SIMPLE AND INSTRUCTIVE

MEASURE THEORY MADE EASY

Clearly explained motivations
Image explanations
less abstract proofs
problem solving techniques

EDITED BY DANIEL LIN

Contents

1 Introduction

This note aims to provide a motivated, easier-to-approach way of learning measure theory, which is written in line with our lecture notes. And I used the books *Introduction to measure and integration* by S.J. Taylor, *An introduction to measure theory* by Terence Tao and lecture notes by P.-F. Rodriguez. You may read this note before reading lecture notes to gain more insights. This note has adjusted the order of some topics for logical coherence, so you may have to change reading order.
For some proofs, I refer to the lecture notes, either because the result is easy or the proof in lecture note is good enough to read easily. For the proofs in lecture notes that are more abstract, an easier proof is provided here. Some results are coming from problem sheets, and are marked, for example, (PS2). Just to extend your insights and proof techniques, some extra contents are included, marked with (extra). You may skip the extra contents.

Before exploring the wonderful world of how the strict concept of "distance" is established, let's first see how to hand write some Mathcal letters.

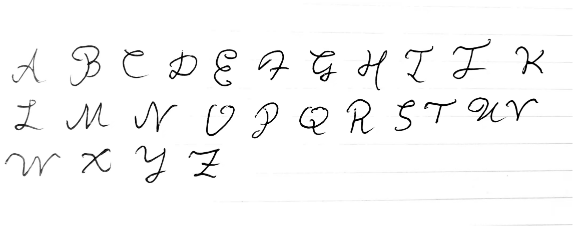

Figure 1: Mathcal handwritten font

These letters are frequently used in set theory and any related subjects to represent collection of sets.

2 Set Theory and Abstract Measure

2.1 Semi-algebra, Algebra, σ-algebra

When measuring the real line, we often consider sets of the form $(a, b]$ (because they can easily cover sets without intersecting each other). And we aim to construct a function $\mu : \mathbb{R} \to [0, \infty]$ (we include ∞ as a measure value) s.t. (1) $\mu((a, b]) = b - a$, (2) μ is additive and (3) $\mu(A + x) = \mu(A) \forall x \in \mathbb{R}$ (This means translation by x will not change the measure). However there are many examples of subsets of $\mathbb{R}$ where the conditions above cannot be all satisfied(e.g. Vitali set given in proposition 1.25 of notes week4.1). So we will use $(a, b]$ to build a collection of sets where such a function μ satisfying all three properties can be defined.

Consider the collection $\mathcal{S} = \{\mathbb{R}, \emptyset, (a, b], (-\infty, b], (a, \infty) \mid a < b, a, b \in \mathbb{R}\}$. We can easily define a measure on it. Observe key properties of this set:

- $\mathbb{R} \in \mathcal{S}$

- If $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$
- $\mathbb{R} \setminus (a, b] = (-\infty, a] \cup (b, \infty)$, $\mathbb{R} \setminus (-\infty, b] = (b, \infty)$, $\mathbb{R} \setminus (a, \infty) = (-\infty, a]$ all of which are union of disjoint sets in $\mathcal{S}$

So we generalise these properties to define Semi-algebra:

Definition 2.1 (Semi-algebra)**.** Given a set X, define $\mathcal{S} \subseteq 2^X = \mathcal{P}(X)$ to be semi-algebra if
(1) $X \in \mathcal{S}$
(2) If $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$
(3) $\forall A \in \mathcal{S}, \exists$ disjoint $E_1, ..., E_n \in \mathcal{S}$ s.t. $A^C = \sum_{i=1}^n E_i$

Remark. We used $\sum_{i=1}^n E_i := \bigcup_{i=1}^n E_i$ to emphasize that E_i are disjoint, we will NOT use Σ for non-disjoint sets.

But there can be more measurable sets, for example (a, b) is not included.

Definition 2.2 (Algebra)**.** Given a set X, define $\mathcal{A} \subseteq 2^X$ to be algebra if
(1) $X \in \mathcal{A}$
(2) If $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$
(3) $\forall A \in \mathcal{A}, A^C \in \mathcal{A}$
$\mathcal{A}$ is $\sigma-$algebra if (2) is replaced by the following:
(4) If given countable set J, $\forall j \in J, A_j \in \mathcal{A}$ then $\cap_{j \in J} A_j \in \mathcal{A}$

Remark. It can be shown that this definition is equivalent to replacing (4) by σ-additivity: If $A_j \in \mathcal{A}$ is a sequence of disjoint sets, then $\sum_j A_j \in \mathcal{A}$. (Use De Morgan's law)

The best example of σ-algebra can be found in probability theory. Say we are tossing a coin twice with possible outcomes H - head and T - tail. So the sample space $X = \{HT, TH, HH, TT\}$. If we only know whether two results are the same or not, a σ-algebra consisting of possible events is $\mathcal{F} = \{\emptyset, \{HT, TH\}, \{HH, TT\}, X\}$. This means we cannot distinguish between HT, TH nor HH, TT with limited knowledge.

.

Remark. 2 Every $\sigma-$algebra algebra is algebra (because countable additivity implies finite additivity by taking the tail to be a sequence of $\emptyset$), every algebra is semi-algebra. But not every algebra is σ-algebra. For example take $X = \mathbb{N}$ and $\mathcal{A} = \{A \subset \mathbb{N} : A \text{ is finite or } A^C = \mathbb{N} \setminus A \text{ is finite.}\}$. This is algebra but not σ-algebra. (Prove it!)

Proposition 2.3. *Intersection(of any number) of $\sigma-$algebra is a $\sigma-$algebra. But union of $\sigma-$algebra may not be $\sigma-$algebra.*

Proof. First sentence is left as exercise. Now we investigate an example of union not being $\sigma-$algebra. Take $X = \{1, 2, 3, 4\}$ and $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, 3, 4\}, X\}$, $\mathcal{F}_2 = \{\emptyset, \{1, 2\}, \{3, 4\}, X\}$. Then the set $\{1, 3, 4\} = \{1\} \cup \{3, 4\} \in \mathcal{F}_1 \cup \mathcal{F}_2$ is not in $\mathcal{F}_1 \cup \mathcal{F}_2$. This is a finite union, so we see even union of algebras is not algebra. ■

.

Given a collection $\mathcal{C} \subseteq 2^X$, we can generate an algebra/$\sigma-$algebra.

Definition 2.4 (Algebra generated by $\mathcal{C}$)**.** $\mathcal{A}(\mathcal{C})$ is a set s.t.
(1) $\mathcal{C} \subseteq \mathcal{A}(\mathcal{C})$
(2) If $\mathcal{B}$ is algebra s.t. $\mathcal{C} \subseteq \mathcal{B}$, then $\mathcal{A}(\mathcal{C}) \subseteq \mathcal{B}$. (i.e. $\mathcal{A}(\mathcal{C})$ is the smallest algebra containing $\mathcal{C}$)

We can define $\sigma-$algebra generated by $\mathcal{C}$, denoted as $\sigma(\mathcal{C})$ using exactly the same definition by replacing "algebra" by "$\sigma-$algebra" . It can be easily shown that this definition is equivalent to upside down approach used in our notes:

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ is a } \sigma-\text{algebra with } \mathcal{C} \subseteq \mathcal{A}} \mathcal{A} \subset 2^X$$

Properties of $\sigma-$algebra generated by $\mathcal{C}$:

1. $\mathcal{C} \subset \sigma(\mathcal{C})$
2. If $\mathcal{C} \subset \mathcal{C}'$, then $\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}')$
3. $\mathcal{C}$ is $\sigma-$algebra $\Leftrightarrow \mathcal{C} = \sigma(\mathcal{C})$
4. $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$
5. $\forall A \in \sigma(\mathcal{C})$, exists a countable sub-collection $\mathcal{D} \subset \mathcal{C}$ s.t. $A \in \sigma(\mathcal{D})$ (PS1)

The following lemma is not in our notes but the insight of proof is worth learning.

Lemma 2.5. *Given a set X and semi-algebra $\mathcal{S}$,*

$$A \in \mathcal{A}(\mathcal{S}) \Leftrightarrow \exists \textit{ disjoint } E_1, ..., E_n \in \mathcal{S} \textit{ s.t. } A = \sum_{j=1}^{n} E_j$$

Proof. Backward arrow is clear. Now assume $A \in \mathcal{A}(\mathcal{S})$ and define $\mathcal{B} := \{\sum_{j=1}^n E_j : E_j \in \mathcal{S} \text{ and are disjoint.}\}$. We aim to prove that $\mathcal{B}$ is algebra, then by $\mathcal{S} \subset \mathcal{B}$ and definition of $\mathcal{A}(\mathcal{S})$, we have $\mathcal{A}(\mathcal{S}) \subset \mathcal{B}$, proving the lemma.
(1) Clearly $X \in \mathcal{B}$
(2) Given $\alpha, \beta \in \mathcal{B}$, we can write $\alpha = \sum_{j=1}^n E_j, \beta = \sum_{k=1}^m F_k$, then

$$\alpha \cap \beta = \sum_{j=1}^{n} \sum_{k=1}^{m} (E_j \cap F_k)$$

. Since E_j are disjoint and F_k are disjoint, $E_j \cap F_k$ are disjoint, so writing $\sum$ is valid and $\alpha \cap \beta \in \mathcal{B}$
(3) For $A = \sum_{j=1}^n E_j \in \mathcal{B}$,

$$A^C = \left(\sum_{j=1}^{n} E_j \right)^C = \bigcap_{j=1}^{n} E_j^C = \bigcap_{j=1}^{n} \left(\sum_{k_j=1}^{l_j} F_{j,k_j} \right)$$

Second equality is De Morgan's law. $F_{j,k_j} \in \mathcal{S}$, and are taken according to (3) of definition of semi-algebra. Now we exchange the order of union and intersection:

$$A^C = \sum_{k_1=1}^{l_1} \sum_{k_2=1}^{l_2} \dots \sum_{k_n=1}^{l_n} (F_{1,k_1} \cap F_{2,k_2} \cap \dots \cap F_{n,k_n}) \in \mathcal{B}$$

So $\mathcal{B}$ is algebra, proof is complete. ∎

2.2 Set Function

A set function is a function whose domain is a collection of sets, and codomain is $\mathbb{R}$. In this course we mostly study set functions with codomain $[0, \infty]$, as intuitively, a "measure" of distance/area/volume should be non-negative.

Definition 2.6. Given $\mu : \mathcal{C} \to [0, \infty]$.

- If for all disjoint $E, F \in \mathcal{C}$, $\mu(E \cup F) = \mu(E) + \mu(F)$, then μ is called *additive*. This generalise to sum of n sets. μ is σ-additive if this hold for every sequence of disjoint sets E_n (countable union).
- If for all $E, F \in \mathcal{C}$ with $E \subseteq F$, $\mu(E) \leq \mu(F)$, we say μ is *monotone.*
- If $\forall E, F \in \mathcal{C}$ satisfying $E \subset F, F \setminus E \in \mathcal{C}$ and $\mu(E) < \infty$, we have $\mu(F \setminus E) = \mu(F) - \mu(E)$, then μ is called *subtractive.*

Proposition 2.7. *For any additive function $\mu : \mathcal{C} \to [0, \infty]$, μ is monotone and subtractive*

Proof is left as exercise.

We define other properties of an (finitely)additive set function μ, but they may not hold for every additive set function.

Definition 2.8 (Continuity)**.** Given $E \in \mathcal{C} \subseteq 2^X$, μ is continuous from below at E if
$\forall (E_n)_{n\geq 1} \subset \mathcal{C}$ with $E_n \uparrow E$, we have $\mu(E_n) \to \mu(E)$.
μ is continuous from above at E if
$\forall (E_n)_{n\geq 1} \subset \mathcal{C}$ s.t. $E_n \downarrow E$ and $\mu(E_N) < \infty$ for some N (note: this implies $\mu(E_n) < \infty \forall n \geq N$), then $\mu(E_n) \to \mu(E)$.

Remark. The condition $\mu(E_N) < \infty$ for some N is necessary, take a measure (though we have not defined a measure, you should be able to understand) μ on set of half-open intervals s.t. $\mu((a, b]) = b - a$, $\mu((-\infty, b]) = \mu((a, +\infty)) = +\infty$. We definitely wish this measure to be continuous, but take $E_n = (n, \infty)$, $\forall n, \mu(E_n) = \infty$ but $E_n \downarrow E = \emptyset$ which has measure 0. So $\mu(E_n) \not\to \mu(E)$.

Definition 2.9 (σ-subadditivity)**.** If $A_k \in \mathcal{A}$, $A \subset \bigcup_{k=1}^{\infty} A_k$, then $\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k)$

Proposition 2.10. *Every σ-additive set function μ is also σ-subadditive.*

Proof: easy exercise.

Proposition 2.11. *Given algebra $\mathcal{A}$ with (finitely)additive function $\mu : \mathcal{A} \to [0, \infty]$,*
(1) μ is σ-additive $\Rightarrow$ μ is continuous from above and below at every $E \in \mathcal{A}$
(2) μ is continuous from below $\Rightarrow$ μ is σ-additive.
(3) μ is continuous from above and μ is finite ($\forall E \in \mathcal{A}, \mu(E) < \infty$) $\Rightarrow$ μ is σ-additive.

Proof: exercise.

Remark. You may find the trick useful in proving above two propositions: Given any sequence of sets $\{F_i\}$, exists pairwise disjoint sequence $\{G_i\}$ with $G_i \subseteq F_i$ and $\bigcup_i G_i = \bigcup_i F_i$. This is done by taking $G_1 = F_1, G_2 = F_2 \setminus F_1, G_3 = F_3 \setminus (F_1 \cup F_2), \ldots$.

Now finally we can define a measure:

Definition 2.12 (Measurable space and measure)**.** $(X, \mathcal{A})$ is measurable space if $\mathcal{A}$ is a σ-algebra over X. And a measure is a function $\mu : \mathcal{A} \to [0, \infty]$ s.t.
(1) $\mu(\emptyset) = 0$
(2) μ is σ-additive. (i.e. countably additive) With a measure defined, $(X, \mathcal{A}, \mu)$ is called a measure space.

So a measure is monotone, subtractive, finitely additive, continuous from above and below, σ-subadditive by the above propositions.

Probability measure is indeed a measure, and it satisfies $\mu(X) = 1$). Lebesgue measure is also an example, it will be introduced later but it is basically what you would expect a measure to be on $\mathbb{R}^n$. For example it assigns $b - a$ to interval $[a, b)$ and assigns 1 to unit square in $\mathbb{R}^2$.

3 Construction of Measure

3.1 Construction of measure

We have seen that a "measure"(actually a pre-pre-measure) can be easily defined on the semi-algebra over $\mathbb{R}$ we stated in the very beginning. This section studies the extension of a pre-pre-measure defined on semi-algebra to a proper measure defined on a σ-algebra containing the semi-algebra.

Road map:
(1) From a pre-pre-measure υ defined on semi-algebra $\mathcal{S}$, we extend it to an pre-measure $\widetilde{\mu}$ over $\mathcal{A}(\mathcal{S})$
(2) Extend again to outer measure μ^* defined over 2^X.
(3) Find a σ-algebra Σ containing $\mathcal{A}(\mathcal{S})$ and let $\mu = \mu^*|_\Sigma$, this is the required extension.

Proposition 3.1. *Given a semi-algebra $\mathcal{S} \subseteq 2^X$ and additive $\upsilon : \mathcal{S} \to [0, \infty]$ with $\upsilon(\emptyset) = 0$. There is a unique $\widetilde{\mu} : \mathcal{A}(\mathcal{S}) \to [0, \infty]$ s.t.*
(1) $\widetilde{\mu}$ is additive with $\widetilde{\mu}(\emptyset) = 0$
(2) $\widetilde{\mu}$ is an extension of υ: $\widetilde{\mu}(A) = \upsilon(A) \forall A \in \mathcal{S}$
(3) $\widetilde{\mu}$ is unique: If $\mu_1, \mu_2 : \mathcal{A}(\mathcal{S}) \to [0, \infty]$ satisfies $\mu_1(A) = \mu_2(A) \forall A \in \mathcal{S}$, then $\mu_1(E) = \mu_2(E) \forall E \in \mathcal{A}$

Proof. Recall by 2.5, $A \in \mathcal{A}(\mathcal{S}) \Leftrightarrow \exists$ disjoint $E_1, ..., E_n \in \mathcal{S}$ s.t. $A = \sum_{j=1}^n E_j$. Since we require $\widetilde{\mu}$ to be additive, it forces $\widetilde{\mu}(A) = \sum_{j=1}^n \upsilon(E_j)$. This makes (1), (2) automatically satisfied. Now we just have to prove that this $\widetilde{\mu}$ is well defined:
If $A = \sum_{j=1}^n E_j = \sum_{k=1}^m F_k$, then

$$\sum_{j=1}^n \upsilon(E_j) = \sum_{j=1}^n \upsilon\left(E_j \cap \sum_{k=1}^m F_k\right)$$

as $E_j \subset A = \sum_{k=1}^m F_k$.

$$\sum_{j=1}^n \upsilon(E_j) = (\text{by additivity of } \upsilon) \sum_{j=1}^n \sum_{k=1}^m \upsilon(E_j \cap F_k) =$$

$$\sum_{k=1}^m \sum_{j=1}^n \upsilon(E_j \cap F_k) = \sum_{k=1}^m \upsilon\left(\left(\sum_{j=1}^n E_j\right) \cap F_k\right) = \sum_{k=1}^m \upsilon(F_k)$$

So $\widetilde{\mu}$ is well-defined. Then (3) is trivial. ■

Remark. The above proposition works with σ-additive set function υ as well. (Proof left as exercise) If we begin with σ-additive function, then $\widetilde{\mu}$ is a pre-measure. (A measure defined on algebra instead of σ-algebra)

One may wish to follow the same route of above proposition and to build a measure on $\sigma(\mathcal{S})$, but we have to deal with infinite sums and worry about convergence. A better way is to first define an outer-measure on collection of all subsets:

Definition 3.2 (Outer-measure)**.** $\mu^* : 2^X \to [0, \infty]$ is outer-measure if $\mu^*(\emptyset) = 0$ and μ^* is σ-subadditive.

In fact, we do not even have to begin with a pre-measure to get outer-measure. As long as we have a set function $\widetilde{\mu}$ defined on $\mathcal{K} \subset 2^X$, cover of X, we can obtain outer-measure. See details below.

Definition 3.3 (Cover)**.** A family $\mathcal{K} \subset 2^X$ is cover of X if:
(1) $\emptyset \in \mathcal{K}$
(2) $\exists (K_n)_{n \geq 1} \subset \mathcal{K}$ s.t. $X = \cup_{n=1}^\infty K_n$

Remark. Any algebra is a cover of X. And there must be a countable sequence of sets covering X.

Proposition 3.4. *If given cover $\mathcal{K}$ of X. $\widetilde{\mu} : \mathcal{K} \to [0, \infty]$ is function s.t. $\widetilde{\mu}(\emptyset) = 0$ (we do not even require it to be additive), then $\mu^* : 2^X \to [0, \infty]$ defined as below is an outer-measure:*

$$\mu^*(A) := \inf\{\sum_{j=1}^\infty \widetilde{\mu}(K_j) : K_j \in \mathcal{K}, A \subset \cup_j K_j\}$$

Proof. Since K is cover, for any $A \subset X$, we can pick $(K_n)_{n \geq 1} \subset \mathcal{K}$ s.t. $A \subset X = \cup_{n=1}^\infty K_n$. So the set $\{\sum_{j=1}^\infty \widetilde{\mu}(K_j) : K_j \in \mathcal{K}, A \subset \cup_j K_j\}$ is non-empty, and we know it is bounded below by 0. So $\mu^*(A)$ is defined for every A.

$\emptyset \in \mathcal{K}$, we can pick $K_j = \emptyset\ \forall j$ and use the fact $\widetilde{\mu}(\emptyset) = 0$, so $\mu^*(\emptyset) = 0$ as.

Given sequence $(A_k)_{k\geq 1}$ and $A \subset \bigcup_k A_k$. We aim to show $\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k)$ (countably subadditivity). For each k and arbitrary $\epsilon > 0$, by definition of $\mu^*(A_k)$, we can find a sequence $(A_{k,j})_{j\geq 1} \subset \mathcal{K}$ s.t. $A_k \subset \bigcup_j A_{k,j}$ and

$$\sum_{j=1}^{\infty} \widetilde{\mu}(A_{k,j}) < \mu^*(A_k) + 2^{-k}\epsilon$$

so $A \subset \bigcap_k \bigcap_j A_{k,j}$. By definition of $\mu^*(A)$, we have:

$$\mu^*(A) \leq \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \widetilde{\mu}(A_{k,j}) < \sum_{k=1}^{\infty} \mu^*(A_k) + \sum_{k=1}^{\infty} 2^{-k}\epsilon < \sum_{k=1}^{\infty} \mu^*(A_k) + \epsilon$$

Since $\epsilon > 0$ is arbitrary, we proved $\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k)$. ■

Finally to step 3: the first mission is to construct a suitable σ-algebra.

Proposition 3.5. *Given outer-measure on set X, the set Σ defined by*

$$A \in \Sigma \Leftrightarrow \forall E \subset X, \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A)$$

is σ-algebra.

Proof. Note: $E \subset (E \cap A) \cup (E \setminus A)$ so by σ-subadditivity, we have $\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \setminus A)$. So the target equality $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A)$ is equivalent to $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \setminus A)$.

$X \in \Sigma$, and Σ closed under complement are clear, left as exercise.

To prove that Σ is closed under countable union, we need to first show it works for finite union, which is equivalent to $A, B \in \Sigma \Rightarrow A \cup B \in \Sigma$:
Given $E \subset X$, $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A)$, we can break second term down again, since $E \setminus A \subset X,\ B \in \Sigma$, $\mu^*(E \setminus A) = \mu^*((E \setminus A) \cap B) + \mu^*((E \setminus A) \setminus B)$. So by substitution we have

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*((E \setminus A) \cap B) + \mu^*(E \setminus (A \cup B))$$

we need $\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B))$ so it remains to show $\mu^*(E \cap A) + \mu^*(E \setminus A \cap B) \geq \mu^*(E \cap (A \cup B))$. By subadditivity of μ^*, suffice to show $(E \cap A) \cup (E \setminus A \cap B)$ contains $E \cap (A \cup B)$. You may draw an Venn diagram and prove this yourself. But here is one way: we divide part of E contained in $A \cup B$ into the part in A and the part not in A.

$$E \cap (A \cup B) = [E \cap (A \cup B) \cap A] \cup \left[E \cap (A \cup B) \cap A^C\right] =$$

$$(E \cap A) \cup (E \cap A^C \cap B) = (E \cap A) \cup ((E \setminus A) \cap B)$$

Now given a sequence $(A_j)_{j\geq 1} \subset \Sigma$, we want to prove $A = \bigcup_j A_j \in \Sigma$:
As we have shown above, the finite sum $\bigcup_{j=1}^n A_j \in \Sigma$, so

$$\mu^*(E) = \mu^*(E \cap (\bigcup_{j=1}^n A_j)) + \mu^*(E \setminus (\bigcup_{j=1}^n A_j))$$

Note that $E \setminus (\bigcup_{j=1}^n A_j)$ is subtracting less terms from E than $E \setminus (\bigcup_{j=1}^\infty A_j) = E \setminus A$, so clearly $E \setminus A \subseteq E \setminus (\bigcup_{j=1}^n A_j)$. Then

$$\mu^*(E) \geq \mu^*(E \cap (\bigcup_{j=1}^n A_j)) + \mu^*(E \setminus A)$$

We can easily obtain a sequence of pairwise disjoint sets F_j from A_j s.t. $\bigcup_j F_j = \bigcup_j A_j$ as defined in Remark**??**. So

$$\mu^*(E) \geq \mu^*(E \cap (\bigcup_{j=1}^n F_j)) + \mu^*(E \setminus A)$$

it can be shown by easy induction that $\mu^*(E \cap (\bigcup_{j=1}^n F_j)) = \sum_{j=1}^n \mu^*(E \cap F_j)$, so

$$\mu^*(E) \geq \sum_{j=1}^n \mu^*(E \cap F_j) + \mu^*(E \setminus A)$$

and we know inequality holds even if $n \to \infty$, so

$$\mu^*(E) \geq \sum_{j=1}^\infty \mu^*(E \cap F_j) + \mu^*(E \setminus A)$$

$$\geq \mu^*(E \cap \bigcup_{j=1}^\infty F_j) + \mu^*(E \setminus A) \text{ by subadditivity of } \mu^*$$

$$\geq \mu^*(E \cap A) + \mu^*(E \setminus A) \text{ by construction of } F_j$$

■

Theorem 3.6 (Hahn-Caratheodory). *Given set X and algebra $\mathcal{A}$ over X with pre-measure $\widetilde{\mu} : \mathcal{A} \to [0, \infty]$. Extend it to outer-measure μ^* by choosing cover $\mathcal{K} = \mathcal{A}$ and define Σ as in Proposition 3.5. Then $\mu = \mu^*|_\Sigma$ is a measure with:*
(1) $\mathcal{A} \subset \Sigma$
(2) μ is an extension of $\widetilde{\mu}$: i.e. $\mu(A) = \widetilde{\mu}(A)\ \forall A \in \mathcal{A}$. (If $A \in \mathcal{A}$, $\mu^*(A) = \mu(A)$, so (2) is equivalent to $\mu^*(A) = \widetilde{\mu}(A)$)

Proof. Prove (1) first:
Given $A \in \mathcal{A}$, we aim to show $\forall E \subset X, \mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \setminus A)$. (So that $A \in \Sigma$):
Note when $\mu^*(E) = \infty$, this is trivially true. So assume WOLG $\mu^*(E) < \infty$
This is essential for finding a covering of E providing a "good" estimation for $\mu^*(E)$ defined as below: By definition of $\mu^*(E)$ as an infimum, for any $\epsilon > 0$ we can always find a covering of $(E_i)_{i\geq 1} \subset \mathcal{A}$ (i.e. $E \subset \bigcup_i E_i$) s.t.

$$\mu^*(E) \leq \sum_{i=1}^\infty \widetilde{\mu}(E_i) \leq \mu^*(E) + \epsilon$$

Since $E_i \cap A \in \mathcal{A}$ and $E \cap A \subseteq \bigcup_i (E_i \cap A), E \setminus A = E \cap A^C \subseteq \bigcup_i (E_i \cap A^C)$, so $(E_i \cap A)_{i \geq 1}$ is a covering of $E \cap A$ and $(E_i \cap A^C)_{i \geq 1}$ is a covering of $E \cap A^C$, by definition of μ^*:

$$\mu^*(E \cap A) \leq \sum_{i=1}^{\infty} \widetilde{\mu}(E_i \cap A)$$

$$\mu^*(E \cap A^C) \leq \sum_{i=1}^{\infty} \widetilde{\mu}(E_i \cap A^C)$$

$$\text{so } \mu^*(E \cap A) + \mu^*(E \cap A^C) \leq \sum_{i=1}^{\infty} \widetilde{\mu}\left((E_i \cap A) \cup (E_i \cap A^C)\right) \text{ since } \widetilde{\mu} \text{ is additive}$$

$$= \sum_{i=1}^{\infty} \widetilde{\mu}(E_i) \leq \mu^*(E) + \epsilon$$

ϵ is arbitrary so we have the required inequality, and $A \in \Sigma$.

(2) For any $A \in \mathcal{A}$, by picking covering $A, \emptyset, \emptyset, ...$, we have $\mu^*(A) \leq \widetilde{\mu}(A)$.
Now given any covering $(E_j)_{j \geq 1} \subset \mathcal{A}$ s.t. $A \subset \bigcup_j E_j$, we can pick disjoint sequence (F_j) s.t. $\bigcup_j F_j = \bigcup_j E_j$ and $F_j \subset E_j$. So

$$\widetilde{\mu}(A) = \widetilde{\mu}(\sum_j (F_j \cap A)) = \sum_j \widetilde{\mu}(F_j \cap A) \text{ by countable additivity of } \widetilde{\mu}$$

And note $F_j \cap A \subset E_j$, so by monoticity(monotone) of $\widetilde{\mu}$,

$$\widetilde{\mu}(A) \leq \sum_j \widetilde{\mu}(E_j)$$

this is true for any covering, so by definition of $\mu^*(A)$, $\widetilde{\mu}(A) \leq \mu^*(A)$. Combining two inequalities, (2) is proved.

Finally we show that μ is countably additive:
Given disjoint sequence $A_j \in \Sigma$, we already know from sub-additivity of μ^* s.t. $\mu^*(\sum_j A_j) \leq \sum_j \mu^*(A_j)$, it remains to show the other direction.
By easy induction one can show that $\sum_{j=1}^n \mu^*(A_j) = \mu^*(\bigcup_{j=1}^n A_j)$, so we have:

$$\mu^*(\sum_j A_j) \geq \mu^*(\bigcup_{j=1}^n A_j) = \sum_{j=1}^n \mu^*(A_j)$$

and sending $n \to \infty$ gives the required inequality $\mu^*(\sum_j A_j) \geq \sum_j \mu^*(A_j)$.
So $\mu^*(\sum_j A_j) = \sum_j \mu^*(A_j)$, but on Σ, this is exactly μ, so $\mu(\sum_j A_j) = \sum_j \mu(A_j)$. i.e. μ is σ-additive. So μ is measure.

■

Definition 3.7 (finite)**.** A set function μ is finite if $\mu(A) < \infty$ for all set A in its domain.

Definition 3.8 ($\sigma-$finite)**.** A set function μ is $\sigma-$finite if there exists sequence of disjoint sets $S_k \in \mathcal{A}$ which covers X (i.e. $\sum_{k=1}^{\infty} S_k = X$) and $\mu(S_k) < \infty \forall k$. Note $\mu(X)$ need not to be finite.

Under the assumption that the pre-measure $\tilde{\mu}$ is $\sigma-$finite, it can be shown that the measure taken by Hahn-Caratheodory theorem is unique. (See theorem 1.14 in notes 3.1)

3.2 Monotone Class (Extra)

In this section, we define monotone class and show an important theorem about relation between monotone class and $\sigma-$algebra generated by an algebra.

Definition 3.9 (Monotone Class)**.** $\mathcal{M} \in 2^X$ is a monotone class if the following conditions hold:
(1) If $A_j \in \mathcal{M}$, $A_j \uparrow A$, then $A \in \mathcal{M}$
(2) If $A_j \in \mathcal{M}$, $A_j \downarrow A$, then $A \in \mathcal{M}$

Note some key properties of monotone class

- Any $\sigma-$field over X (with operations intersection and union) is monotone class
- If $\mathcal{M}_\alpha \subseteq 2^X\big|_{\alpha\in I}$ are monotone classes, then $\mathcal{M} = \bigcap_{\alpha \in I} \mathcal{M}_\alpha$ is monotone class.

So for any collection $\mathcal{C}$ of subsets of X, we can define the monotone class generated by $\mathcal{C}$:

$$\mathcal{M}(\mathcal{C}) = \bigcap_{\text{Monotone class}\mathcal{M}, \mathcal{M} \supseteq \mathcal{C}} \mathcal{M}$$

i.e. the smallest monotone class containing $\mathcal{C}$. Note $\mathcal{C} \subset \mathcal{M}(\mathcal{C})$.

Theorem 3.10. *Given any algebra $\mathcal{A} \subset 2^X$, $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$. ($\sigma-$algebra generated by an algebra is the same as the monotone class generated by the algebra)*

Proof. Since $\sigma(\mathcal{A})$ is $\sigma-$algebra, it must be monotone class and $\mathcal{A} \subset \sigma(\mathcal{A})$. By definition, $\mathcal{M}(\mathcal{A})$ is the smallest monotone class containing $\mathcal{A}$, so $\sigma(\mathcal{A}) \supseteq \mathcal{M}(\mathcal{A})$.

Now aim to show $\mathcal{M}(\mathcal{A})$ is a $\sigma-$algebra so that we can proceed in similar way to obtain $\mathcal{M}(\mathcal{A}) \supseteq \sigma(\mathcal{A})$. Given $E \in \mathcal{M}(\mathcal{A})$, define

$$g(E) := \{F \in \mathcal{M}(\mathcal{A}) \;:\; E \setminus F, E \cap F, F \setminus E \in \mathcal{M}(\mathcal{A})\}$$

Claim 1: $E \in \mathcal{A} \Rightarrow g(E)$ contains $\mathcal{M}(\mathcal{A})$

- First we need to show $g(E)$ contains $\mathcal{A}$. Given $H \in \mathcal{A}$, since $\mathcal{A}$ is algebra, $E \setminus H, E \cap H, H \setminus E$ are in $\mathcal{A} \subset \mathcal{M}(\mathcal{A})$. So $H \in g(E)$ by definition.
- We also need to show that $g(E)$ is a monotone class. Assume $H_k \uparrow H$ and $H_k \in g(E)$

– $(E \setminus H_k) \downarrow (E \setminus H)$, since $\mathcal{M}(\mathcal{A})$ is monotone class and $(E \setminus H_k) \in \mathcal{M}(\mathcal{A})$ by definition of $g(E)$, $E \setminus H \in \mathcal{M}(\mathcal{A})$

– By similar arguments one can obtain $H \setminus E, E \cap H \in \mathcal{M}(\mathcal{A})$.

So $H \in g(E)$. Similar arguments shows if $H_k \downarrow H$ and $H_k \in g(E)$, then $H \in g(E)$. So $g(E)$ is a monotone class.

Combining the arguments, since $\mathcal{M}(\mathcal{A})$ is the smallest monotone class containing $\mathcal{A}$, $g(E)$ contains $\mathcal{M}(\mathcal{A})$.

Claim 2: $E \in \mathcal{M}(\mathcal{A}) \Rightarrow g(E)$ contains $\mathcal{M}(\mathcal{A})$
Proof of claim 2 is left as exercise. Works in similar way as claim 1.

One can easily show that $\mathcal{M}(\mathcal{A})$ is an algebra by claim 2 and the definition of $g(E)$ (left as exercise)
Now we show that it is closed under countable union of disjoint sets. Take disjoint sequence $A_j \in \mathcal{M}(\mathcal{A})$, let $A := \sum_{j=1}^{\infty} A_j$. Since it is algebra, $\sum_{j=1}^{n} A_j \in \mathcal{M}(\mathcal{A})$. $\sum_{j=1}^{n} A_j \uparrow A$, and $\mathcal{M}(\mathcal{A})$ is monotone class, so $A \in \mathcal{M}(\mathcal{A})$. We have shown that $\mathcal{M}(\mathcal{A})$ is $\sigma-$algebra.
Since $\sigma(\mathcal{A})$ is the smallest $\sigma-$algebra containing $\mathcal{A}$ and $\mathcal{A} \subset \mathcal{M}(\mathcal{A})$, $\mathcal{M}(\mathcal{A}) \supseteq \sigma(\mathcal{A})$.

■

4 Lebesgue Measure

We may now return to the semi-algebra defined in the very beginning and construct Lebesgue measure on $\mathbb{R}$ using the road map given as above.

$$\mathcal{S} = \{\mathbb{R}, \emptyset, (a,b], (-\infty, b], (a, \infty) \mid a < b,\ a, b \in \mathbb{R}\}$$

We define a pre-pre-measure ν as follows:

- $\nu(\emptyset) = 0, \nu(\mathbb{R}) = +\infty$
- $\nu((-\infty, b]), \nu((a, \infty)) = +\infty$
- $\nu((a,b]) = b - a$

It is easy to verify that ν is additive. In order to build a $\sigma-$additive pre-measure $\tilde{\mu}$ according to σ-version of Proposition3.1, we need to show ν is $\sigma-$additive.

Proof. Given $A := \sum_{j=1}^{\infty} A_j$ where $A_j \in \mathcal{S}$. Assume $A \in \mathcal{S}$, we aim to prove $\sum_{j=1}^{\infty} \nu(A_j) = \nu(A)$
$\nu(A) \geq \nu(\sum_{j=1}^{n} A_j) = \sum_{j=1}^{n} \nu(A_j)$ by finite additivity of ν. Inequality holds even if adding limit $n \to \infty$. So it remains to show $\nu(A) \leq \sum_{j=1}^{\infty} \nu(A_j)$.

- Finite Case: $A = (a, b], A = \sum_{j=1}^{\infty} A_j$ where A_j are disjoint. So we know that $A_j = (a_j, b_j]$ for some finite a_j, b_j.

$$\nu(A) \leq \sum_{j=1}^{\infty} \nu(A_j) \Leftrightarrow b - a \leq \sum_{j=1}^{\infty} (b_j - a_j)$$

For any $\epsilon > 0$,

$$[a + \epsilon, b] \subseteq (a, b] = \sum_{j=1}^{\infty} (a_j, b_j] \subseteq \bigcup_{j \geq 1} (a_j, b_j + \frac{\epsilon}{2^j})$$

Note $(a_j, b_j + \frac{\epsilon}{2^j})$ may not be disjoint, so symbol $\sum$ is not used. These two set inequalities give us a compact set $[a + \epsilon, b]$ and its open cover $\bigcup_{j \geq 1} (a_j, b_j + \frac{\epsilon}{2^j})$. So we know there is a finite cover,

$$[a + \epsilon, b] \subseteq \bigcup_{k=1}^{m} (a_{j_k}, b_{j_k} + \frac{\epsilon}{2^{j_k}}) \subseteq \bigcup_{k=1}^{m} (a_{j_k}, b_{j_k} + \frac{\epsilon}{2^{j_k}}]$$

ν is additive and monotone, so

$$\nu((a + \epsilon, b]) \leq \nu\left(\bigcup_{k=1}^{m} (a_{j_k}, b_{j_k} + \frac{\epsilon}{2^{j_k}}]\right) = \sum_{k=1}^{m} \nu\left((a_{j_k}, b_{j_k} + \frac{\epsilon}{2^{j_k}}]\right)$$

then by definition of ν,

$$b - a - \epsilon \leq \sum_{k=1}^{m} b_{j_k} - a_{j_k} + \frac{\epsilon}{2^{j_k}} \leq \sum_{j=1}^{\infty} b_j - a_j + \frac{\epsilon}{2^j} = \sum_{j=1}^{\infty} (b_j - a_j) + \epsilon$$

The choice ϵ is arbitrary, so $b - a \leq \sum_{j=1}^{\infty} (b_j - a_j)$.

- General case: For $A = \sum_{j=1}^{\infty} A_j$, where $A, A_j \in \mathcal{S}$, one can consider $E_n = (-n, n] \uparrow \mathbb{R}$. $A \cap E_n$ must be of the form $(a, b]$ where $a < b$ for any $A \in \mathcal{S}$. So we can somehow use the conclusion of finite case. The proof is left as exercise. (Hint: first prove $\nu(A) = \lim_{n\to\infty} \nu(A \cap E_n)$, then note $A \cap E_n = \sum_{j=1}^{\infty} (A_j \cap E_n)$)

■

Before we complete the pre-measure obtained above(to a measure), let's consider the general case: $\mathbb{R}^n$. Intervals on $\mathbb{R}^n$ are defined as below:
If $a = (a_1, a_2, ..., a_n), b = (b_1, b_2, ..., b_n)$, then interval (a, b) is defined as

$$\prod_{k=1}^{n} (a_k, b_k)$$

where $\prod$ denotes consecutive Cartesian products of sets. Other intervals $(a, b], [a, b), [a, b]$ are defined similarly for $a < b$ (defined as $a_k < b_k \, \forall k$). If we collect all intervals (where we allow a_k, b_k to be $\pm\infty$) and the empty set, we have a semi-algebra $\mathcal{S}$. Collecting all finite union of sets in $\mathcal{S}$ gives us an algebra $\mathcal{A}$. (Elements of $\mathcal{A}$ are called *elementary figure* in $\mathbb{R}^n$) A set function $\tilde{\lambda} : \mathcal{A} \to \mathbb{R}^+ \cup \{\infty\}$ is defined as below:

- $\tilde{\lambda}(\emptyset) = 0$
- $\tilde{\lambda}((a,b)) = \tilde{\lambda}((a,b]) = \tilde{\lambda}([a,b)) = \tilde{\lambda}([a,b]) := \prod_{k=1}^{n}(b_k - a_k) \leq \infty$
- $\tilde{\lambda}\left(\sum_{k=1}^{m} I_k\right) := \sum_{k=1}^{m} \tilde{\lambda}(I_k)$. where I_k are disjoint intervals.

The proof of $\tilde{\lambda}$ is $\sigma-$additive is a generalisation to the proof for $\mathbb{R}$ presented above. (See notes 3.2 lemma 1.16 for detailed proof)

$\tilde{\lambda}$ is $\sigma-$finite. As taking the countable collection of disjoint "cubes" $(\xi, \xi+1], \xi \in \mathbb{Z}^n$, each with measure $1 < \infty$, we can cover $\mathbb{R}^n$. So construction given in Hahn-Caratheodory theorem is a measure(called Lebesgue measure) on Σ:

$$\Sigma := \left\{A \in 2^X : \forall E \subset X, \lambda^*(E) = \lambda^*(E \cap A) + \lambda^*(E \setminus A)\right\}$$

recall that outer-measure λ^* is defined as

$$\lambda^*(A) := \inf \left\{ \sum_{j=1}^{\infty} \widetilde{\lambda}(K_j) : K_j \in \mathcal{K}, A \subset \bigcup_j K_j \right\}$$

Definition 4.1 (Borel set and Borel σ-algebra)**.** A Borel set in a topological space (X, τ) is a set formed by open sets under countable union or countable intersection or relative complement (to the space X). The collection of all Borel sets is called Borel σ-algebra $\mathcal{B}(X)$(as $\mathcal{B}(X) = \sigma(\tau)$, proof is left as exercise.). Any measure defined on Borel σ-algebra is called Borel measure.

Lemma 4.2. $\mathcal{B}(\mathbb{R}^n) \subset \Sigma$

Proof. Denote the set of all open sets in $\mathbb{R}^n$ as τ.
If we are able to show $\tau \subset \Sigma$, then $\mathcal{B}(\mathbb{R}^n) = \sigma(\tau) \subset \sigma(\Sigma) = \Sigma$.
Given open set $A \in \tau$, it can be written as countably many disjoint half-open "cubes" as described in lemma 1.17 of note 3.2) so $A \in \sigma(\mathcal{A}) \subset \Sigma$. ∎

Definition 4.3 (Restriction of measure)**.** Given measure space $(X, \mathcal{F}, \mu)$, and set $A \in \mathcal{F}$. $\mathcal{F}|_A := \{A \cap B : B \in \mathcal{F}\} \subset \mathcal{F}$. And

$$\mu|_A(B) := \mu(B)$$

where $\mu|_A : \mathcal{F}|_A \to \overline{\mathbb{R}}$ is indeed a measure, called restriction of μ to A.
In fact, restriction measure can also be defined on $\mathcal{F}$, namely

$$1_A\mu(B) := \mu(B \cap A)$$

Remark. It is not difficult to prove that $\{A \cap B : B \in \mathcal{F}\} = \{C \in \mathcal{F} : C \subseteq A\}$

Proposition 4.4. *Restriction of a measure is indeed a measure. (i.e. $\mathcal{F}|_A$ is a $\sigma-$algebra on A)*

Proof left as exercise.

Definition 4.5 (Probability measure)**.** It is a measure on space X s.t. $\mu(X) = 1$. Since μ is monotone, this implies μ is a function from X to $[0,1]$.

4.1 Complete Measure

Definition 4.6 (Complete Measure)**.** Given a measure μ on $\sigma-$algebra $\mathcal{F}$. $(X, \mathcal{F}, \mu)$ is *complete* if for all $A \in \mathcal{F}$ such that $\mu(A) = 0$: if $E \subseteq A$, then $E \in \mathcal{F}$.
Note: since a measure is monotone, such E must satisfies $\mu(E) = 0$. And sets E are called *negligible sets* (w.r.t. A). Such a $\sigma-$algebra $\mathcal{F}$ is called $\mu-$complete.

Recall the outer measure μ^* and $\sigma-$algebra Σ from Hahn-Caratheodory theorem. We now prove $(\mu^*|_\Sigma, \Sigma)$ is complete.

Proof. Given $A \subset B$ where $B \in \Sigma$ with $\mu^*(B) = 0$. We need to show $A \in \Sigma$. i.e. $\forall E \subset X, \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^C)$. And as μ^* is monotone (as it is subadditive), we only have to show $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^C)$.

This is easy. As A is negligible, so $E \cap A \subseteq A \subseteq B$ is also negligible. $\mu^*(E \cap A) \leq \mu^*(B) = 0$ so $\mu^*(E \cap A) = 0$.
$\mu^*(E \cap A^C) \leq \mu^*(E)$ as μ^* is monotone. So $\mu^*(E \cap A) + \mu^*(E \cap A^C) \leq 0 + \mu^*(E) = \mu^*(E)$. ∎

Then we know that Lebesgue measure is complete.

The following are extra contents.

In fact, even if we are given an general measure, we can extend it to a complete measure.

Theorem 4.7 (Completion of a measure)**.** *Given a measure space $(\mathcal{F}, \mu)$, there exists $\sigma-$algebra $\overline{\mathcal{F}}$ and an unique measure extension $\overline{\mu}$ s.t. $(\overline{\mathcal{F}}, \overline{\mu})$ is complete. Unique extension means $\overline{\mu}|_{\mathcal{F}} = \mu$, and if any measure $\nu : \overline{\mathcal{F}} \to \mathbb{R}^+ \cup \{\infty\}$ satisfies $\nu(A) = \overline{\mu}(A)\ \forall A \in \mathcal{F}$, then $\nu(B) = \overline{\mu}(B)\ \forall B \in \overline{\mathcal{F}}$.*

Proof. Let $\overline{\mathcal{F}} = \{A \cup N : A \in \mathcal{F}, N \subseteq E \in \mathcal{F} \text{ s.t. } \mu(E) = 0\}$. Here N stands for negligible (also called μ-null set).
Claim: $\overline{\mathcal{F}}$ is $\sigma-$algebra.

- $X = X \cup \emptyset$ where $\emptyset \subseteq \emptyset$, $\mu(\emptyset) = 0$. So $\emptyset$ is negligible, $X \in \overline{\mathcal{F}}$.
- pick $A \cup N \in \overline{\mathcal{F}}$ where $N \subseteq E$ with $\mu(E) = 0$. We want to show that $(A \cup N)^C$ is also a union of negligible set and a set in $\mathcal{F}$. $(A \cup N)^C \cap E$ is obviously negligible. And we note that $(A \cup N)^C = \big((A \cup N)^C \cap E\big) + \big((A \cup N)^C \cap E^C\big)$. The second set $(A \cup N)^C \cap E^C = A^C \cap N^C \cap E^C$. As $N \subseteq E$, $N^C \cap E^C = E^C$. So second set is simply $(A \cup E)^C \in \mathcal{F}$. Therefore, $(A \cup N)^C \in \overline{\mathcal{F}}$.
- Closed under countable union is easy to prove. left as an exercise.

Now we attempt to define measure $\overline{\mu}$. Note $\overline{\mu}(A \cup N) \geq \overline{\mu}(A) = \mu(A)$, as measure must be monotone, and the new measure must coincide with μ on $\mathcal{F}$. Also we have $\overline{\mu}(A \cup N) \leq \overline{\mu}(A \cup E) = \mu(A \cup E) \leq \mu(A) + \mu(E) = \mu(A)$. As $A \cup N \subseteq A \cup E$. So $\overline{\mu}(A \cup N)$ is forced to be $= \mu(A)$.

- This is well-defined as if we have $A \cup N = B \cup M$ where $A, B \in \mathcal{F}$ and $N \subseteq E$, $M \subseteq F$ with $\mu(E) = \mu(F) = 0$. Then since $A \subseteq A \cup E = B \cup M \subseteq B \cup F$, we have $\mu(A) \leq \mu(B \cup F) = \mu(B) + 0$. By symmetry we can also get $\mu(B) \leq \mu(A)$, so $\mu(A) = \mu(B)$. Then both sets $A \cup N, B \cup M$ are assigned to the same value.

- It is easy to prove $\overline{\mu}$ is $\sigma-$additive. Proof is left as exercise.

Finally, we have to prove $(\overline{\mathcal{F}}, \overline{\mu})$ is complete.
Given $A \subseteq E \in \overline{\mathcal{F}}$ s.t. $\overline{\mu}(E) = 0$. Say $E = B \cup N$ where $B \in \mathcal{F}, N \subseteq H$ s.t. $\mu(H) = 0$. Then $\mu(B) = \overline{\mu}(E) = 0$ by definition of $\overline{\mu}$.
$A \subseteq E = B \cup N \subseteq B \cup H$, and $\mu(B \cup H) \leq \mu(B) + \mu(H) = 0$, so $\mu(B \cup H) = 0$, A is negligible.
$A = \emptyset \cup A$ where $\emptyset \in \mathcal{F}$, so $A \in \overline{\mathcal{F}}$.

For uniqueness, we take another measure $\nu : \overline{\mathcal{F}} \to \mathbb{R}^+ \cup \{\infty\}$ s.t. $\nu(A) = \overline{\mu}(A)\ \forall A \in \mathcal{F}$. Now for any $B \in \overline{\mathcal{F}}$, say $B = A \cup N$ where $A \in \mathcal{F}$ and $N \subseteq E$ with $\mu(E) = 0$.
$\overline{\mu}(B) = \mu(A) = \nu(A) \leq \nu(B)$ as ν is monotone and $A \subseteq B$.
Note $B \subseteq A \cup E$, so $\nu(B) \leq \nu(A \cup E) = \mu(A \cup E) \leq \mu(E) + 0 = \overline{\mu}(B)$.
Combining two inequalities give $\overline{\mu}(B) = \nu(B)$. So the extension is unique. ∎

Note the extension depends on the measure μ, so we write $\overline{\mathcal{F}_\mu}$ instead unless the measure used is clear from the content.

4.2 Approximation and regular measure

For Σ taken by Hahn-Caratheodory, though we cannot explicitly write members of Σ, measure of each member can be estimated by a set in $\mathcal{A}$.

Theorem 4.8 (Approximation theorem). *Given algebra $\mathcal{A} \subseteq 2^X$, let (Σ, μ) be taken as in Hahn-Caratheodory theorem. Given $A \in \Sigma$ with $\mu(A) < \infty$. $\forall \epsilon > 0$, $\exists E \in \mathcal{A}$ s.t.*

$$\mu(E \setminus A) + \mu(A \setminus E) < \epsilon$$

you may recognise that $E \setminus A, A \setminus E$ are components of $E \triangle A$ (symmetric difference of E, A)

Proof. Fix $\epsilon > 0$.
Recall that $\mu^* := \inf_{\{A_i\}} \sum_{i \geq 1} \tilde{\mu}(A_i)$ where $\{A_i\} \subseteq \mathcal{A}$ is any covering of A. By this definition, there exists $\{A_i\}$ s.t.

$$\mu^*(A) \leq \sum \tilde{\mu}(A_i) \leq \mu^*(A) + \epsilon \quad (1)$$

Since $\mu(A) = \mu^*(A) < \infty$, the infinite sum is finite. So we can find n_0 s.t.

$$\sum_{i=n_0+1}^{\infty} \tilde{\mu}(A_i) \leq \epsilon \quad (2)$$

Let $E := \bigcup_{i=1}^{n_0} A_i \in \mathcal{A}$, then

$$\mu^*(E \setminus A) = \mu^*(\bigcup_{i=1}^{n_0} A_i \setminus A) \leq \mu^*(\bigcup_{i=1}^{\infty} A_i \setminus A)$$

$$= \mu^*(\bigcup_{i=1}^{\infty} A_i) - \mu^*(A) \leq \sum_{i=1}^{\infty} \mu^*(A_i) - \mu^*(A) \leq \epsilon$$

where last inequality is consequence of (1).
For the other difference $\mu^*(A \setminus E)$:

$$\mu^*(A \setminus E) \leq \mu^*(\bigcup_{i=1}^{\infty} A_i \setminus \bigcup_{i=1}^{n_0} A_i) = \mu^*(\bigcup_{i=n_0+1}^{\infty} A_i)$$

$$\leq \sum_{i=n_0+1}^{\infty} \mu^*(A_i) \leq \epsilon \quad \text{by (2)}$$

■

Proposition 1.22 of lecture notes can be proved in similar way to above theorem: Given algebra $\mathcal{A}$ and σ-finite measure μ on $\sigma(\mathcal{A})$. $\forall A \in \sigma(\mathcal{A})$, $\epsilon > 0$, there are pairwise disjoint sets $A_1, A_2, ... \in \mathcal{A}$ s.t. $A \subset \bigcup_{n=1}^{\infty} A_n$ and $\mu\left(\bigcup_{n=1}^{\infty} A_n \setminus A\right) < \epsilon$.

Definition 4.9 (Regular measure)**.** Given topological space X, let $\mathcal{B} := \mathcal{B}(X)$, the smallest $\sigma-$algebra containing all open sets.(i.e. Borel $\sigma-$algebra) Now given measure $\mu : \mathcal{F} \to \mathbb{R}^+ \cup \{\infty\}$ where $\mathcal{F}$ contains $\mathcal{B}$. μ is regular if $\forall A \in \mathcal{F}, \epsilon > 0$, $\exists F \subseteq A \subseteq G$ where F is closed, G is open and $\mu(G \setminus F) \leq \epsilon$.
Note: we do not restrict to A s.t. $\mu(A) < \infty$ here, compared to approximation theorem.

Remark. Regularity has two parts: upper regularity (for all $\epsilon > 0$, there is open set $G \supseteq A$ s.t. $\mu(G \setminus A) \leq \epsilon$) and lower regularity (for all $\epsilon > 0$, there is closed set $F \subseteq A$ s.t. $\mu(A \setminus F) \leq \epsilon$). Upper regularity is equivalent to

$$\forall A \in \mathcal{F},\ \lambda(A) = \inf_{G \supseteq A,\text{ open } G} \{\lambda(G)\}$$

similarly lower regularity is equivalent to

$$\forall A \in \mathcal{F},\ \lambda(A) = \sup_{F \subseteq A,\text{ closed } F} \{\lambda(F)\}$$

(proof of equivalence is left as exercise)

Proposition 4.10. *If μ is regular, then $\mathcal{F} \subseteq \overline{\mathcal{B}_\mu}$, the completion of Borel-algebra.*

Proof. Pick $A \in \mathcal{F}, \epsilon > 0$, there exists sequences of closed F_n and open G_n s.t. $F_n \subseteq A \subseteq G_n$ and $\mu(G_n \setminus F_n) \leq \frac{1}{n}$. Let $F := \bigcup_{n \geq 1} F_n, G := \bigcap_{n \geq 1} G_n$. Then

$$\mu(G \setminus F) \leq \mu(G_n \setminus F) \text{ as } G_n \supseteq G$$

$$\leq \mu(G_n \setminus F_n) = \frac{1}{n}$$

This is true for any n, so $\mu(G \setminus F) = 0$. Write $A = F \cup (A \setminus F)$, where $F \in \mathcal{B}$. $A \setminus F \subset G \setminus F$ so $A \setminus F$ is negligible. Therefore, $A \in \overline{\mathcal{B}_\mu}$. ■

Theorem 4.11. *Lebesgue measure $\mu : \Sigma \to \mathbb{R}^+ \cup \{\infty\}$ is regular.*

Proof. Fix $A \in \Sigma, \epsilon > 0$. We first aim to find open set G s.t. $A \subseteq G$ and $\mu(G \setminus A) \leq \epsilon$. Then we find another F s.t. $F \subseteq A$ and $\mu(A \setminus F) \leq \epsilon$. So we know we can bound $\mu(G \setminus F) \leq \mu(G \setminus A) + \mu(A \setminus F) \leq 2\epsilon$.

We pick $E_n = [-n, n]$. Let $A_n := A \cap E_n$. Note $\mu(A_n) \leq \mu(E_n) = 2n < \infty$. So by definition of Lebesgue measure, there is a sequence $\{B_{n,k}\}_{k\geq 1} \subset \mathcal{A}$ (the set of elementary figures) s.t. $A_n \subseteq \bigcup_{k\geq 1} B_{n,k}$ and

$$\mu(A_n) \leq \sum_{k=1}^{\infty} \mu(B_{n,k}) \leq \mu(A_n) + \frac{\epsilon}{2^n} \qquad (2)$$

Though $B_{n,k}$ are not open, they are in $\mathcal{A}$. So we can write it as union of disjoint half-cubes $B_{n,k} = \sum_j I_{n,k,j}$ where $I_{n,k,j} = (a_{n,k,j}, b_{n,k,j}]$. Half-cubes are yet not open, so we define $J_{n,k,j} = (a_{n,k,j}, b_{n,k,j} + \delta_{n,k,j}) \supseteq I_{n,k,j}$. We will determine how to bound $\delta_{n,k,j} > 0$ later. Define $G_{n,k} := \bigcup_j J_{n,k,j} \supseteq B_{n,k}$.

$$\mu(G_{n,k}) \leq \sum_j \mu(J_{n,k,j}) = \sum_j (b_{n,k,j} - a_{n,k,j} + \delta_{n,k,j})$$

since we forced $I_{n,k,j}$ to be disjoint, $\sum_j (b_{n,k,j} - a_{n,k,j}) = \mu(B_{n,k})$ and so

$$\mu(G_{n,k}) \leq \mu(B_{n,k}) + \sum_j \delta_{n,k,j}$$

now we say choose $J_{n,k,j}$ s.t. $\sum_j \delta_{n,k,j} \leq \frac{\epsilon}{2^k 2^n}$.

$$A_n \subseteq \bigcup_{k\geq 1} B_{n,k} \subseteq \bigcup_{k\geq 1} G_{n,k} =: G_n$$

with this definition of G_n, we have

$$\mu(G_n) \leq \sum_{k\geq 1} \mu(G_{n,k}) \leq \sum_{k\geq 1} \mu(B_{n,k}) + \sum_{k=1}^{\infty} \frac{\epsilon}{2^n 2^k} \leq \mu(A_n) + 2\frac{\epsilon}{2^n} \text{ by (2)}$$

By defining $G := \cup_{n\geq 1} G_n$,

$$\mu(G \setminus A) = \mu(\bigcup_n G_n \setminus \bigcup_k A_k) \leq \mu\left(\bigcup_n (G_n \setminus A_n)\right)$$

The last inequality is because for $\bigcup_n (G_n \setminus A_n)$ we only subtract A_n from each G_n instead of subtracting all A_k as in $\bigcup_n G_n \setminus \bigcup_k A_k$ .Then

$$\mu(G \setminus A) \leq \sum_{n=1}^{\infty} \mu(G_n \setminus A_n) \leq 2 \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = 2\epsilon$$

In order to prove the other side of regularity (the closed set $F \subseteq A$ s.t. $F \subseteq A$ and $\mu(A \setminus F) \leq \epsilon$), use the fact that A^C is also in Σ. So we are able to find open H s.t. $A^C \subseteq H$ and $\mu(H \setminus A^C) \leq \epsilon$, it is easy to prove that closed set $F := H^C$ works. (left as exercise.) ∎

Note that if we let $\mathcal{F}_\sigma$ to be collection of all countable unions of closed sets of $\mathbb{R}^n$ and $\mathcal{G}_\sigma$ to be collection of all countable unions of open sets of $\mathbb{R}^n$. According to the fact that Lebesgue measure is regular, one can always find $F \in \mathcal{F}_\sigma$ and $G \in \mathcal{G}_\sigma$ s.t. $F \subseteq A \subseteq G$ and $\mu(G \setminus F) = 0$. (Proof: exercise.)

Using regularity of Lebesgue measure, we can prove the last important property that we wished our measure on $\mathbb{R}^n$ can satisfy: translation invariance.

Theorem 4.12 (translation invariance of Lebesgue)**.** *Given Lebesgue measure λ and fix $x_0 \in \mathbb{R}^n$, define $\Phi(x) = x_0 + x$, $\Phi(A) := \{\Phi(x) : x \in A\}$. Then $\lambda(\Phi(A)) = \lambda(A) \forall A \in \mathcal{B}(\mathbb{R}^n)$*

Proof. This is trivial if $A = (a, b]$ (a half-cube) This proves the case where A is open since we can represent A as a countable union of disjoint half-cubes (see Lemma4.2 for details).
For arbitrary A, upper regularity is required to reduce it to open case. If $A \subset G$ and G is open, then $\Phi(A) \subset \Phi(G)$ and $\Phi(G)$ is open, so by regularity of λ:

$$\lambda(\Phi(A)) = \inf_{\text{open } G,\, A \subset G} \lambda(\Phi(G)) = \inf_{\text{open } G,\, A \subset G} \lambda(G) = \lambda(A)$$

■

5 Lebesgue Integration

Recall the concept of Riemann integral for a function $X \to \mathbb{R}$ is first picking intervals I_k on x-axis and points x_k as representatives, then calculate the area of the rectangles (the red one in figure below) with height $f(x_k)$ and length $|I_k|$. Where $|I_k|$ means "length" of interval I_k.

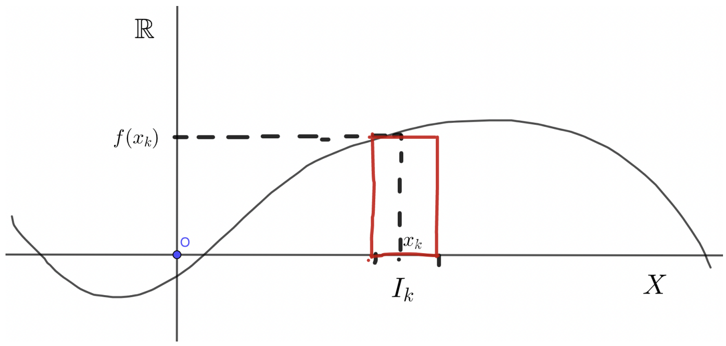

Figure 2: Riemann Integral

So the integral is written as $\sum f(x_k)|I_k|$.
But we can do the other way around. First pick intervals J_k on y-axis. Let $I_k := f^{-1}(J_k)$, then pick representatives y_k in intervals J_k. As shown in figure below:

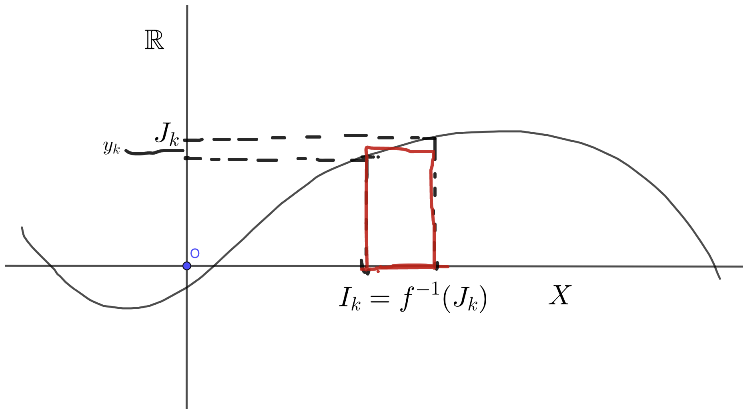

Figure 3: Lebesgue integral

Now $I = \sum y_k \mu(f^{-1}(J_k))$. This is the idea of Lebesgue measure. But this definition requires an important property: $f^{-1}(J_k) \subseteq X$ should be in the $\sigma-$algebra where μ is defined. So we need a definition for measurable function:

Definition 5.1 ($\mathcal{F}-$measurable function($\mathbb{R}$ version))**.** Given measurable space $(X, \mathcal{F}, \mu)$, and function $f : X \to \mathbb{R}$. f is said to be $\mathcal{F}-$measurable if $\forall A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \mathcal{F}$.

Definition 5.2 (Borel-Measurable)**.** Given topological spaces X, Y, $f : X \to Y$ is Borel-measurable if f is $\mathcal{B}(X)$-$\mathcal{B}(Y)$ measurable. i.e. for any open set $B \subset Y$, $f^{-1}(B)$ is a Borel set in X.

Remark. Many functions are $\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R})$ measurable, for example, all continuous functions, and all monotone functions. (PS3)

This definition can be generalised to the case $f : X \to \overline{\mathbb{R}}$ where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. But we need to extend the Borel-algebra:

$$\overline{\mathcal{B}} := \{A \cup B \ : \ A \in \mathcal{B}(\mathbb{R}), B \subseteq \{-\infty, \infty\}\}$$

. (Exercise: prove $\overline{\mathcal{B}}$ is a $\sigma-$algebra).

Definition 5.3 ($\mathcal{F}-$measurable function($\overline{\mathbb{R}}$ version))**.** Given measurable space $(X, \mathcal{F}, \mu)$, and function $f : X \to \overline{\mathbb{R}}$. f is said to be $\mathcal{F}-$measurable if $\forall A \in \overline{\mathcal{B}},\ f^{-1}(A) \in \mathcal{F}$.

Even more generally,

Definition 5.4 ($\mathcal{A}$-$\mathcal{A}'$ measurable function)**.** $f : X \to Y$, and given $\sigma-$algebras $\mathcal{A}, \mathcal{A}'$ over X, Y respectively. f is $\mathcal{A}$-$\mathcal{A}'$ measurable if

$$\forall A' \in \mathcal{A}', f^{-1}(A') \in \mathcal{A}$$

Definition 5.5. Given space X and measurable space $(Y, \mathcal{F})$, function $f : X \to Y$.

$$\sigma(f) := \left\{f^{-1}(A) : A \in \mathcal{F}\right\}$$

It is $\sigma-$algebra (PS1) and this is the smallest $\sigma-$algebra on which f is measurable.

Back to $\overline{\mathbb{R}}$ case, we have the following lemma which allows us to check only specific sets in $\overline{\mathcal{B}}$ instead of every set. It was mentioned in MATH50010 Probability for Statistics.

Lemma 5.6. *Given measurable space $(X, \mathcal{F}, \mu)$, and function $f : X \to \overline{\mathbb{R}}$. The following are equivalent to f being $\mathcal{F}-$measurable:*

- $f^{-1}((-\infty, x]) \in \mathcal{F}$ *for all* $x \in \mathbb{R}$.
- $f^{-1}((-\infty, x)) \in \mathcal{F}$ *for all* $x \in \mathbb{R}$.
- $f^{-1}([x, \infty)) \in \mathcal{F}$ *for all* $x \in \mathbb{R}$.
- $f^{-1}((x, \infty)) \in \mathcal{F}$ *for all* $x \in \mathbb{R}$.

Proof. Exercise. ∎

This is because $\mathcal{B}$ is generated by open sets, and open sets can be generated by $\mathcal{G} = \{(-\infty, x] : x \in \mathbb{R}\}$ (or other forms of intervals) so these intervals generates $\mathcal{B}$.

Lemma 5.7 (Generating set and measurable)**.** *If $\sigma(\mathcal{C}) = \mathcal{A}'$, then f is $\mathcal{A}$-$\mathcal{A}'$ measurable $\Leftrightarrow$*

$$\forall A' \in \mathcal{C}, f^{-1}(A') \in \mathcal{A}$$

Proof. $\Leftarrow$ is obvious. Now assume $\forall A' \in \mathcal{C}, f^{-1}(A') \in \mathcal{A}$. Define the following set:

$$\mathcal{G} := \left\{A' \subset X' : f^{-1}(A') \in \mathcal{A}\right\}$$

this is a $\sigma-$algebra.(Proof: exercise) And it contains $\mathcal{C}$ by assumption. Since $\sigma(\mathcal{C}) = \mathcal{A}'$, $\mathcal{G}$ contains $\mathcal{A}'$. ∎

Remark. The proof is basic but this idea is very useful. The proof of Lemma2.5 also uses this idea.

Examples of measurable functions:

- If $X = Y, \mathcal{A} = \mathcal{A}'$, then $id : X \to X$ must be measurable.
- $X = Y = \mathbb{R}, \mathcal{A} = \mathcal{A}' = \mathcal{B}(\mathbb{R})$. If f is continuous, it must be measurable.
- Indicator function $1_A(x)$ where $A \subset X$ is measurable iff $A \in \mathcal{A}$. It is an example of simple function.

Lemma 5.8. *If $(X, \mathcal{A}), (Y, \mathcal{A}'), (Z, \mathcal{A}'')$ are measurable spaces, $f : X \to Y$ is $\mathcal{A}$-$\mathcal{A}'$ measurable and $g : Y \to Z$ is $\mathcal{A}'$-$\mathcal{A}''$ measurable. Then $h = g \circ f$ is $\mathcal{A}$-$\mathcal{A}''$ measurable.*

Theorem 5.9 (Algebraic rules of measurable functions)**.** *$(X, \mathcal{F}, \mu)$ is a measurable space and $f, g : X \to \mathbb{R}$ are ($\mathcal{A}$-$\mathcal{B}(\mathbb{R})$)measurable functions. Then*
(1) $f + g, fg, |f|, f \wedge g := \min\{f, g\}, f \vee g := \max\{f, g\}$ are also measurable. If $g \neq 0$, then $\frac{1}{g}$ is also measurable.
(2) If $f_k : X \to \mathbb{R}$ are measurable. Then

$$\inf_{k\geq 1} f_k, \sup_{k\geq 1} f_k, \liminf_{k\to\infty} f_k, \limsup_{k\to\infty} f_k$$

are also measurable.

Proof. Proof of (1)
If $(f + g)(x) < a$, then possible values of $f(x), g(x)$ are: $f(x) < r, g(x) < s$ where $r, s \in \mathbb{Q}$ and $r + s < a$. Any pair of such (r, s) would work. So

$$(f + g)^{-1}((-\infty, a)) = \bigcup_{r,s\in\mathbb{Q},r+s<a} f^{-1}((-\infty, r)) \cap g^{-1}((-\infty, s)) \in \mathcal{A}$$

then by lemma5.6, $f + g$ is measurable.
$s \to s^2, s \to -s, s \to s/2$ are all continuous, so they are measurable. By composition rule (lemma5.8),

$$fg = \frac{1}{2}\left[(f + g)^2 - f^2 - g^2\right]$$

so fg is also measurable.

Note if given $f : X \to \mathbb{R}$, we can represent it using two functions $f^+, f^- : X \to \mathbb{R}^+$, where $f^+ := \max\{f, 0\}, f^- := \min\{-f, 0\}$. Clearly f^+, f^- are measurable if f is measurable.
then we have:

- $f = f^+ - f^-$
- $|f| = f^+ + f^-$
- $f \wedge g = f - (g-f)^-$
- $f \vee g = f + (g-f)^+$

So these functions are all measurable.
Finally for $\frac{1}{g}$, consider $(\frac{1}{g})^{-1}((-\infty, a))$

- If $a < 0$, $\frac{1}{g} < a \Rightarrow 1/a < g < 0$.
- If $\frac{1}{g} < 0$, then $g < 0$.
- If $a > 0$, $\frac{1}{g} < a \Rightarrow \frac{1}{g} < 0$ or $0 < \frac{1}{g} < a \Rightarrow g < 0$ or $g > \frac{1}{a}$.

$$\Rightarrow (\frac{1}{g})^{-1}((-\infty, a)) = \begin{cases} g^{-1}((\frac{1}{a}, 0)) & a < 0 \\ g^{-1}((-\infty, 0)) & a = 0 \\ g^{-1}((-\infty, 0) \cup (\frac{1}{a}, \infty)) & a > 0 \end{cases}$$

As g is measurable, for each case of a, pre-image of $1/g$ is in $\mathcal{A}$. Therefore, $\frac{1}{g}$ is measurable.

(2)
Aim to prove:

$$(\inf_k f_k)^{-1}((-\infty, a)) = \bigcup_{k=1}^{\infty} f_k^{-1}((-\infty, a))$$

$\subseteq$: Assume $L := \inf_k f_k(x) < a$, define $L' := \frac{L+a}{2}$. This is not lower bound for the set $\{f_k(x)\}$, so exists k s.t. $f_k(x) < L' < a$. That means $x \in \bigcup_{k=1}^{\infty} f_k^{-1}((-\infty, a))$. So $(\inf_k f_k)^{-1}((-\infty, a)) \subseteq \bigcup_{k=1}^{\infty} f_k^{-1}((-\infty, a))$
$\supseteq$: If for some m, $f_m(x) < a$. Then $\inf_k f_k(x) \leq f_m(x) < a$.
So $\inf_k f_k$ is measurable. Since $\sup_k f_k = -\inf_k(-f_k)$, $\sup_k f_k$ is also measurable.

$\limsup$ and $\liminf$ are just compositions of $\sup, \inf$ so are also measurable. ■

5.1 Simple Function

A simple set function only takes one of the finite choices of value.

Definition 5.10 (Simple function)**.** $(X, \mathcal{F})$ is a measure space. $f : X \to \overline{\mathbb{R}}$ is simple if there exists a finite partition $\{E_j\}_{1 \leq j \leq n}$ of X and

$$f = \sum_{j=1}^{n} c_j 1_{E_j}, \text{ for some } c_j \in \mathbb{R}$$

Lemma 5.11. *Simple function is $\mathcal{F}-$measurable if $E_j \in \mathcal{F}$ for each j.*

Proof. This is because if $E_j \in \mathcal{F}$, then indicator 1_{E_j} is measurable. ■

We wish to define integration $I(f)$ where $f : X \to \overline{\mathbb{R}}$ which satisfies the following properties:

- $I(\alpha f + g) = \alpha I(f) + I(g)$ where α is constant.
- $I(f) \geq 0$ if $f \geq 0$
- If $f_n \uparrow f$ and $f_n \geq 0$, then $I(f_n) \uparrow I(f)$. (Convergence of f_n is point-wise)

This can be easily defined on simple function $f = \sum_{j=1}^{n} c_j 1_{E_j}$ as follows:

$$I(f) = \sum_{j=1}^{n} c_j \mu(E_j)$$

We can restrict it to a subset $A \subseteq X$:

$$\int_A f d\mu := \sum_{j=1}^{n} c_j \mu(E_j \cap A)$$

However, we require $c_j \geq 0$. Otherwise, if $\mu(E_i) = \mu(E_j) = \infty$ and $c_i = 1, c_j = -1$, it is difficult to determine the value of $\infty - \infty$.
$I(f)$ defined for simple function f is well-defined.

Proof. If $f = \sum_{k=1}^{n} d_k \mu(F_k)$ is another way to express f. Then note that if $E_j \cap F_k \neq \emptyset$, pick $w \in E_j \cap F_k$. By definition of f,

$$f(w) = c_j = d_k \quad (1)$$

For each j, since $\{F_k\}$ is a partition,

$$\mu(E_j) = \sum_{k=1}^{m} \mu(E_j \cap F_k)$$

So

$$\sum_j c_j \mu(E_j) = \sum_{j,k \; s.t. \; E_j \cap F_k \neq \emptyset} c_j \mu(E_j \cap F_k) = \sum_{j,k \; s.t. \; E_j \cap F_k \neq \emptyset} d_k \mu(E_j \cap F_k) \quad \text{by (1)}$$

$$= \sum_{1 \leq j \leq n, 1 \leq k \leq m} d_k \mu(E_j \cap F_k) = \sum_{1 \leq k \leq m} d_k \mu(F_k) \text{ as} \{E_j\} \text{ is a partition}$$

■

Here is an example of Lebesgue integral (though we have not defined Lebesgue measure, but we can do Lebesgue integration on simple functions). $X = (0, 1]$, $E = \mathbb{Q} \cap X$. $f = 1_{E^C}$ is simple(takes 1 at irrational x and takes 0 at rational x). So by definition $I(f) = \lambda(E^C) = 1$. Note the riemann integral $\int_0^1 f(x)\, dx$ does not exists as the function is not piece-wise continuous.

Proposition 5.12 (Monotonicity of Integral)**.** *If f, g are simple functions and $f \leq g \Rightarrow I(f) \leq I(g)$.*

Proof. Assume $f = \sum_{j=1}^{n} c_j 1_{E_j}, g = \sum_{k=1}^{m} d_k 1_{F_k}$. Then if $E_j \cap F_k \neq \emptyset$, by condition $f \leq g$, $c_j \leq d_k$. And as in proof of $I(f)$ is well-defined, $f = \sum_{j,k} c_j 1_{E_j \cap F_k}, g = \sum_{j,k} d_k 1_{E_j \cap F_k}$. So

$$I(f) = \sum_{j,k} c_j \mu(E_j \cap F_k) \leq \sum_{j,k} d_k \mu(E_j \cap F_k) = I(g)$$

. ■

Also, integral defined for simple function is linear. The proof is easy, left as exercise.

5.2 Defining the Integral

Now in order to extend this definition to general functions, we need the following proposition:

Proposition 5.13. *If $f : X \to \overline{\mathbb{R}^+}$ is measurable, where $\overline{\mathbb{R}^+} = \mathbb{R}^+ \cup \{\infty\}$. There exists set of functions $(f_n)_{n \geq 1}$ s.t. f_n are simple functions with $f_n \geq 0$ and $f_n \uparrow f$.*

Then we can simply define $I(f) := \lim_{n\to\infty} I(f_n)$.
For general $f : X \to \overline{\mathbb{R}}$, we use the trick used in theorem5.9: define $f^+ := \max\{f, 0\}, f^- := \max\{-f, 0\}$, both have codomain $\overline{\mathbb{R}^+}$. Then $f = f^+ - f^-$, we can define $I(f) := I(f^+) - I(f^-)$ if $I(f^+), I(f^-) < \infty$. And we define $I(f) = \infty$ if $I(f^+) = +\infty, I(f^-) < +\infty$. (They can not be simultaneously $+\infty$ as that raise the problem of $\infty - \infty$) The integrals on subset $\int_A f d\mu$ is defined similarly.

Proof. Since simple functions must be bounded, we first set $f_n(x) := n$ if x is s.t. $f(x) \geq n$. For $f(x) \in [0, n]$, we divide it into 2^n half-open intervals. If $k/2^n \leq f(x) < (k+1)/2^n$ for some $k \in \{1, 2, ..., n2^n - 1\}$, then $k \leq 2^n f(x) < k+1$, which means $k = \lfloor 2^n f(x) \rfloor$. So define $f_n(x) := \frac{\lfloor 2^n f(x) \rfloor}{2^n}$ when $f(x) < n$. f_n is a simple function, and $f_n \leq f_{n+1}$.
Then we have to prove $f_n \to f$. By algebraic rules of measurable functions, limit of f_n is measurable if exists, so this statement make sense.
If $f(x) = +\infty$, then $f_n(x) = n \; \forall n$. So $f_n(x) \to f(x) = \infty$.
If $f(x) < \infty$, pick n_0 s.t. $f(x) < n_0$. $\forall n \geq n_0$, we have $f_n(x) = \frac{\lfloor 2^n f(x) \rfloor}{2^n}$. Since $a - 1 < \lfloor a \rfloor \leq a$,

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

so as $n \to \infty$, $f_n(x) \to f(x)$. ■

We have to prove $I(f) := \lim I(f_n)$ is well-defined ((f_n) is the sequence of simple functions taken according to the proposition above)

Proposition 5.14 (Well-definedness of integral)**.** *If f_n, g_n are non-negative simple functions s.t. $f_n, g_n \uparrow f$. Then $\lim I(f_n) = \lim I(g_n)$.*

Proof. It is enough to show for any non-negative simple function g s.t. $g \leq \lim f_n$, we have $I(g) \leq \lim_n I(f_n)$ (*). Why? If this statement is true, we have $g_k \leq f = \lim f_n$ so $I(g_k) \leq \lim_n I(f_n)$. Taking limit over k will not change inequality as RHS is independent of k, so $\lim_k I(g_k) \leq \lim_n I(f_n)$). By symmetry, $\lim_k I(f_k) \leq \lim_n I(g_n)$). So the proposition follows.

- One value case: if $g = c1_E \leq \lim f_n$ where $E \in \mathcal{F}$. If $c = 0$, $I(g) = 0$. The inequality (*) is trivially true as f_n is non-negative.
  Now assume $c > 0$, fix $\epsilon \in (0, c)$. Now $1_E f_n := 1_E(x) f_n(x)$ is a new simple function that is non-negative, and satisfies
  $$1_E f_n \leq f_n \quad \lim_n 1_E f_n \geq g$$
  second inequality is true because: if $x \in E$, $\lim_n 1_E f_n(x) = \lim_n f_n(x) \geq g(x)$. Otherwise, LHS and RHS are both 0 as g is supported on E.
  Now define
  $$A_n := \{x \in E : f_n(x) \geq c - \epsilon\}$$
  It satisfies
  $$A_n \subseteq A_{n+1} \quad \bigcup_n A_n = E$$
  proof is left as exercise. Now we have for all n,
  $$\begin{aligned} I(f_n) \geq I(f_n 1_{A_n}) &\geq I((c-\epsilon)1_{A_n}) \quad \text{as } f_n 1_{A_n} \geq (c-\epsilon)1_{A_n} \\ &= (c-\epsilon)\mu(A_n) \end{aligned}$$
  taking limit $n \to \infty$,
  $$(c-\epsilon)\mu(E) = (c-\epsilon)\lim_n \mu(A_n) \leq \lim_n I(f_n)$$
  this is true for any ϵ, so send ϵ to 0 yields
  $$I(g) = c\mu(E) \leq \lim_n I(f_n)$$
- General Case: if $g = \sum_{k=1}^m c_k 1_{E_k} \leq \lim f_n$, then $I(g) = \sum_{k=1}^m I(c_k 1_{E_k})$. Note $\lim_n f_n 1_{E_k} \geq c_k 1_{E_k}$, so using the result in the one-value case, $I(c_k 1_{E_k}) \leq \lim_n I(f_n 1_{E_k})$. So
  $$\begin{aligned} I(g) &\leq \sum_{k=1}^m \lim_n I(f_n 1_{E_k}) = \lim_n \sum_{k=1}^m I(f_n 1_{E_k}) \quad \text{the sum is finite, so can exchange it with limit.} \\ &= \lim_n I(\sum_{k=1}^m f_n 1_{E_k}) = \lim_n I(f_n) \quad \text{by linearity of integral} \end{aligned}$$

■

Definition 5.15 ($\mu-$almost everywhere)**.** If $(X, \mathcal{A}, \mu)$ is a measure space. Property $P(x), x \in X$ hold $\mu-$almost everywhere if
$$\mu\left(\{x \in X : \neg P(x)\}\right) = 0$$
.

5.3 Properties of Integral

In this section, we use $\int f\,d\mu$ for $I(f)$. $\int_A f\,d\mu$ means $I(f1_A)$

Proposition 5.16. *If $E \subset X$ has measure 0, then for any function f,*

$$\int_E f\,d\mu = 0$$

Proof. Left as execise. ∎

Proposition 5.17 (Linearity)**.** *I is a linear functional. i.e. $I(f+g) = I(f)+I(g), I(cf) = cI(f)$ if f,g are measurable functions and c is a constant.*

Proof. Left as exercise. For simple functions linearity is clear. Then use definition of general integral to prove linearity. ∎

Proposition 5.18 (Positive Semi-definite)**.** *If $g \geq 0$, then $I(g) \geq 0$*

Proof. Left as exercise. ∎

Note from the above two propositions, if $f \geq g \geq 0$, we can write $f = g + (f-g)$. By linearity, $I(f) = I(g) + I(f-g)$. $f - g \geq 0$ so $I(f-g) \geq 0$ by positive semi-definiteness, so $I(f) \geq I(g)$. Monotonicity of I can also be proved directly from the definition:

Proposition 5.19 (Monotone)**.** *I is monotone.*

Proof. Given $0 \leq g \leq f$, pick simple functions f_n, g_n s.t. $f_n \uparrow f, g_n \uparrow g$. Define $h_n := f_n \wedge g_n \leq f_n$. We see $\lim_{n\to\infty} h_n = f \wedge g = g$, and h_n must also be simple (Exercise). So

$$I(g) = \lim_{n\to\infty} I(h_n) \leq \lim_{n\to\infty} I(f_n) = I(f)$$

inequality is due to monotonicity of I on simple functions. ∎

Remark. By Proposition5.16, this proposition can be loosen to:

$$g \leq f \quad \mu\text{-almost everywhere } \Rightarrow \int g\,d\mu \leq \int f\,d\mu$$

We need to build a collection of functions whose integral does not blow up, as defined below:

Definition 5.20 (Integrable)**.** A measurable function f is integrable if $I(f)$ is properly defined and $I(f) < \infty$ (equivalent to $I(f^-), I(f^+) < \infty$ or equivalent to $I(|f|) < \infty$). The set of integrable functions on measure space $(X, \mathcal{F}, \mu)$ is called $L^1((X, \mathcal{F}, \mu))$. Note in general

$$L^p((X, \mathcal{F}, \mu)) := \left\{ f \text{ measurable } : \left(\int |f|^p\,d\mu \right)^{1/p} < \infty \right\}$$

More about L^p space will be discussed in chapter 6.

Lemma 5.21. *Given a set $E \subset X$ and function $f : X \to \mathbb{R}$. If f is integrable, then $1_E f$ is integrable.*

Proof. $(1_E f)^+ = 1_E f^+ \leq f^+$, $(1_E f)^- = 1_E f^- \leq f^-$. So

$$I((1_E f)^+), -I((1_E f)^-) < \infty$$

by monotonicity. ■

Lemma 5.22. *If f is integrable, then cf is integrable for any $c \in \mathbb{R}$.*

Proof. The result is clear for $c \geq 0$. If $c < 0$, $(cf)^+ = -cf^-$ so $I((cf)^+) = -cI(f^-) < \infty$ by linearity. Similarly, $(cf)^- = -cf^+$, so $I((cf)^-) = -cI(f^+) < \infty$. So cf is integrable. ■

Lemma 5.23. *If f, g are integrable, then $f + g$ is integrable.*

Proof. Exercise. ■

Proposition 5.24. *If $A, B \in \mathcal{F}$ where $(X, \mathcal{F}, \mu)$ is a measure space s.t. $A \cap B = \emptyset$. $f, g : X \to \mathbb{R}$ are measurable functions, then*

$$\int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu$$

Proof. It is easy to see that $(1_{A\cup B} f)^+ = 1_{A\cup B} f^+ = 1_A f^+ + 1_B f^+$. If we change $+$ to $-$, the same works. So

$$\int_{A\cup B} f \, d\mu = \int 1_{A\cup B} f \, d\mu = \int (1_{A\cup B} f)^+ \, d\mu - \int (1_{A\cup B} f)^- \, d\mu$$

$$= \int (1_A f)^+ \, d\mu + \int (1_B f)^+ \, d\mu - \int (1_A f)^- \, d\mu - \int (1_B f)^- \, d\mu = \int 1_A f \, d\mu + \int 1_B f \, d\mu$$

■

Note, by this proposition, $\int_A f \, d\mu \leq \int_B f \, d\mu$ if $A \subseteq B$.

Proposition 5.25. *If f is integrable, then $|f| < \infty$ μ-almost everywhere*

Proof. We define $A_n := \{x : |f(x)| \geq n\}$, note $\bigcap_n A_n = \{x : |f(x)| = \infty\}$. Our target is to find a bound on $\mu(A_n)$.
Note $\int_{A_n} |f| \, d\mu \leq \int |f| \, d\mu$ by last proposition, and

$$\int_{A_n} |f| \, d\mu \geq \int n 1_{A_n} \, d\mu = n\mu(A_n)$$

so $\mu(A_n) \leq \frac{1}{n} \int |f| \, d\mu$. f is integrable so $\int |f| \, d\mu < \infty$

$$\mu(\{x : |f(x)| = \infty\}) = \lim_{n\to\infty} \mu(A_n) = 0$$

■

Proposition 5.26. *If $\int f \, d\mu < \infty$, then $f < \infty$ μ-almost everywhere.*

Proof. exactly the same as the above proposition. ■

Proposition 5.27 (Triangle inequality)**.**

$$\left|\int f \, d\mu\right| \leq \int |f| \, d\mu$$

Proof.

$$\left|\int f \, d\mu\right| = \left|\int f^+ \, d\mu - \int f^- \, d\mu\right| \leq \int f^+ \, d\mu + \int f^- \, d\mu$$

$$= \int f^+ + f^- \, d\mu = \int |f| \, d\mu$$

we do not have to worry about integrability of f here. If $\left|\int f \, d\mu\right| = \infty$, then $\int |f| \, d\mu = \infty$. ■

Proposition 5.28. *If $f \geq 0$, $\int f \, d\mu = 0 \Leftrightarrow f = 0$ almost everywhere*

Proof. $\Leftarrow$ is an easy consequence of Proposition5.16
($\Rightarrow$):
Define $A_n := \{x : f(x) \geq 1/n\}$, note $\bigcup_n A_n = \{x : f(x) > 0\}$. Again we aim to bound $\mu(A_n)$, note $1_{E_n} f \geq \frac{1}{n} 1_{E_n}$. So

$$\frac{1}{n}\mu(A_n) = \int \frac{1}{n} 1_{E_n} \, d\mu \leq \int f(x) 1_{E_n} \, d\mu \leq \int f(x) d\mu = 0$$

therefore, $\mu(A_n) = 0$. So $\mu(\{x : f(x) > 0\}) = 0$. Similarly one can show $\mu(\{x : f(x) < 0\}) = 0$. So $\mu(\{x : f(x) \neq 0\}) = 0$. ■

Proposition 5.29. *$f = g$ a.e. $\Rightarrow \int f \, d\mu = \int g \, d\mu$*

Proof. Let $E := \{x : f(x) = g(x)\}$ we have $\mu(E^C) = 0$

$$\int f \, d\mu = \int (1_E + 1_{E^C}) f \, d\mu = \int 1_E f \, d\mu + \int 1_{E^C} f \, d\mu$$

By definition of E, f is the same as g on E, so

$$= \int 1_E g \, d\mu + \int 1_{E^C} f \, d\mu = \int 1_E g \, d\mu = \int 1_E g \, d\mu + (0 =) \int 1_{E^C} g \, d\mu$$

where the last two equalities are because of $\mu(E^C) = 0$ and Proposition5.16. ■

Proposition 5.30. *If $h : X \to \overline{\mathbb{R}}$ is measurable, and $|h| \leq f$ where f is integrable, then h is integrable.*

Proof. Left as exercise. Note $|h| = h^+ + h^-$. ■

Corollary 5.31. *If f is measurable and $|f| \leq c$(a constant) on some subset $E \in \mathcal{F}$ s.t. $\mu(E) < \infty$, and $f = 0$ on E^C. Then f is integrable.*

Proof. Left as exercise. ■

Proposition 5.32 (Markov's inequality)**.** *Given measure space $(X, \mathcal{A}, \mu)$, and $f \geq 0$ is measurable. $\forall M > 0$,*

$$\mu(\{f \geq M\}) \leq \frac{\int f \, d\mu}{M}$$

Proof. See PS4. The proof is based on the simple fact that $f \geq f 1_{f \geq M}$ and monotonicity of integral. ■

Theorem 5.33 (Monotone Convergence Theorem(Beppo Levi theorem))**.** *If $f_n \uparrow f$ and $f_n \geq 0$ are integrable, then*

$$\int f_n \, d\mu \uparrow \int f \, d\mu$$

Note, the equation allows $\infty = \infty$ case.
If further given $\sup \int f_n \, d\mu < \infty$, then f is integrable. i.e. $f \in L^1$.

Proof. Since $f_n \leq f$, $\int f_n \, d\mu \leq \int f \, d\mu$, we have

$$\limsup_n \int f_n \, d\mu \leq \int f \, d\mu$$

the other direction is more difficult and we need to construct a sequence of simple functions $g_k \uparrow f$.
For each n, we take sequences of simple functions $g_{n,k} \uparrow f_n$. So we have

$$\begin{array}{lllll} g_{1,1} & g_{1,2} & g_{1,3} & \cdots & \uparrow f_1 \\ g_{2,1} & g_{2,2} & g_{2,3} & \cdots & \uparrow f_2 \\ g_{3,1} & g_{3,2} & g_{3,3} & \cdots & \uparrow f_1 \\ \cdots & & & & \end{array}$$

Let $g_1 := g_{1,1}$. It seems great to take $g_2 = g_{2,2}$, but since they are on different lines, we cannot guarantee $g_1 \leq g_2$. So we take $g_2 := \max(g_{1,2}, g_{2,2})$. Similarly, take $g_k := \max_{n \leq k}(g_{n,k})$.
It is easy to see that g_k are simple and $g_k \leq g_{k+1}$. Now we have to prove $g_k \uparrow f$:
Note $g_k \leq f_k$ as for any $n \leq k$, $g_{n,k} \leq f_n \leq f_k$.
Fix m and take $k \geq m$, we have $g_k \geq g_{m,k}$, so

$$g_{m,k} \leq g_k \leq f_k$$

sending k to ∞ we have

$$f_m \leq \lim_k g_k \leq f$$

then send m to infinity we see $\lim_k g_k = f$.

Now

$$\int f \, d\mu = \lim_k \int g_k \, d\mu \leq \liminf_k \int f_k \, d\mu \quad \text{as } g_k \leq f_k$$

so

$$\limsup_n \int f_n \, d\mu \leq \int f \, d\mu \leq \liminf_k \int f_k \, d\mu$$

the result follows. To prove f is integrable, we need Fatou's lemma. (see below lemma5.36) ■

Remark. Only pointwise convergence of f_n is required, this does not hold for Riemann integral. (where uniform convergence is required) This is one of the powerful properties of Lebesgue integrals.

Corollary 5.34. *If $f \geq 0$, define $\mu_f(A) := \int_A f \, d\mu$. μ_f is a measure.*

Proof. Left as exercise. ■

Remark. This corollary is important in probability theorem. Here f is called a density, or denoted as $\frac{d\mu}{d\mu_f}$. And if μ is a probability measure, μ_f can be called a distribution with density f.

Corollary 5.35. *If $f_n \downarrow f$ and f_n are integrable, then*

$$\int f_n \, d\mu \downarrow \int f \, d\mu$$

If further given $\sup \int f_n \, d\mu < \infty$*, then f is integrable. i.e. $f \in L^1$.*

Proof. Direct consequence if we take $g_n := -f_n$ and use Monotone convergence theorem. ■

Lemma 5.36 (Fatou's lemma)**.** *If $f_n \geq 0$, then*

$$\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu$$

Proof. Let $g_k := \inf_{n \geq k} f_n \leq f_n \ \forall n \geq k$. So by monotonicity,

$$\int g_k \, d\mu \leq \inf_{n \geq k} \int f_n \, d\mu \quad (1)$$

Note $\lim_k g_k = \liminf_n f_n$ by definition and $g_k \leq g_{k+1}$ for all k. So $g_k \uparrow \liminf_n f_n$ and clearly, $g_k \geq 0$. Then by monotone convergence theorem and (1),

$$\int \liminf f_n \, d\mu = \lim_k \int g_k \, d\mu \leq \lim_k \inf_{n \geq k} \int f_k \, d\mu$$

■

Remark. Note, if we have $h_n \leq 0$, define $f_n := -h_n \geq 0$. Then by Fatou's lemma,

$$\int \liminf(-h_n) \, d\mu \leq \liminf \int -h_n \, d\mu$$

$$\Rightarrow -\int \limsup(h_n) \, d\mu \leq -\limsup \int h_n \, d\mu$$

$$\Rightarrow \int \limsup(h_n) \, d\mu \geq \limsup \int h_n \, d\mu$$

More generally, if $f_n \geq g$ where g is integrable,

$$\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu \quad (F1)$$

and if $f_n \leq g$ where g is integrable,

$$\int \limsup(h_n) \, d\mu \geq \limsup \int h_n \, d\mu \quad (F2)$$

Fatou's lemma has an important consequence:

Theorem 5.37 (Dominated Convergence)**.** *If $f_n \to f$ and $|f_n| \le g$ where g is integrable, then f is integrable and*

$$\int f_n \, d\mu \to \int f \, d\mu$$

Proof. It is easy to see that f_n are integrable as it is bounded by integrable function. f is the point-wise limit of f_n, so $|f| \le g$. Then f is integrable. Note $f_n \le g$, $f_n \ge -g$. So we can use two versions of general Fatou's lemma:

$$\limsup \int f_n \, d\mu \le \int \limsup f_n \, d\mu \qquad \text{by (F2)}$$

$$= \int f \, d\mu = \int \liminf f \, d\mu \qquad \text{as } f_n \to f$$

$$\text{by (F1)} \qquad \le \liminf \int f_n \, d\mu \le \limsup \int f_n \, d\mu$$

So we have $\int f \, d\mu = \limsup \int f_n \, d\mu$. The result follows. ■

Remark. The condition $f_n \to f$ can be loosen to $f_n \to f$ μ-a.e. And with assumption $\mu(X) < \infty$, the condition $f_n \to f$ a.e. can be replaced with $f_n \xrightarrow{\mu} f$ (convergence in measure, see chapter 6), the conclusion still holds. Unfortunately, the condition $|f_n| \le g$ for some integrable g cannot be removed. Otherwise, theorem fails. (counterexample: $f_n = \frac{1}{n} 1_{[0,n]} \downarrow 0$)

5.4 Lebesgue Integral

Definition 5.38 (Lebesgue Integral)**.** Given measure space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$ where λ is Lebesgue measure, $\int f \, d\lambda$ for $f \in L^1(\lambda)$ is the Lebesgue Integral. We usually complete the Borel σ-algebra and define Lebesgue integral on $(\mathbb{R}^n, \overline{\mathcal{B}(\mathbb{R}^n)}, \lambda)$. Recall from Theorem 4.7 that $\overline{\mathcal{B}(\mathbb{R}^n)} = \mathcal{B}(\mathbb{R}^n) \cup \{\lambda\text{-null sets}\}$.

A function is Lebesgue integrable if $\int |f| \, d\lambda < \infty$

Lebesgue integral is the generalisation of Riemann integral:

Proposition 5.39. *If bounded function $f : I \to R$ where $I = [a, b]$ is Riemann integrable. (This means $U(f, a, b) = L(f, a, b)$ with $|\int_a^b f(x) \, dx| < \infty$), then $1_I f$ is Lebesgue integrable and*

$$\int_I f \, d\lambda = \int_a^b f(x) \, dx$$

where LHS is Lebesgue integral and RHS is Riemann integral.

But a Lebesgue integrable function is not necessarily Riemann integrable.

Proof. Let Partition $\mathcal{P}$ of $[a,b]$ be $a =: x_0, x_1, ..., x_n := b$.
Let

$$M := \sup_{x\in[a,b]} |f(x)|, \quad m := \inf_{x\in[a,b]} |f(x)|$$
$$M_j := \sup_{x\in[x_{j-1},x_j]} |f(x)|, \quad m_j := \inf_{x\in[x_{j-1},x_j]} |f(x)|$$

and define simple functions

$$g_{\mathcal{P}} := \sum_{j=1}^{n} m_j 1_{(x_{j-1},x_j]}, \quad h_{\mathcal{P}} := \sum_{j=1}^{n} M_j 1_{(x_{j-1},x_j]}$$

Since f is Riemann integrable, we can pick increasing partitions $\mathcal{P}_n$ s.t.

$$\lim_{n\to\infty} U(f,\mathcal{P}_n) = \lim_{n\to\infty} L(f,\mathcal{P}_n) = \int_a^b f(x)\,dx$$

let $g_n := g_{\mathcal{P}_n}, h_n := h_{\mathcal{P}_n}$. By definition of Lebesgue integral:

$$U_n := U(f,\mathcal{P}_n) = \int_{[a,b]} h_n d\lambda, \quad L_n := L(f,\mathcal{P}_n) = \int_{[a,b]} g_n d\lambda$$

and we also have $g_n \le f \le h_n$. (g_n) is increasing and is bounded above by M, so g_n converges point-wise to some measurable function g. Similarly h_n converges to some measurable function h. And $g \le f \le h$.
Since $\max_{n\in\mathbb{N}}\{|g_n|, |h_n|\} \le M$, by dominated convergence theorem, g, h are Lebesgue integrable on $[a,b]$ and

$$\int_{[a,b]} g\,d\lambda = \lim_{n\to\infty} L_n(f) = \int_a^b f(x)\,dx = \lim_{n\to\infty} U_n(f) = \int_{[a,b]} h\,d\lambda$$

where for the first and last equality, we used monotone convergence theorem to exchange limit and integral.
So $\int_{[a,b]}(h-g)\,d\lambda = 0$ by linearity of Lebesgue integral. Therefore, $h(x) = g(x)$ almost everywhere. So $f = g$ a.e. This implies f is measurable(Note: this requires the property that $(\mathbb{R}^n, \overline{\mathcal{B}(\mathbb{R}^n)}, \lambda)$ is complete, this is not true for arbitrary measure) and integrable. And

$$\int_{[a,b]} f\,d\lambda = \int_{[a,b]} g\,d\lambda = \int_a^b f(x)\,dx$$

■

5.5 Product measure and Fubini's theorem

Our goal that given two measure spaces $(X_j, \mathcal{F}_j, \mu_j), j \in \mathbb{N}$, we want to find a proper measure on the countable product of X_j. (Call it X)

First consider the simple case, defining measure on $X := X_1 \times X_2$ s.t.

$$\mu(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2) \quad \forall\, E_1 \in \mathcal{F}_1,\, E_2 \in \mathcal{F}_2$$

Now we define $\mathcal{F}_1 \times \mathcal{F}_2$ as the set of "rectangles" i.e. $\mathcal{F}_1 \times \mathcal{F}_2 := \{E_1 \times E_2 : E_1 \in \mathcal{F}_1,\, E_2 \in \mathcal{F}_2\}$.

Proposition 5.40. *$\mathcal{F}_1 \times \mathcal{F}_2$ is a semi-algebra*

Proof.
- It is clear that X is in $\mathcal{F}_1 \times \mathcal{F}_2$ as $X = X_1 \times X_2$.
- If $A = E_1 \times E_2, B = F_1 \times F_2$ with $E_1, F_1 \in \mathcal{F}_1,\ E_2, F_2 \in \mathcal{F}_2$, then $A \cap B = (E_1 \cap F_1) \times (E_2 \cap F_2) \in \mathcal{F}_1 \times \mathcal{F}_2$.
- It is left as an exercise to show that if $A := E_1 \times E_2$, $A^C = (E_1^C \times E_2) \cup (E_1 \times E_2^C) \cup (E_1^C \times E_2^C)$. So A^C is a finite union of sets from $\mathcal{F}_1 \times \mathcal{F}_2$.

■

A nature definition of a pre-pre measure $\nu : \mathcal{F}_1 \times \mathcal{F}_2 \to \overline{\mathbb{R}^+}$ is

$$\nu(E_1 \times E_2) := \mu_1(E_1)\mu_2(E_2)$$

define $\mathcal{F} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ and we will use extension of measure to construct a measure μ on $\mathcal{F}$. We just need to show that it is σ-additive.

Proposition 5.41. *ν is σ-additive on $\mathcal{F}_1 \times \mathcal{F}_2$*

Proof. Given $A \in \mathcal{F}$. First we need the concept of slice: given $x \in X_1$, define A_x or written as $A(x)$ as $\{y \in X_2 : (x, y) \in A\} \subseteq X_2$. It is a slice of set A along X_2. Similarly we can define $A(y)$ given $y \in X_2$.
Claim. $A_x \in \mathcal{F}_2,\ A_y \in \mathcal{F}_1$.
Let

$$\mathcal{C} = \{A \in \mathcal{F} : \forall x \in X_1, A_x \in \mathcal{F}_2\}$$

this technique has been used for many times and it should be clear now that we have two goals: $\mathcal{C}$ contains $\mathcal{F}_1 \times \mathcal{F}_2$, $\mathcal{C}$ is σ-algebra.

- We note that if $A = E_1 \times E_2 \in \mathcal{F}_1 \times \mathcal{F}_2$,

$$A_x = \begin{cases} E_2 & \text{if } x \in E_1 \\ \emptyset & \text{if } x \notin E_1 \end{cases}$$

  so clearly $A_x \in \mathcal{F}_2$, then $\mathcal{C} \supseteq \mathcal{F}_1 \times \mathcal{F}_2$.
- Now we prove $\mathcal{C}$ is σ-algebra
  - $(A_x)^C = (A^C)_x$ by definition. So if $A_x \in \mathcal{F}_2$, $(A^C)_x \in \mathcal{F}_2$.
  - $X(x) = X_2 \in \mathcal{F}_2$ so $X \in \mathcal{C}$.
  - Given disjoint $A_j \in \mathcal{C}$, fix $x \in X_1$,

$$(\bigcup_{j\geq 1} A_j)_x = \left\{ y \in X_2 : (x, y) \in \bigcup_{j\geq 1} A_j \right\}$$

$$= \bigcup_{j\geq 1} \{y \in X_2 : (x, y) \in A_j\} = \bigcup_{j\geq 1} (A_j)_x \in \mathcal{F}_2$$

Therefore, $\mathcal{C}$ contains $\mathcal{F} := \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$.

Claim proved $\square$

We first show that ν is additive. The same arguments apply to show that it is σ-additive. Now given $A = E \times F$ where $E \in \mathcal{F}_1, F \in \mathcal{F}_2$ and $A = \sum_{j=1}^n A_j$ (A_j are disjoint) s.t. $A, A_j \in \mathcal{F}_1 \times \mathcal{F}_2$. Assume $A_j = E_j \times F_j$, we can show that

$$A_x = \sum_{j=1}^n (A_j)_x$$

note if $j \neq k$, $(A_j)_x \cap (A_k)_x = \emptyset$, so using $\sum$ is valid. Now we can write $(A_j)_x$ using F_j as below.

$$A_x = \sum_j^n F_j 1_{E_j}(x)$$

which is a simple function, so must be measurable. So if $x \in E$,

$$\mu_2(F) = \sum_{j=1}^n \mu_2(F_j) 1_{E_j}(x)$$

we can write this as $1_E(x)\mu_2(F)$ for general $x \in X_1$.

$$\Rightarrow \int 1_E \mu_2(F) \, d\mu_1 = \int \sum_{j=1}^n \mu_2(F_j) 1_{E_j}(x) \, d\mu_1$$

$$\Rightarrow \mu_1(E)\mu_2(F) = \sum_{j=1}^n \int \mu_2(F_j) 1_{E_j}(x) \, d\mu_1$$

we can exchange the sum because sum is finite, for infinite sum, use dominated convergence theorem. So we have

$$\mu(A) = \sum_{j=1}^n \mu_1(E_j)\mu_2(F_j) = \sum_{j=1}^n \mu(A_j)$$

$\blacksquare$

So we define the product measure μ(written as $\mu_1 \otimes \mu_2$) on X as the measure taken from Hahn-Caratheodory theorem based on ν. Note that uniqueness product measure requires measure spaces $(X_j, \mathcal{F}_j, \mu_j)$, $j = 1, 2$ to be σ-finite.

Define coordinate maps $Z_i : X \to X_i$ to be $x = (x_1, x_2) \mapsto x_i$ for $i = 1, 2$. Then $\mathcal{F} = \sigma(\{Z_i^{-1}(A_i) \, : \, A_i \in \mathcal{F}_i, \ i = 1, 2\})$

If we define λ^n to be the Lebesgue measure of $\mathbb{R}^n$, then

$$\lambda^n \otimes \lambda^m = \lambda^{m+n}$$

Product measure on countable spaces

Now attempt to generalise this to countable product of measure spaces. Denote $X := \prod_{j\geq 1} X_j$ and let $\mathcal{C} = \{E_1 \times E_2 \times \ldots \times E_n \times X_{n+1} \times \ldots : E_j \in X_j,\, n \in \mathbb{N}\}$ (called set of cylinders) and let $\mathcal{F} = \sigma(\mathcal{C})$. We want to find a measure μ on $\mathcal{F}$ s.t.

$$\mu(E) = \prod_{j=1}^{n} \mu_j(E_j) \prod_{j>n} \mu_j(X_j) \text{ for } E \in \mathcal{C}$$

for convenience assume that $\mu_j(X_j) = 1\ \forall j$. Our goal becomes

$$\mu(E) = \prod_{j=1}^{n} \mu_j(E_j)$$

this is quite similar to previous case, and it turns out that similar procedures prove $\mathcal{C}$ is a semi-algebra. We can not take the set of cylinders to be $\prod_{j=1}^{\infty} E_j$ as in that case $\mathcal{C}$ fails to be a semi-algebra.
Again we can define $\nu(E) = \prod_{j=1}^{n} \mu_j(E_j)$ on $\mathcal{C}$, and similar proof yields ν is additive. However, we cannot prove that ν defined on $\mathcal{C}$ is σ-additive easily. The solution is to extend ν to $\tilde{\mu}$ on $\mathcal{A}(\mathcal{C})$ instead and prove that this is continuous from above. Then by Proposition2.11, $\tilde{\mu}$ is σ-additive. So then we extend to a measure μ on $\mathcal{F} = \sigma(\mathcal{C}) = \Sigma(\mathcal{A}(\mathcal{C}))$ using Hahn-Caratheodory theorem.

The proof of $\tilde{\mu}$ is continuous from above is elementary but very lengthy, so I will omit it here. Again the "slice" of A, A_x, is used, the proof shares idea of integration with proof of Proposition5.41.

Fubini's theorem

Given two measure spaces $(X_j, \mathcal{F}_j, \mu_j), j = 1, 2$, use the product measure and product σ-algebra we used above, μ, $\mathcal{F} := \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. We aim to show that we can swap integrals $d\mu_1, d\mu_2$.

Theorem 5.42 (Fubini's theorem)**.** *Assume two measure spaces are σ-finite. Given $f : X_1 \times X_2 \to \overline{\mathbb{R}}$, fix $x \in X_1$, we will denote $f(x,y)$ as $f_x(y)$. If f is integrable, then we have*

$$\int \left[\int f_x(y)\, d\mu_2(y)\right]\, d\mu_1(x) = \int f\, d\mu$$

note, $d\mu_2(y)$ emphasises that we are integrating w.r.t. y, but the definition is the same as $d\mu_2$.
And we need to first prove that $f_x(y)$ is $\mathcal{F}_2$-measurable for fixed f and $\int f_x(y)\, d\mu_2(y)$ is $\mathcal{F}_1$-measurable.

Remark. There is another half of Fubini's theorem, which is

$$\int \left[\int f_y(x)\, d\mu_1(x)\right]\, d\mu_2(y) = \int f\, d\mu$$

but the proof is completely symmetric, so there is no need to worry about this side during the proof.

The following lemma ensures integration is valid.

Lemma 5.43. *$f_x(y)$ is $\mathcal{F}_2$-measurable for any $x \in X_1$.*

Proof. Given Borel set $B \in \overline{\mathcal{B}}$,

$f_x^{-1}(B) = \{y \in X_2 : f(x, y) \in B\} = \{y \in X_2 : (x, y) \in f^{-1}(B)\} = (f^{-1}(B))_x$

since $f^{-1}(B) \in \mathcal{F}$, $(f^{-1}(B))_x \in \mathcal{F}_2$ as proved in a claim before. ∎

Now we prove a helper proposition:

Proposition 5.44. *Assume two measure spaces are σ-finite, given $E \in \mathcal{F}$, then the functions*

$$x \mapsto \mu_2(E_x),\ y \mapsto \mu_1(E_y)$$

are measurable. And

$$\int \mu_2(E_x)\, d\mu_1 = \mu(E) = \int \mu_1(E_y)\, d\mu_2$$

Proof. First assume $\mu_1(X_1), \mu_2(X_2) < \infty$, we will remove these assumptions later.
We will prove the first part of proposition, i.e. $x \mapsto \mu_2(E_x)$ is measurable, step by step

- **Rectangular case** If $E = A \times B$ for $A \in \mathcal{F}_1, B \in \mathcal{F}_2$, then we have $\mu_2(E_x) = 1_A(x)\mu_2(B)$ which is a simple function. So it must be measurable.
- **Finite union of rectangles** We could have directly go to $E \in \mathcal{F} := \sigma(F_1 \times F_2)$ case, but we will see in the next part why this is difficult. Consider $\mathcal{A} := \mathcal{A}(\mathcal{C})$, where $\mathcal{C}$ denotes set of rectangles(which is a semi-algebra). Given $E \in \mathcal{A}$, then we know $E = \sum_{j=1}^{n} E_j$ for some $E_j = A_j \times B_j \in \mathcal{C}$.

  $$\mu_2(E_x) = \mu_2(\left(\sum_{j=1}^{n} E_j\right)_x) = \mu_2(\sum_{j=1}^{n} (E_j)_x) = \sum_{j=1}^{n} \mu_2((E_j)_x) = \sum_{j=1}^{n} 1_{A_j}(x)\mu_2(B_j)$$

  this is a simple function in x, so must be measurable.
- **General Case** If we attempt to prove

  $$\mathcal{G} = \{E \in \mathcal{F} : x \mapsto \mu_2(E_x) \text{ is measurable}\}$$

  is σ-algebra, then note $\mu_2((E \cup F)_x) = \mu_2(E_x \cup F_x)$, there is no neat way to write the term using $\mu_2(E_x), \mu_2(F_x)$. So we prove the set class $\mathcal{G}$ is monotone class instead. Then because $\mathcal{G} \supseteq \mathcal{A}$, we have $\mathcal{G} \supseteq \mathcal{F} = \sigma(A) = \mathcal{M}(A)$, proving the proposition.

  This is quite straightforward, if we have $E^{(n)} \uparrow E$ where $E^{(n)} \in \mathcal{G}$, then we know $\mu_2(E_x^{(n)})$ are measurable. And we have $E_x^{(n)} \uparrow E_x$ so by σ-additivity of μ_2 (so μ_2 is continuous from below),

  $$\mu_2(E_x) = \lim_{n \to \infty} \mu_2\left(E_x^{(n)}\right)$$

limit of measurable functions is also measurable, so $x \mapsto \mu_2(E_x)$ is measurable, $E \in \mathcal{G}$. Same works for $E^{(n)} \downarrow E$.

Now remove the assumption $\mu_1(X_1), \mu_2(X_2) < \infty$. As two measures are σ-finite, there exists $A_n \in \mathcal{F}_1, B_n \in \mathcal{F}_2$ s.t. $X_1 = \bigcup A_n, X_2 = \bigcup B_n$ where $\mu_1(A_n), \mu_2(B_n) < \infty \; \forall n$. Define $F_n := A_n \times B_n$, we know $\mu(F_n) = \mu_1(A_n)\mu_2(B_n) < \infty$, and $\bigcup F_n = X := X_1 \times X_2$. $x \mapsto \mu_2((E \cap F_n)_x)$ is measurable by first part of proof (by treating $E \cap F_n$ as a set in space $X|_{F_n}$, which is a finite measure space), and we have $(E \cap F_n)_x \uparrow E_x$, so by continuity of μ_2,

$$\mu_2(E_x) = \lim_{n \to \infty} \mu_2((E \cap F_n)_x)$$

so $x \mapsto \mu_2(E_x)$ is measurable.

The integral identity is left as exercise. (Hint: follow the same idea as above: add extra condition $\mu_1(X_1), \mu_2(X_2) < \infty$, prove $\mathcal{G}$ (set of E satisfying the integral identity) is monotone class and then remove the assumption using σ-finiteness). ■

Now we prove the reduced case of Fubini's theorem, Tonelli's theorem, which only consider positive-valued functions.

Theorem 5.45 (Tonelli's theorem)**.** *Assume two measure spaces are σ-finite. Given $f : X_1 \times X_2 \to \overline{\mathbb{R}^+}$. If f is $\mathcal{F}$-measurable, then we have*

$$\int \left[\int f_x(y) \; d\mu_2 \right] \; d\mu_1 = \int f \; d\mu$$

Note here we do not require f to be integrable.

Proof. The proof process is very clear, first prove this for $f = c1_E$ where $E \in \mathcal{F}$, and then prove this for general simple functions. Finally use monotone convergence theorem to prove this for any positive measurable function.

First two parts of the proof are left as exercise (hint: use helper proposition. Remember that you have to justify the inner integral $\int f_x(y) \; d\mu_2(y)$ as a function in x is $\mathcal{F}_1$-measurable so that outer integral make sense)

General case. Given $f \geq 0$ that is $\mathcal{F}$-measurable. We can find sequence of simple functions $f^{(j)}$ s.t. $f^{(j)} \uparrow f$. So we know $f_x^{(j)} \uparrow f_x$ and using monotone convergence theorem(MCT),

$$\int f_x^{(j)}(y) \; d\mu_2 \uparrow \int f_x(y) \; d\mu_2$$

the first integral as a function of x is $\mathcal{F}_1$-measurable by previous two parts of proof. So the limit(RHS) is also $\mathcal{F}_1$-measurable. Now we can use MCT on $\mathcal{F}_1$ to get

$$\int \left[\int f_x^{(j)}(y) \; d\mu_2(y) \right] \; d\mu_1(x) \uparrow \int \left[\int f_x(y) \; d\mu_2(y) \right] \; d\mu_1(x)$$

LHS is the simple function case, so we have proved that LHS $= \int f^{(j)}\, d\mu$, by MCT on $\mathcal{F}$, this converges to $\int f\, d\mu$. But the limit is unique, so

$$\int f\, d\mu = \int \left[\int f_x(y)\, d\mu_2(y)\right]\, d\mu_1(x)$$

proof complete. ■

Finally we can prove Fubini's theorem.

Proof. We have to be very careful. Given f integrable, then $f = f^+ - f^-$ and we have $\int f^+\, d\mu,\ \int f^-\, d\mu < \infty$. And using Tonelli's theorem we have

$$\int f^+\, d\mu = \int \left[\int f_x^+(y)\, d\mu_2\right]\, d\mu_1 < \infty$$

but this only indicates $\int f_x^+(y)\, d\mu_2 < \infty$ μ_1-a.e, this integral may be ∞ somewhere. Same applies to f^-. So when attempting to prove Fubini's theorem:

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu = \int \left[\int f_x^+(y)\, d\mu_2\right]\, d\mu_1 - \int \left[\int f_x^-(y)\, d\mu_2\right]\, d\mu_1 \quad (*)$$

you cannot combine two integrals on RHS because the integrand will be $\int f_x^+(y)\, d\mu_2 - \int f_x^-(y)\, d\mu_2$ and if both integrals happens to be ∞ at some y, this expression is invalid. Luckily we are dealing with integrals here (two functions that are equal except for a set of measure 0 has the same integral value), so we can define equivalent functions g^+ which is always finite and satisfies

$$\int g^+(x)\, d\mu_1 = \int \left[\int f_x^+(y)\, d\mu_2\right]\, d\mu_1$$

this can be done by defining g^+ as below:

$$g(x) := \begin{cases} \int f_x^+(y)\, d\mu_2 & \text{if } \int f_x^+(y)\, d\mu_2 < \infty \\ 0 & \text{otherwise} \end{cases}$$

clearly $g^+(x) = \int f_x^+(y)\, d\mu_2$ μ_1-a.e. Similarly, one can define $g^-(x)$. Let $g(x) := g^+(x) - g^-(x)$, we can proceed from (*) now.

$$(*) = \int g^+(x)\, d\mu_1 - \int g^-(x)\, d\mu_1 = \int g^+ - g^-\, d\mu_1 = \int g\, d\mu_1$$

it is not hard to show that $g(x) = \int f_x(y)\, d\mu_2$ μ_1-a.e., so

$$\int g\, d\mu_1 = \int \left[\int f_x(y)\, d\mu_2\right]\, d\mu_1 = (*)$$

proof complete. ■

Remark. We can improve Fubini's theorem, by replacing assumption f is integrable by

$$\int \left[\int |f_x(y)|\, d\mu_2\right]\, d\mu_1 < \infty$$

this is because $f_x^+, f_x^- \le |f_x|$, so using Tonelli's theorem,

$$\int f_x^+ \, d\mu = \int \left[\int f_x^+(y) \, d\mu_2 \right] \, d\mu_1 \le \int \left[\int |f_x(y)| \, d\mu_2 \right] \, d\mu_1 < \infty$$

same works for f^- giving $\int f_x^- \, d\mu < \infty$. So f is integrable, we can use original Fubini's theorem.

It is mentioned in the lecture notes a theorem relating π-system and λ system. So I'll introduce something about these two systems here.

Definition 5.46 (π-system)**.** A collection $\mathcal{P}$ of subsets of X is called a π-system if

- $\mathcal{P}$ is not empty
- $\mathcal{P}$ is closed under finite intersections

Examples. Collection of intervals of form $(-\infty, a]$ is a π-system, and collection of $(a, b]$ is π-system if $\emptyset$ is included. Every topology is a π-system. And every σ-algebra is a π-system.

π-systems are useful as if two measures agree on a π-system $\mathcal{P}$, then they also agree on $\sigma(\mathcal{P})$. Other properties like integral equalities can also carry from $\mathcal{P}$ to $\sigma(\mathcal{P})$. π-system is closely related to λ-system (weakened version of σ-algebra)

Definition 5.47 (λ-system)**.** Collection $\mathcal{D}$ of subsets of X is called λ-system if

- $X \in \mathcal{D}$
- If $A \in \mathcal{D}$, then $A^C := X \setminus A \in \mathcal{D}$
- Given sequence of pairwise disjoint sets A_k in $\mathcal{D}$, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{D}$.

Remark. λ-system is indeed a weaker definition than σ-algebra. Consider $X = \{a, b, c, d\}$, and the collection $\mathcal{D} = \{\Omega, \emptyset, \{a, b\}, \{b, c\}, \{c, d\}, \{a, d\}, \{b, d\}, \{a, c\}\}$. Prove $\mathcal{D}$ is λ-system but not σ-algebra.

By definition, $\mathcal{P}$ is λ-system and π-system $\Leftrightarrow$ $\mathcal{P}$ is σ-algebra.

Theorem 5.48 (Dynkin's π-λ Theorem)**.** *Given set X and π-system $\mathcal{P}$, λ-system $\mathcal{D}$, then*

$$\mathcal{P} \subset \mathcal{D} \Rightarrow \sigma(\mathcal{P}) \subset \mathcal{D}$$

Proof. Omitted. Beyond our level. ■

6 Convergence

Definition 6.1 (σ-complete)**.** Given $f : X \to \overline{\mathbb{R}}$ that is $\mathcal{F} - \overline{\mathcal{B}(\mathbb{R})}$ measurable(simply denote this as $\mathcal{F}$ measurable), and function g s.t. $f = g$ μ-a.e. If $(X, \mathcal{F}, \mu)$ ensure that all such g will also be $\mathcal{F}$ measurable, then the space is called σ-complete

For this chapter assume that the measure space $(X, \mathcal{F}, \mu)$ is σ-finite and σ-complete.

6.1 Modes of Convergence

Note if we define relation $f \sim g$ as $f = g$ μ-a.e., this relation is equivalence relation. Define $\mathcal{M}$ as the collection of equivalent class i.e. $\mathcal{M} := \{[f] : f : X \to \overline{\mathbb{R}}\}$ where $[f]$ is equivalence class under relation $\sim$.

Definition 6.2 (point-wise convergence)**.** $f, f_n : E \to \overline{\mathbb{R}}$ where $A \subseteq X$, we say $f_n \to f$ on A if $\forall x \in A$, $f_n(x) \to f(x)$.

We can also define convergence almost everywhere on X. It means there is a set $E \in \mathcal{F}$ with $\mu(X \setminus E) = 0$ and $f_n \to f$ point-wise on E. We denote this as $f_n \to f$ μ-a.e or simply a.e. if the measure used is clear. Point-wise convergence a.e. corresponds to almost-surely convergence in probability theory.

This is well defined, as

Proposition 6.3. *If $f_n \to f$ a.e., and $f_n \sim g_n$, $f \sim g$, then $g_n \to g$ a.e.*

Proof. Pick $E \in \mathcal{F}$ with $\mu(E^C) = 0$ and $f_n \to f$ $\forall x \in E$,
pick $F_n \in \mathcal{F}$ with $\mu(F_n^C) = 0$ and $f_n = g_n$ $\forall x \in F_n$,
pick $F \in \mathcal{F}$ with $\mu(F^C) = 0$ and $f = g$ $\forall x \in F$,
define $H := E \cap F \cap \bigcap_{n \geq 1} F_n$, then $\mu(H^C) = \mu(E^C \cup F^C \cup \bigcup_{n \geq 1} F_n) = \mu(E^C) + \mu(F^C) + \sum_{n \geq 1} \mu(F^C) = 0$ and $g_n \to g$ on H. ■

So no matter what representative of the equivalence class you pick, convergence a.e. still holds.

Proposition 6.4. *If $f_n \to f$ a.e., $f_n \to g$ a.e., then $f \sim g$. This means limit is unique up to equivalence class.*

Proof. Easy proof left as exercise. ■

Recall the definition of uniform convergence

Definition 6.5 (Uniform Convergence)**.** f_n converges to f uniformly on E, denote as $f_n \to f$ unif, if

$$\forall \epsilon > 0, \exists N, \forall n \geq N, x \in E, \sup_{x \in E} |f_n(x) - f(x)| < \epsilon$$

Definition 6.6 (Uniform Convergence a.e.)**.** If there exists $E \subseteq X$ with $\mu(E^C) = 0$ and $f_n \to f$ unif on E.

Definition 6.7 (Essential supremum)**.** If there is $a > 0$ s.t. $\mu\{|f| \geq a\} = 0$, then essential supremum of f is defined as

$$\operatorname{ess\,sup} |f| := \inf\{a > 0 : \mu\{|f| \geq a\} = 0\}$$

It is obvious that $\operatorname{ess\,sup} |f| \leq \sup |f|$. And if $c := \operatorname{ess\,sup} |f|$, $\mu(\{x : |f(x)| > c\}) = 0$. This can be seen by taking increasing sets $\{x : |f(x)| > c + 1/n\}$, which all has measure 0 by definition of c.

Proposition 6.8. *Essential supremum is class property. i.e. if $f \sim g$, then $ess \sup |f| = ess \sup |g|$*

Proof. Take E s.t. $f = g$ on E and $\mu(E^C) = 0$. Let $c := \operatorname{ess\,sup} |f|$ and consider the set $A := \{x : |g(x)| > c\}$. We aim to show that $\mu(A) = 0$. Note $\mu(A) \leq \mu(A \cap E) + \mu(A \cap E^C)$ by subadditivity, and $\mu(A \cap E^C) \leq \mu(E^C) = 0$, so

$$\mu(A) \leq \mu(A \cap E) = \mu(\{x : |f(x)| > c\} \cap E) \leq \mu(\{x : |f(x)| > c\}) = 0$$

So using definition of $\operatorname{ess\,sup} |g|$ as an infimum, $\operatorname{ess\,sup} |g| \leq c = \operatorname{ess\,sup} |f|$. By symmetry, $\operatorname{ess\,sup} |f| \leq \operatorname{ess\,sup} |g|$ so they are equal. ■

Essential supremum can be used to define a distance between two equivalent classes.

Definition 6.9 (distance)**.** $d(f, g) := \operatorname{ess\,sup} |f - g|$.

Proposition 6.10. *d is a metric on $\mathcal{M}$. i.e.*

- $d(f, g) = 0 \Rightarrow f \sim g$
- $d(f, g) = d(g, f)$
- $d(f, h) \leq d(f, g) + d(g, h)$

Proof. First two properties are quite obvious. Now assume $a := \operatorname{ess\,sup} |f - g|, b := \operatorname{ess\,sup} |g - h|$, consider the set $\mu(\{x : |f(x) - h(x)| > a + b\})$, then we know

$$a + b < |f - h| \leq |f - g| + |g - h| \;\Rightarrow\; |f - g| > a \text{ or } |g - h| > b$$

so

$$\mu(\{x : |f(x)-h(x)| > a+b\}) \leq \mu(\{x : |f(x)-g(x)| > a\})+\mu(\{x : |g(x)-h(x)| > b\}) = 0$$

that means $\operatorname{ess\,sup} |f - h| \leq a + b$. ■

The point of defining essential supremum is to give a metric that is equivalent to uniform convergence a.e. Just like the infinite norm $\|f - f_n\|_\infty \to 0 \Leftrightarrow f_n \to f$ unif.

Theorem 6.11. *$f_n \to f$ unif a.e. $\Leftrightarrow d(f_n, f) \to 0$*

Proof. ($\Rightarrow$) Assume $f_n \to f$ unif a.e., there is $E \in \mathcal{F}$ with $\mu(E^C) = 0$ and $f_n \to f$ unif. on E, fix $\epsilon > 0$, there is N s.t. $\forall n \geq N$, $\sup_{x \in E} |f_n(x) - f(x)| < \epsilon$. So the set of x satisfying $|f_n - f| \geq \epsilon$ is completely outside E, i.e. it is contained in E^C, so

$$\forall n \geq N, \mu(\{x : |f_n(x) - f(x)| \geq \epsilon\}) \leq \mu(E^C) = 0$$

that means $\operatorname{ess\,sup} |f_n - f| \leq \epsilon$. So we found for arbitrary $\epsilon > 0$ there is N s.t. $\forall n \geq N$, $d(f_n, f) = \operatorname{ess\,sup} |f_n - f| \leq \epsilon$, which means $d(f_n, f) \to 0$.

($\Leftarrow$) Assume $d(f_n, f) = \operatorname{ess\,sup} |f_n - f| \to 0$, then for any $k \in \mathbb{N}$, there is a N_k s.t. for all $n \geq N_k$, $\operatorname{ess\,sup} |f_n - f| \leq \epsilon$, i.e. $\mu(\{x : |f_n(x) - f(x)| > 2^{-k}\}) = 0$.

$$\Rightarrow \mu\left(\bigcup_{n \geq N_k} \{x : |f_n(x) - f(x)| > 2^{-k}\}\right) = 0$$

as this holds for any k,

$$\Rightarrow \underbrace{\mu\left(\bigcup_{k\geq 1}\bigcup_{n\geq N_k}\{x\,:\,|f_n(x)-f(x)|>2^{-k}\}\right)}_{=:E^C}=0$$

we define this set as E^C and prove $f_n \to f$ unif on E.
But this is obvious after we write down expression of E:

$$E=\bigcap_{k\geq 1}\bigcap_{n\geq N_k}\{x\,:\,|f_n(x)-f(x)|\leq 2^{-k}\}$$

which reads as $\forall k>0$, there is N_k s.t. $\forall n \geq N_k, |f_n(x)-f(x)|\leq 2^{-k}$, which is equivalent to uniform convergence as $2^{-k}\to 0$ when $k\to\infty$. ∎

Here is an example of f_n that converges a.e. but not converge uniformly a.e.. $f_n(x)=x^n$, $X=[0,1]$. $f_n \to f=0$ a.e. (actually it converges except at $x=1$) but around any small interval at $x=1$, say $(1-\epsilon,1]$, uniform convergence fails, and $\mu((1-\epsilon,1])=\epsilon>0$, so f_n does not converge uniformly to f a.e.

Now we define $\mathcal{L}_\infty=\{f\in\mathcal{M}\,:\,\operatorname{ess\,sup}|f|<\infty\}$, and attempt to show that it is linear space. i.e. if $f,g\in\mathcal{L}_\infty,\alpha\in\mathbb{R}$, then $\alpha f+g\in\mathcal{L}_\infty$.

Proposition 6.12. *$\mathcal{L}_\infty$ is linear space*

Proof. Proof is easy, just need to show

$$\operatorname{ess\,sup}|\alpha f|=|\alpha|\operatorname{ess\,sup}|f|,\quad \operatorname{ess\,sup}|f+g|\leq\operatorname{ess\,sup}|f|+\operatorname{ess\,sup}|g|$$

left as exercise. ∎

In order to include a larger family of functions, we define convergence almost uniformly,

Definition 6.13 (almost uniform convergence)**.** $f_n\to f$ a. unif if
$\forall\epsilon>0$, there is set $E_\epsilon\in\mathcal{F}$ .s.t $\mu(E_\epsilon^C)\leq\epsilon$ and $f_n\to f$ unif. on E_ϵ.

Note we no longer require condition like $\mu(E^C)=0$. Now one can show easily that $f_n=x^n$ converges a. unif to $f=0$ on $[0,1]$.

Let us summarise three convergences discussed so far, we have

- (1) Convergence a.e.
- (2) Uniform convergence a.e.
- (3) almost Uniform convergence

It is quite easy to show (2) $\Rightarrow$ (3) and (2) $\Rightarrow$ (1), what about (1) and (3)?

Proposition 6.14. *$f_n\to f$ a.unif $\Rightarrow$ $f_n\to f$ a.e.*

Proof. Left as exercise.

Hint: take E_k to be s.t. $f_n \to f$ unif on E_k and $\mu(E_k^C) \le \frac{1}{k}$. Then consider $E := \bigcup_{k\ge 1} E_k$. ∎

If we add additional assumption that $\mu(X) < \infty$, surprisingly, point-wise convergence implies almost uniform convergence.

Theorem 6.15 (Egorov)**.** *Given μ s.t. $\mu(X) < \infty$, if $f_n, f : X \to \mathbb{R}$ are measurable $f_n \to f$ a.e. $\Rightarrow f_n \to f$ a.unif*

Proof. $f_n \to f$ a.e. means

$$\mu\left(\left(\bigcap_{k\ge 1}\bigcup_{N\ge 1}\bigcap_{n\ge N}\left\{x : |f_n(x) - f(x)| \le \frac{1}{k}\right\}\right)^C\right) = 0$$

$$= \mu\left(\bigcup_{k\ge 1}\bigcap_{N\ge 1}\bigcup_{n\ge N}\left\{x : |f_n(x) - f(x)| > \frac{1}{k}\right\}\right)$$

$$\Rightarrow \forall k, \mu\left(\bigcap_{N\ge 1}\bigcup_{n\ge N}\left\{x : |f_n(x) - f(x)| > \frac{1}{k}\right\}\right) = 0$$

and here we use assumption $\mu(X) < \infty$ and continuity from above,

$$\Rightarrow \forall k,\ \mu\left(\bigcup_{n\ge N}\left\{x : |f_n(x) - f(x)| > \frac{1}{k}\right\}\right) \downarrow 0 \ \text{ as } N \to \infty$$

so there is $N_{\epsilon,k}$ s.t. $\forall N \ge N_{\epsilon,k}$,

$$\mu\left(\bigcup_{n\ge N}\left\{x : |f_n(x) - f(x)| > \frac{1}{k}\right\}\right) \le \frac{\epsilon}{2k}$$

so

$$\underbrace{\mu\left(\bigcup_{k\ge 1}\bigcup_{n\ge N_{\epsilon,k}}\left\{x : |f_n(x) - f(x)| > \frac{1}{k}\right\}\right)}_{:=E_\epsilon^C} \le \sum_{1}^{k}\frac{\epsilon}{2k} = \epsilon$$

note $E_\epsilon = \bigcap_{k\ge 1}\bigcap_{n\ge N_{\epsilon,k}}\left\{x : |f_n(x) - f(x)| \le \frac{1}{k}\right\}$ which is exactly the set where $f_n \to f$ uniformly. ∎

The next theorem is a special case of Egorov on Lebesgue measure space, so there is actually no need to prove it. But the proof contains many useful ideas, so I include it here.

Theorem 6.16 (Egorov(Lebesgue measure version))**.** *If $\Omega \subset \mathbb{R}^n$ is s.t. $\lambda(\Omega) < \infty$ where λ is Lebesgue measure. Measurable $f_k : \Omega \to \mathbb{R}$ $k \in \mathbb{N}, f : \Omega \to \mathbb{R}$*

satisfies $f_k \to f$ $\lambda-$almost everywhere. Then given $\delta > 0$, exists compact $F \subset \Omega$ s.t.

$$\lambda(\Omega \setminus F) < \delta, \quad \sup_{x \in F} |f_k(x) - f(x)| \to 0$$

i.e. $f_k \to f$ a. unif

Proof. We take out the set of x s.t. $f_k(x)$ deviates "too much" from $f(x)$, i.e. $|f_k(x) - f(x)| > 2^{-i}$. But we are looking at asymptotic behaviour of f_k, so we require this deviation to be true for all $k > j \in \mathbb{N}$. So define

$$C_{i,j} := \bigcup_{k \geq j} \left\{x \in \Omega : |f_k(x) - f(x)| > 2^{-i}\right\}$$

Since f_k, f are measurable, $|f_k - f|$ is measurable by algebraic rule of measurable functions. So $C_{i,j}$, as a union of pre-images of open sets $(2^{-i}, \infty)$, is measurable. And $C_{i,j+1} \subseteq C_{i,j}$, so by continuity of measure,

$$\lim_j \lambda(C_{i,j}) = \lambda(\bigcap_{j \geq 1} C_{i,j})$$

$x \in \bigcap_{j \geq 1} C_{i,j} \Leftrightarrow \forall N > 1, \exists n \geq N$ s.t. $|f_n(x) - f(x)| > 2^{-i} \Leftrightarrow f_k(x) \not\to f(x)$. So

$$\lim_j \lambda(C_{i,j}) = \lambda(\{x : \Omega : f_k(x) \not\to f(x)\}) = 0$$

So for any i, we can pick $J(i)$ s.t. $\lambda(C_{i,J(i)}) < \delta 2^{-i-1}$. Let $A := \Omega \setminus \left(\bigcup_{i=1}^{\infty} C_{i,J(i)}\right)$. We have

$$\lambda(\Omega \setminus A) = \lambda\left(\bigcup_{i=1}^{\infty} C_{i,J(i)}\right) \leq \sum_{i=1}^{\infty} \lambda(c_{i,J(i)}) < \frac{\delta}{2}$$

By regularity of Lebesgue measure, there is a closed set $F \subset A$ s.t. $\lambda(A \setminus F) < \frac{\delta}{2}$. Then we have $\lambda(\Omega \setminus F) < \delta$.
And by definition of A, for all i, $F \subset C_{i,J(i)}^C$, so $\forall x \in F$,

$$|f_k(x) - f(x)| < 2^{-i} \ \forall k \geq J(i), i \geq 1$$

i.e. $\sup_{x \in F} |f_k(x) - f(x)| \to 0$. ■

Theorem 6.17 (Lusin)**.** *If $\Omega \subset \mathbb{R}^n$ is s.t. $\lambda(\Omega) < \infty$ where λ is Lebesgue measure. $f : \Omega \to \mathbb{R}$ is measurable, then given $\delta > 0$, exists compact $F \subset \Omega$ s.t.*

$$\lambda(\Omega \setminus F) < \delta \quad f|_F \text{ is continuous}$$

Proof. This proof technique has been used many times. Begin with simple functions, and use regularity of Lebesgue measure.
Then apply Egorov theorem and Proposition5.13 to prove this for general case. See details in Notes 5.2 Theorem 2.11. ■

6.2 Convergence in measure

Similar to convergence in probability, we can define convergence in measure

Definition 6.18 (Convergence in measure)**.** $f_n \xrightarrow{\mu} f$ if for any $\epsilon > 0$,

$$\mu\left(\{x : |f_n(x) - f(x)| > \epsilon\}\right) \to 0 \quad \text{as } n \to \infty$$

This is a very weak convergence as we will see in the example below:
Example. (typewriter function) Pick $X = [0,1)$Construct f_n using the following steps:

$$\text{Step 1.} \quad f_1 = 1_{[0,1)}$$

$$\text{Step 2.} \quad f_2 = 1_{[0,\frac{1}{2})},\ f_3 = 1_{[\frac{1}{2},1)}$$

$$\text{Step 3.} \quad f_4 = 1_{[0,\frac{1}{3})},\ f_5 = 1_{[\frac{1}{3},\frac{2}{3})}\ f_6 = 1_{[\frac{2}{3},1)}$$

We claim that this converges to 0 in measure. If $\epsilon \geq 1$, then $\mu\left(\{x : |f_n(x)| > \epsilon\}\right) = 0$. If $0 < \epsilon < 1$, we can see at step k, $\mu\left(\{x : |f_n(x)| > \epsilon\}\right) = \frac{1}{k}$, so it converges to 0 as $n \to \infty$. However for any $x \in X$, at any step k there is an f_n s.t. $f_n(x) = 1$. So $f_n(x)$ does not converge to 0 everywhere.

Let's observe some properties of convergence in measure

Proposition 6.19 (Limit is unique)**.** *If $f_n \xrightarrow{\mu} f,\ f_n \xrightarrow{\mu} g$, then $f \sim g$.*

Proof. Fix $\delta > 0$, consider $\mu\left(\{x : |f(x) - g(x)| > \delta\}\right) =: d_\delta$. Since

$$\delta < |f(x)-g(x)| \leq |f_n(x)-f(x)|+|f_n(x)-g(x)| \Rightarrow |f_n(x)-f(x)| > \frac{\delta}{2} \text{ or } |f_n(x)-g(x)| > \frac{\delta}{2}$$

we have

$$d_\delta \leq \mu\left(\{x : |f_n(x) - f(x)| > \frac{\delta}{2}\}\right) + \mu\left(\{x : |f_n(x) - g(x)| > \frac{\delta}{2}\}\right)$$

two terms on RHS, by assumption, converges to 0 as $n \to \infty$ so $d_\delta \to 0$. But d_δ is independent of n, so $d_\delta = 0$. Choice of δ is arbitrary so

$$\mu\left(\{x : f(x) \neq g(x)\}\right) = 0$$

■

Proposition 6.20. *If $f_n \xrightarrow{\mu} f$ and $f_n \sim g_n,\ f \sim g$, then $g_n \xrightarrow{\mu} g$*

Proof. Fix $\epsilon > 0$, consider $d_n := \mu\left(\{x : |g_n(x) - g(x)| > \epsilon\}\right)$.
By assumption, if $E_n := \{x : f_n(x) = g_n(x)\}$, $E_0 := \{x : f(x) = g(x)\}$, then $\mu(E_j^C) = 0$ for $j \geq 0$. So define $E := \bigcap_{n\geq 0} E_n$, then $\mu(E^C) = 0$. So

$$d_n \leq \mu\left(\{x : |g_n(x) - g(x)| > \epsilon\} \cap E\right) + \mu\left(\{x : |g_n(x) - g(x)| > \epsilon\} \cap E^C\right)$$

the second term is just 0 as $\mu(E^C) = 0$. As $f_n = g_n,\ f = g$ on E,

$$d_n \leq \mu\left(\{x : |f_n(x) - f(x)| > \epsilon\} \cap E\right) \leq \mu\left(\{x : |f_n(x) - f(x)| > \epsilon\}\right)$$

RHS converges to 0 as $n \to 0$, so $d_n \to 0$. ■

Although convergence in measure does not give point-wise convergence a.e., there can be a subsequence that converges point-wise a.e.

Proposition 6.21. *If $f_n \xrightarrow{\mu} f$, exists subsequence $f_{n_k} \to f$ a.e. as $k \to \infty$.*

Recall our typewriter example, if we take the first set from each step and form a sequence, i.e. $f_{n_k} = 1_{[0,\frac{1}{k})}$, then clearly this subsequence converges to 0.

Proof. Define $g_n(x) := f_n(x) - f(x)$. For any $\epsilon > 0$, we have

$$\mu\left(\{x : |g_n(x)| > \epsilon\}\right) \to 0$$

so there is N_k s.t. for any $n \geq N_k$

$$\mu\left(\{x : |g_n(x)| > 1/k\}\right) < \frac{1}{2^{k+1}}$$

so we can pick n_k in such a way that $n_{k+1} > n_k$ and

$$\mu\left(\{x : |g_{n_k}(x)| > 1/k\}\right) < \frac{1}{2^{k+1}}$$

so for any $K \in \mathbb{N}$,

$$\mu\left(\bigcup_{k\geq K}\{x : |g_{n_k}(x)| > 1/k\}\right) \leq \sum_{k=K}^{\infty}\frac{1}{2^{k+1}} = \frac{1}{2^K}$$

then we have

$$\underbrace{\mu\left(\bigcap_{K\geq 1}\bigcup_{k\geq K}\{x : |g_{n_k}(x)| > 1/k\}\right)}_{E^C} = 0$$

Then we translate the meaning of $x \in E$.

$$E = \bigcup_{K\geq 1}\bigcap_{k\geq K}\{x : |g_{n_k}(x)| \leq 1/k\}$$

so if $x \in E$, there is $K \in \mathbb{N}$ s.t. for all $k \geq K$, $|g_{n_k}(x)| \leq 1/k$, so $g_{n_k} \to 0$ on E, i.e. $f_{n_k} \to f$ as $k \to \infty$ on E. ■

We might expect $f_n \to f$ a.e. implies $f_n \xrightarrow{\mu} f$, but this is not true. Consider $f_n = 1_{[n,\infty)}$, then for any $\epsilon < 1$,

$$\mu\left(\{x : |f_n(x)| > \epsilon\}\right) = \infty$$

so f_n does not converge to 0 in measure. But it clearly converges point-wise to 0. However, using Egorov theorem, just with an additional assumption, point-wise convergence a.e. implies convergence in measure.

Proposition 6.22. *If $\mu(X) < \infty$, then $f_n \to f$ a.e. $\Rightarrow f_n \xrightarrow{\mu} f$*

Proof. Given $\epsilon > 0$, now fix $\delta > 0$, our target is to find n_δ s.t. $\forall n \geq n_\delta$,

$$d_{\delta,n} := \mu\left(\{x : |f_n(x) - f(x)| > \epsilon\}\right) \leq \delta$$

using Egorov, we have $f_n \to f$ a. unif, so there is set E_δ with $\mu(E_\delta^C) \leq \delta$ and $f_n \to f$ unif. on E_δ. i.e. exists $N_{\epsilon,\delta}$(N depends on the set E_δ chosen and our ϵ fixed in the beginning) s.t. $\forall n \geq N_{\epsilon,\delta}$, $\sup_{x\in E_\delta} |f_n(x) - f(x)| \leq \epsilon$. So if $n \geq N_{\epsilon,\delta}$,

$$d_{\delta,n} \leq \mu\left(\{x : |f_n(x) - f(x)| > \epsilon\} \cap E_\delta\right) + \mu\left(\{x : |f_n(x) - f(x)| > \epsilon\} \cap E_\delta^C\right)$$

the first term is 0, as on E_δ, we cannot have $|f_n(x) - f(x)| > \epsilon$. And the second term is bounded by $\mu(E_\delta^C)$, so

$$d_{\delta,n} \leq \delta$$

which is our target. ■

Here we can introduce another convergence

Definition 6.23 (Convergence in $\mathcal{L}^1$)**.** $f_n \to f$ in L^1 if $\int |f_n - f|\, d\mu \to 0$ as $n \to \infty$.

there are other L^p convergence, but those will be discussed in later section.

Proposition 6.24. *Convergence in L^1 implies convergence in measure. If $\mu(X) < \infty$ i.e. measure is finite.*

Proof. See PS4 Q6a ■

Proposition 6.25. *Convergence almost uniformly implies convergence in measure*

Proof. Easy manipulation of definitions, left as exercise. ■

Let us summarise again the five modes of convergence we learnt

- Point-wise convergence a.e.
- Uniform convergence a.e.
- Almost uniform convergence
- Convergence in measure
- Convergence in L^1

If one modifies f_n or f on a set of measure 0, then any convergence above is not affected. (This is already proved for convergence in measure in Proposition6.20, and it is clear for first two modes. You are suggested to prove it for third and fifth modes)

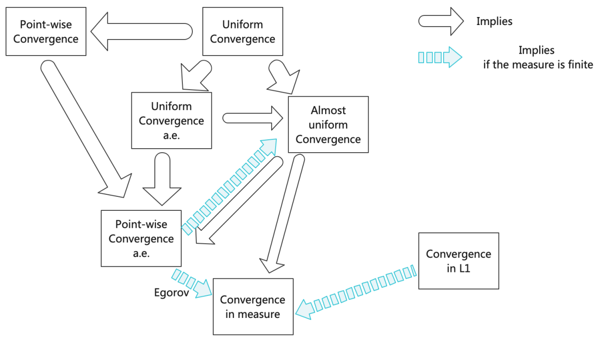

Figure 4: Implications between Modes of Convergence

All the five modes of convergences share some important properties

Proposition 6.26. *If f_n converges to f in one mode, $|f_n - f|$ converges to 0 in that mode.*

If f_n converges to f, g_n converges to g in a mode, then $f_n + g_n$ converges to $f + g$ in that mode. And if c is a constant in the codomain, then cf_n converges to cf in that mode.

If f_n converges to 0 in a mode, then for any g_n with $|g_n| \leq f_n$, g_n also converges to 0 in that mode. (Squeeze test)

Convergence is compatible. i.e. if f_n converges to f and g under a mode, then $f \sim g$. (We have proved this for convergence in measure and point-wise a.e.

Proof. Proof is long but elementary, left as exercise. ■

Here are some more examples to distinguish these convergences

Example 1.

$$f_n := \frac{1}{n} 1_{[0,n]}$$

it converges in all modes but L^1 to 0.

Example 2.

$$f_n := n 1_{[\frac{1}{n}, \frac{2}{n}]}$$

converges to 0 almost uniformly (so converges in measure), point-wise a.e., but not uniformly a.e., not in L^1.

From the typewriter example, we can see that even with assumption $\mu(X) < \infty$, $\mathcal{L}^1$ convergence cannot imply point-wise a.e. convergence nor almost uniform convergence. But there is a modification (called fast $\mathcal{L}^1$ convergence) that can imply point-wise a.e. convergence and almost uniform convergence.

Definition 6.27 (fast L^1 convergence)**.** If $\mu(X) < \infty$ and f_n converges to f in fast L^1 if

$$\sum_{n\geq 1} \int |f_n(x) - f(x)| \, d\mu < \infty$$

this is stronger than $\int |f_n(x) - f(x)| \, d\mu \to 0$.

6.3 Hahn-Jordan Theorem

In this section we are interested in signed measures, i.e. measures that allow negative values.

Definition 6.28 (Signed measure)**.** Given measure space $(X, \mathcal{F})$, a set function $\nu : \mathcal{F} \to \overline{\mathbb{R}}$ is called signed measure if the following holds

1. $\nu(\emptyset) = 0$
2. ν is σ-additive.

Remark. Though not explicitly mentioned, if $E, F \in \mathcal{F}$ are disjoint, then

$$\nu(E \cup F) = \nu(E) + \nu(F)$$

so for ν to be well-defined, we require there are no disjoint E, F s.t. $\nu(E) = \infty, \nu(F) = -\infty$.

Also consider disjoint sequence $(E_j)_{j\geq 1} \subset \mathcal{F}$. Then

$$\nu(\sum_j E_j) = \sum_j \nu(E_j)$$

so for ν to be well-defined, we require the infinite series to be independent of order, i.e.

- $\sum_j \nu(E_j)$ is absolutely convergent
- or one of $\sum_{j;\, \nu(E_j)\geq 0} \nu(E_j)$, $\sum_{j;\, \nu(E_j)<0} \nu(E_j)$ is finite.

We will include these two assumptions in definition of ν, but will not state them explicitly.

Let's list some important properties of signed measure

Lemma 6.29. *If ν is signed measure on $(X, \mathcal{F})$ and $E, F \in \mathcal{F}$ are s.t. $E \subseteq F$, then*

- *If $|\nu(E)| < \infty$, then $\nu(F \setminus E) = \nu(F) - \nu(E)$*
- *If $\nu(E) = \infty$, then $\nu(F) = \infty$*
- *If $\nu(E) = -\infty$, then $\nu(F) = -\infty$*

Proof. (1) $|\nu(E)| < \infty$
$F = (F \cap E) \cup (F \setminus E)$ but $E \subseteq F$ so $F \cap E = E$. Then

$$\nu(F) = \nu(E) + \nu(F \setminus E)$$

by additivity of ν. Note $\nu(E)$ is assumed to be finite, so we can subtract it on two sides. (**Warning: You should never add $\pm\infty$ on two sides of equation**)

$$\nu(F) - \nu(E) = \nu(F \setminus E)$$

(2) $\nu(E) = \infty$.
Again we have

$$\nu(F) = \nu(E) + \nu(F \setminus E)$$

by hidden assumption mentioned in remark, we require $\nu(\nu(F \setminus E)) > -\infty$, so $\nu(F \setminus E)$ is real number or ∞, but either case gives $\nu(F) = +\infty$.
Case (3) can be handled similarly. ■

Lemma 6.30. *If $\exists E \in \mathcal{F}$ s.t. $\nu(E) = +\infty$, then $\nu(F) \neq -\infty$ for all $F \in \mathcal{F}$. If $\exists E \in \mathcal{F}$ s.t. $\nu(E) = -\infty$, then $\nu(F) \neq +\infty$ for all $F \in \mathcal{F}$.*

Proof. Second part can be proved from first part using the fact that if ν is signed measure, then $-\nu$ is also signed measure.

So given set E, F s.t. $\nu(E) = +\infty$. If $E \cap F = \emptyset$ then result is clear by hidden assumption. Therefore for general E, F we consider $\nu(E \cap F)$. If it is $-\infty$, since $E \cap F \subseteq E$, this would imply $\nu(E) = -\infty$ by last lemma, contradiction.
If $\nu(E \cap F) = +\infty$, then $\nu(F) = +\infty$ by last lemma.
If $\nu(E \cap F) = c \in \mathbb{R}$, then $+\infty = \nu(E) = \nu(E \cap F) + \nu(E \setminus F)$. So this requires $\nu(E \setminus F) = +\infty$. F, $E \setminus F$ are disjoint, so by hidden assumption, $\nu(F) \neq -\infty$. ■

Remark. By the above lemma, we can study signed measure of the first type. i.e. $\nu(E) \in (-\infty, +\infty]$. Then properties can be moved to the second type by taking signed measure $\nu' = -\nu$.

An important property for signed measure is that it is continuous from above and below (Definitions are the same as those for measure.)

Proposition 6.31. *If ν is signed measure, then it is continuous from above and below. i.e.*

- *If $E_k \uparrow E$, then $\nu(E_k) \to \nu(E)$. Note since ν can be negative, we no longer have guarantee that $\nu(E_k)$ also is increasing. (proof of monotonicity of unsigned measure used the fact that $\mu \geq 0$)*
- *If $E_k \downarrow E$, and exists $N \in \mathbb{N}$ s.t. $\nu(E_N) < \infty$, then $\nu(E_k) \to \nu(E)$*

Proof. Almost the same as the proof for unsigned measure, left as exercise. ■

Now we state the Hahn-Jordan theorem, also called Hahn decomposition theorem. It basically allow us to separate any signed measure to positive and negative parts.

Theorem 6.32 (Hahn-Jordan)**.** *Given signed measure ν (assume $\nu(F) \in (-\infty, +\infty]$ for all $F \in \mathcal{F}$), there is a set $P \in \mathcal{F}$ s.t.*

- *If $E \subseteq P$, then $\nu(E) \geq 0$. (P stands for positive)*
- *If $F \subseteq P^C =: N$, then $\nu(F) \leq 0$. (N stands for negative)*

this decomposition is called Hahn decomposition.

Remark. N, P may not be unique. The easiest example is function ν defined on real line s.t. $\nu \equiv 0$. Then basically any set P works.

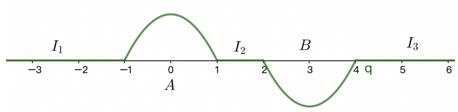

Figure 5: Example of Hahn decomposition

In the above example, we can take set $P = A \cup I$ where I is any combination of I_1, I_2, I_3.

Now we prove Hahn-Jordan theorem

Proof. First we define $\alpha := \inf_{A \in \mathcal{F}} \nu(A)$. Note although $\nu(A) > -\infty$, α can be $-\infty$. (Which is a useless case) So first step of proof is to show $\alpha > -\infty$. And for second step, we find a set $N \in \mathcal{F}$ s.t $\nu(N) = \alpha$.
Note if we have such a set N, then let $P = N^C$ and given $E \subseteq P$,
If $\nu(E) < 0$, then consider $\nu(N \cup E) = \nu(N) + \nu(E) = \alpha + \nu(E) < \alpha$. But we defined α as infimum, so contradiction.
As exercise, try to show given $F \subseteq N$, and if $\nu(F) > 0$, there is a contradiction. (Hint: consider $\nu(N \setminus F)$) So the set N indeed satisfies the conditions. Now we prove the two claims.

Claim 1. $\alpha > -\infty$.
For contradiction, assume $\alpha = -\infty$. Then we wish to construct $A_k \downarrow$ (this means A_k is decreasing sequence) and construct B_k s.t. $\nu(B_k) \leq -k$. Then $\nu(\bigcap B_k) = -\infty$ by continuity but this contradicts with assumption that $\nu(A) > -\infty$ for any set $A \in \mathcal{F}$.
First we let $A_0 = X$, and define $\lambda(C) := \inf_{E \subseteq C} \nu(E)$. So $\lambda(A_0) = \lambda(X) = \alpha = -\infty$. So there is $B_1 \subseteq A_0$ s.t. $\nu(B_1) \leq -1$. And now we define

$$A_1 := \begin{cases} B_1 & \text{if } \lambda(B_1) = -\infty \\ A_0 \setminus B_1 & \text{otherwise} \end{cases}$$

Now we show $\lambda(A_1) = -\infty$ as well. Given $E \subseteq A_0$,

$$\nu(E) = \nu(E \cap B_1) + \nu(E \cap (A_0 \setminus B_1)) \geq \lambda(B_1) + \lambda(A_0 \setminus B_1)$$

take infimum on LHS, we have

$$-\infty = \lambda(A_0) \geq \lambda(B_1) + \lambda(A_0 \setminus B_1)$$

so either $\lambda(B_1) = -\infty$ or $\lambda(A_0 \setminus B_1) = -\infty$. Which means $\lambda(A_1) = -\infty$ by definition.
This process can be repeated to construct decreasing sequence A_k with $\lambda(A_k) = -\infty$ and sequence B_k s.t. $\nu(B_k) \leq -k$.

Note for each generation (each A_k), the second case $A_k = A_{k-1} \setminus B_k$ is called a bifurcation. As shown in the graph below
Note for all $k > 2$, B_k and B_2 will not intersect as there is a bifurcation for

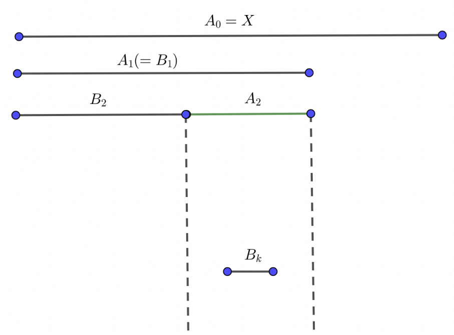

Figure 6: Bifurcation

the second generation (i.e. when we decide A_2)
The rigorous statement of the above observation is if bifurcation happens at stage k, i.e. $A_k = A_{k-1} \setminus B_k$ then $\forall j > k$, $B_j \cap B_k = \emptyset$. Proof is left as exercise.

Now we wish to construct B s.t. $\nu(B) = -\infty$ to give contradiction.
Case 1. Finite number of bifurcations
i.e. there exists $N > 0$ s.t. $\forall n \geq N$, no bifurcation exists. So $B_n = A_n$. That means B_n is also a decreasing sequence for $n \geq N$. By construction $\nu(B_n) \leq -n < \infty$, so using continuity from above, if $B_n \downarrow B$, then

$$\nu(B) = \lim_{n\to\infty} \nu(B_n) = -\infty$$

Case 2. Infinite bifurcations
Call those generations of bifurcation $n_0, n_1, ...$ so

$$B_{n_j} \cap B_{n_k} = \emptyset \quad \text{If } j \neq k$$

So let $B := \sum^{k\geq 0} B_{n_k}$, then $\nu(B) = \sum_{k\geq 0} \nu(B_{n_k})$ by σ-additivity of ν. So $\nu(B) \leq \sum_{k\geq 0}(-n_k) = -\infty$, i.e. $\nu(B) = -\infty$ contradiction.

$\square_{\text{Claim 1}}$

Claim 2. we can find a set $N \in \mathcal{F}$ s.t $\nu(N) = \alpha$
There are C_n s.t. $\alpha \leq \nu(C_n) \leq \alpha + \frac{1}{2^n}$ by definition of α. And note

$$\nu(C_n \cup C_{n+1}) = \nu(C_n) + \nu(C_{n+1}) - \nu(C_n \cap C_{n+1}) \leq \alpha + \frac{1}{2^n} + \alpha + \frac{1}{2^{(n+1)}} - \alpha = \alpha + \frac{1}{2^n} + \frac{1}{2^{(n+1)}}$$

Similarly one can show

$$\nu(\bigcup_{k=n}^{n+q} C_n) \leq \alpha + \frac{1}{2^n} + \cdots \frac{1}{2^{n+q}}$$

as $q \to \infty$, LHS approaches $\nu(\bigcup_{k\geq n} C_n)$ we define $D_n := \bigcup_{k\geq n} C_n$. Then

$$\nu(D_n) \leq \alpha + \sum_{k\geq n} \frac{1}{2^k} = \alpha + \frac{2}{2^n}$$

$D_n \downarrow$ so let $D := \bigcap_{n\geq 1} D_n$, by continuity from above,

$$\nu(D) = \lim_{n\to\infty} \nu(D_n) \leq \alpha$$

but $\nu(D) \geq \alpha$ by definition of α as infimum, so $\nu(D) = \alpha$ as required. ■

Remark. With Hahn-Jordan theorem, one can define

$$\nu_+(E) := \nu(E \cap P) \geq 0$$

$$\nu_-(F) := -\nu(F \cap N) \geq 0$$

then $\nu = \nu_+ - \nu_-$

6.4 Absolute Continuity

In this section we aim to investigate what implies L^1 convergence, and grab some insights of differentiation.

Recall an important corollary of monotone convergence theorem.

Corollary 6.33. *If $f \geq 0$, define $\mu_f(A) := \int_A f \, d\mu$. μ_f is a measure.*

Definition 6.34 (Absolute continuity)**.** μ, ν are measures on $(X, \mathcal{F})$. $\mu \ll \nu(\mu$ is absolute continuous w.r.t. $\nu)$ if

$$\nu(A) = 0 \Rightarrow \mu(A) = 0$$

Note if $B \in \mathcal{F}$, $\mu(B) = 0 \Rightarrow \mu_f(B) = 0$, so $\mu_f \ll \mu$. There is a theorem proving the converse of this statement:

Theorem 6.35 (Radon-Nykodim)**.** *For two σ-finite measures μ, ν, if $\mu \ll \nu$, then $\exists g$(denoted as $\frac{d\mu}{d\nu}$) s.t.*

$$\mu(B) = \int_B g \, d\nu$$

Proof. Proof will be given in the next section. ■

Proposition 6.36 (Absolutely uniformly continuous integrability)**.** *Assume $f \in \mathcal{L}^1(\mu)$ i.e. f is integrable, then for any $\epsilon > 0$, there is $\delta > 0$ s.t. if $\mu(A) < \delta$, $\int_A |f| \, d\mu < \epsilon$.*

Proof. See PS5 Q4 ■

The above proposition is generalised to a family of functions

Definition 6.37 (Uniform integrability(uniformly absolutely continuous integrals))**.** Given measure space $(X, \mathcal{A}, \mu)$, a family $\mathcal{F}$ of functions has absolutely uniformly continuous integrals if

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall f \in \mathcal{F},\ A \in \mathcal{A}, \mu(A) < \delta \Rightarrow \mu_{|f|}(A) = \int_A |f|\ d\mu < \epsilon$$

So by above proposition, any family with only one integrable function has absolutely uniformly continuous integrals. So this generalises to family of finite integrable functions.

Corollary 6.38. *If $\nu \ll \mu$, $\nu(X) < \infty$, then*

$$\forall \epsilon > 0, \exists \delta > 0 \ s.t.\ \forall A \in \mathcal{F} \text{ with } \mu(A) < \delta \Rightarrow \nu(A) < \epsilon$$

so $\nu \ll \mu$ is called absolute continuous.

Proof. Direct result from Randon-Nykodim and Proposition6.36. ■

Remark. We can have two measures that are not absolute continuous w.r.t. each other. So $\ll$ is just a partial order.
Let $f > 0$ be a function and fix $x \in \mathbb{R}$, let $\delta_x(B) = 1$ if B contains x, and $\delta_x(B) = 0$ otherwise.

$$\mu_f(\{x\}) = \int_{\{x\}} f\ d\mu = 0 \quad \text{ as } \mu(\{x\}) = 0$$

but $\delta_x(\{x\}) = 1$. So $\delta_x \not\ll \mu_f$. If $[a, b)$ does not contain x, we have

$$\delta_x([a, b)) = 0 \quad \mu_f(\{x\}) = \int_{[a,b)} f\ d\mu > 0$$

so $\mu_f \not\ll \delta_x$

Theorem 6.39 (Vitali's theorem)**.** *If $\mu(X) < \infty$, f, f_n are integrable and measurable, let $\mathcal{F} = \{f_n : n \in \mathbb{N}\}$, we have*

$$f_n \xrightarrow{\mu} f \text{ and } \mathcal{F} \text{ has uniformly absolute continuous integrals. } \Leftrightarrow f_n \to f \text{ in } \mathcal{L}^1$$

Proof. $(\Leftarrow)$ $f_n \xrightarrow{\mu} f$ by Proposition6.24. To prove $\mathcal{F}$ has uniformly absolutely continuous integrals, we note

$$\int_A |f_n|\ d\mu \leq \int_A |f|\ d\mu + \int_A |f_n - f|\ d\mu$$

Fix $\epsilon > 0$, aim to bound second term first. Using the condition $\int |f_n - f|\ d\mu \to 0$, we can find N_ϵ s.t. for any $n \geq N_\epsilon$, $\int |f_n - f|\ d\mu < \epsilon/2$. For terms before N_ϵ and the first term $\int_A |f|\ d\mu$, consider the family of functions $\mathcal{F}' := \{f, f_1, ..., f_{N_\epsilon}\}$. It is finite and its members are all integrable, so $\mathcal{F}'$ has uniformly absolute continuous integrals. i.e. we can find $\delta > 0$ s.t. for any $A \in \mathcal{A}$ with $\mu(A) < \delta$, we have

$$\int_A |f|\ d\mu < \epsilon/2, \quad \max_{n \leq N_\epsilon} \int_A |f_n|\ d\mu < \epsilon/2$$

so using these conclusions, we have that if $\mu(A) < \delta$

$$\int_A |f_n| < \epsilon/2 + \epsilon/2 = \epsilon$$

($\Rightarrow$) As $f_n \xrightarrow{\mu} f$, using Proposition 6.21, we can find subsequence $f_{n_k} \to f$ a.e. as $k \to \infty$. Let us assume that $\limsup \int |f_n - f| \, d\mu > 0$ for contradiction. ($\limsup$ is used as limit may not exist, but by integrability of f_n, f, integral $\int |f_n - f| \, d\mu$ is bounded so $\limsup$ exists) So for the convergent subsequence,

$$(*) = \int |f_{n_k} - f| \, d\mu > 0$$

we aim to contradict by proving this integral is smaller than ϵ for any $\epsilon > 0$. First using the fact that $\mathcal{F}$ has uniform absolutely continuous integrals, and the fact that f is integrable, $\mathcal{F}' := \{f_n, f : n \in N\}$ also has uniform absolutely continuous integrals. So we can pick $\delta > 0$ s.t. if $A \in \mathcal{A}$, $\mu(A) < \delta$, then

$$\int_A |f| \, d\mu < \frac{\epsilon}{3}, \int_A |f_n| \, d\mu < \frac{\epsilon}{3}$$

By Egorov, one can pick measurable F s.t. $\mu(F^C) < \delta$ and $f_{n_k} \to f$ uniformly on F. i.e. we can pick K s.t. if $k \geq K$,

$$\sup_{x \in F} |f_{n_k}(x) - f(x)| < \frac{\epsilon}{3\mu(X)}$$

so if $k \geq K$, we integrate (*) on F, where uniform convergence holds and set F^C with small measure respectively,

$$\int |f_{n_k} - f| \, d\mu = \int_F |f_{n_k} - f| \, d\mu + \int_{F^C} |f_{n_k} - f| \, d\mu$$

$$\leq \int_F \sup_{x \in F} |f_{n_k} - f| \, d\mu + \int_{F^C} |f| \, d\mu + \int_{F^C} |f_n| \, d\mu < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

■

Remark. It is possible to generalise this theorem to L^p convergence if we change the uniform absolutely continuous condition to $\mathcal{F} := \{|f_n|^p : n \in \mathbb{N}\}$ has uniform absolutely continuous integrals.

If unfortunately we can not assume $\mu(X) < \infty$, then Vitali's theorem still hold if we require f_n to decay "at infinity", i.e.

$$\forall \epsilon > 0, \exists X_\epsilon \in \mathcal{A} \text{ s.t. } \mu(X_\epsilon) < \infty \text{ and } \sup_{n \geq 1} \int_{X_\epsilon^C} |f_n|^p \, d\mu < \epsilon$$

6.5 Radon Nykodim theorem

Definition 6.40 (Singular measure)**.** Given two measures ν, μ on measure space $(X, \mathcal{F})$. ν is singular w.r.t. μ, written as $\nu \perp \mu$, if there is a set $A \in \mathcal{F}$ s.t. $\mu(A) = 0$ and $\nu(A^C) = 0$. This means the set A contains the support of ν and set A^C contains the support of μ. This relationship is actually symmetric.

Example. Take Lebesgue measure λ and find enumeration of rational numbers $\{q_j : j \in \mathbb{N}\}$. Define $\nu := \sum c_j \delta_{q_j}$ where $c_j \geq 0$. For $x \in \mathbb{R}$, δ_x is the Dirac measure defined as below

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Take $A = \mathbb{Q}$, it is easy to see that $\nu(A^C) = 0$ (as of ν is concentrated on $\mathbb{Q}$) and $\lambda(A) = 0$. So $\nu \perp \lambda$.

Theorem 6.41 (Lebesgue decomposition)**.** *For σ-finite measure μ and σ-finite signed measure ν, there is a unique decomposition $\nu = \nu_1 + \nu_2$ s.t. $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$. And further, there is a $\mathcal{F}$-measurable function f s.t.*

$$\nu_1(A) = \int_A f \, d\mu$$

Note Radon Nykodim theorem is a corollary of this decomposition.

Proof. We first impose two restrictions and we will later drop these two conditions.

1. ν, μ are finite
2. ν is a measure. i.e. $\nu(A) \geq 0$.

If we indeed have such $\nu_1(A) = \int_A f \, d\mu \ \nu \geq 0$ so we require $\nu(A) \geq \nu_1(A)$. So we define the following set

$$\mathcal{H} := \left\{ f \geq 0 : \int_A f \, d\mu \leq \nu(A) \ \forall A \in \mathcal{F} \right\}$$

and we aim to maximise $\int_A f \, d\mu$ now. Let

$$\alpha := \sup_{f \in \mathcal{H}} \int_X f \, d\mu$$

(If not specified, $\int f \, d\mu$ means $\int_X f \, d\mu$.) Since $\int_X f \, d\mu \leq \nu(X) < \infty$ for all $f \in \mathcal{H}$, $\alpha < \infty$. So now we try to find $g \in \mathcal{H}$ s.t. $\int g \, d\mu = \alpha$. By definition of α, we can find sequence of functions $f_n \in \mathcal{H}$ s.t.

$$\alpha - \frac{1}{n} \leq \int f_n \, d\mu \leq \alpha$$

In order to use MCT(monotone Convergence Theorem), we need a monotone sequence, so define

$$g_n := \max\{f_1, ..., f_n\}$$

clearly g_n is increasing and we call the limit g and now we prove that $g_n, g \in \mathcal{H}$ and further $\int g \, d\mu = \alpha$

- For $1 \leq k \leq n$, we define

$$E_{n,k} := \{x : g_n(x) = f_k(x)\}$$

clearly $(E_{n,k})_{1\le k\le n}$ gives a partition of X, so we have

$$\int_A g_n \, d\mu = \sum_{k=1}^{n} \int_{A\cap E_{n,k}} f_k \, d\mu \le \sum_{k=1}^{n} \nu(A\cap E_{n,k}) = \nu(A)$$

where the inequality follows from the fact that $f_k \in \mathcal{H}$.

- For all $n \in \mathbb{N}, A \in \mathcal{F}$, $\int_A g_n \, d\mu \le \nu(A)$ as proved above. So by MCT, LHS converges to $\int_A g \, d\mu$. So $\int_A g \, d\mu \le \nu(A)$.
- By definition of α, $\int g \, d\mu \le \alpha$. Now for all n,

$$\int g \, d\mu \ge \int g_n \, d\mu \ge \int f_n \, d\mu \ge \alpha - \frac{1}{n}$$

  so $\int g \, d\mu \ge \alpha$.

So define $\nu_1(A) := \int_A g \, d\mu \le \nu(A)$ (as $g \in \mathcal{H}$). Let $\nu_2(A) := \nu(A) - \nu_1(A) \ge 0$. Clearly, $\nu_1 \ll \mu$. So now the target is $\nu_2 \perp \mu$.

Informally consider the inequality $\frac{\nu_2}{\mu} \ge \frac{1}{n}$. We expect $\nu_2 \perp \mu$, so if $\nu_2 > 0$, $\mu = 0$; if $\mu > 0$, then $\nu_2 = 0$. So for the sets s.t. $\nu_2 > 0$, this inequality should be true. Define signed measure

$$\sigma_n := \nu_2 - \frac{1}{n}\mu$$

note this measure can take negative values. Using Hahn-Jordan decomposition, we can find sets P_n, N_n ($P_n^C = N_n$) and if

$$E \subseteq P_n \Rightarrow \sigma_n(E) \ge 0$$

$$F \subseteq N_n \Rightarrow \sigma_n(F) \le 0$$

Now we aim to show $\mu(P_n) = 0$. Consider function $g + \frac{1}{n}1_{P_n}$, it is in $\mathcal{H}$ as

$$\int_A \left[g + \frac{1}{n}1_{P_n}\right] \, d\mu = \nu_1(A) + \frac{1}{n}\mu(P_n \cap A)$$

since $P_n \cap A \subseteq P_n$, $\sigma_n(P_n\cap A) = \nu_2(P_n \cap A) - \frac{1}{n}\mu(P_n\cap A) \ge 0$, so $\nu_2(P_n\cap A) \ge \frac{1}{n}\mu(P_n \cap A)$. Therefore,

$$\int_A \left[g + \frac{1}{n}1_{P_n}\right] \, d\mu \le \nu_1(A) + \nu_2(P_n\cap A) \le \nu_1(A) + \nu_2(A) = \nu(A)$$

But

$$\int \left[g + \frac{1}{n}1_{P_n}\right] \, d\mu = \alpha + \frac{1}{n}\mu(P_n)$$

if $\mu(P_n) > 0$, $\alpha + \frac{1}{n}\mu(P_n) > \alpha$ but this contradicts definition of α. So $\mu(P_n) = 0$. Let $P := \bigcup_n P_n$, $\mu(P) = 0$. Let $N = P^C = \bigcap_n P_n^C = \bigcap_n N_n$. We aim to prove that $\nu_2(N) = 0$.
$\nu_2(N) \le \nu_2(N_n)$ for any n, and we have

$$\sigma_n(N_n) = \nu_2(N_n) - \frac{1}{n}\mu(N_n) \le 0$$

so

$$\nu_2(N) \leq \nu_2(N_n) \leq \frac{1}{n}\mu(N_n) \leq \frac{1}{n}\mu(X) < \infty$$

the above inequality is true for any n, so $\nu_2(N) = 0$. This proves $\nu_2 \perp \mu$ using the sets P, N.

Firstly, we try to remove the assumption μ, ν are finite. Assume μ, ν are σ-finite instead, we can find sequences $E_n \uparrow, F_n \uparrow$ s.t. $X = \bigcup E_n = \bigcup F_n$ and $\mu(E_n), \nu(F_n) < \infty$. Let $G_n = E_n \cap F_n$, it is easy to show that $G_n \uparrow$ and $X = \bigcup G_n$. So we define

$$H_1 = G_1, H_2 = G_2 \setminus G_1, H_3 = G_3 \setminus G_2, ...$$

H_j are all disjoint and $X = \sum H_j$ and $\mu(H_j), \nu(H_j) < \infty$. Now define $\mu_j(A) := \mu(H_j \cap A), \quad \nu_j(A) = \nu(H_j \cap A)$ which are finite measures on space H_j, so using the conclusion from restircted case, we have

$$v_j = v_j^1 + v_j^2 \quad \text{s.t. } v_j^1 \ll \mu_j, v_j^2 \perp \mu_j$$

define $\nu^1 := \sum_j \nu_j^1$, $\nu^2 := \sum_j \nu_j^2$ it is left as exercise to show that $\nu^1 \ll \mu$ and $\nu^2 \perp \mu$.

Hint: on each H_j, we know there is a set B_j s.t. support of ν_j^2 is inside B_j and support of μ_j is inside B_j^C. So taking $B := \bigcup_j B_j$, then clearly ν^2 will only take non-zero values on B.

As for proving $\nu^1 \ll \mu$, use the fact that there is a measurable function g_j s.t. for all $A \subseteq H_j$, $\nu_j^1(A) = \int_A g_j \, d\mu_j$. Consider combining these functions and prove that ν^1 is also an integral.

Finally we remove the assumption $\nu \geq 0$. For general ν, using Hahn-Jordan decomposition, $\nu = \theta_1 - \theta_2$ where θ_1, θ_2 are measures and θ_1 only take non-zero values on a set P and θ_2 only take non-zero values on $N := P^C$. And using the restricted version of proof above, we have decompositions $\theta_1 = \theta_1^1 + \theta_1^2, \theta_2 = \theta_2^1 + \theta_2^2$ s.t. $\theta_1^1, \theta_2^1 \ll \mu$ and $\theta_1^2, \theta_2^2 \perp \mu$. There are measurable f_1^1, f_1^2 s.t.

$$\theta_1^1(A) = \int_A f_1^1 \, d\mu, \quad \theta_1^2(A) = \int_A f_1^2 \, d\mu$$

and $f_1^1 = 0$ a.e. on N, $f_2^1 = 0$ a.e. on P. So let $f^1 = f_1^1 - f_2^1$, and let $\nu^1(A) := \int_A f^1 \, d\mu = \theta_1^1 - \theta_2^1$, we can see $\nu^1 \ll \mu$.

Now let $\nu^2 = \theta_1^2 - \theta_2^2$, it is easy to show $\nu^2 \perp \mu$ (using the technique in the hint above) and $\nu = \nu^1 + \nu^2$.

Uniqueness Assume we have two Lebesgue decompositions

$$\nu = \nu_1 + \nu_2 = \overline{\nu_1} + \overline{\nu_2}$$

with $\nu_1, \overline{\nu_1} \ll \mu$, and $\nu_2, \overline{\nu_2} \perp \mu$. We have $\nu_1 - \overline{\nu_1} = \overline{\nu_2} - \nu_2$ (*).

There is a set A s.t. $\mu(A) = 0, \nu_2(A^C) = 0$ and a set B s.t. $\mu(B) = 0, \overline{\nu_2}(B^C) = 0$. Let $C := A \cup B$, then $\mu(C) = 0$. Take set $E \in \mathcal{F}$ and consider part of E in C and not in C.

$$\mu(E \cap C) \leq \mu(C) = 0, \Rightarrow \nu_1(E \cap C) = \overline{\nu_1}(E \cap C) = 0$$

by absolute continuity. Since $E \cap C^C = E \cap (A \cup B)^C = E \cap A^C \cap B^C$

$$\nu_2(E \cap C^C) \le \nu_2(A^C) = 0, \quad \text{similarly } \overline{\nu_2}(E \cap C^C) = 0$$

so we see

$$(\nu_1 - \overline{\nu_1})(E) = (\nu_1 - \overline{\nu_1})(E \cap C) + (\nu_1 - \overline{\nu_1})(E \cap C^C) \overset{\text{by } (*)}{=} 0 + (\overline{\nu_2} - \nu_2)(E) = 0 + 0 = 0$$

so $\nu_1 = \overline{\nu_1}$, and it follows $\overline{\nu_2} = \nu_2$. ∎

Corollary 6.42 (Radon-Nykodim)**.** *If μ, ν are two σ-finite measures(unsigned) on $(X, \mathcal{F})$, then the following are equivalent:*

1. $\nu \ll \mu$
2. $\nu = \mu_f$ *for some $\mathcal{F}$-measurable* $f : X \to \overline{\mathbb{R}^+}$

Proof. Clearly 2 implies 1.
We have Lebesgue decomposition $\nu = \nu_1 + \nu_2$ where $\nu_1 \ll \mu$, $\nu_2 \perp \mu$ and $\nu_1 = \mu_f$ for some $\mathcal{F}$-measurable f. So for $1 \Rightarrow 2$, we just need to show $\nu_2 = 0$. There is a set $A \in \mathcal{F}$ s.t. $\nu_2(A^C) = 0, \mu(A) = 0$. So $\nu(A) = 0$ as $\nu \ll \mu$. $\nu_1 \ge 0$ so $\nu_2(A) \le \nu(A) = 0$. Therefore, $\nu_2(X) = \nu_2(A) + \nu_2(A^C) = 0$. This means $\nu_2 \equiv 0$ ∎

6.6 L-p space

Recall that L^p norm of a function is defined as

$$\|f\|_{L^p(\mu)} := \left(\int |f|^p \, d\mu \right)^{1/p}$$

and we can induce a metric $d_p(f, g) = \|f - g\|_p$. When the measure used is clear, L^p norm can be written as $\|f\|_p$ instead.

The infinite norm is defined using $\operatorname{ess\,sup} |f|$:

$$\|f\|_{L^\infty(\mu)} := \operatorname{ess\,sup} |f| = \inf\{a \ge 0 \,:\, \mu\{|f| > a\} = 0\} = \inf\{a \ge 0 \,:\, |f| \le a \ \mu\text{-a.e.}\}$$

Remark. I defined in previous section $\operatorname{ess\,sup} |f| := \inf\{a > 0 \,:\, \mu\{|f| \ge a\} = 0\}$. But actually it is the same as $\inf\{a \ge 0 \,:\, \mu\{|f| > a\} = 0\}$.

The space $\mathcal{L}_p(\mu)$ is defined as

$$\left\{ f : X \to \mathbb{R} \,:\, \|f\|_{L^p(\mu)} < \infty \right\}$$

There are two important inequalities relating to L^p norms, Hölder and Minkowski inequalities.

Definition 6.43 (Conjugate)**.** Given $p > 1$, we say q is conjugate of p if $\frac{1}{p} + \frac{1}{q} = 1$. And if $p = 1$, we take $q = \infty$. (vice versa)

Proposition 6.44 (Hölder's inequality)**.** *If p, q are conjugate, and given $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$, we have*

$$\int |fg| \, d\mu \le \|f\|_p \|g\|_q < \infty$$

We use a helper inequality for the proof

Proposition 6.45 (Young's inequality)**.** *Given* $1 < p, q < \infty$ *s.t.* p, q *are conjugate, then for any* $a, b \geq 0$*,*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

Proof. The proof is not difficult, but I will introduce a new perspective here. This clearly hold if $a = 0$ or $b = 0$. So assume $a, b > 0$.
Define function $F(x) := \frac{x^p}{p}$, as $p > 1$, it is convex. Legendre transformation of a convex function is defined as

$$(TF)(\theta) := \sup_{x>0}\{\theta x - F(x)\}$$

so we can see that $\theta x \leq F(x) + (TF)(\theta)$. So the only mission left is to show $(TF)(\theta) = \frac{\theta^q}{q}$. This is just a regular calculus question, so left as exercise. ■

Now we can go back to prove Hölder's inequality

Proof. **Case 1.** $p > 1$
From Young's inequality we have

$$\int |fg| \, d\mu \leq \int \frac{|f|^p}{p} \, d\mu + \int \frac{|g|^q}{q} \, d\mu < \infty$$

Alternatively, pick $A > 0$, we can apply Young's inequality on $\frac{f}{A}, gA$

$$\int |\frac{f}{A} gA| \, d\mu \leq \int \frac{|f|^p}{A^p p} \, d\mu + \int \frac{|g|^q}{q} A^q \, d\mu$$

define $\alpha := \int |f|^p \, d\mu$, $\beta := \int |g|^q \, d\mu$, and we note that LHS of inequality is independent of A, so we can take inf on RHS:

$$\int |\frac{f}{A} gA| \, d\mu \leq \inf_{A>0} \left\{ \alpha \frac{1}{pA^p} + \beta \frac{A^q}{q} \right\}$$

again using calculus one can find $A = \left(\frac{\alpha}{\beta}\right)^{\frac{1}{p+q}}$ minimises RHS, so

$$\int |fg| \, d\mu \leq \alpha^{\frac{q}{p+q}} \beta^{\frac{p}{p+q}} \left(\frac{1}{p} + \frac{1}{q} \right) = \alpha^{\frac{q}{p+q}} \beta^{\frac{p}{p+q}}$$

Using the fact that p, q are conjugate, divide nominator and denominator by pq,

$$\frac{q}{p+q} = \frac{\frac{1}{p}}{\frac{1}{q} + \frac{1}{p}} = \frac{1}{p}$$

similarly $\frac{p}{p+q} = \frac{1}{q}$, so

$$\int |fg| \, d\mu \leq \alpha^{\frac{1}{p}} \beta^{\frac{1}{q}}$$

which is the inequality required.

Case 2. $p = 1, q = \infty$.
Denote $\operatorname{ess\,sup} |g| =: c$, by definition,

$$\mu(\{x : |g(x)| > c\}) = 0$$

so

$$\int |fg| \, d\mu = \int_{\{x : |g(x)| \le c\}} |fg| \, d\mu$$

$$\le \int_{\{x : |g(x)| \le c\}} c|f| \, d\mu \le c \int |f| \, d\mu = \|g\|_\infty \|f\|_1$$

■

A consequence of Hölder's inequality is triangular inequality for L^p norms,

Proposition 6.46 (Minkowski inequality)**.** *If f, g are s.t. $\|f\|_p, \|g\|_p < \infty$ where $1 \le p \le \infty$, then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p$$

Note we have already proved the case $p = \infty$, see Proposition 6.10.

Proof. Case $p = 1$ is easy.
Consider $1 < p < \infty$,

$$\int |f+g|^p \, d\mu = \int |f+g|^{p-1}|f+g| \, d\mu \le \int |f+g|^{p-1}|f| \, d\mu + \int |f+g|^{p-1}|g| \, d\mu$$

using Hölder inequality,

$$\int |f + g|^{p-1}|f| \, d\mu \le \left(\int |f + g|^{(p-1)q} \right)^{\frac{1}{q}} \|f\|_p$$

$$\int |f + g|^{p-1}|g| \, d\mu \le \left(\int |f + g|^{(p-1)q} \right)^{\frac{1}{q}} \|g\|_p$$

where q is conjugate of p. And note $1 = \frac{1}{p} + \frac{1}{q} \Rightarrow pq = p + q \Rightarrow pq - q = (p-1)q = p$. So

$$\int |f + g|^p \, d\mu \le (\|f\|_p + \|g\|_p) \left(\int |f + g|^p \right)^{\frac{1}{q}}$$

The inequality is trivial if $\int |f + g|^p \, d\mu = 0$, so we assume it is not 0 now,

$$\left(\int |f + g|^p \right)^{1 - \frac{1}{q}} \le \|f\|_p + \|g\|_p$$

LHS is exactly $\|f + g\|_p$. ■

The following are corollaries of Hölder and Minkowski inequalities.

Corollary 6.47. *$\mathcal{L}_p(\mu)$ is linear space for any $1 \le p \le \infty$. i.e. take constant $\alpha \in \mathbb{R}$ and $f, g \in \mathcal{L}_p(\mu)$, then $\alpha f + g \in \mathcal{L}_p(\mu)$.*
Note we already proved the case $p = \infty$ in the section 6.1

Proof. Left as exercise. ∎

Corollary 6.48. *L^p norms are indeed norms on space $\mathcal{L}_p(\mu)$. Recall this means*

- *Given $\alpha \in \mathbb{R}, f \in \mathcal{L}_p(\mu)$, then $\|\alpha f\|_p = |\alpha|\|f\|_p$*
- *$\|f\|_p = 0 \Rightarrow f = 0$ a.e. (or $f \sim 0$)*
- *$\|f+g\|_p \leq \|f\|_p + \|g\|_p$*

Now let's do some topologies on $\mathcal{L}_p$ space.

Definition 6.49 (Convergence in $\mathcal{L}_p$)**.** Given $f_n \in \mathcal{L}_p$, we say it converges to some $f \in \mathcal{L}^p$ if

$$\|f_n - f\|_p \to 0 \text{ as } n \to \infty$$

Consider the following example
$f_n := n^{1/p} 1_{[0,\frac{1}{n}]}$, clearly it converges to 0 a.e. and converges to 0 in measure, but

$$\|f_n\|_p = \left(\int |f_n|^p \, d\mu \right)^{1/p} = 1$$

so f_n does not converge to 0 in $\mathcal{L}_p$.

What if we change the condition to uniform convergence a.e.?
Let $f_n := \frac{1}{n^{1/p}} 1_{[0,n]}$, it convergence to 0 unif a.e. and converges to 0 in measure, but again $\|f_n\|_p = 1$.

Before investigating the relationships between convergences, we will first prove that $\mathcal{L}_p$ is complete.

Theorem 6.50. *$\mathcal{L}_p$ is complete.*

Proof. **Step 1.** Given Cauchy sequence $(f_n)_{n\geq 1}$, we aim to find subsequence f_{n_k} and a set F with $\mu(F^C) = 0$ s.t. $(f_{n_k}(x))_{k\geq 1}$ is Cauchy for $x \in F$.

Since $(f_n)_{n\geq 1}$ is Cauchy, given $\epsilon > 0$, exists N s.t. $\forall n, m \geq N$, $\int |f_n - f_m|^p \, d\mu = \|f_n - f_m\|_p^p \leq \epsilon$. Let $\epsilon_k := \frac{1}{2^{kp}} \frac{1}{2^k}$ (this choice will make sense later), then there is n_k s.t. $\forall n, m \geq n_k$, $\int |f_n - f_m|^p \, d\mu \leq \epsilon_k$, pick $n = n_k$, $m = n_{k+1}$,

$$\int |f_{n_k} - f_{n_{k+1}}|^p \, d\mu \leq \frac{1}{2^{kp}} \frac{1}{2^k} \qquad (*)$$

Consider the set $F_k := \left\{ x : |f_{n_k}(x) - f_{n_{k+1}}(x)| \geq \frac{1}{2^k} \right\}$, for $x \in F_k$, $1^p \leq 2^{kp} |f_{n_k}(x) - f_{n_{k+1}}(x)|^p$, so

$$\mu(F_k) = \int_{F_k} 1 \, d\mu \leq \int_{F_k} 2^{kp} |f_{n_k}(x) - f_{n_{k+1}}(x)|^p \, d\mu \leq \frac{1}{2^k} \quad \text{by } (*)$$

Define $F^C := \bigcap_{m\geq 1} \bigcup_{k\geq m} F_k$, so

$$\mu(F^C) \leq \sum_{k\geq m} \frac{1}{2^k} = 2\frac{1}{2^m} \; \forall m \geq 1$$

so $\mu(F^C) = 0$.

Now we prove that $(f_{n_k}(x))_{k\geq 1}$ is Cauchy for $x \in F$. First translate $x \in F$:

$$F = \bigcup_{m\geq 1} \bigcap_{k\geq m} \left\{ x \, : \, |f_{n_k}(x) - f_{n_{k+1}}(x)| < \frac{1}{2^k} \right\}$$

so $x \in F$ means $\exists m \geq 1,$ s.t. $\forall k \geq m, |f_{n_k}(x) - f_{n_{k+1}}(x)| < \frac{1}{2^k}$. So given any $m \leq k < j$,

$$|f_{n_k}(x) - f_{n_j}(x)| \leq |f_{n_k}(x) - f_{n_{k+1}}(x)| + ... + |f_{n_{j-1}}(x) - f_{n_j}(x)| < \sum_{i=k}^{\infty} \frac{1}{2^i} = 2\frac{1}{2^k}$$

it is easy to see that this implies $(f_{n_k}(x))_{k\geq 1}$ is Cauchy.

Step 2. Construct $f(x)$, the limit
Since $\mathbb{R}$ is complete, $f_{n_k}(x) \to f(x)$ as $k \to \infty$ for $x \in F$, and we define $f(x) = 0$ if $x \in F^C$. Since $\mu(F^C) = 0$, $f_{n_k} \to f$ a.e. so f is measurable. Then we aim to prove $f \in \mathcal{L}_p$.
First note $\|f_n\|_p \leq c \; \forall n \geq 1$ (c is constant), this is a direct consequence of f_n being Cauchy in $\mathcal{L}_p$, please try to prove it by yourself.
Now apply Fatou's lemma,

$$\int |f|^p \, d\mu \leq \liminf \int |1_F f_{n_k}|^p \, d\mu = \liminf \int |f_{n_k}|^p \, d\mu < \infty$$

Step 3. Prove that $f_n \xrightarrow{\mathcal{L}_p} f$.
Note $f_{n_k} \to f$ a.e., so apply Fatou's lemma

$$\int |f - f_n|^p \, d\mu \leq \lim_k \int |f_{n_k} - f_n|^p \, d\mu \qquad (**)$$

using the fact that f_n is Cauchy, for any $\epsilon > 0$ we can find n_0 s.t. for all $a, b \geq n_0$, $\int |f_a - f_b|^p \, d\mu < \epsilon$, so pick n s.t.$n, n_k \geq n_0$, then RHS of (**) is less than ϵ, which finishes step 3. ∎

Now we investigate what conditions gives us Convergence in measure $\Rightarrow$ convergence in $\mathcal{L}^p$.

Definition 6.51 (Equicontinuous)**.** A set of measures $\{\nu_\alpha \, : \, \alpha \in I\}$ is equicontinuous at $\emptyset$ if for any sequence of sets $B_k \downarrow \emptyset$, $\forall \epsilon > 0, \exists k_0$ s.t. if $k \geq k_0$, then $\sup_{\alpha\in I} \nu_\alpha(B_k) \leq \epsilon$.

Definition 6.52 (Uniformly absolutely continuous w.r.t μ)**.** If for any $\epsilon > 0, \exists \delta > 0$, s.t.

$$\mu(B) \leq \delta \Rightarrow \sup_{\alpha\in I} \nu_\alpha(B) \leq \epsilon$$

then $\{\nu_\alpha \, : \, \alpha \in I\}$ is said to be uniformly absolutely continuous w.r.t μ.

In this section, we usually study a sequence of functions $f_n \in \mathcal{L}^p$, and we define the set of measures $\nu_n(A) := \int_A |f_n|^p \, d\mu$ so $\nu_n < \infty$. Also we see $\nu_n \ll \mu$. Recall the example we studied before

$$f_n = n^{1/p} 1_{[0, \frac{1}{n}]}$$

this sequence converges to 0 point-wise a.e. and in measure but not in $\mathcal{L}^p$. Now we show that the measures induced by this sequence $\{\nu_n : n \in \mathbb{N}\}$ is not equicontinuous.

Let $B_k = (0, \frac{1}{k}) \downarrow \emptyset$. If n is large enough that $n \geq k$, then

$$\nu_n(B_k) = \int_{(0,\frac{1}{n})} |f_n|^p \, d\lambda = 1$$

so $\sup_{n\in\mathbb{N}} \nu_n(B_k) \to 1 \neq 0$ as $k \to \infty$.
You can show, by picking a sequence of sets decreasing to $\emptyset$ s.t. the other example

$$g_n := \frac{1}{n^{1/p}} 1_{[0,n]}$$

which converges to 0 uniformly and in measure but not in $\mathcal{L}_1$, that the measures $\{\nu_n : n \in \mathbb{N}\}$ is not equicontinuous.

Theorem 6.53. *Given $(f_n) \in \mathcal{L}_p$, and measure space $(X, \mathcal{F}, \mu)$ that is σ-finite. Define $\nu_n(A) := \int_A |f_n|^p \, d\mu$.*
If $\{\nu_n : n \in \mathbb{N}\}$ is equicontinuous at $\emptyset$ and $f_n \xrightarrow{\mu} f$, then $f_n \to f$ in $\mathcal{L}^p$.

Before proving main theorem of this section, we need a helper lemma.

Lemma 6.54. *If $\{\nu_\alpha : \alpha \in I\}$ is equi-cont. at $\emptyset$, and for any $\alpha \in I$, $\nu_\alpha \ll \mu$, then $\{\nu_\alpha\}$ is uniform. abs. cont.*

Proof. Assume for contradiction that $\{\nu_\alpha\}$ is not uniform. abs. cont. then reversing the definition we have

$$\exists \epsilon > 0 \text{ s.t. } \forall \delta_n := \frac{1}{2^n}, \exists B_n \text{ with } \mu(B_n) \leq \frac{1}{2^n} \text{ s.t. } \exists \alpha_n \text{ s.t. } \nu_{\alpha_n}(B_n) \geq \epsilon$$

Let $A_k := \bigcup_{n\geq k} B_n$ so $\mu(A_k) \leq \sum_{n\geq k} \frac{1}{2^n} = \frac{2}{2^k}$. We see

$$A_k \downarrow \bigcap_{j\geq 1} \bigcup_{n\geq j} B_n =: A$$

Since $A \subseteq A_k$, $\mu(A) \leq \mu(A_k) \leq \frac{1}{2^{k-1}}$ for any $k \geq 1$. This means $\mu(A) = 0$. Using absolute continuity, we have $\nu_\alpha(A) = 0$ for any $\alpha \in I$.

Now we need to construct a sequence going down to $\emptyset$ so that we can contradict. Since $A_k \downarrow A$, $A_k \setminus A \downarrow \emptyset$,

$$\begin{aligned} \nu_\alpha(A_k \setminus A) &= \nu_\alpha(A_k) \quad \text{as } A \subseteq A_k \text{ and } \nu_\alpha(A) = 0 \\ &\geq \nu_\alpha(B_k) \end{aligned}$$

picking $\alpha = \alpha_k$, we get $\nu_{\alpha_k}(A_k \setminus A) = \nu_{\alpha_k}(A_k) \geq \epsilon$ as $\nu_{\alpha_k}(B_k) \geq \epsilon$ by our choice. So we see $\sup_{\alpha\in I} \nu_\alpha(A_k \setminus A)$ cannot converge to 0 as $k \to \infty$. Contradiction to the fact that $\{\nu_\alpha : \alpha \in I\}$ is equi-cont. at $\emptyset$. ∎

Now we can prove the theorem

Proof. **Step 1.** Use the fact that μ is σ-finite.
We can find $E_k \uparrow X$ with $\mu(E_k) < \infty$ so $E_k^C \downarrow \emptyset$. So by equicontinuity,

$$\sup_{n\geq 1} \nu_n(E_k^C) \to 0 \qquad \text{as } k \to \infty$$

Now aim to prove that $(f_n)_{n\geq 1}$ is Cauchy in $\mathcal{L}^p$.

Step 2. Fix $\epsilon > 0$, we want to find $n_0 > 0$ s.t. $\forall n, m \geq n_0$, $\int |f_n - f_m|^p \, d\mu \leq \epsilon$. First pick k_0 s.t. if $k \geq k_0$, $\sup_{n\geq 1} \nu_n(E_k^C) \leq \epsilon$ (*). We divide the integral into three parts:

$$\text{(I)} \quad \int_{E_{k_0}^C} |f_n - f_m|^p \, d\mu$$

using binomial expansion one can see that $|f_n - f_m|^p = |(f_n - f_m)^p|$ can be expanded into a sum of 2^p terms of the form $(f_n)^j(-f_m)^{p-j}$. And since $\max\{f_n, f_m\} \leq |f_n| + |f_m|$, $|f_n - f_m|^p \leq 2^p|f_n|^p + 2^p|f_m|^p$.

$$\int_{E_{k_0}^C} |f_n - f_m|^p \, d\mu \leq 2^p \left(\int_{E_{k_0}^C} |f_m|^p \, d\mu + \int_{E_{k_0}^C} |f_n|^p \, d\mu \right)$$

RHS, by definition of ν_n, is

$$2^p(\nu_n(E_{k_0}^C) + \nu_m(E_{k_0}^C)) \leq 2^{1+p}\epsilon \qquad \text{by (*)}$$

the above inequality is true for any $n, m \geq 1$ as no requirement is imposed on m, n.

Now consider integral on E_{k_0}, to use the conditions equicontinuity and the fact that $\nu_n \ll \nu$ for all n (due to definition), which imply unif. abs. cont. continuity by helper lemma, we divide E_{k_0} further into two parts

$$\text{(II)} \quad \int_{E_{k_0} \cap \left\{|f_n - f_m|^p \leq \epsilon/\mu(E_{k_0})\right\}} |f_n - f_m|^p \, d\mu$$

$$\leq \frac{\epsilon}{\mu(E_{k_0})} \mu\left(E_{k_0} \cap \{|f_n - f_m|^p \leq \epsilon/\mu(E_{k_0})\}\right) \leq \epsilon$$

Now last part,

$$\text{(III)} \quad \int_{E_{k_0} \cap \underbrace{\left\{|f_n - f_m|^p > \frac{\epsilon}{\mu(E_{k_0})}\right\}}_{=:B_{n,m}}} |f_n - f_m|^p \, d\mu$$

using the same trick as above,

$$\leq 2^p(\nu_n(B_{n,m}) + \nu_m(B_{n,m}))$$

we will only consider the first term as second term can be dealt in the same way.

$$2^p\nu_n(B_{n,m}) \leq 2^p\nu_n\left(|f_n - f|^p > \frac{\epsilon}{2\mu(E_{k_0})}\right) + 2^p\nu_n\left(|f_m - f|^p > \frac{\epsilon}{2\mu(E_{k_0})}\right)$$

by unif. abs. cont. we can find $\delta > 0$ s.t. if $\mu(B) \leq \delta$, $\nu_k(B) \leq \epsilon \forall k$. Since $f_n \xrightarrow{\mu} f$,

$$\mu \left\{ |f_j - f| > \frac{\epsilon}{2\mu(E_{k_0})} \right\} \to 0 \quad \text{as } j \to \infty$$

so we can find n_0 s.t. if $j \geq n_0$, $\mu \left\{ |f_j - f| > \frac{\epsilon}{2\mu(E_{k_0})} \right\} \leq \delta$, so if $n, m \geq n_0$,

$$\nu_n \left\{ |f_n - f| > \frac{\epsilon}{2\mu(E_{k_0})} \right\} \leq \epsilon \quad \nu_n \left\{ |f_m - f| > \frac{\epsilon}{2\mu(E_{k_0})} \right\} \leq \epsilon$$

similarly one can bound the term with ν_m. So if $n, m \geq n_0$,

$$\text{(III)} \leq 2^{p+2}\epsilon$$

So we see the integral $\int |f_n - f_m|^p \, d\mu$ can be bounded by a constant times ϵ, which implies claim of step 2.

Step 3.
Since $\mathcal{L}^p$ is complete, and (f_n) is Cauchy, there is a function g s.t. $f_n \to g$ in $\mathcal{L}^p$. This implies $f_n \xrightarrow{\mu} g$ (Though we have not proved convergence in $\mathcal{L}^p$ implies convergence in measure, but the proof is very similar to the case $p = 1$) Also by condition given, $f_n \xrightarrow{\mu} f$, so $f = g$ a.e. That means $f_n \to f$ in $\mathcal{L}^p$. ∎

Now we discuss two practical criteria that implies equicontinuity.

Proposition 6.55. *If $(f_n) \in \mathcal{L}_p$ and $\exists h \in \mathcal{L}^1$ s.t. $|f_n|^p \leq h$, then $\{\nu_n\}$(as defined in Theorem 6.53) is equicontinuous at $\emptyset$.*

Proof.

$$\nu_n(A) = \int_A |f_n|^p \, d\mu \leq \int_A h \, d\mu =: \mu_n(A) < \infty$$

Since measure ν_n is finite, if $B_k \downarrow \emptyset$, by continuity from above,

$$\sup_{n \geq 1} \nu_n(B_k) \leq \sup_{n \geq 1} (\mu_n(B_k)) \to \sup_{n \geq 1} (\mu_n(\emptyset)) = 0 \text{ as } k \to \infty$$

∎

Proposition 6.56. *If $(f_n) \in \mathcal{L}_p$ with $\mu(X) < \infty$, and*

$$\sup_{n \geq 1} \int_{|f_n| \geq a} |f_n|^p \, d\mu \to 0 \text{ as } a \to \infty$$

then $\{\nu_n\}$ is equicontinuous at $\emptyset$.

Proof. Pick $B_k \downarrow \emptyset$,

$$\nu_n(B_k) = \int_{B_k} |f_n|^p \, d\mu \leq \underbrace{a^p \mu(B_k)}_{\text{when we restrict integral to } \{|f_n| < a\}} + \int_{\{|f_n| \geq a\}} |f_n|^p \, d\mu$$

note the second integral should be on the set $B_k \cap \{|f_n| > a\}$ but to use the condition we drop B_k.
Fix $\epsilon > 0$, exists $A \in \mathbb{R}$ s.t. if $a \geq A$,

$$\int_{|f_n| \geq a} |f_n|^p \, d\mu \leq \frac{\epsilon}{2}$$

so

$$\nu_n(B_k) \leq A^p \mu(B_k) + \frac{\epsilon}{2}$$

since $B_k \downarrow \emptyset$ and measure is finite, by continuity from above, $\mu(B_k) \downarrow 0$. So as k grows large enough, $A^p \mu(B_k) < \frac{\epsilon}{2}$. ■

6.7 $\mathcal{L}^p$ as Linear Space (Extra)

In this section we study the dual space of $\mathcal{L}^p$, first we note that $B = \mathcal{L}^p$ is a linear space. i.e. if $\alpha \in \mathbb{R}, f, g, \in B$, then $\alpha f + g \in B$. We can define norms $\|f\|$ on B (e.g. $\|f\|_p$) Recall that norm $\|\dot{}\| : B \to \mathbb{R}$ has to satisfy the following three properties

- If $f = 0$, then $\|f\| = 0$
- If $\alpha \in \mathbb{R}, f \in B$, then $\|\alpha f\| = |\alpha| \|f\|$.
- $\|f + g\| \leq \|f\| + \|g\|$

Remark. One needs to be careful with the first property, because L^p norm is only a norm if we define $\mathcal{L}^p$ as set of equivalent classes, where $f \sim g$ iff $f = g$ a.e. So first property of norm actually includes all functions s.t. $f = 0$ a.e.

Definition 6.57 (Linear transformation)**.** A map/transformation $T : B \to \mathbb{R}$ is called linear if

$$T(\alpha f + g) = \alpha T(f) + T(g)$$

Sometimes we abbreviate $T(f)$ as Tf.

For example, let $B = \mathcal{L}^p$, pick $g \in \mathcal{L}^q$ where q is conjugate of $p(\frac{1}{p} + \frac{1}{q} = 1)$. Define $T_g : \mathcal{L}^p \to \mathbb{R}$ to be

$$T_g(f) := \int fg \, d\mu$$

this integral is always finite by Hölder's inequality.

Definition 6.58 (Bounded Linear Transformation)**.** Linear map $T : B \to \mathbb{R}$ is called bounded if $\exists c < \infty$ s.t. $|T(f)| \leq c\|f\|$ for all $f \in B$.

We always define the operator norm as $\|T\| := \sup_{f \neq 0} \frac{|T(f)|}{\|f\|}$, so if T is bounded, $\|T\| \leq c < \infty$ (alternative definition of bounded is $\|T\| \leq c$ for some $c \in \mathbb{R}$). Also note that due to linearity,

$$\sup_{f \neq 0} \frac{|T(f)|}{\|f\|} = \sup_{f \neq 0} \left| \frac{T(f)}{\|f\|} \right| = \sup_{f \neq 0} \left| T\left(\frac{f}{\|f\|} \right) \right| = \sup_{\|f\|=1} |T(f)|$$

you are encouraged to prove that $|T(f)| \leq \|T\| \|f\|$ on your own.

Now recall the example given above where $T_g(f) := \int fg \, d\mu$. We will study this important example for the rest of section

Proposition 6.59. *If $T_g : \mathcal{L}^p \to \mathbb{R}$ is defined as in above example, where $g \in \mathcal{L}^q$. Then $\|T_g\| = \|g\|_q < \infty$.*

Proof. One direction is easy,

$$|T_g f| = \left|\int fg \, d\mu\right| \leq \|f\|_p \|g\|_q \text{ by Hölder's inequality}$$

take $\sup_{\|f\|_p=1}$ on both sides will not change inequality, so $\|T_g\| \leq \|g\|_q$.

To prove the other direction, we need to find $\hat{f} \neq 0$ s.t.

$$|T_g \hat{f}| = \|\hat{f}\|_p \|g\|_q$$

then by definition of $\|T_g\| := \sup_{f\neq 0} \frac{|T_g(f)|}{\|f\|_p}$, we have $\|T_g\| \geq \|g\|_q$.

Case 1. $p > 1$.

$$T_g(f) = \int gf \, d\mu \qquad \text{and } \|g\|_q = \left(\int |g|^q \, d\mu\right)^{\frac{1}{q}}$$

so we need to build $|g|$ first, by adding factor $sign(g)$ to f, where

$$sign(g) = \begin{cases} 1 & \text{if } g \geq 0 \\ -1 & \text{if } g < 0 \end{cases}$$

But f needs another term $|g|^a$ to build $|g|^q$ so let's try to decide the value of a. Let $f := sign(g)|g|^a$

$$T_g(f) = \int sign(g) g |g|^a \, d\mu = \int |g|^{1+a} \, d\mu$$

since $\frac{1}{p} + \frac{1}{q} = 1$, $\frac{q}{p} + 1 = q$. So picking $a = \frac{q}{p}$ gives us $T_g(\hat{f}) = \int |g|^q \, d\mu = \|g\|_q^q$. We need to check that $f \in \mathcal{L}^p$ now,

$$\int |f|^p \, d\mu = \int \left(|g|^{\frac{q}{p}}\right)^p \, d\mu = \int |g|^q \, d\mu < \infty \quad (*)$$

as $g \in \mathcal{L}^q$. Now prove f satisfies our requirement

$$|T_g(\hat{f})| = \|g\|_q^q = \frac{\|g\|_q^q}{\|\hat{f}\|_p} \|\hat{f}\|_p = \|g\|_q^{q-\frac{q}{p}} \|\hat{f}\|_p$$

where the last equality is due to (*). It is not hard to derive from $\frac{1}{p} + \frac{1}{q} = 1$ that $q - \frac{q}{p} = 1$. So $\hat{f}$ satisfies requirement.

Case 2. $p = 1, q = \infty$
$\|g\|_\infty = \operatorname{ess\,sup} |g| = \inf\{a \geq 0 : \mu(\{|g| > a\}) = 0\}$ This case is similar to above case, construct a function $\hat{f}$ s.t. $|T_g \hat{f}| = \|\hat{f}\|_1 \|g\|_\infty = \|\hat{f}\|_1 \operatorname{ess\,sup} |g|$. Proof left as exercise. ■

Surprisingly, the converse of above proposition holds

Theorem 6.60. *Fix $1 < p < \infty$ and linear map $T : \mathcal{L}^p \to \mathbb{R}$ which is bounded. Then exists $g \in \mathcal{L}^q$ (q is conjugate of p) s.t.*

$$T(f) = \int fg \, d\mu \quad \forall f \in \mathcal{L}^p$$

Proof. Make the hypothesis (H) $\mu(X) < \infty$. Remove it later.
The only theorem constructing a new function within integral is Radon-Nykodim(R-N) theorem. But we have to first construct a measure

Step 1. Construct a measure
Let set function $\nu : \mathcal{F} \to \mathbb{R}$ be defined as $\nu(A) := T(1_A)$. Please note ν takes negative values as T can send positive values to negative ones. So we aim to prove ν is a signed measure.
Claim 1. ν is additive
Easy exercise.
Claim 2. ν is σ-addtive
Given sequence of disjoint sets $(A_j)_{j\geq 1}$, we aim to show $\left|\nu(\sum_{j\geq 1} A_j) - \sum_{j=1}^{n} \nu(A_j)\right| \to 0$. This term equals

$$\left|T\left(1_{\sum_{j\geq 1} A_j}\right) - T\left(1_{\sum_{j=1}^{n} A_j}\right)\right| = \left|T\left(1_{\sum_{j\geq n} A_j}\right)\right| \text{ by linearity of } T$$

$$\leq c \left\| 1_{\sum_{j\geq n} A_j} \right\|_p \text{ as } T \text{ is bounded}$$

It is not hard to show $\left\| 1_{\sum_{j\geq n} A_j} \right\|_p = \mu(\sum_{j\geq n} A_j) \to 0$ as $n \to \infty$. Since $\mu(X) < \infty$, $\sum_{j\geq n} A_j \downarrow \emptyset$ and μ is continuous from above.
So $\sum_{j\geq 1} A_j) = \lim_{n\to\infty} \sum_{j=1}^{n} \nu(A_j) = \sum_{j=1}^{\infty} \nu(A_j)$.

Claim 3. ν is σ-finite and $\nu \ll \mu$

$$|\nu(A)| = |T(1_A)| \leq c\|1_A\|_p = c\left(\int |1_A|^p \, d\mu\right)^{\frac{1}{p}} = c\mu(A)^{\frac{1}{p}} \text{ (**)}$$

By Hahn-Jordan, there are disjoint sets P, N s.t. $P = N^C$ and if $A \subseteq P$, $\nu(A) \geq 0$, if $B \subseteq N$, $\nu(B) \leq 0$.

- If $A \subseteq P$, then by (**), $|\nu(A)| \leq c\mu(P)^{\frac{1}{p}} < \infty$.
- If $B \subseteq N$, then by (**), $|\nu(B)| \leq c\mu(N)^{\frac{1}{p}} < \infty$.

And since $\nu = \nu^+ - \nu^-$ where ν^+ is only defined for sets inside P and ν^- is only defined for sets inside N. We see from the above two inequalities that ν^+, ν^- are finite, so ν is finite. (implies σ-finite)
And the above two inequalities also shows absolute continuity: if $\mu(A) = 0$

$$|\nu(A)| \leq c\mu(A \cap N)^{1/p} + c\mu(A \cap P)^{1/p} = 0$$

so $\nu(A) = 0$.

Step 2. Construct g and prove it satisfies the theorem

By R-N theorem , $\exists g$ s.t. $\nu(A) = \int_A g\, d\mu$. Note

$$\int_A g\, d\mu = \nu(A) = T(1_A) = \int 1_A g\, d\mu$$

and $1_A \in \mathcal{L}_p$ as $\int |1_A|^p\, d\mu = \int 1_A\, d\mu = \mu(A) < \mu(X) < \infty$.
So the required equation $T(f) = \int fg\, d\mu$ holds for single indicator function. Now it is clear that we should extend this to simple functions, positive functions, then all functions.

Case 1. Simple function easy exercise.
Case 2. For $f \in \mathcal{L}^p$, $f \geq 0$, pick $0 \leq f_n \uparrow f$ where f_n are all simple. Decompose g into $g = g^+ - g^-$. We first prove $\int fg^+\, d\mu < \infty$.
Consider $E^+ := \{g \geq 0\}$, $f_n 1_{E^+} \uparrow f 1_{E^+}$. Since f_n are simple, by case 1,

$$T\left(f_n 1_{E^+}\right) = \int f_n 1_{E^+} g\, d\mu = \int f_n g^+\, d\mu \uparrow \int f g^+\, d\mu$$

Now consider the difference between $T(f_n 1_{E^+})$ and $T(f 1_{E^+})$:

$$|T(f_n 1_{E^+}) - T(f 1_{E^+})| \leq c \|1_{E^+}(f_n - f)\|_p = c\left(\int |f_n - f|^p\, d\mu\right)^{1/p}$$

since $f_n \uparrow f$, $|f_n - f| \leq |f|$, and $f \in \mathcal{L}^p$ so by dominated convergence theorem, the last term of above inequality approaches 0. So $T\left(f_n 1_{E^+}\right)$ approaches to both $T(f 1_{E^+})$ and $\int fg^+\, d\mu$, that means $\int fg^+\, d\mu = T(f 1_{E^+}) < \infty$ (since codomain of T is defined as $\mathbb{R}$). Similarly one can show $T(f 1_{E^-}) = \int fg^-\, d\mu$ where $E^- := \{g < 0\}$, so $\int fg^-\, d\mu < \infty$. Now we have

$$T(f_n) = \int f_n g\, d\mu = \int f_n g^+\, d\mu - \int f_n g^-\, d\mu \xrightarrow{\text{by MCT}} \int fg^+\, d\mu - \int fg^-\, d\mu$$

since two terms in the last expressions are finite,

$$= \int f(g^+ - g^-)\, d\mu = \int fg\, d\mu$$

Case 3. General f. Easy exercise.

Step 3. Prove $g \in \mathcal{L}^q$.

We use boundedness of T and construct a sequence of functions f_n to bound $\|g\|_q$. First define a function $F_n(g)$ that only takes bounded part of g:

$$F_n(g(x)) := \begin{cases} n & \text{if } g(x) > n \\ g(x) & \text{if } |g(x)| \leq n \\ -n & \text{if } g(x) < -n \end{cases}$$

now we define $f_n := sign(g)|F_n(g)|^{q/p}$, just like the proof of previous proposition $sign(g)$ is to create $|g|$ and exponent q/p is to create q-norm of g. Note f_n is bounded by $n^{q/p}$ and $\mu(X) < \infty$ so $f_n \in \mathcal{L}^p$. Since T is bounded on $\mathcal{L}^p$, $|T(f_n)| \leq c\|f_n\|_p$ where $c = \|T\|$.

$$\Rightarrow \left|\int |g||F_n(g)|^{q/p}\, d\mu\right| \leq c\left(\int |F_n(g)|^q\right)^{\frac{1}{p}} \leq c\left(\int |F_n(g)|^{q-1}|g|\right)^{\frac{1}{p}}$$

where the last inequality is due to $|F_n(g)| \le |g|$.
Note $q/p = q-1$ as p, q are conjugate, so we have two same terms $\int |g||F_n(g)|^{q/p}\, d\mu$ on ends of the inequality. And since $F_n(g)$ is bounded, $|g| \in \mathcal{L}_1$ (as $\nu(A) = \int_A g\, d\mu$ and μ, ν are finite, so g must be finite) Function $0 \in \mathcal{L}^q$ so we can assume $g \neq 0$, then $F_n(g) \neq 0$, so the term $0 < \int |g||F_n(g)|^{q/p}\, d\mu < \infty$. Therefore, we can divide it on both sides of inequality without changing sign of inequality, that means:

$$\left(\int |g||F_n(g)|^{q-1}\right)^{1-\frac{1}{p}} \le c = \|T\|$$

$|F_n(g)| \uparrow |g|$ so by MCT,

$$\left(\int |g||F_n(g)|^{q-1}\right)^{\frac{1}{q}} \uparrow \left(\int |g|^q\right)^{\frac{1}{q}}$$

so we have $\|g\|_q \le \|T\| < \infty$. This finish the proof of theorem with hypothesis (H).

Step 4. Remove assumption (H)
If μ is σ-finite, $X = \sum_{j\ge 1} E_j$ for some sets E_j with $\mu(E_j) < \infty$. Consider the measure spaces $(E_j, \mathcal{F}_j, \mu_j)$ where $\mathcal{F}_j := \mathcal{F}|_{E_j} = \{A \cap E_j \,:\, A \in \mathcal{F}\}$ and $\mu_j(B) := \mu(B \cap E_j)$. On each of these measure space, by previous steps, we can find g_j s.t. $T(f1_{E_j}) = \int_{E_j} fg_j\, d\mu$. Though g_j is only defined on E_j, we can extend it by defining $g_j(x) = 0$ for $x \notin E_j$. Now define $g := \sum_{j\ge 1} g_j 1_{E_j}$, we have

$$T(f1_{E_j}) = \int f1_{E_j} g\, d\mu$$

so the target equation $T(f) = \int fg\ d\mu$ holds for functions with support completely in one of E_j. Using linearity of T and integral, if we define $F_n := \sum_{j=1}^n E_j$, then

$$T(f1_{F_n}) = \int f1_{F_n} g\, d\mu$$

Also, since $g_j \in \mathcal{L}^q$ and $\mu(E_j) < \infty$, it is not hard to show that $g \in \mathcal{L}^q$.

Now consider a general function $f \in \mathcal{L}^p$ s.t. $f \ge 0$, decompose $g = g^+ - g^-$. One can show $\int fg^+\, d\mu < \infty$ using exactly the same arguments presented above (using $E^+ := \{g \ge 0\}$ and $E^- := \{g < 0\}$) Now consider

$$T\left(f1_{F_n \cap E^\pm}\right) = \int 1_{F_n} fg^\pm\, d\mu \to \int fg^\pm\, d\mu$$

as $F_n \uparrow X$. Note the second term should use $1_{F_n \cap E^\pm}$ but since $g^\pm$ is 0 outside $E^\pm$, we can remove $E^\pm$. But $T\left(f1_{F_n \cap E^\pm}\right)$ actually also converge to $T(f1_{E^\pm})$, as

$$|T(f1_{E^\pm}) - T\left(f1_{F_n \cap E^\pm}\right)| = \left|T\left(f1_{\sum_{j\ge n} E_j \cap E^\pm}\right)\right| \le \|T\| \|f1_{\sum_{j\ge n} E_j \cap E^\pm}\|_p$$

but $\sum_{j\ge n} E_j \downarrow \emptyset$, so it is not hard to see that RHS converges to 0 as $n \to \infty$. So

$$T(f1_{E^\pm}) = \int fg^\pm\, d\mu$$

then the result $T(f) = \int fg \, d\mu$ follows by linearity. One can extend the result to general functions using linearity of T. ■

Remark. This theorem holds for $p = 1, q = \infty$. The proof is very similar, in fact you can take the same g in each step, but you have to prove $\operatorname{ess\,sup} |g| < \infty$ so that $g \in \mathcal{L}^\infty$ instead of using an integral.

7 Differentiation

7.1 Differentiation theorem

Definition 7.1 (Differentiable)**.** Recall that a function $f : [a, b] \to \mathbb{R}$ is called differentiable at point x if the limit

$$\lim_{y \to x, y \in [a,b] \setminus \{x\}} \frac{f(y) - f(x)}{y - x}$$

exists. And the limit is called the strong derivative/derivative.

Proposition 7.2. *If $f : [a, b] \to \mathbb{R}$ is differentiable everywhere on $[a, b]$, then f is continuous f' is measurable.*
If f is differentiable λ-a.e., f' is measurable, but may not be continuous.

Proof. First statement is easy. But if we have a null set s.t. f is not differentiable, one needs completeness of Lebesgue measure to show f' is measurable. i.e. this may not generalise to other function spaces with non-complete measure space.

Proof of this proposition is left as exercise. Remember to give an example of an almost everywhere differentiable which is not continuous ■

Recall two important theorems from real analysis:

Theorem 7.3 (Fundamental theorem of calculus (1))**.** *If $f : [a, b] \to \mathbb{R}$ is differentiable and f' is Riemann integrable, then*

$$\int_a^b f'(x) = f(b) - f(a)$$

proof of the theorem is based on mean value theorem.

Theorem 7.4. *If $f : [a, b] \to \mathbb{C}$ and define $F(x) := \int_a^x f(t)\, dt$, then F is differentiable on (a, b) with $F'(x) = f(x)$, and F is continuously differentiable.*

Proof. First consider the right limit

$$\lim_{h \downarrow 0} \frac{F(x + h) - F(x)}{h}$$

and we aim to prove it equals f. Using substitution, it is not hard to see

$$\frac{F(x + h) - F(x)}{h} = \int_0^1 f(x + ht)\, dt$$

Given fixed $x \in [a, b)$ and small enough $h > 0$, since f is continuous, $t \mapsto f(x+ht)$ converges uniformly to $f(x)$ on $[0, 1]$ as $h \downarrow 0$. Also $[0, 1]$ is bounded so integral will be finite, thus

$$\int_0^1 f(x + ht)\, dt \xrightarrow{h \downarrow 0} \int_0^1 f(x)\, dt = f(x)$$

The left limit

$$\lim_{h \uparrow 0} \frac{F(x + h) - F(x)}{h}$$

can be dealt similarly. ■

Corollary 7.5 (Differentiation theorem for Riemann integrals)**.** *Given continuous $f : [a, b] \to \mathbb{C}$,*

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{x}^{x+h} f(t)\, dt = f(x) \qquad \forall x \in [a, b)$$

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{x-h}^{x} f(t)\, dt = f(x) \qquad \forall x \in (a, b]$$

$$\lim_{h\downarrow 0} \frac{1}{2h} \int_{x-h}^{x+h} f(t)\, dt = f(x) \qquad \forall x \in (a, b)$$

Lebesgue generalised this to Lebesgue integrals

Theorem 7.6 (Lebesgue differentiation theorem (1D case))**.** *Given $f : \mathbb{R} \to \mathbb{C}$ that is absolutely integrable and define $F(x) := \int_{(-\infty,x]} f(t)\, d\lambda(t)$. Then F is continuous and differentiable a.e., with $F' = f$ a.e.*

As exercise, prove continuity of F. To prove $F' = f$ a.e., we need the following theorem

Theorem 7.7 (Lebesgue differentiation theorem (1D case, alternative formulae))**.** *Given $f : \mathbb{R} \to \mathbb{C}$ that is absolutely integrable*

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{[x,x+h]} f(t)\, dt = f(x) \qquad \forall x \in \mathbb{R}$$

and

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{[x-h,x]} f(t)\, dt = f(x) \qquad \forall x \in \mathbb{R}$$

It is easy to see that first theorem follows from the alternative formulae. And to prove Theorem7.7, proving case $f : \mathbb{R} \to \mathbb{R}$ is enough. The alternative formulae is also important for generalisation to higher dimensions:

Theorem 7.8 (Lebesgue differentiation theorem(general))**.** *Given $f \in L^1_{loc}(\mathbb{R}^n)$ i.e. $f \in L^1(\Omega)$ for every bounded open $\Omega \subset \mathbb{R}^n$.*

$$f(x) = \lim_{r\to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y)\, dy$$

for λ-a.e. points $x \in \mathbb{R}^n$.

Lebesgue differentiation theorem is a convergence theorem of the form $T_h f(x)$ converges to given limit in some way (can be one of the modes of convergences we discussed in Chapter 6). We use density argument for this kind of theorem. Note T_h must be linear operator on f, and in the case of 1-D Lebesgue differentiation (alternative formulae), $T_h f(x) = \frac{1}{h} \int_{[x,x+h]} f(t)\, dt$.

Density argument:

1. Prove the result for a dense sub-class of functions. Dense means every function f can be approximated in some sense by functions in this subclass. For example, simple functions is a dense subclass of measurable functions. This technique is commonly applied in functional analysis.

2. Bound the fluctuation of $T_h f$ quantitatively using "size" of f, for example, L^p norm of f.

We focus on the 1-D case for now. The dense subclass is already hinted in Corollary 7.5. Define

$$C_C^0(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R} \, : \, f \text{ is continuous and is compactly supported}\}$$

compactly supported means $f(x) = 0$ outside a compact set. So Theorem 7.7 holds for $f \in C_C^0(\mathbb{R})$ because of Corollary 7.5. We will prove $C_C^0(\mathbb{R}^n)$ is dense in Ł$^1(\mathbb{R}^n)$ for any n later. (This is also a consequence of Littlewood's second principle, which says every measurable function agrees with a continuous function except for a set with arbitrarily small measure)

And the quantitative upper bound for fluctuation of T_h is given by Hardy-Littlewood maximal inequality(one-sided),

Definition 7.9 (Hardy-Littlewoood maximal function)**.** Given measurable function $f : \mathbb{R}^n \to \mathbb{R}$ or $\mathbb{C}$, the maximal function is defined as

$$f^*(x) := \sup_{r>0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f(y)| \, dy$$

this is the supremum of average of $|f|$ for all balls centred at x.

Remark. Now we quickly run through properties of maximal functions:
First if $f = c$ (constant), it is not hard to show that $f^*(x) = |c|$.
Given $f(x) = \frac{1}{x} 1_{(0,\infty)}$, the indicator guarantees f is measurable, then $f^*(x) = +\infty$. As for every $x \in \mathbb{R}$, the ball $B_{2|x|}(x)$ contains the point 0, so $\int_{B_{2|x|}(x)} |f(y)| \, dy = \infty$. This example illustrates that maximal function can take $+\infty$.

Note if general Lebesgue differentiation theorem is true, for a.e. $x \in \mathbb{R}^n$,

$$|f(x)| = \lim_{r\downarrow 0} \left| \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y) \, dy \right| \le \sup_{r>0} \left| \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y) \, dy \right|$$

$$\le \sup_{r>0} \left| \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y) \, dy \right| \le \sup_{r>0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f(y)| \, dy = f^*(x)$$

so it should not be surprising that maximal function takes ∞. But what is surprising is that even if f is integrable, f is not integrable unless $\|f\|_1 = 0$. Pick $r_0 > 0$ s.t. $\int_{B_{r_0}(0)} |f(y)| \, dy = c_1 > 0$, for all x with $|x| > r_0$,

$$f^*(x) \ge \frac{1}{\lambda(B_{|x|+r_0}(x))} \int_{B_{|x|+r_0}(x)} |f(y)| \, dy$$

note $\lambda(B_{|x|+r_0}(x)) = \frac{1}{(|x|+r_0)^n} \ge |x|^{-n}$ and using substitution, one can show that $\int_{B_{|x|+r_0}(x)} |f(y)| \, dy = c \int_{B_{r_0}(0)} |f(y)| \, dy \ge cc_1$ for some constant c, so $f^*(x) \ge c'|x|^{-n}$ for some constant c'. But $|x|^{-n}$ is not integrable on $\mathbb{R}^n$, so f^* is not integrable.

The maximal inequality relates quantitatively maximal function and $\|f\|_1$

Proposition 7.10 (Hardy-Littlewood maximal inequality(One-sided))**.** *Given $f : \mathbb{R} \to \mathbb{C}$ that is absolutely integrable, for all* $\mathbf{a} > 0$*,*

$$\lambda\left(\{x \in \mathbb{R} \,:\, f^*(x) \geq a\}\right) \leq \frac{1}{a}\int_{\mathbb{R}} |f(t)|\; dt = \frac{1}{a}\|f\|_1$$

we will leave the proof later and let's see how density argument can be used to prove 1-D Lebesgue differentiation theorem:

Proof. Given arbitrary $\epsilon, a > 0$, we can find function $g : \mathbb{R} \to \mathbb{C}$ that is continuous and compactly supported with

$$\int_{\mathbb{R}} |f(x) - g(x)|\; dx \leq \epsilon$$

by one-sided Hardy-Littlewood maximal inequality,

$$\lambda\left(\left\{x \in \mathbb{R} \,:\, \sup_{h>0} \frac{1}{h}\int_{[x,x+h]} |f(t) - g(t)|\; dt \geq a\right\}\right) \leq \frac{\epsilon}{a}$$

and by Markov's inequality, we also have

$$\lambda\left(\{x \in \mathbb{R} \,:\, |f(t) - g(t)| \geq a\}\right) \leq \frac{\epsilon}{a} \qquad (1)$$

if we collect the union of above two sets written inside $\lambda(\dot{})$, and call it E, then $\mu(E) \leq \frac{2\epsilon}{a}$ and we have for $x \in \mathbb{R} \setminus E$,

$$\int_{[x,x+h]} |f(t) - g(t)|\; dt < a, \quad |f(x) - g(x)| < a \qquad \forall h > 0$$

now apply Corollary7.5 to continuous function g, for small enough h, we have

$$\left|\frac{1}{h}\int_x^{x+h} g(t)\; dt - g(x)\right| < a$$

Then using two inequalities in (1), and triangular inequality,

$$\left|\frac{1}{h}\int_x^{x+h} f(t)\; dt - f(x)\right| < 3a$$

so

$$\limsup_{h\to 0}\left|\frac{1}{h}\int_x^{x+h} f(t)\; dt - f(x)\right| < 3a \qquad (2)$$

for $x \in E^C$ where $\mu(E) = \frac{2\epsilon}{a}$. Fix a and send ϵ to 0, we have (2) holds for a.e. $x \in \mathbb{R}$. And then since a is arbitrary, we send it to 0:

$$\limsup_{h\to 0}\left|\frac{1}{h}\int_x^{x+h} f(t)\; dt - f(x)\right| = 0$$

for a.e. on $\mathbb{R}$. which proves the theorem. ■

Now let's study the maximal inequality. One-sided Hardy-Littlewood maximal inequality is proved using the following lemma:

Lemma 7.11 (Rising sun lemma)**.** *Given $f : [a,b] \to \mathbb{R}$, one can find a finite/countable family of disjoint open intervals $I_n = (a_n, b_n)$ with $a_n < b_n$ and are all contained in $[a,b]$, the following properties hold for I_n:*

- *$f(a_n) = f(b_n)$ or if $a_n = a$, $f(b_n) \geq f(a_n)$*
- *for region in $[a,b]$ outside intervals I_n, the function is monotone decreasing.*

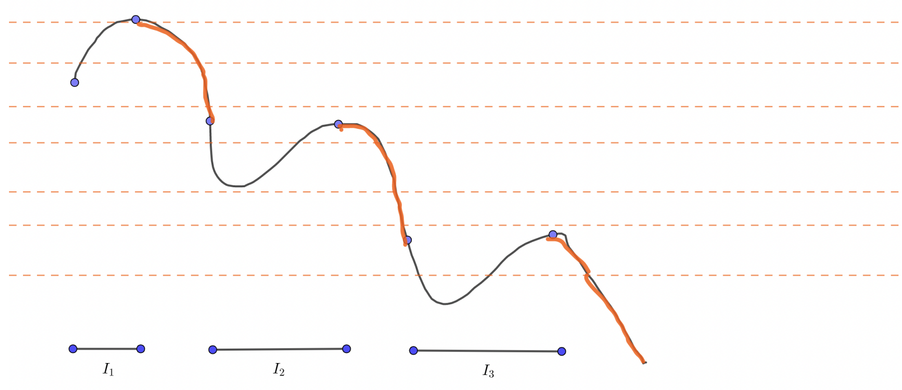

Figure 7: Interpretation of rising sun lemma

Remark. Imagine a continuous function like a hill, and the sun is rising from the east. $(+\infty)$ The intervals I_n are the shadows and any region outside I_n are shined by the sun. (f monotone decreasing)

Unfortunately the proof of this lemma depends heavily on nature of $\mathbb{R}$ that is not extended to $\mathbb{R}^n$, and one-sided Hardy-Littlewood maximal inequality does not hold for higher dimensions. So we will not investigate the proof here but rather work towards a more general inequality.

From now on, we aim to prove the general case. A stronger statement implies the equation given in general Lebesgue differentiation theorem,

$$\lim_{r\to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| \, dy = 0 \qquad (\dagger)$$

$$\Rightarrow f(x) = \lim_{r\to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y) \, dy$$

this can be easily seen by triangle inequality. And all points $x \in \mathbb{R}^n$ satisfying $(\dagger)$ are called Lebesgue points of f.

We prove the theorem with stronger condition $f \in L^1(\mathbb{R}^n)$. Since locally integrable ($f \in L^1_{\text{loc}}(\mathbb{R}^n)$) means

$$\int_{B_r(x)} |f(x)|\, dx < \infty \qquad \forall r > 0, x \in \mathbb{R}^n$$

$f 1_{B_r(x)} \in L^1(\mathbb{R}^n)$. So using the theorem with stronger condition, we have

$$\lim_{r\to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y) 1_{B_r(x)}(y)\, dy = f(x) 1_{B_r(x)}(x) = f(x) \text{ as } x \in B_r(x)$$

and this implies the required result directly.

Beginning with first step of density argument: find a dense subclass s.t. result hold

Lemma 7.12. *If $f \in C^0_C(\mathbb{R}^n)$, then for any $x \in \mathbb{R}^n$,*

$$f(x) = \lim_{r\to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y)\, dy$$

Proof. Easy proof left as exercise. Try to generalise what we argued in 1-D case. ■

Now it is time to prove the "dense" property

Lemma 7.13. *$\forall 1 \leq p \leq +\infty$, and $f \in L^p(\mathbb{R}^n)$, given $\epsilon > 0$, there is function $g \in C^0_C(\mathbb{R}^n)$ s.t. $\|f - g\|_p < \epsilon$*

Proof. **Step 1.** Restrict f to a bounded region s.t. distance(in terms of L^p induced metric) of restricted function to f is not too far.

$$\|f - f 1_{B_r(0)}\|_p \leq 2\|f\|_p < \infty$$

by Minkowski inequality and the fact that $f \in \mathcal{L}^p$. For $n \in \mathbb{N}$, $f 1_{B_r(0)} \to f$ so by dominated convergence theorem, there is $r \in \mathbb{N}$ s.t.

$$\|f - f 1_{B_r(0)}\|_p < \frac{\epsilon}{4}$$

we divide ϵ by 4 because we need 3 other approximations.

Step 2. Use simple function to approximate $f 1_{B_r(0)}$
Note f is measurable so $f 1_{B_r(0)}$ is measurable, and $\|f 1_{B_r(0)}\|_p \leq \|f\|_p < \infty$ so $f 1_{B_r(0)} \in \mathcal{L}^p$. Using Proposition 5.13, one can find simple function $\tilde{g} = \sum_i a_i 1_{E_i}$ where $\{E_i\}$ is a partition of X s.t.

$$\|f 1_{B_r(0)} - \tilde{g}\|_p < \frac{\epsilon}{4}$$

Step 3. Approximate $\tilde{g}$ by $\tilde{\tilde{g}}$ which consists of indicator functions defined on rectangles (in $\mathbb{R}^n$).
By regularity of Lebesgue measure, for each E_i, there is an open set $G_i \supseteq E_i$ s.t. $\lambda(G_i \setminus E_i)$ is arbitrarily small. (We may have to expand support $B_r(0)$

so that G_i are all included) Then we use the technique of lemma 1.17 of note 3.2 (approximate open set using countable half-open disjoint rectangles) say $(Q_{i,k})_{k\geq 1}$ estimates G_i with arbitrary small error. So if we can make sure $\lambda\left(E_i \setminus \left(\bigcup_{k\geq 1} Q_{i,k}\right)\right) < \frac{2^{-ip}}{|a_i|^p}(\frac{\epsilon}{4})^p$, then define

$$\tilde{\tilde{g}} := \sum_i a_i 1_{\bigcup_{k\geq 1} Q_{i,k}} = \sum_{i,k} a_i 1_{Q_{i,k}}$$

second inequality holds as $Q_{i,k}$ are all disjoint. Then we have

$$\|\tilde{g} - \tilde{\tilde{g}}\|_p \leq \sum_i \left\| a_i 1_{E_i} - a_i 1_{\bigcup_{k\geq 1} Q_{i,k}} \right\|_p$$

$$= \sum_i \left(\int |a_i|^p \left| 1_{E_i \setminus \bigcup_{k\geq 1} Q_{i,k}} \right|^p \, d\lambda \right)^{\frac{1}{p}}$$

$$= \sum_i \left(|a_i|^p \lambda(E_i \setminus \bigcup_{k\geq 1} Q_{i,k}) \right)^{\frac{1}{p}} < \sum_i \frac{\epsilon}{4}(2^{-ip})^{\frac{1}{p}} = \frac{\epsilon}{4} \sum_i 2^{-i} = \frac{\epsilon}{4}$$

Step 4. Finally we approximate $\tilde{\tilde{g}}$ using a continuous function g supported on $B_r(0)$ (may have to extend to $B_{2r}(0)$, but g is still compactly supported)
The basic idea is easy, suppose we have a step function $\tilde{\tilde{g}}$, we can construct a continuous function g as below Red horizontal lines are steps of $\tilde{\tilde{g}}$ and the

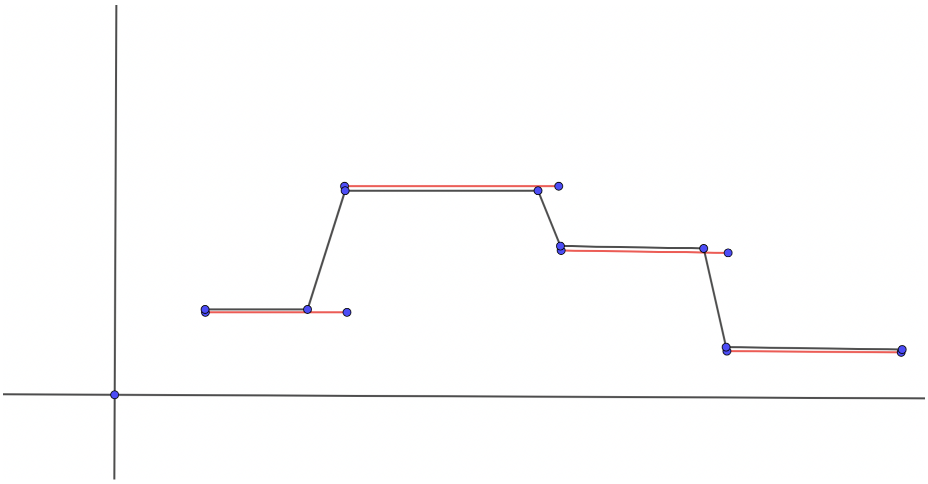

Figure 8: Continuous estimation to step function

black function is g. Note I draw a small gap between black and red line for demonstration but they are supposed to be overlapping. I will not write the algebraic calculations but as you can see if we make the oblique parts of g steep enough, we can make $\|\tilde{\tilde{g}} - g\|_p < \frac{\epsilon}{4}$.

Combining everything from step 1 to step 4, we have $\|f - g\|_p < \epsilon$ as required. ∎

Remark. Note I haven't prove the case $p = \infty$ (where the norm is not an integral), you can follow the same steps to handle this case.

Then we proceed to second step of density argument: construct a quantitative bound for fluctuation of $T_h f$.

Theorem 7.14 (Hardy-Littlewood maximal inequality)**.** *Let $f : \mathbb{R}^n \to \mathbb{C}$ be absolutely integrable, and given $a > 0$,*

$$\lambda\left(\{x \in \mathbb{R}^n \,:\, f^*(x) \geq a\}\right) \leq \frac{C_n}{a}\|f\|_1$$

for some constant C_n depending on n only.

This can be proved using Vitali's covering lemma. First, let's introduce a new notation. For a ball $B = B_r(x)$, we denote $cB := B_{cr}(x)$, the ball with the same centre but c times radius. Note $\lambda(cB) = c^n\lambda(B)$. However, cB is different from $\{cy \,:\, y \in B\}$.

Lemma 7.15 (Vitali Covering lemma(finite version))**.** *Given open balls $B_1, ..., B_k$ that may not be disjoint, then there is a sub-collection $B'_1, ..., B'_m$ (all disjoint) s.t.*

$$\bigcup_{i=1}^{k} B_i \subset \bigcup_{j=1}^{m} 3B'_j$$

then by sub-additivity,

$$\lambda\left(\bigcup_{i=1}^{k} B_i\right) \leq 3^n \sum_{j=1}^{m} \lambda(B'_j)$$

The following proof is not strict mathematically, but it captures why number 3 is used in the inequality and this idea of "covering".

Proof. Use greedy algorithm, selecting balls as large as possible while keeping them disjoint.

- Step 0. let $m = 0$
- Step 1. pick largest ball(the ball with largest radius) among B_i that does not intersect one of $B'_1, ..., B'_m$, and let this ball be B'_{m+1}. So in this step $m = 0$, we just let B'_1 be the largest ball.
- Step 2. Repeat step 1, and keep repeating until there is no more balls disjoint from $B'_1, ..., B'_m$. Or stop when $m = k$

Note this process must stop in finite time as there are only finite balls. Then let's assume the process stop at step m and argue the lemma is satisfied.
If $m = k$, result is trivial. Otherwise, if ball B_i is not picked, there must be ball B'_j picked s.t. $B_i \cap B'_j \neq \emptyset$. Pick the smallest j satisfying this, i.e. B_i is disjoint from $B'_1, ..., B'_{j-1}$. Then by choice of B'_j, it has largest radius among all the balls not picked at step j, including B_i. So we have

- $B_i \cap B'_j \neq \emptyset$
- $\lambda(B_i) \leq \lambda(B'_j)$

triangular inequality gives us $B_i \subseteq 3B'_j$, which proves the lemma. ∎

Remark. There is actually an infinite version of Vitali's covering lemma, mentioned in Lecture Notes 10.1. Since we do not have to use it to prove maximal inequality, I will state and prove it in the end of this section.

Now we can go back to prove Hardy-Littlewood maximal inequality

Proof. Looking at Vitali's covering lemma(finite case), $c_n = 3^n$ is a reasonable choice. We can prove the inequality when replacing $f^*(x) \geq a$ by $f^*(x) > a$, then using limits we can recover the original inequality.

Given any $\epsilon > 0$, using regularity of Lebesgue measure, we can pick compact set $K \subseteq A := \{x \in \mathbb{R}^n : f^*(x) > a\}$ s.t. $\lambda(A \setminus K) < \epsilon$. So given $x \in K$, by definition of A, there is open ball $B_r(x)$ s.t.

$$\frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f(y)|\, dy > a \qquad (*)$$

Consider the cover $\{B_r(x) : x \in \mathbb{R}\}$, since K is compact, we can find $B_1, ..., B_n$ s.t. $K \subseteq \bigcup_{i=1}^n B_i$, then by Vitali's covering lemma, we can find disjoint sub-cover s.t.

$$\lambda\left(\bigcup_{i=1}^k B_i\right) \leq 3^n \sum_{j=1}^m \lambda(B'_j)$$

Then using (*) we have for each j,

$$\lambda(B'_j) < \frac{1}{a} \int_{B'_j} |f(y)|\, dy$$

put sums in j to the above inequality, we get

$$3^n \sum_{j=1}^m \lambda(B'_j) \leq \frac{3^n}{a} \int_{\mathbb{R}^n} |f(y)|\, dy$$

so the required inequality follows. ∎

Then using similar arguments to 1D case, we can prove Lebesgue differentiation theorem using density argument. It is worth noting that there are other differentiation theorems like Rademacher differentiation theorem, but we will not study it here.

Remark. In fact, $c_n = 3^n$ is not the best bound. One can consider the distance between centres of balls in Vitali's covering lemma (finite version) instead, the centres are contained in $\bigcup_{j=1}^m 2B'_j$. So c_n can be improved to 2^n. The bound use in our lecture is 5^n. While improving bound does not really affect the proof of Lebesgue differentiation theorem, it was proved by Stein and Strömberg in 1983 that c_n can be chosen s.t. it grows linearly in d.

Now we prove the infinite version of Vitali's covering lemma. It is worth noting first an important definition and a lemma.

Definition 7.16 (Partial order)**.** A relation R defined on set $\mathcal{F}$ is called partial order if following hold

- For any $a \in \mathcal{F}$, aRa. (Self-reflexive)
- For any $a, b \in \mathcal{F}$, if aRb and bRa, then $a = b$. (Anti-symmetric)
- For $a, b, c \in \mathcal{F}$, if aRb, bRc, then aRc. (Transitive)

if we have additionally for any $a, b \in \mathcal{F}$, aRb, bRa or $a = b$. Then relation R is called total order.

Zorn's lemma is very important and it is equivalent to the axiom of choice.

Lemma 7.17 (Zorn's lemma)**.** *If $\mathcal{P}$ is a set with partial order $\leq$ defined and every chain of $\mathcal{P}$ is bounded above (in the sense of $\leq$), then there is a maximal element $s \in \mathcal{P}$. (i.e. $a \leq s\ \forall a \in \mathcal{P}$)*

Remark. In our intuition, a chain is something like $B_1 \subseteq B_2 \subseteq \cdots$. But in fact a chain can have uncountably infinite length. General definition of chain is a collection of sets $\mathcal{C}$ s.t. $\forall B_i, B_j \in \mathcal{C}$, $B_i \leq B_j$ or $B_j \leq B_i$.

Lemma 7.18 (Vitali's covering lemma(infinite version))**.** *Let $\mathcal{F}$ be a family of balls in $\mathbb{R}^n$ satisfying $d_0 := \sup_{B \in \mathcal{F}} diam(B) < \infty$. Where $diam(B) := \sup_{x,y \in B} \|x - y\|$. Then there is a countable family $\mathcal{G} \subset \mathcal{F}$ of pairwise disjoint balls s.t.*

$$\bigcup_{B \in \mathcal{F}} B \subset \bigcup_{B \in \mathcal{G}} 5B$$

Proof. We follow exactly the same guideline as in finite case. But dealing with infinite case requires picking balls with decreasing radius while keeping them disjoint to cover the whole set $\bigcup_{B \in \mathcal{F}} B$.

We define $\mathcal{P}$ as a collection of subsets $\mathcal{S}$ of $\mathcal{F}$ with order $\subseteq$, satisfying the assumption defined as below

$$\text{(H)}\quad \text{sets in } \mathcal{S} \text{ are disjointand if a ball } B \in \mathcal{F} \text{ meets some ball in } \mathcal{S},$$
$$\text{then } B \text{ meets another ball } B' \in \mathcal{S} \text{ s.t. } diam(B') \geq \frac{1}{2} diam(B)$$

First we prove $\mathcal{P} \neq \emptyset$. Pick any $B_0 \in \mathcal{F}$ with $diam(B_0) \geq \frac{1}{2} d_0$. Then we can ensure $\mathcal{S} := \{B_0\} \in \mathcal{P}$. And given chain $(\mathcal{C}_i)_{i \in I} \subset \mathcal{P}$, let $\omega := \bigcup_{i \in I} \mathcal{C}_i$. It is easy to check $\omega \in \mathcal{P}$ and clearly ω is an upper bound for the chain. So by Zorn's lemma, there is a maximal element in $\mathcal{P}$, we call it $\mathcal{G}$.

If there is a ball in $\mathcal{F}$ disjoint from $\bigcup \mathcal{G} := \bigcup_{B \in \mathcal{G}} B$. Pick $B_0 \in \mathcal{F}$ s.t. $diam(B_0) \geq \frac{1}{2} diam(B) \forall B \in \mathcal{F}$ not meeting $\bigcup \mathcal{G}$ (this is possible as $d_0 < \infty$), then it is easy to see $\mathcal{G}' := \mathcal{G} \cup \{B_0\}$ also satisfies (H), contradicting maximality of $\mathcal{G}$.

So all balls in $\mathcal{F}$ intersects a ball $B' \in \mathcal{G}$ with $diam(B) \leq 2 diam(B')$.(using the contradiction and property (H)) It is left as exercise to show Vitali's covering lemma follows. (Hint: use triangular inequality) ∎

7.2 Vitali Covering theorem (Extra)

Now we aim to prove the covering theorem, which covers $E \subseteq \mathbb{R}$ except for a null set using a countable sub-collection of a collection of intervals called Vitali covering. Two lemmas we proved before are just basic results.

Definition 7.19 (Vitali covering)**.** Given a set $E \subseteq \mathbb{R}$ (may not be measurable), a collection of intervals

$$\mathcal{V} = \{I_\alpha \,:\, \alpha \in J\}$$

where J is a finite, countable or uncountable index set, is called Vitali covering of E if for all $x \in E$, given arbitrary $\epsilon > 0$, exists $I_\alpha \in \mathcal{V}$ s.t. $x \in I_\alpha$ and $0 < |I_\alpha| < \epsilon$. $|I_\alpha|$ means length of the interval, we require it to be positive to avoid cheating by picking $I_\alpha = [x, x]$.

Remark. It is easy to see that if $\mathcal{V}$ is Vitali covering of E, then for any subset $E' \subseteq E$, $\mathcal{V}$ is also a Vitali covering of E'.

the following theorem is also called the Vitali covering theorem.

Theorem 7.20 (Vitali covering lemma (measure version))**.** *Given $E \subseteq \mathbb{R}$ and Vitali covering $\{I_\alpha \,:\, \alpha \in J\}$ of E. Though the lemma does not require I_α to be closed, we mainly use this lemma on closed covering so we will assume that I_α are closed. Exists disjoint $(I_j)_{j\in\mathbb{N}}$ for $I_j \in J$ s.t.*

$$\lambda^*\left(E \setminus \bigcup_{j\geq 1} I_j\right) = 0$$

where λ^* is the Lebesgue outer measure taken using Hahn-Caratheodory extension theorem. (E may not be measurable so we cannot use Lebesgue measure directly)

Remark. Recall that if a set is closed and bounded, then for any open covering of this set, there is a finite sub-cover. Vitali covering lemma is similar to this result. We do not require closed and boundedness anymore, but "finite" sub-cover is eased to "countable" sub-cover. The word sub-cover is meant to be cover almost everywhere.

Proof. **Step 1.** Reduce the case to $(k, k+1)$ $(k \in \mathbb{Z})$
Given $E \subseteq \mathbb{R}$, for all $k \in \mathbb{Z}$, let $E_k := E \cap (k, k+1)$. Suppose one can find $I_j^{(k)} \subseteq (k, k+1)$ s.t.

$$\lambda^*\left(E_k \setminus \bigcup_{j\geq 1} I_j^{(k)}\right) = 0$$

then clearly the family $\left\{I_j^{(k)} \,:\, j \in \mathbb{N}, k \in \mathbb{Z}\right\}$ is disjoint and consider

$$E \setminus \bigcup_{k\in\mathbb{Z}} \bigcup_{j\geq 1} I_j^{(k)} \subset \mathbb{Z} \cup \bigcup_{k\in\mathbb{Z}} \left[E_k \setminus \bigcup_{j\geq 1} I_j^{(k)}\right]$$

as any integer is E is not included in any E_k, we need $\mathbb{Z}$ on RHS. Outer measure if RHS is 0 as $\lambda^*(\mathbb{Z}) = 0$ and $\lambda^*\left(E_k \setminus \bigcup_{j\geq 1} I_j^{(k)}\right) = 0$. So outer measure of

LHS is 0 by monotonicity.

Now aims to reduce $\mathcal{V}$ to $(k, k+1)$. WOLG assume $k = 0$, for $E' = E \cap (0,1)$. Since $E' \subseteq E$, $\{I_\alpha : \alpha \in J\}$ is a Vitali covering for E'. Now define

$$\overline{\mathcal{V}} := \{I_\alpha : \alpha \in J, I_\alpha \subseteq (0,1)\}$$

$\overline{\mathcal{V}}$ is also Vitali covering for E'. Since given $x \in E, \epsilon > 0$, pick $\delta := \min\{\epsilon, x, 1-x\}$. $\mathcal{V}$ is Vitali covering of E', so exists $I_\alpha \in \mathcal{V}$ s.t. $x \in I_\alpha$ and $0 < |I_\alpha| < \delta$. So $I_\alpha \subseteq (0,1)$.

Step 2. Prove the reduced case.
Now assume $E \subseteq (0,1), \mathcal{V} := \{I_\alpha : \alpha \in J\}$ where $I_\alpha \in (0,1)$ and are closed. $\mathcal{V}$ is a Vitali covering of E.

If E is empty, any family (I_j) works. Otherwise, $E \neq \emptyset$, define $s_1 := \sup\{|I_\alpha| : \alpha \in J\}$. Since there is $x \in E$, we can find $I_\alpha \in \mathcal{V}$ s.t. $|I_\alpha| > 0$ so $s_1 > 0$. Also $s_1 \leq 1$ as $I_\alpha \in (0,1)$.
Find set $I_1 \in \mathcal{V}$, s.t. $|I_1| \geq \frac{s_1}{2} > 0$.

Now we can repeat the process. So for $2 \leq n \in \mathbb{N}$, assume we have already defined $I_1, ..., I_{n-1}$. If $E \setminus \bigcup_{j=1}^{n-1} I_j = \emptyset$, I_n can be chosen to be any set. Otherwise, if exists $x \in E \setminus \bigcup_{j=1}^{n-1} I_j$, then we define

$$s_n := \sup \left\{ |I_\alpha| : I_\alpha \in \mathcal{V}, I_\alpha \cap \bigcup_{j=1}^{n-1} I_j = \emptyset \right\} \leq 1$$

by similar arguments as before, we can claim $s_n > 0$. So exists I_n disjoint from $\bigcup_{j=1}^{n-1} I_j$ s.t. $|I_n| \geq \frac{s_n}{2}$.

This process either ends when some $E \setminus \bigcup_{j=1}^{n-1} I_j = \emptyset$, or it goes on forever. By construction, I_n are disjoint and $|I_n| \geq \frac{s_n}{2}$. Note since $\bigcup_{j \geq 1} I_j \subseteq (0,1)$,

$$1 \geq \lambda^* \left(\bigcup_{j \geq 1} I_j \right) = \sum_{j \geq 1} \lambda^*(I_j)$$

so the sum converges, i.e.

$$\sum_{j \geq n} |I_j| \to 0 \text{ as } n \to \infty$$

please remember this important result as we will use later.
If $I_n = [a_n, b_n]$, then define $K_n := [a_n - 2|I_n|, b_n + 2|I_n|]$.
Claim:

$$E \setminus \bigcup_{j=1}^{N} I_j \subseteq \bigcup_{j > N} K_j$$

pick $x \in E \setminus \bigcup_{j=1}^{N} I_j$, since I_α are all closed, there is a distance between x and $\bigcup_{j=1}^{N} I_j$. $\mathcal{V}$ is Vitali covering, so for $J \in \mathcal{V}$ containing x, we can make its length arbitrarily small(for this case, make it smaller than the distance between x and closed intervals $I_j, j = 1, 2, ..., N$), so we can pick J s.t. $J \cap \bigcup_{j=1}^{N} I_j = \emptyset$ and $|J| > 0$. That means $|J| \leq s_{N+1}$ by definition of s_{N+1} as a supremum.

Assume $J \cap I_m = \emptyset$ for all m with $N < m \leq N'$, using similar argument we also have $|J| \leq s_{N'+1} \leq 2|I_{N'+1}|$. Second inequality is due to construction of I_n. $\sum_{j \geq n} |I_j| \to 0$, so $|I_n| \to 0$ as $n \to \infty$. But $|J| > 0$, so assumption

$$J \cap I_m = \emptyset \text{ for all } m \text{ with } N < m \leq N'$$

cannot hold for all $N' > N$. i.e. there is $M > N$ s.t. $J \cap I_M \neq \emptyset$. (There may be multiple M's satisfying this requirement, we pick the smallest M) So $J \cap \bigcup_{j=1}^{M-1} I_j = \emptyset$, that means $|J| \leq s_M \leq 2|I_M|$.
It is easy to see that then $J \subseteq K_M$ using the following facts

- $I_M \cap J \neq \emptyset$
- $|J| \leq 2|I_M|$.
- K_M is formed by extending I_M at two ends with length $2|I_M|$.

draw a graph to check this. So this means $x \in K_M$ for some $M > N$ and that proves the claim

$\square_{\text{claim}}$.

Using the claim,

$$\lambda^* \left(E \setminus \bigcup_{j=1}^{N} I_j \right) \leq \lambda^* \left(\bigcup_{j>N} K_j \right) \leq \sum_{j>N} \lambda^*(K_j) = \sum_{j>N} 5|I_j| \xrightarrow{N\to\infty} 0$$

second inequality follows from sub-additivity of outer measure. So we have

$$\lambda^* \left(E \setminus \bigcup_{j=1}^{\infty} I_j \right) = 0$$

■

Remark. Now we show that assumption I_α are all closed can be removed. Assume $\mathcal{V} := \{I_\alpha : \alpha \in J\}$ is a collection of intervals (not necessarily closed) that is a Vitali covering of $E \subseteq \mathbb{R}$. Define $J_\alpha = \overline{I_\alpha}$. And define the new family $\overline{\mathcal{V}} := \{J_\alpha : \alpha \in J\}$, since $|I_\alpha| = |J_\alpha|$, it is not hard to see $\overline{\mathcal{V}}$ is also a Vitali covering of E. By lemma prove above, there is a countable sub-cover, i.e. exists $(J_k)_{j\geq 1}$ s.t. $\lambda^* \left(E \setminus \bigcup_{k\geq 1} J_k \right) = 0$. Assume $J_k = [a_k, b_k]$, then

$$E \setminus \bigcup_{k\geq 1} I_k \subseteq E \setminus \bigcup_{k\geq 1} J_k \cup \{a_k, b_k : k \in \mathbb{N}\}$$

outer measure of RHS is 0, so that of LHS is 0.

Corollary 7.21. *Assume E is bounded, i.e. $E \subseteq (-L, L)$ for some $L \in \mathbb{R}$ and given Vitali covering of E: $\{I_\alpha : \alpha \in J\}$. Given $\epsilon > 0$, there is a finite family $(I_k)_{k=1,\ldots,N}$ s.t.*

$$\lambda^*\left(E \setminus \bigcup_{k=1}^{N} I_k\right) \leq \epsilon$$

Proof. We proved that there is a countable family $(I_k)_{k\geq 1}$ s.t.

$$\lambda^*\left(E \setminus \bigcup_{k\geq 1} I_k\right) = 0$$

we showed in the proof that any I_α not contained in $(-L, L)$ can be removed and Vitali covering property is preserved.

$$E \setminus \bigcup_{k=1}^{N} I_k \subseteq \left(E \setminus \bigcup_{k=1}^{\infty} I_k\right) \cup \left(\bigcup_{k>N} I_k\right)$$

outer measure of first term of RHS is 0, so

$$\lambda^*\left(E \setminus \bigcup_{k=1}^{N} I_k\right) \leq \sum_{k>N} \lambda^*(I_k) = \sum_{k>N} |I_k| \xrightarrow{N\to\infty} 0$$

as proved in the proof of Vitali covering lemma. That proves the corollary. ■

7.3 Differentiation for monotone functions(Extra)

Definition 7.22 (limsup and liminf for series)**.** For sequences x_n, limsup and liminf are defined as below

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} (\inf_{m\geq n} x_m) = \sup_{n\geq 0} \inf_{m\geq n} x_m$$

the second equality is because $\inf_{m\geq n} x_m$ is an increasing sequence.

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} (\sup_{m\geq n} x_m) = \inf_{n\geq 0} \sup_{m\geq n} x_m$$

Remark. If $a = \liminf_{n\to\infty} x_n$, then given $\epsilon > 0$, we can find $N \in \mathbb{N}$ s.t. $x_n > a - \epsilon$ if $n \geq N$. So only finite terms of sequence are less than $a - \epsilon$. So any value below a will become lower bound of sequence if we cut off enough (but finite) terms at the head. Similar interpretation can be made for $\limsup$.

Proposition 7.23. *In general*

$$\inf_{n\geq 1} x_n \leq \liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n \leq \sup_{n\geq 1} x_n$$

The idea of liminf and limsup can be extended to functions:

Definition 7.24 (limsup and liminf for functions)**.**

$$\limsup_{x\to a} f(x) := \lim_{\epsilon\to 0} (\sup\{f(x) : x \in E \cap B_\epsilon(a) \setminus \{a\}\})$$

$$\liminf_{x\to a} f(x) := \lim_{\epsilon\to 0} (\inf\{f(x) : x \in E \cap B_\epsilon(a) \setminus \{a\}\})$$

Again, any value below $\liminf_{x\to a} f(x)$ will become lower bound of $f(x)$ on a small enough disc around a. (Not including a) and any value above $\limsup_{x\to a} f(x)$ will become an upper bound become lower bound of $f(x)$ on a small enough disc around a. (Not including a)

If domain of f is in $\mathbb{R}$, This definition can be extended to $x \downarrow a$ and $x \uparrow a$, simply by restricting the open balls $B_\epsilon(a)$ to one side of a.

Definition 7.25 (Derivatives)**.** There are four types of derivatives for function $f : X \to \mathbb{R}$ where $X \subseteq \mathbb{R}$

$$D^+ f(x) := \limsup_{h\downarrow 0} \frac{f(x+h) - f(x)}{h}$$

$$D^- f(x) := \liminf_{h\downarrow 0} \frac{f(x+h) - f(x)}{h}$$

$$D_+ f(x) := \limsup_{h\downarrow 0} \frac{f(x) - f(x-h)}{h}$$

$$D_- f(x) := \liminf_{h\downarrow 0} \frac{f(x) - f(x-h)}{h}$$

f is called differentiable at x if $D^+ f(x) = D^- f(x) = D_+ f(x) = D_- f(x)$. And we define $f'(x) := D^+ f(x)$ if f is differentiable at x.

We now define the set of differentiable points:

$$E_f := \{x \in \mathbb{R} \, : \, f \text{ is differentiable at } x\}$$

if the function f is unambiguous, we abbreviate E_f as E.

Theorem 7.26 (Differentiability of monotone functions)**.** *Given monotone increasing $f : [a, b] \to \mathbb{R}$, then*

$$\lambda^*(E_f^C) = 0$$

Same applies to monotone decreasing functions, but for this prove we assume f is monotone increasing.

Proof. E^C can be decomposed into union of $24C_2$ sets of the form

$$\{x \in (a, b) \, : \, D_1 f(x) < D_2 f(x)\}$$

where D_1, D_2 are two different derivatives chosen from $D^\pm, D_\pm$. It is enough to show that outer measure of each set is 0. We only consider $E_1 := \{x \in (a, b) \, : \, D_- f(x) < D^+ f(x)\}$ and other sets and be dealt in similar ways. Note

$$E_1 = \bigcup_{s<t, s,t\in\mathbb{Q}} \underbrace{\{x \in (a, b) \, : \, D_- f(x) < s < t < D^+ f(x)\}}_{=:E_{s,t}}$$

it is enough to show $\lambda^*(E_{s,t}) = 0$ for every pair of s, t. We abbreviate this as λ.

Since Lebesgue measure is regular and $\lambda \leq \lambda^*((a, b)) = b - a < \infty$, fix $\epsilon > 0$, ,

there is an open set G s.t. $E_{s,t} \subseteq G$ and $\lambda^*(G) \leq \lambda + \epsilon$.
Given $x \in E_{s,t}$, since $\liminf_{h\downarrow 0} \frac{f(x)-f(x-h)}{h} < s$, we can find $h_k \downarrow 0$ s.t.

$$f(x) - f(x - h_k) \leq sh_k$$

so we will use $[x - h_k, x]$ to build Vitali covering for $E_{s,t}$, but we only want to keep those contained in G for convenience. (This is possible as G is open, and $x \in E_{s,t} \subseteq G$) So we define

$$\mathcal{V} := \left\{ [x - h_k, x] \, : \, \begin{cases} x \in E_{s,t} \\ f(x) - f(x - h_k) \leq sh_k \\ [x - h_k, x] \subseteq G \end{cases} \right\}$$

this is Vitali covering for $E_{s,t}$. Since given $x \in E_{s,t}, \delta > 0$, there is an interval $[z - h_k, z] \in \mathcal{V}$ s.t. $0 < |h_k| < \delta$ as $h_k \downarrow 0$. Clearly $z \in [z - h_k, z]$. So using the corollary of Vitali covering lemma, we can find disjoint $I_1, ..., I_M \in \mathcal{V}$ s.t.

$$\lambda^*(E_{s,t} \setminus \bigcup_{j=1}^{M} I_j) \leq \epsilon$$

By definition of $\mathcal{V}$, $I_j = [x_j - h_j, x_j]$ for some x_j, h_j and it satisfies two properties defined in $\mathcal{V}$. For convenience we re-order I_j s.t. $x_{j-1} \leq x_j - h_j$ i.e. put these disjoint intervals in order, I_1 to the left, I_M to the right.

$$\sum_{j=1}^{M} f(x_j) - f(x_j - h_j) \leq s \sum_{j=1}^{M} |I_j| = s\lambda^* \left(\bigcup_{j=1}^{M} I_j \right)$$

since $I_j \subseteq G$ for all j, $\bigcup_{j=1}^{M} I_j \subseteq G$, and $\lambda^*(G) \leq \lambda + \epsilon$ by choice of G, so

$$\sum_{j=1}^{M} f(x_j) - f(x_j - h_j) \leq s(\lambda + \epsilon) \qquad (\dagger)$$

Now we have to deal with the other inequality $t < D^+ f(x)$. Define $I_j^0 = (x_j - h_j, x_j)$ i.e. interior of I_j, $B := E_{s,t} \cap \bigcup_{j=1}^{M} I_j^0$. So

$$E_{s,t} = B \cup \left(E_{s,t} \setminus \bigcup_{j=1}^{M} I_j^0 \right)$$

and since

$$E_{s,t} \setminus \bigcup_{j=1}^{M} I_j^0 = \left(E_{s,t} \setminus \bigcup_{j=1}^{M} I_j \right) \cup \{x_j - h_j, x_j \, : \, j = 1, ..., M\}$$

and by construction of I_j, outer measure of first term is less than ϵ, and outer measure of second term is 0, so

$$\lambda = \lambda^*(E_{s,t}) \leq \lambda^*(B) + \epsilon$$

that means

$$\lambda^*(B) \geq \lambda - \epsilon \qquad (*)$$

Now we can attempt to build Vitali covering for $B = E_{s,t} \cap H$ where $H := \bigcup_{j=1}^{M} I_j$. This process is very similar to the process of building first $\mathcal{V}$. For $y \in B$ we can find $r_k \downarrow 0$ s.t.

$$f(y + r_k) - f(y) \geq tr_k$$

We define

$$\mathcal{V} := \left\{ [y, y + r_k] \, : \, \begin{cases} y \in B \\ f(y + r_k) - f(y) \geq tr_k \\ [y, y + r_k] \subseteq H \end{cases} \right\}$$

Please do not confuse this with the first Vitali covering we defined for $E_{s,t}$. And using exactly the same arguments as before, $\mathcal{V}$ is Vitali covering for B. So by Vitali covering lemma, we can find disjoint $J_1, ..., J_N$ s.t.

$$\lambda^*(B \setminus \bigcup_{k=1}^{N} J_k) \leq \epsilon$$

Claim.

$$\sum_{k \geq 1}^{N} |J_k| \geq (\lambda - 2\epsilon) \qquad (**)$$

$$B = B \cap \left(\bigcup_{k=1}^{N} J_k \right) \cup \left(B \setminus \bigcup_{k=1}^{N} J_k \right)$$

so

$$\lambda^*(B) \leq \lambda^* \left(B \cap \left(\bigcup_{k=1}^{N} J_k \right) \right) + \epsilon$$

because of the construction of J_k, and since J_k are disjoint intervals

$$\lambda^*(B) \leq \sum_{k=1}^{N} |J_k| + \epsilon$$

combining this with (*), we proved the claim. $\square_{\text{claim}}$

Call $J_k = [y_k, y_k + r_k]$. Again we order these intervals s.t. J_1 is at the left of all intervals and J_N is to the right of all intervals.

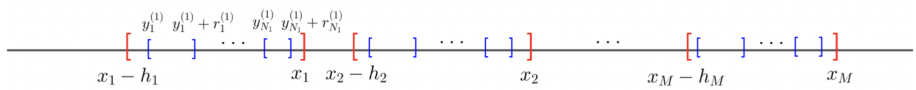

Figure 9: Intervals I_j and J_k

Figure9 shows the distribution of intervals(red for I_j, blue for J_k). Note $J_k \subseteq I_j^0$ for some j by construction. The intervals J_k has be re-numbered, so $[y_j^{(1)}, y_j^{(1)} + r_j^{(1)}], \cdots, [y_{N_j}^{(1)}, y_{N_j}^{(1)} + r_{N_j}^{(1)}]$ are intervals J_k that are inside I_j^0. Note since f is monotone, the increment of f on red intervals is more than total increment of f on blue intervals, i.e.

$$\sum_{j=1}^{M} f(x_j) - f(x_j - h_j) \geq \sum_{j=1}^{M} \sum_{i=1}^{N_j} \left[f(y_j^{(i)} + r_j^{(i)}) - f(y_j^{(i)}) \right]$$

but RHS is just an enumeration for blue intervals, so

$$\sum_{j=1}^{M} f(x_j) - f(x_j - h_j) \geq \sum_{k=1}^{N} f(y_k + r_k) - f(y_k) \geq \sum_{k=1}^{N} t r_k$$

second inequality follows from definition of $\mathcal{V}$. And note $r_k = |J_k| = y_k + r_k - y_k$, so

$$\sum_{j=1}^{M} f(x_j) - f(x_j - h_j) \geq \sum_{k=1}^{N} t|J_k| \geq t(\lambda - 2\epsilon) \qquad \text{by our claim}$$

combining the above inequality with (†), we have

$$t(\lambda - 2\epsilon) \leq s(\lambda + \epsilon)$$

this is true for arbitrary $\epsilon > 0$, so $t\lambda \leq s\lambda$. But $t > s$, so this is only possible when $\lambda = 0$. ■

Lemma 7.27. *Following the theorem above, given $f : [a, b] \to \mathbb{R}$ that is monotone increasing, it is differentiable a.e. with $f'(x)$ being derivative. Then f' is Lebesgue measurable, non-negative and*

$$\int_a^b f' \, d\lambda \leq f(b) - f(a)$$

Remark.

$$\int_a^b f' \, d\lambda := \int_{[a,b]} f' \, d\lambda$$

The inequality is necessary as we will see in the example: given c between a, b.

$$f = \begin{cases} 0 & \text{if } a \leq x < c \\ 1 & \text{if } c \leq x \leq b \end{cases}$$

the function is differentiable except at c, and it is constant a.e. so $\int_a^b f' \, d\lambda = 0$, but $f(b) - f(a) = 1$.

Proof. Define

$$f_n(x) := \frac{f(x + \frac{1}{n}) - f(x)}{\frac{1}{n}}$$

$f_n \to f'$ a.e. as f is differentiable a.e., so f' is Lebesgue measurable as it is a limit(almost everywhere) of measurable function. This holds because Lebesgue measure is complete, otherwise we require $f_n \to f'$ point-wise instead of a.e.

Now we prove the integral equation. First extend $f(x)$ s.t. $f(x) = f(b)$ if $x \geq b$ for f_n to be defined on $[a, b]$. By Fatou's lemma

$$\int_a^b f' \, d\lambda \leq \liminf \int_a^b f_n \, d\lambda$$

$$= \liminf n \int_a^b f(x + \frac{1}{n}) - f(x) \, d\lambda$$

using substitution,

$$= \liminf n \left[\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} f \, d\lambda - \int_a^b f \, d\lambda \right] = \liminf n \left[\int_b^{b+\frac{1}{n}} f \, d\lambda - \int_a^{a+\frac{1}{n}} f \, d\lambda \right]$$

as removing $[a, b]$ from $[a + \frac{1}{n}, b + \frac{1}{n}]$ gives $[b, b + \frac{1}{n}]$ and $[a + \frac{1}{n}, a]$(opposite direction).

$$\leq f(b) - f(a)$$

as f is monotone increasing, so on $[a, a + \frac{1}{n}]$, $f(x) \geq f(a)$. That gives $f(a)$ for the last line. And since f is constant by our extended definition on $[b, b + \frac{1}{n}]$, $f(b)$ is given for the last line. ∎

8 Problem Solving Strategies

Here I will give some problem solving techniques for real analysis/measure theory problems. Some techniques have been used for multiple times in this note and hopefully you have spotted them.

Splitting equality

Proving $a = b$ can be split into two parts: $a \leq b, b \leq a$ if such partial order is defined. For example, you can use this technique for proving additivity of measures. This applies to proving two sets are equal. Though for numbers a, b, another way of proving $a = b$ is $a - b = 0$.

Creating epsilon of room

Showing $a \leq b$ can be done by proving $a \leq b + \epsilon$ for any $\epsilon > 0$. Combining with splitting equality technique, you can accomplish great proofs.
There are some variants of this strategy

- Showing $a = 0$ is equivalent to showing $|a| \leq \epsilon$ for every $\epsilon > 0$. This can be used to show measure of a set is 0.
- Proving $f = g$ a.e. is equivalent to showing $|f(x) - g(x)| \leq \epsilon$ a.e. for all $\epsilon > 0$. Or even weaker, $|f(x) - g(x)| \leq \epsilon$ holds just outside a small set of measure $\leq \epsilon$. This can be used when studying the various modes of convergence.
- showing $x_n \to 0$ is equivalent to showing $\limsup_{n\to\infty} |x_n| \leq \epsilon$. This was used to prove Lebesgue differentiation theorem.

Of course, choosing such ϵ of room is very flexible, as long as the error bound can be controlled by making parameters arbitrarily small, it will work. For example 5ϵ works, or even with two parameters: $\frac{\epsilon}{\delta} + 3\delta$. This can be bounded by choosing small enough δ, then choose $\epsilon = \epsilon(\delta)$ small enough. And most of the time one do not have to crack an expression for upper bound in the beginning, but deduce it along the way.

Warning. But using epsilon of room requires $a, b \neq \pm\infty$, please carefully check this. For example signed measure can take both values $\pm\infty$.

Decomposing a theorem to simpler case

- If something has to be proved for unbounded sets/infinite measure, one can first prove it for bounded sets/finite measure, and then remove the finiteness assumption. We used this several times in the last few chapters.
- Prove something for measurable sets can be achieved by proving for open/closed/compact/bounded/basic sets first. Where basic sets means cubes or balls.
- Prove something for measurable function, first try to prove it for continuous/bounded/compactly supported/simple/integrable functions. This is frequently used in integration, convergence and differentiation.

- Prove something for infinite sum, first try to prove it for finite sum. But for any bounds or key values, try to keep it independent of number of terms in finite sum.
- Prove something for complex-valued function can be done by proving that for real-valued function first. And one can further reduce the case to positive-valued functions.
- The order of considering functions is: indicator, simple functions, positive functions, general functions.

And to go back to general case, one may need to decompose or approximate using objects in the reduced case. (For example, approximating positive measurable functions by simple functions) And please first try to prove theorem without any assumption, and if you feel stuck and really need that additional assumption to proceed, try this decomposition technique.

Flipping upper bound to lower bound

If one need to show lower bound, but only know how to easily obtain an upper bound, one can seek for "reflection". For example, use the set $X \setminus E$ instead of E, or $-f$ instead of f, or $F - f$ where F is called a dominating function. One can use this technique to prove monotone convergence theorem from the reduced version with assumption monotone increasing.

Dealing with Uncountable

Uncountable unions gives unsatisfactory properties, for example, measure of uncountable union of null sets may not be 0. But we can usually break uncountable into countable. For example, statement with $\epsilon > 0$ (uncountably many ϵ) can be replaced by countable sequence $\frac{1}{n}, n = 1, 2, ...$ Other sequences converging to 0 also work.

Similarly one can reduce statements for all cubes (uncountably many) to just dyadic cubes w.r.t to a countable partition of the space. This can be all cubes of the form $[c, c+1]$ where $c \in \mathbb{Z}^n$.

And with the aid of compact set, we can even reduce uncountable unions by finite ones. This is used when prove the Hardy-Littlewood maximal inequality. Similarly, Vitali's covering lemma replace uncountable unions by countable ones.

Global to local

We have already seen finite measures are easy to deal with. So when studying $\mathbb{R}^n$, one can just work on a large ball $B_R(0)$, or even just small ball $B_\epsilon(x)$ first. Compactness can be used to patch local properties together.

Throw away small exceptional set

The core concept of measure theory is to throw away small sets of measure 0 that are irrelevant to our statement. So one can throw away bad points like singularities, infinite, sequence not converging. Sometimes, throwing a set of measure small enough e.g. $\lambda(E) < \epsilon$ is also necessary. Egorov and Lusin's theorem are using this spirit.

Drawing pictures

Pictures are very important especially for Euclidean spaces. But please do not draw a picture satisfying conclusion. Try to make conclusion fail, though it could be right. You would see why you cannot make conclusion fail when drawing pictures in this way. And hopefully while reading this note, you are also drawing many pictures to understand concepts.

Zeno's trick

Epsilon can be split using geometric series

$$\epsilon = \sum_{i=1}^{\infty} \frac{\epsilon}{2^i}$$

this facts can be used to deal with countable additivity and subadditivity.

understanding control type

If one can control a function at every point/ or at almost every point (but not uniformly), this is called pointwise control. We also have uniform control and integrated control. It is worth noting that there could be implications between these types of controls, for example, $f \leq g$ a.e. implies $\int f \, d\mu \leq \int g \, d\mu$. Egorov's theorem is important for converting point-wise convergence to local uniform convergence after throwing away a null set, and Markov's inequality allow converting integral bounds to pointwise bounds after throwing away a null set.

Importance of subsequence

Passing properties to subsequence gives you better convergence.

- If $x_n \to x$ slowly, then a subsequence can make it faster. For example, pick n_j s.t. $d(x_{n_j}, x) \leq 2^{-j}$, then one can have exponential speed of convergence for the subsequence chosen.
- Any sequence of functions converging in $\mathcal{L}^1$ or in measure have a subsequence converging pointwise a.e.
- Sequence in compact space always have a subsequence converging.
- (Pigeonhole) A sequence taking only finitely many values has a constant subsequence. And in general, sequence living in union of finitely many sets $(E_i)_{1 \leq i \leq}$ has a subsequence that lives in only one E_i.

Two techniques recover convergence for original sequence from convergence of subsequence

- For metric spaces, if x_n is Cauchy, and a subsequence converges to x, then x_n also converge to x.
- (Urysohn subsequence principle) In topological space, if every sub-sequence of x_n has a sub-sub-sequence converging to x, then $x_n \to x$.

Limit viewed as limsup and liminf meeting

Limit of a sequence may not exist, but limit superior and limit inferior of a sequence always exists, they are defined as below

$$\limsup_{n\to\infty} x_n := \inf_{N\geq 1} \sup_{n\geq N} x_n$$

$$\liminf_{n\to\infty} x_n := \sup_{N\geq 1} \inf_{n\geq N} x_n$$

they are already explained in section 7.3 Differentiation for monotone functions.

They are convenient as for example, when showing x_n converges, it is sufficient with

$$\limsup_{n\to\infty} x_n \leq \liminf_{n\to\infty} x_n + \epsilon$$

for all $\epsilon > 0$. Or if one wants to show $x_n \to 0$, just show

$$\limsup_{n\to\infty} |x_n| \leq \epsilon$$

for all $\epsilon > 0$.
Warning. $\limsup$ is subadditive but may not be additive. $\liminf$ is superadditive but may not be additive. i.e. $\liminf_{n\to\infty}(x_n + y_n) \geq \liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n$.

8.1 Measure theory and Probability

Measure theory has a close connection with probability theory. First, we will illustrate one of the connections by defining a measure using a probability density function

$$\mu_h(A) := \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-h)^2}{2}}$$

where the PDF is for distribution $\mathrm{N}(h, 1)$. CDF of this distribution is restriction of μ to sets of the form $(-\infty, x]$, so μ is like generalisation of CDF.

Definition 8.1 (Probability space)**.** A probability space is a measure space $(\Omega, \mathcal{F}, P)$ where $P(\Omega) = 1$. Then P is usually called a probability measure.

Remark. Though the mathematical meaning of the underlying set Ω, σ-algebra $\mathcal{F}$ and measure P are the same as $X, \mathcal{F}, \mu$. But they have different interpretations. Recall that Ω is the sample space, $\mathcal{F}$ is the event space and $P(A)$ is probability of event A occurring. Note under this definition, Kolmogorov axioms are automatically satisfied.

Definition 8.2 (Probability distribution)**.** If Ω is a non-empty set (can be infinite) with σ-algebra 2^Ω and $(p_\omega)_{\omega\in\Omega}$ is a set of real numbers in $[0,1]$ s.t. $\sum_{\omega\in\Omega} p_\omega = 1$. Then

$$P(E) := \sum_{\omega\in E} p_\omega$$

is a probability measure and the function $\omega \mapsto p_\omega$ is called discrete probability distribution.

Similarly we can define continuous probability distribution If $\int_\Omega f(x)\, dx = 1$, the following indeed defines a probability measure

$$P(E) := \int_E f\, d\mu$$

and f is called probability density.

Proposition 8.3. *If P is the continuous probability distribution defined as*

$$P(E) := \int_E f\, d\mu$$

then

$$\int_A dP = \int_A f\, dx, \qquad \int_A g\, dP = \int_A fg\, dx$$

the first integral is defined as expectation if we replace A by Ω.

There are other relations to probability theory, for example, almost everywhere is exactly the concept of almost surely in probability. And measurable functions are now called random variables.

End of note, thank you for reading. For any typo or questions, please contact me via Teams, or WhatsApp with +86 130 0362 5205 or via WeChat using ID: DanielLin_kansaki.