, \ldots$, so this is the main difficulty of generalization.

Definition 5.7.1. If a function $p(x, A)$ statisfies $p(x, \cdot)$ is a probability on $\mathscr{F}$, and $p(\cdot, A) = P(A|\mathscr{G})$, then we say p is a **regular conditional probability** on $\mathscr{G}$, denoted by $P_{\mathscr{G}}(x, A)$.

Since the regular conditional probability may not exist, we need to study it on a simpler σ-algebra, say $\sigma(f)$ for some r.v. f.

$$p(x, \{f \in B\}) = \mu(x, B) \to F(x, a)$$

This means we only need to find a distribution $F(x, \cdot)$.

Definition 5.7.2. Let f be a r.v., if $F(x, a)$ statisfies $F(x, \cdot)$ is a distribution, and $F(\cdot, a) = P(f \le a|\mathscr{G}), a.s.$, we call it the **regular conditional distribution function** of f with respect to $\mathscr{G}$, denoted by $F_{f|\mathscr{G}}(\cdot, \cdot)$.

Theorem 5.7.3

Let f be a r.v., then the regular conditional distribution function always exists.

Proof. For all $r \in \mathbb{Q}$, we can take a r.v. $G(\cdot, r)$ s.t.

$$G(\cdot, r) = P(f \le r|\mathscr{G}), a.s.$$

We get a function $G(\cdot, \cdot)$ on $X \times \mathbb{Q}$.

Recall that distribution statisfies: monotonicity, right continuity and normality (range is $[0, 1]$).

Let N_1, N_2, N_3 be subsets of X where the above condition doesn't hold, respectively. Let $N = N_1 \cup N_2 \cup N_3$.

For fixed r_1, r_2, the set $A_{r_1, r_2} := \{x \mid G(x, r_1) > G(x, r_2)\}$ is null because of the properties conditional expectation. Thus $N_1 = \bigcup_{r_1, r_2 \in \mathbb{Q}} A_{r_1, r_2}$ is null.

By similar techniques, we can prove N_2, N_3 are null as well. (Note that here we can consider them in N_1^c, which means $G(x,\cdot)$ is increasing)

Hence $P(N) = 0$, let

$$F(x,a) = \inf\{G(x,r) : r \in \mathbb{Q}, r > a\}.$$

Then $F(x,\cdot)$ is right continuous on $X \backslash N \times \mathbb{R}$. In fact we can also check the other two requirements, so F is indeed a regular conditional d.f..

For $\forall a \in \mathbb{R}$, let

$$F_{f|\mathscr{G}}(x,a) := \begin{cases} F(x,a), & x \notin N; \\ H(a), & x \in N. \end{cases}$$

where $H(a)$ is an arbitary distribution function. We've already proved that $F_{f|\mathscr{G}}(x,\cdot)$ is a d.f.; For fixed a, by Levi's theorem,

$$F_{f|\mathscr{G}} = \lim_{r\in\mathbb{Q}, r\to a^+} G(\cdot, r) = \lim_{r\in\mathbb{Q}, r\to a^+} P(f \le r|\mathscr{G}) = P(f \le a|\mathscr{G}), a.s.$$

So $F_{f|\mathscr{G}}$ is the desired regular conditional d.f.. □

Similarly we can define a **regular conditional distribution** $\mu(x,B)$ for a r.v. f.

Theorem 5.7.4

Let h be a function,

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

In particular, $f^*(x) = \int_{\mathbb{R}} a\mu(x, \mathrm{d}a)$.

Let $g : (X, \mathscr{F}) \to (Y, \mathscr{S})$ be a measurable map, $\mathscr{G} = \sigma(g)$. Then $f^* \in \mathscr{G} \iff f^* = \varphi(g), a.s.$, where $\varphi : (Y, \mathscr{S}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.

Definition 5.7.5. We say $\varphi(\cdot)$ is the conditional expectation of f under a **given value** of g, denoted by $E(f|g = \cdot)$. It's a real-valued function on Y.

Definition 5.7.6. If a function $\nu(y,B)$ statisfies: $\nu(y,\cdot)$ is a distribution on $\mathscr{B}_{\mathbb{R}}$, and $\nu(y,B) = P(f \in B|g = y), a.s.$ in $\mathscr{L}(g)$ (the measure on Y induced by g), then we call it the regular conditional distribution of f under **given value** of g, we denote this by $\mu_{f|g}(y,B)$.

Corollary 5.7.7

$\nu(y,B)$ exists, and

$$E(h(f)|g = y) = \int_{\mathbb{R}} h(a)\mu(y, \mathrm{d}a), \mathscr{L}(g)\text{-}a.s.$$

Example 5.7.8

Consider a continuous random vector on $\mathbb{R}^2$. Let λ_2 be the Lebesgue measure on $\mathbb{R}^2$.

Recall that (f,g) is continuous iff there exists $p(x,y)$ s.t.

$$P((f,g)\in B)=\iint_B p(x,y)\,\mathrm{d}\lambda_2, \forall B\in\mathscr{B}_2.$$

Let $p_g(y)=\int_{\mathbb{R}} p(x,y)\lambda(\mathrm{d}x)$, in probability we learned

$$p_{f|g}(x|y)=\begin{cases}\dfrac{p(x,y)}{p_g(y)}, & p_g(y)>0;\\ 0, & p_g(y)=0.\end{cases}$$

By our corollary we get $\mu_{f|g}(y,B)=\int_B p_{f|g}(x|y)\lambda(\mathrm{d}x)$.

§6 Product spaces

§6.1 Finite dimensional product spaces (skipped)

This section is almost covered in real variable functions.

Let $X_1,\dots,X_n$ be original spaces, $X=\prod_{k=1}^n X_k$. We're going to build measurable structure on X.

Let

$$\mathscr{Q}:=\{\prod_{k=1}^n A_k : A_k\in\mathscr{F}_k, k=1,\dots,n\}$$

denote the measurable rectangles, we can check $\mathscr{Q}$ is a semi-ring, and $X\in\mathscr{Q}$. Let

$$\mathscr{F}=\prod_{k=1}^n \mathscr{F}_k:=\sigma(\mathscr{Q})$$

be the **product σ-algebra**.

Let π_k be the projection map onto the k-th component, we have

Proposition 6.1.1

For each k, π_k is a measurable map $(X,\mathscr{F})\to(X_k,\mathscr{F}_k)$, and

$$\mathscr{F}=\sigma\left(\bigcup_{k=1}^n \pi_k^{-1}\mathscr{F}_k\right).$$

Theorem 6.1.2

Let $f=(f_1,\dots,f_n):\Omega\to X$, then $f:(\Omega,\mathscr{S})\to(X,\mathscr{F})$ measurable iff each f_k is measurable.

A **section** is to fix some components of a subset of X.

Definition 6.1.3. A function $p(x_1, A_2)$ is called a **transform function** from X_1 to X_2 if $p(x_1, \cdot)$ is a measure on $\mathscr{F}_2$, and $p(\cdot, A_2)$ is measurable in $\mathscr{F}_1$.

If $X_2 = \sum_n A_n$ and $p(x, A_n) < \infty$ for all n and x, then we say $p(\cdot, \cdot)$ is σ-finite. Note that this partition is independent of x. If each $p(x, \cdot)$ is a probabilty, we say p is a **probabilty transform function**.

Let $X = X_1 \times X_2, \hat{X} = X_2 \times X_1, \mathscr{F} = \mathscr{F}_1 \times \mathscr{F}_2$.

Theorem 6.1.4

Let $p(x_1, A_2)$ be a σ-finite transform function from X_1 to X_2.

- For all σ-finite measure μ_1 on X_1, $\exists!$ measure μ s.t.

$$\mu(A_1 \times A_2) = \int_{A_1} p(x_1, A_2)\mu_1(\mathrm{d}x_1),$$

- If $f : X \to \mathbb{R}$'s integral exists, then

$$\int_X f \,\mathrm{d}\mu = \int_{X_1} \mu_1(\mathrm{d}x_1) \int_{X_2} f(x_1, x_2)p(x_1, \mathrm{d}x_2).$$

Proof. See proof of Fubini's theorem in analysis. □

Hence given a measure on X_1 and a transform function, we can get a measure on the product space.

If we start from the conditional probabilty, let $g(x) = x_1, f(x) = x_2$, we have

$$E(h_2(x_2)|x_1) = \varphi(x_1), \quad \phi(x_1) = \int_{X_2} h_2(x_2)\nu(x_1, \mathrm{d}x_2).$$

Multiplying a function of x_1, (i.e. $h_1(x_1)$) taking the integral we get

$$E(h_1(x_1)h_2(x_2)) = \int_{X_1} \mu_1(x_1) \int_{X_2} h_1(x_1)h_2(x_2)\nu(x_1, \mathrm{d}x_2).$$

Thus by typical method we can generalize $h_1(x_1)h_2(x_2)$ to any function $f(x_1, x_2)$. Hence the transform function p is nothing but the regular conditional probabilty.

Corollary 6.1.5 (Fubini's theorem)

If $p(x_1, \cdot) \equiv \mu_2$, denote μ as $\mu_1 \times \mu_2$, if the integral of f exists,

$$\int_X f \,\mathrm{d}\mu_1 \times \mu_2 = \int_{X_1} \mu_1(\mathrm{d}x_1) \int_{X_2} f(x_1, x_2)\mu_2(\mathrm{d}x_2) = \int_{X_2} \mu_2(\mathrm{d}x_2) \int_{X_1} f(x_1, x_2)\mu_1(\mathrm{d}x_1).$$

Remark 6.1.6 — The integral of f exists means that the integral of f exists in the product space, i.e. the LHS must exist. It's not true we only have the RHS exists.

Example 6.1.7

Let $X_1 = X_2 = \mathbb{R}$, we use the Lebesgue measure λ. Let $f(x,y) = \mathbf{I}_{\{0<y\le 2\}} - \mathbf{I}_{\{-1<y\le 0\}}$.

It's easy to see the integral of f doesn't exist, but $\iint f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \infty$, while $\iint f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ does not exist.

By induction we can reach product space of finitely many spaces:

Theorem 6.1.8

Let p_k be the transform function from $\prod_{i=1}^{k-1} X_i$ to X_k, for any σ-finite measure μ_1 on X_1, $\exists!$ measure μ, such that ...TODO

§6.2 Countable dimensional product space

Again let π_n be the projection onto X_n, and $\pi_{(n)}$ be the projection onto $X_{(n)} := \prod_{i=1}^n X_i$.

Let $\mathscr{F}_{(n)} := \prod_{i=1}^n \mathscr{F}_i = \sigma(\mathscr{Q}_{(n)})$, and define

$$\mathscr{Q}_{[n]} = \left\{ \prod_{k=1}^n A_k \times \prod_{k=n+1}^\infty X_k \mid A_k \in X_k \right\} = \pi_{(n)}^{-1} \mathscr{Q}_{(n)}.$$

Proposition 6.2.1

$\mathscr{Q} = \bigcup_{n=1}^\infty \mathscr{Q}_{[n]}$ is a semi-ring, and $X \in \mathscr{Q}$. Similarly, $\mathscr{A} = \bigcup_{n=1}^\infty \mathscr{F}_{[n]}$ is an algebra.

Theorem 6.2.2 (Tulcea)

Let p_k be probabilty transform functions $\prod_{i=1}^{k-1} X_i \to X_k$, then for all probabilty P_1 on X_1, there exists unique probabilty P on $\prod_{k=1}^\infty X_k$ s.t.

$$P\left(\prod_{k=1}^n A_k \times \prod_{k=n+1}^\infty X_k\right) = \int_{A_1} P_1(\mathrm{d}x_1) \int_{A_2} p_2(x_1, \mathrm{d}x_2) \cdots \int_{A_n} p_n(x_1, \ldots, x_{n-1}, \mathrm{d}x_n).$$

Proof. By results in previous section, we can define P_n on $\mathscr{F}_{[n]}$.

Since $P_{n+1}\big|_{\mathscr{F}_{[n]}} = P_n$, we can get a function P on the algebra $\mathscr{A} = \bigcup_{n=1}^\infty \mathscr{F}_{[n]}$. (By transfinite induction)

At last we'll prove P is a measure on $\mathscr{A}$, thus it can be uniquely extended to $\mathscr{F} = \sigma(\mathscr{A})$.

Claim 6.2.3. $P_n = P_{n+1}\big|_{\mathscr{F}_{[n]}}$.

Proof. Some abstract nonsense. Just note that $A_{(n+1)} = A_{(n)} \times X_{n+1}$ for $A \in \mathscr{F}_{(n)}$, and just compute the $(n+1)$-th integral to get the equality. $\square$

Claim 6.2.4. P is countablely additive on $\mathscr{A}$.

Proof. It's easy to see that P has finite additivity, so it suffices to prove P is continuous at empty set.

Let $A_1, A_2, \dots \in \mathscr{A}$, $A_n \downarrow \emptyset$, if $P(A_n) \not\to 0$, let $\varepsilon :=\downarrow \lim_{n\to\infty} P(A_n) > 0$.

There exist $1 \leq m_1 < m_2 < \cdots$ s.t. $A_n \in \mathscr{F}_{[m_n]}$. WLOG $m_n = n$ (otherwise add more sets in the sequence, i.e. $B_k = A_n$ when $m_n \leq k < m_{n+1}$).

Therefore we have $A_{(n)} = \pi_{(n)}^{-1} A_{(n)}$,

$$A_n = \pi_{(n+1)}^{-1}(A_{(n)} \times X_{n+1}) \supseteq A_{n+1} \implies A_{(n)} \times X_{n+1} \supseteq A_{(n+1)}.$$

Equivalently,

$$\mathbf{I}_{A_{(n+1)}}(x_1, \dots, x_{n+1}) \leq \mathbf{I}_{A_{(n)}}(x_1, \dots, x_n).$$

Therefore, we have $0 \leq \phi_{1,n+1}(x_1) \leq \phi_{1,n}(x_1) \leq 1$, where

$$\phi_{1,n}(x_1) := \int_{X_2} p_2(x_1, \mathrm{d}x_2) \cdots \int_{X_n} \mathbf{I}_{A_{(n)}}(x_1, \dots, x_n) p_n(x_1, \dots, x_{n-1}, \mathrm{d}x_n).$$

Note that $P(A_{[n]}) = P_n(A_{[n]}) = \int_{X_1} \phi_{1,n} P_1(\mathrm{d}x_1)$.

Let $\phi_1 :=\downarrow \lim_{n\to\infty} \phi_{1,n}$, by dominated convergence theorem,

$$\int_{X_1} \phi_1 \, \mathrm{d}P_1 =\downarrow \lim_{n\to\infty} \int_{X_1} \phi_{1,n} \, \mathrm{d}P_1 = \varepsilon > 0.$$

Hence $\exists \tilde{x}_1 \in X_1$ s.t. $\phi_1(\tilde{x}_1) > 0$. We must have $\tilde{x}_1 \in A_{(1)}$, otherwise

$$\mathbf{I}_{A_{(n)}}(\tilde{x}_1, x_2, \dots, x_n) \leq \mathbf{I}_{A_{(1)}}(\tilde{x}_1) = 0,$$

which gives $\phi_{1,n}(\tilde{x}_1) = 0$, $\forall n$, contradiction!

By the same process we can take $\phi_2(x_2) =\downarrow \lim_{n\to\infty} \phi_{2,n}(x_2)$, where $\phi_{2,n}(x_2)$ is defined as

$$\int_{X_3} p_3(\tilde{x}_1, x_2, \mathrm{d}x_3) \cdots \int_{X_n} \mathbf{I}_{A_{(n)}}(\tilde{x}_1, x_2, \dots, x_n) p_n(\tilde{x}_1, x_2, \dots, x_{n-1}, \mathrm{d}x_n).$$

We'll get $\tilde{x}_2$ s.t. $(\tilde{x}_1, \tilde{x}_2) \in A_{(2)}$, and $\phi_2(\tilde{x}_2) > 0$.

By induction we get $(\tilde{x}_1, \tilde{x}_2, \dots) \in \bigcap_{n=1}^{\infty} A_{[n]}$, which contradicts with $A_n \downarrow \emptyset$! □

Hence the conclusion holds. □

Theorem 6.2.5 (Kolmogorov)

Let P_k be a probability on $(X_k, \mathscr{F}_k)$, then there exists a unique measure P on $(\prod X_k, \prod \mathscr{F}_k)$, such that

$$P\left(\prod_{k=1}^{n} A_k \times \prod_{k=n+1}^{\infty} X_k\right) = \prod_{k=1}^{n} P_k(A_k).$$

Proof. This is immediate by Tulcea's theorem. □

Let's make a summary of Tulcea's theorem. To get a measure on $\mathscr{F}$, we need:

- Measures P_n on $\mathscr{F}_{[n]}$, which is induced by measures on $\mathscr{F}_{(n)}$.
- Compatibility, i.e. $P_{n+1}|_{\mathscr{F}_{[n]}} = P_n$. Hence we'll get a function P on the algebra $\bigcup \mathscr{F}_{[n]}$.
- At last to prove P is a measure, we need the continuity at $\emptyset$.

Tulcea's theorem tells us that the measure induced by the probability transform functions statisfies above conditions.

§6.3 Arbitary infinite dimensional product space

Let $\{X_t, t \in T\}$ be a collection of sets, where T is uncountable. Let $X = \prod_{t \in T} X_t$ be the product space.

Let $U \subset S \subset T$, where $|S| < \infty$, define the projection

$$\pi_S : X \to X_S := \prod_{t \in S} X_t, \quad \pi_{S \to U} : X_S \to X_U, \quad \pi_{S \to U} \circ \pi_S = \pi_U.$$

Similarly, we can define the cylinder set:

$$\mathscr{Q}_S = \left\{ \pi_S^{-1} \left(\prod_{t \in S} A_t \right) : A_t \in \mathscr{F}_t, \forall t \in S \right\}; \quad \mathscr{F}_S = \sigma(\mathscr{Q}_S).$$

Proposition 6.3.1

We have $\mathscr{Q}_S$, $\mathscr{Q} := \bigcup_{|S| < \infty} \mathscr{Q}_S$ are semi-rings containing X.

Proposition 6.3.2

$\mathscr{A} := \bigcup_{|S| < \infty} \mathscr{F}_S$ is an algebra containing $\mathscr{Q}$.

Proposition 6.3.3

Let $\mathscr{F} := \sigma(\mathscr{Q}) = \sigma(\mathscr{A})$, we have

$$\mathscr{F} = \sigma(\{\pi_t, t \in T\}) = \{\pi_S^{-1} A : A \in \mathscr{F}_S, |S| \le \omega\}.$$

Remark 6.3.4 — To prove the equality, first note $LHS = \sigma(\bigcup_{t \in T} \pi_t^{-1} \mathscr{F}_t)$, and RHS is a σ-algebra.

In random process, $(\Omega, \mathscr{S})$ is the sample space, the index set T is regarded as time, for each time $t \in T$, there's a random variable $f_t : \Omega \to X_t$. Thus $f := \{f_t, t \in T\}$ is a map $\Omega \to \prod_{t \in T} X_t$.

Theorem 6.3.5

Let $\mathscr{F} = \prod_{t \in T} \mathscr{F}_t$,

$$f : (\Omega, \mathscr{S}) \to (X, \mathscr{F}) \iff f_t : (\Omega, \mathscr{S}) \to (X_t, \mathscr{F}_t), \forall t.$$

If $(X_t, \mathscr{F}_t) \equiv (S, \mathscr{S}_0)$, then we say f is a random process; S is said to be the range space, and $f(\omega) = \{f_t(\omega) : t \in T\} \in S^T$ is an orbit.

For any probability Q on $(\Omega, \mathscr{S})$, $Q \circ f^{-1}$ is the distribution of f, by previous proposition, we only need all the countably dimensional joint distribution of f.

From Tulcea's theorem, we only need to study *finite dimensional joint distribution* $P_{t_1, \ldots, t_n}$ where $t_1, \ldots, t_n \in T$.

Similarly we require the probability to have some compatibility:

- Let $t(1), \dots, t(n)$ be a permutation of $t_1, \dots, t_n$. We require

$$P_{t_1,\dots,t_n}\left(\prod_{i=1}^n A_{t_i}\right) = P_{t(1),\dots,t(n)}\left(\prod_{i=1}^n A_{t(i)}\right).$$

- Let $t_{n+1} \in T$,

$$P_{t_1,\dots,t_{n+1}}\left(\prod_{i=1}^n A_{t_i} \times X_{t_{n+1}}\right) = P_{t_1,\dots,t_n}\left(\prod_{i=1}^n A_{t_i}\right).$$

Theorem 6.3.6 (Kolmogorov)
If $\mathbf{P}$ is compatible, then $\exists! P$ on $(\mathbb{R}^T, \mathscr{B}^T)$ s.t.

$$P(\pi_S^{-1} A) = P_S(A), \quad \forall |S| < \infty, A \in \mathscr{B}^S.$$

Sketch of the proof. Let $\mathscr{F}_0 = \{\pi_{T_0}^{-1}(A) : A \in \mathscr{F}_{T_0}, |T_0| \leq \omega\}$.

Step 1, fix a countable $T_0 \subset T$, by Tulcea's theorem, we can define $P(\pi_{T_0}^{-1} A) = P_{T_0}(A)$.

Step 2, P is well-defined in different permutations of T_0.

Step 3, if T_1, T_2 countable, and $\pi_{T_1}^{-1}(A_1) = \pi_{T_2}^{-1}(A_2)$, we have $P_{T_1}(A_1) = P_{T_2}(A_2)$. This can be done by looking at $T_0 = T_1 \cup T_2$.

Step 4, check P statisfies countable additivity. $\square$

Example 6.3.7 (Brownian motion)
Let $\mathbf{B} = \{B_t, t \in T\}$, $T = \mathbb{R}_+$. Let $(\Omega, \mathscr{S}, \hat{P})$ be the sample space, $(\mathbb{R}^T, \mathscr{B}^T)$ be the orbit space, where $\varphi : T \to \mathbb{R}$ is an orbit.

$$\mathbf{B}(\omega) := \varphi : t \mapsto \varphi(t) = B_t(\omega).$$

Initially, let $B_0 = 0$, define the transformation density

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right).$$

Starting from finite dimensional orbit distribution, we can get countable dimensional orbit distribution.

TODO

§7 Review

- λ-π theorem, monotone class theorem, typical method.
- σ-finite measures on semi-ring can be uniquely extended to the σ-algebra. The uniqueness only requires π-system.
- Different convergence of functions: a.u., a.e./a.s., μ, L^p, $(w)L^p$.
- The construction of integrals, check if the integral exists.

- The linearity, monotonicity of integrals and their conditions.
- The three limit theorems and their proofs, the countable additivity of indefinite integrals.
- Substitution formula of integrals, expectations of random variables.
- L^p space is a Banach space, Holder's & Minkowski's inequality.
- Equivalent conditions of L^p convergence and weak L^p convergence.
- In probability space, the inckusion relations of L^p's, definition and equivalent conditions of uniformly integrable.
- Definitions of signed measures, absolute continuity and mutual singularity.
- Hahn decomposition, Jordan decomposition, the "maximum" sets.
- R-N derivatives $\frac{\mathrm{d}\varphi}{\mathrm{d}\mu}$, the "maximum" function.
- Absolutely continuous signed measures = indefinite integrals, R-N derivatives = functions being integrated.
- Lebesgue decomposition, taking R-N derivatives with respect to $\varphi + \mu$.
- Conditional expectations $f^* = E(f|\mathscr{G})$, where the integral of f exists.
- Two properties: (1) $f^* \in \mathscr{G}$, (2) $Ef^*\mathbf{I}_A = Ef\mathbf{I}_A, \forall A \in \mathscr{G}$.
- Linearity of conditional expectations, $f \in \mathscr{G}$ vs. f and $\mathscr{G}$ are independent.
- The use of λ- π theorem and typical method!
- Multiple expectation formula, limit theorems of conditional expectations.
- Regular conditional distribution of given values, $\nu(x, B) = \nu_x(B) = P(\eta \in B|\xi = x)$,

$$E(h(\eta)|\xi) = \psi(\xi), \quad \psi(x) = \int_{\mathbb{R}} h(y)\,\mathrm{d}\nu_x(y).$$

- Transformation functions in product space, $\mathrm{d}\mu(x_1, x_2) = \mu_1(\mathrm{d}x_1)p(x_1, \mathrm{d}x_2)$.
- Transformation functions = regular conditional distribution of given values.
- Fubini's theorem, computations (pay attention to the condition).
- Tulcea's theorem, the statement and applications.