Notes on Measure Theory

Tianyu Du

June 25, 2020

Contents

1 Sigma Algebra

Definition 1.1. For a set X, a set $\mathcal{A} \subseteq \mathcal{P}(X)$ is a σ**-algebra** if it satisfies the following properties:

1. $\varnothing, X \in \mathcal{A}$;
2. for all $A \in \mathcal{A}$, $A^c \in \mathcal{A}$ as well;
3. for a sequence in $\mathcal{A}$, $\{A_i\}_{i \in \mathbb{N}}$, the union $\bigcup_{i \in \mathbb{N}} \in \mathcal{A}$ as well.

An element $A \in \mathcal{A}$ is called a $\mathcal{A}$**-measurable set**.

Remark 1.1. It's easy to show that the largest σ-algebra of set X is the power set $\mathcal{P}(X)$, and the smallest σ-algebra is $\{\varnothing, X\}$.

Theorem 1.1. Let $\{\mathcal{A}_i\}_{i \in I}$ be the collection of all σ-algebra on X. Then, $\bigcap_{i \in I} \mathcal{A}_i$ is also a σ-algebra on X.

Proof. Clearly, $\varnothing, X \in \bigcap_{i \in I} \mathcal{A}_i$ given that every $\mathcal{A}_i$ is a σ-algebra.

For $A \in \bigcap_{i \in I} \mathcal{A}_i$, $A \in \mathcal{A}_i$ for all $i \in I$. Hence $A^c \in \mathcal{A}_i$ for all $i \in I$. Therefore, $A^c \in \bigcap_{i \in I} \mathcal{A}_i$.

Let $\{F_j\}_{j \in \mathbb{N}}$ be a sequence such that $F_j \in \bigcap_{i \in I} \mathcal{A}_i$ for every j. Then $F_j \in A_i$ for all i, j since A_i's are σ-algebra. Hence, $\bigcup_{j \in \mathbb{N}} F_j \in A_i$ for all $i \in I$, and $\bigcup_{j \in \mathbb{N}} F_j \in \bigcap_{i \in I} \mathcal{A}_i$. ∎

Remark 1.2. The union of σ-algebra are not necessarily a σ-algebra. For example, consider

$$X = \{a, b, c\} \tag{1.1}$$
$$\mathcal{A}_1 = \{\varnothing, \{a\}, \{b, c\}, X\} \tag{1.2}$$
$$\mathcal{A}_2 = \{\varnothing, \{b\}, \{a, c\}, X\} \tag{1.3}$$
$$\mathcal{A}_1 \cup \mathcal{A}_2 = \{\varnothing, \{a\}, \{b\}, \{b, c\}, \{a, c\}, X\} \tag{1.4}$$

Both $\mathcal{A}_1$ and $\mathcal{A}_2$ are σ-algebra, but $\mathcal{A}_1 \cup \mathcal{A}_2$ is not a σ-algebra because $\{a\} \cup \{b\} \notin \mathcal{A}_1 \cup \mathcal{A}_2$.

Definition 1.2. For $\mathcal{M} \subseteq \mathcal{P}(X)$ ($\mathcal{M}$ is not necessarily a σ-algebra), the smallest σ-algebra (by taking intersections) containing $\mathcal{M}$ is defined as the **σ-algebra generated by** $\mathcal{M}$. The generated σ-algebra is simply the intersection of all σ-algebra that are supersets of $\mathcal{M}$.

$$\sigma(\mathcal{M}) = \bigcap_{\mathcal{A} \supseteq \mathcal{M} \ s.t. \ \mathcal{A} \text{ is } \sigma\text{-algebra}} \mathcal{A} \tag{1.5}$$

The σ-algebra generated by $\mathcal{M}$ is therefore the smallest σ-algebra containing $\mathcal{M}$.

Definition 1.3. Let (X, τ) be a topological space, then the **Borel algebra** is σ-algebra generated by the collection of open sets τ.

$$\mathcal{B}(X) := \sigma(\tau) \tag{1.6}$$

Theorem 1.2. Let $\mathcal{O}$ and $\mathcal{C}$ denote collections of open and close sets in $\mathbb{R}$, by definition, $\mathcal{B}(\mathbb{R}) \equiv \sigma(\mathcal{O})$. Moreover, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ as well.

Proof. For every open interval $(a, b) \subseteq \mathbb{R}$,

$$(a, b) = \bigcup_{n=1}^{\infty} [a + 1/n, b - 1/n] \in \sigma(\mathcal{C}) \tag{1.7}$$

Let $A \in \mathcal{B}(\mathbb{R})$, then A is resulted from a sequence of (countable) intersection, (countable) union and complement operations on open sets. Because every open sets can be written as a countable union of open intervals, A can also be constructed using a sequence of above-mentioned operations from open intervals. Hence A is resulted from a sequence of operations on elements in $\sigma(\mathcal{C})$ as well, therefore $A \in \sigma(\mathcal{C})$.

Similarly, for every closed interval $[a, b] \subseteq \mathbb{R}$,

$$[a, b] = \bigcup_{n=1}^{\infty} ((-\infty, a - 1/n) \cup (b + 1/n, \infty))^c \in \sigma(\mathcal{O}) \tag{1.8}$$

Therefore, for every $F \in \sigma(\mathcal{O})$, $F \in \sigma(\mathcal{C})$ and $\sigma(\mathcal{O}) = \sigma(\mathcal{C})$. ■

Theorem 1.3. Let $\mathcal{H}$ denote the collection of all half-open intervals in $\mathbb{R}$:

$$\mathcal{H} := \{[a, b) \mid a, b \in \mathbb{R}, a \leq b\} \tag{1.9}$$

then $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H})$.

Proof. For every half-open interval $[a, b)$, it can be expressed as

$$[a, b) = ((-\infty, a) \cup [b, \infty))^c \in \mathcal{B}(\mathbb{R}) \tag{1.10}$$

and every open interval can be written as

$$(a, b) = \bigcup_{n=1}^{\infty} [a + 1/n, b) \in \sigma(\mathcal{H}) \tag{1.11}$$

The proof is similar to Theorem (1.2). ■

Remark 1.3. We do not use the entire power set for analysis because it's too large to construct a sensible measure on (see Theorem 1.4).

Definition 1.4. For a measurable space $(X, \mathcal{A})$, a map $\mu : \mathcal{A} \to [0, \infty]$ is a **measure** if μ satisfies

1. $\mu(\varnothing) = 0$.
2. (σ-addivitity)$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$, where $A_i \in \mathcal{A}$ for all i and $A_i \cap A_j = \varnothing$ for all $i \neq j$.

Example 1.1. For an element $x \in X$, the **Dirac measure**,δ_x, on a measurable space $(X, \mathcal{A})$ is defined as

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{1.12}$$

Definition 1.5. For a measurable space $(X, \mathcal{A})$ and a measure μ defined on it, the triple $(X, \mathcal{A}, \mu)$ is a **measure space**.

Theorem 1.4. There is no measure μ on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ satisfying the following two conditions: (i) $\mu((a, b]) = b - a$ for every $a < b$ and (ii) $\mu(x + A) = \mu(A)$ for all $a \in \mathbb{R}$ and $A \in \mathcal{P}(\mathbb{R})$.

Proof. Suppose, for contradiction, there exists such a measure μ, then $\mu((0, 1]) = 1 < \infty$.

Claim: the only measure on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ satisfying $\mu((0, 1]) < \infty$ and $\mu(x + A) = \mu(A)$ is the zero measure.

To prove the claim, let $I := (0, 1]$ and defien the following equivalence relation on I:

$$x \sim y \iff x - y \in \mathbb{Q} \tag{1.13}$$

the corresponding equivalence class of x on I can be written as

$$[x] = \{x + r : r \in \mathbb{Q} \land x + r \in I\} \tag{1.14}$$

The collection of all such equivalence classes, $\mathcal{A}$, is a disjoint decomposition of I. (for every $x \in I$, $[x]$ must in $\mathcal{A}$ and $x \in [x]$ trivially. If there exists different $[x] \neq [y]$ but $[x] \cap [y] \neq \varnothing$, take $z \in [x] \cap [y]$, by the transitivity of equivalence relation, $x \sim z \sim y$. Therefore, $[x] = [y]$, contradiction.)

For each $[x] \in \mathcal{A}$, take exactly one $a_x \in [x]$ and define set $A := \{a_x : [x] \in \mathcal{A}\}$. As a result, set A satisfies the following two properties:

1. $\forall x \in I, \exists a_x \in A \ s.t. \ a_x \in [x]$.

2. $\forall x, y \in A$, $x \sim y \implies x = y$.

Since $\mathbb{Q} \cup (-1, 1]$ is countable, let $(r_n)_{n\in\mathbb{N}}$ be an enumeration of all elements in it.

For each $n \in \mathbb{N}$, define $A_n := r_n + A$.

Note that for any m, n such that $A_m \cap A_n \neq \varnothing$, take $x \in A_m \cap A_n$. By definition,

$$x = r_n + a_n \tag{1.15}$$

$$x = r_m + a_m \tag{1.16}$$

where $a_n, a_m \in A$ and $r_n, r_m \in \mathbb{Q}$. Consequently,

$$a_n - a_m = r_m - r_n \in \mathbb{Q} \tag{1.17}$$

Therefore, $a_n \sim a_m$. By the second property of A, $a_n = a_m$. Thus, $r_m = r_n$ and $m = n$.

Take the counterposition of what we just proved, $m \neq n \implies A_m \cap A_n = \varnothing$.

Let $z \in (0, 1]$, there exists some $a \in A$ such that $z \in [x]$. That is, $z = x + r$ for some $r \in \mathbb{Q} \cap (-1, 1]$. There must exist some $m \in \mathbb{N}$ such that $r_m = r$, and consequently, $z \in A_m$.

Therefore, $(0, 1] \subseteq \bigcup_{n\in\mathbb{N}} A_n \subseteq (-1, 2]$ (the second relation is obvious). Moreover,

$$\mu((0,1]) \leq \mu(\bigcup_{n\in\mathbb{N}} A_n) \leq \mu((-1,2]) = \mu((-1,0]) + \mu((0,1]) + \mu((1,2]) = 3\mu((0,1]) \tag{1.18}$$

Note that we just proved $\bigcup_{n\in\mathbb{N}} A_n$ is a disjoint union, hence,

$$\mu((0,1]) \leq \sum_{n=1}^{\infty} \mu(A_n) \leq 3\mu((0,1]) \tag{1.19}$$

$$\implies ((0,1])\mu \leq \sum_{n=1}^{\infty} \mu(A + r_n) \leq 3\mu((0,1]) \tag{1.20}$$

$$\implies \mu((0,1]) \leq \sum_{n=1}^{\infty} \mu(A) \leq 3\mu((0,1]) \tag{1.21}$$

Since $\mu((0, 1])$ is finite, the only value $\mu(A)$ can take is zero, and $\mu(I) = 0$ as well. Consequently, for any set $S \in \mathcal{P}(\mathbb{R})$, if $S \subseteq I$, then $\mu(S) \leq \mu(I)$ and $\mu(S) = 0$. Otherwise, let $l = \lfloor \inf(S) \rfloor$ and $u = \lceil \sup(S) \rceil$.

$$I \subseteq S \subseteq \bigcup_{n=l}^{u} (n, n+1] \tag{1.22}$$

Therefore,

$$0 \leq \mu(S) \leq \sum_{n=l}^{u} \mu(n + (0, 1]) = \sum_{n=l}^{u} \mu((0, 1]) = 0 \tag{1.23}$$

It's shown that $\mu(S) = 0$ for every $S \subseteq \mathcal{P}(\mathbb{R})$.

This leads to a contradiction to the first property required ($\mu((a, b]) = b - a$). ■

2 Measurable Spaces and Measurable Maps

Definition 2.1. Let $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$ be two measurable spaces. A function $f : X_1 \to X_2$ is a **measurable map** with respect to $\mathcal{A}_1$ and $\mathcal{A}_2$ (sometimes written as $f : (X_1, \mathcal{A}_1) \to (X_2, \mathcal{A}_2)$) if

$$f^{-1}(A_2) \in \mathcal{A}_1 \quad \forall A_2 \in \mathcal{A}_2 \tag{2.1}$$

That is, the pre-image of every set in $\mathcal{A}_2$ is an element in $\mathcal{A}_1$ as well.

Theorem 2.1. Let $(X, \mathcal{A})$ be a measurable space, then the indicator (characteristic) function for any $A \in \mathcal{A}$, $\mathcal{X}_A : X \to \mathbb{R}$, is measurable with respect to $\mathcal{A}$ and $\mathcal{B}(\mathbb{R})$.

$$\mathcal{X}_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{2.2}$$

Proof. Since $\mathcal{X}_A$ can only take values from $\{0, 1\}$, the pre-image of any set $\not\subseteq \{0, 1\}$ is undefined. We only need to consider pre-images of subsets of $\{0, 1\}$:

$$\mathcal{X}_A^{-1}(\varnothing) = \varnothing \tag{2.3}$$

$$\mathcal{X}_A^{-1}(\{0, 1\}) = X \tag{2.4}$$

$$\mathcal{X}_A^{-1}(\{0\}) = A^c \tag{2.5}$$

$$\mathcal{X}_A^{-1}(\{1\}) = A \tag{2.6}$$

Therefore, $\mathcal{X}_A$ is measurable. ■

Theorem 2.2. The composition of measurable maps is measurable.

Proof. For measurable spaces $(X_1, \mathcal{A}_1)$, $(X_2, \mathcal{A}_2)$, and $(X_3, \mathcal{A}_3)$, let $f : (X_1, \mathcal{A}_1) \to (X_2, \mathcal{A}_2)$ and $g : (X_2, \mathcal{A}_2) \to (X_3, \mathcal{A}_3)$ be two measurable functions.

Let $A_3 \in \mathcal{A}_3$, $A_2 := g^{-1}(A_3) \in \mathcal{A}_2$. Similarly, $A_1 := f^{-1}(A_2) \in \mathcal{A}_1$ as well. Note that $A_1 = (g \circ f)^{-1}(A_3)$, therefore, $g \circ f$ is measurable. ■

Theorem 2.3. For measurable spaces $(X, \mathcal{A})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and measurable maps $f, g : \Omega \to \mathbb{R}$, $f + g$, $f - g$ and $|f|$ are measurable.

Proof. ■

3 Lebesgue Measures and Lebesgue Integrals

Definition 3.1. For measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R})))$, the **Lebesgue measure** $\mu : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ is defined as

$$\mu([a, b)) = b - a \tag{3.1}$$

Definition 3.2. Let $(X, \mathcal{A}, \mu)$ be a measure space and for any $A \in \mathcal{A}$, the **Lebesgue integral** of indicator function $\mathcal{X}_A$ on X is defined to be $\mu(A) \in [0, \infty]$.

$$\int_X \mathcal{X}_A \; d\mu := \mu(A) \tag{3.2}$$

Definition 3.3. A function $f : (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a **simple function** (also termed step function and stair-case function) if there exists finitely many $A_1, \cdots, A_n \in \mathcal{A}$ and $c_1, \cdots, c_n \in \mathbb{R}$ such that

$$f = \sum_{i=1}^{n} c_i \mathcal{X}_{A_i} \tag{3.3}$$

That is, a function f is simple if it can be expressed as a linear combination of *finitely* many indicators.

Let $\mathbb{S}^+$ denote the set of non-negative simple functions.

$$\mathbb{S}^+ := \{f : (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \mid f \text{ is simple and } f \geq 0\} \tag{3.4}$$

Since simple functions only take finitely many values, every $f \in \mathbb{S}^+$ can be written as

$$f = \sum_{t \in f(X)} t \mathcal{X}_{\{x \in X \mid f(x) = t\}} = \sum_{i=1}^{n} c_i \mathcal{X}_{A_i}, \quad c_i \geq 0 \tag{3.5}$$

Theorem 3.1. Simple functions are measurable.

Definition 3.4 (Lebesgue integral for $\mathbb{S}^+$)**.** For $f \in \mathbb{S}^+$ such that $f = \sum_{i=1}^{n} c_i \mathcal{X}_{A_i}$ with $c_i \geq 0$, the **Lebesgue integral** of f with respect to μ is

$$I(f) = \int_X f \; d\mu := \sum_{i=1}^{n} c_i \mu(A_i) \in [0, \infty] \tag{3.6}$$

Theorem 3.2. The Lebesgue integral of $f, g \in \mathbb{S}^+$ satisfies

1. $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for $\alpha, \beta \geq 0$,

2. $f \leq g \implies I(f) \leq I(g)$.

Proof. ∎

Definition 3.5 (Lebesgue integral for non-negative functions)**.** For $f \geq 0$ be a measurable function, the **Lebesgue integral** of f with respect to measure μ is

$$I(f) = \int_X f \; d\mu := \sup\left\{\int_X s \; d\mu : s \in \mathbb{S}^+ \text{ and } s \leq f\right\} \tag{3.7}$$

Definition 3.6. A function f is μ-**integrable** if $\int_X f \; d\mu < \infty$.

Theorem 3.3. Let $f, g : (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable functions, if $0 \leq f = g$ except a μ-measure-zero set, that is,

$$\mu\left(\{x \in X \mid f(x) \neq g(x)\}\right) = 0 \tag{3.8}$$

then $\int_X f \; d\mu = \int_X g \; d\mu$.

Lemma 3.1. Let $h : X \to [0, \infty)$ be a simple function, for any $\tilde{X} \subseteq X$ such that $\mu(\tilde{X}^c) = 0$, $\int_X h \; d\mu$ is independent from the value of h on $\tilde{X}^c$.

Proof. of Lemma 3.1. Since h is a simple function, it takes only finitely many values and can be written as

$$h = \sum_{t \in h(X)} t\mathcal{X}_{\{x \in X | h(x) = t\}} = \sum_{t \in h(X) \backslash \{0\}} t\mathcal{X}_{\{x \in X | h(x) = t\}} \tag{3.9}$$

define

$$\tilde{h} = \begin{cases} h(x) & \text{if } x \in \tilde{X} \\ a & \text{if } x \in \tilde{X}^c \end{cases} \tag{3.10}$$

where $a \in [0, \infty)$ takes an arbitrary value, and $\tilde{h} \in \mathbb{S}^+$ as well.

$$\begin{aligned}
\int_X \tilde{h} \; d\mu &= \sum_{t \in \tilde{h}(X)} t\mu(\{x \in X | \tilde{h}(x) = t\}) && (3.11) \\
&= a \underbrace{\mu(\tilde{X}^c)}_{=0} + \sum_{t \in h(\tilde{X}) \backslash \{0\}} t\mu(\{x \in \tilde{X} | h(x) = t\}) && (3.12) \\
&= \sum_{t \in h(\tilde{X}) \backslash \{0\}} t\mu(\{x \in \tilde{X} | h(x) = t\}) && (3.13) \\
&= \sum_{t \in h(\tilde{X}) \backslash \{0\}} t\mu(\{x \in \tilde{X} | h(x) = t\}) + \sum_{t \in h(\tilde{X}) \backslash \{0\}} t \underbrace{\mu(\{x \in \tilde{X}^c | h(x) = t\})}_{=0} && (3.14) \\
&= \sum_{t \in h(\tilde{X}) \backslash \{0\}} t\mu(\{x \in \tilde{X} | h(x) = t\} \cup \{x \in \tilde{X}^c | h(x) = t\}) && (3.15) \\
&= \sum_{t \in h(\tilde{X}) \backslash \{0\}} t\mu(\{x \in X | h(x) = t\}) + \sum_{t' \in h(X) \backslash (h(\tilde{X}) \cup \{0\})} t'\mu(\{x \in X | h(x) = t'\}) && (3.16)
\end{aligned}$$

Note that t's are values that are attained in $\tilde{X}^c$ only, therefore, $\{x \in X | h(x) = t'\} \subseteq \tilde{X}^c$ and have measure zero.

$$(3.16) = \sum_{t \in h(X)\setminus\{0\}} t\mu(\{x \in X | h(x) = t\}) = \int_X h \ d\mu \tag{3.17}$$

Hence, the value of $\int_X h \ d\mu$ is the same no matter how we change h's values on $\tilde{X}^c$. ∎

Proof. of Theorem 3.3. Let $\tilde{X} := \{x \in X : f(x) \neq g(x)\}$, for each simple function h in above set, define

$$\tilde{h} = \begin{cases} h(x) & \text{if } x \in \tilde{X} \\ 0 & \text{if } x \in \tilde{X}^c \end{cases} \tag{3.18}$$

Then,

$$\int_X f \ d\mu = \sup\left\{I(h) \mid h \in \mathbb{S}^+, h \leq f \text{ on } X\right\} \tag{3.19}$$

$$= \sup\left\{I(\tilde{h}) \mid h \in \mathbb{S}^+, h \leq f \text{ on } X\right\} \tag{3.20}$$

$$= \sup\left\{I(\tilde{h}) \mid h \in \mathbb{S}^+, h \leq f \text{ on } \tilde{X}\right\} \tag{3.21}$$

$$= \sup\left\{I(\tilde{h}) \mid h \in \mathbb{S}^+, h \leq g \text{ on } \tilde{X}\right\} \tag{3.22}$$

$$= \int_X g \ d\mu \tag{3.23}$$

Where equation (5.11) holds because the value of h on $\tilde{X}^c$ does not affect $I(\tilde{h})$. ∎

Theorem 3.4. Let $f, g : (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable functions, if $0 \leq f \leq g$ except a μ-measure-zero set, then $\int_X f \ d\mu \leq \int_X g \ d\mu$.

Proof. By definition of Lebesgue integral,

$$\int_X f \ d\mu = \sup\left\{I(h) \mid h \in \mathbb{S}^+, h \leq f \text{ on } X\right\} \tag{3.24}$$

Let $\tilde{X} := \{x \in X : f(x) \leq g(x)\}$, for each simple function h in above set, define

$$\tilde{h} = \begin{cases} h(x) & \text{if } x \in \tilde{X} \\ 0 & \text{if } x \in \tilde{X}^c \end{cases} \tag{3.25}$$

Then $h \leq f \iff \tilde{h} \leq f$, and $I(h) = I(\tilde{h})$ by Lemma 3.1.

$$\sup\left\{I(h) \mid h \in \mathbb{S}^+, h \leq f \text{ on } X\right\} = \sup\left\{I(\tilde{h}) \mid h \in \mathbb{S}^+, h \leq f \text{ on } \tilde{X}\right\} \tag{3.26}$$

$$\leq \sup\left\{I(\tilde{h}) \mid h \in \mathbb{S}^+, h \leq g \text{ on } \tilde{X}\right\} \tag{3.27}$$

$$= \int_X g \ d\mu \tag{3.28}$$

Therefore,

$$\int_X f \, d\mu \leq \int_X g \, d\mu \tag{3.29}$$

■

Theorem 3.5. Let $f : (X, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable functions, $f = 0$ except a μ-measure-zero set if and only if $\int_X f \, d\mu = 0$.

Proof. Similar to previous proofs. ■

Theorem 3.6 (Monotone Convergence Theorem)**.** For measure space $(X, \mathcal{A}, \mu)$, let $(f_n : X \to [0, \infty))_{n\in\mathbb{N}}$ be a sequence of measurable functions such that

1. $f_n \leq f_{n+1}$ except for a μ-measure-zero set,
2. $\lim_{n\to\infty}$ converges point-wisely to f except for a μ-measure-zero set.

Then,

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X \lim_{n\to\infty} f_n \, d\mu = \int_X f \, d\mu \tag{3.30}$$

Proof. Since $f_n \leq f_{n+1}$ almost everywhere, and $f_n \to f$ point-wisely almost everywhere, therefore,

$$f_n \leq f_{n+1} \leq f \quad \forall n \in \mathbb{N} \text{ except a set with zero measure} \tag{3.31}$$

Consequently,

$$\int_X f_n \, d\mu \leq \int_X f_{n+1} \, d\mu \leq \int_X f \, d\mu \quad \forall n \in \mathbb{N} \tag{3.32}$$

As a result,

$$\lim_{n\to\infty} \int_X f_n \, d\mu \leq \int_X f \, d\mu \tag{3.33}$$

Let h be a simple function such that $0 \leq h \leq f$, let $\varepsilon > 0$, define

$$X_n := \{x \in X \mid f_n(x) \geq (1 - \varepsilon)h(x)\} \tag{3.34}$$

$$\tilde{X} := \bigcup_{n=1}^{\infty} X_n \tag{3.35}$$

Note that $f_{n+1} \geq f_n$ implies $X_n \subseteq X_{n+1}$ and $\lim_{n\to\infty} X_n = \tilde{X}$. Moreover, because the monotonicity and point-wise convergence happen almost everywhere in X, almost all $x \in X$ are in some X_n with n sufficiently large, hence $\mu(\tilde{X}^c) = 0$ (this holds for the limit $\tilde{X}^c$ only but not necessarily for X_n^c).

Because $X_n \subseteq X$ and $f_n \geq 0$, for any $n \in \mathbb{N}$,

$$\int_X f_n \, d\mu \geq \int_{X_n} f_n \, d\mu \geq \int_{X_n} (1-\varepsilon)h \, d\mu \tag{3.36}$$

$$\implies \lim_{n\to\infty} \int_X f_n \, d\mu \geq \lim_{n\to\infty} \int_{X_n} (1-\varepsilon)h \, d\mu \tag{3.37}$$

$$= \int_{\tilde{X}} (1-\varepsilon)h \, d\mu \tag{3.38}$$

$$= \int_X (1-\varepsilon)h \, d\mu \tag{3.39}$$

Where the last equality holds because $\mu(\tilde{X}^c) = 0$.

Since this inequality holds for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \int_X f_n \, d\mu \geq \lim_{\varepsilon\to 0^+} \int_X (1-\varepsilon)h \, d\mu = \int_X h \, d\mu \tag{3.40}$$

moreover, this inequality holds for all $0 \leq h \leq f$,

$$\lim_{n\to\infty} \int_X f_n \, d\mu \geq \int_X f \, d\mu \tag{3.41}$$

Results (3.33) and (3.41) lead to the conclusion. ■

Corollary 3.1. Let $(g_n)_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions, $g_n : X \to [0,\infty]$, then the integral of series equals the series of integrals:

$$\sum_{n=1}^{\infty} g_n : X \to [0,\infty] \tag{3.42}$$

is measurable, and

$$\int_X \sum_{n=1}^{\infty} g_n \, d\mu = \sum_{n=1}^{\infty} \int_X g_n \, d\mu \tag{3.43}$$

Proof. Let $g_k := \sum_{n=1}^{k} g_n$ and $g = \lim_{k\to\infty} g_k$. Since $g_n \geq 0$, $g_k \leq g_{k+1}$ for every k. By the monotone convergence theorem,

$$\int_X \lim_{k\to\infty} g_k = \lim_{k\to\infty} \int_X g_k \; d\mu \tag{3.44}$$

$$\implies \int_X \sum_{n=1}^{\infty} g_n \; d\mu = \lim_{k\to\infty} \int_X \sum_{n=1}^{k} g_n \; d\mu \tag{3.45}$$

$$= \lim_{k\to\infty} \sum_{n=1}^{k} \int_X g_n \; d\mu \tag{3.46}$$

$$= \sum_{n=1}^{\infty} \int_X g_n \; d\mu \tag{3.47}$$

■

Lemma 3.2 (Fatou's Lemma)**.** For a measure space $(X, \mathcal{A}, \mu)$, let $(f_n)_{n\in\mathbb{N}}$ be a sequence of <u>measurable</u> (note this is the only requirement) functions with range $[0, \infty]$, then

$$\int_X \liminf_{n\to\infty} f_n \; d\mu \leq \liminf_{n\to\infty} \int_X f_n \; d\mu \tag{3.48}$$

Proposition 3.1. Infimum of measurable functions is measurable.

Proposition 3.2. Limit of measurable functions is measurable.

Proof of Lemma 3.2. Define

$$g_n(x) := \inf_{k\geq n} f_k(x) \tag{3.49}$$

Note that (g_n) is a non-decreasing sequence of measurable functions, the Monotone convergence theorem suggests

$$\int_X \lim_{n\to\infty} g_n \; d\mu = \lim_{n\to\infty} \int_X g_n \; d\mu \tag{3.50}$$

Since g_n's are non-decreasing, $(\int_X g_n \; d\mu)_{n\in\mathbb{N}}$ is non-decreasing as well. Therefore, $\int_X g_n \; d\mu = \inf_{k\geq n} \int_X g_k \; d\mu$ for every $n \in \mathbb{N}$ and consequently,

$$\lim_{n\to\infty} \int_X g_n \; d\mu = \liminf_{n\to\infty} \int_X g_n \; d\mu \tag{3.51}$$

Moreover, for every $n \in \mathbb{N}$, $g_n \leq f_n$ by the definition, therefore,

$$\int_X g_n \; d\mu \leq \int_X f_n \; d\mu \tag{3.52}$$

This inequality is preserved under $\liminf$, hence

$$\liminf_{n\to\infty} \int_X g_n \; d\mu \leq \liminf_{n\to\infty} \int_X f_n \; d\mu \tag{3.53}$$

Put everything together,

$$\int_X \liminf_{n\to\infty} f_n \, d\mu \equiv \int_X \lim_{n\to\infty} g_n \, d\mu = \lim_{n\to\infty} \int_X g_n \, d\mu = \liminf_{n\to\infty} \int_X g_n \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu \tag{3.54}$$

$\blacksquare$

Definition 3.7. For a measure space $(X, \mathcal{A}, \mu)$, define the $\mathcal{L}^p$ **space** of measurable functions to be

$$\mathcal{L}^p(X, \mathcal{A}, \mu) := \left\{ f : X \to \mathbb{R} \;\middle|\; f \text{ is measurable and } \left(\int_X |f|^p \, d\mu \right)^{1/p} < \infty \right\} \tag{3.55}$$

where $\left(\int_X |f|^p \, d\mu\right)^{1/p}$ is called the p**-norm** of f, denoted as $||f||_p$. For simplicity, the $\mathcal{L}^p$ space is often denoted as $\mathcal{L}^p(\mu)$.

Definition 3.8. Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f \in \mathcal{L}^1(\mu)$ be an arbitrary function. f can be expressed as the sum of two non-negative functions f^+ and f^-. In particular,

$$f^+(x) := \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad f^-(x) := \begin{cases} -f(x) & \text{if } f(x) \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.56}$$

Then, the **Lebesgue integral** of function f is defined as

$$\int_X f \, d\mu := \int_X f^+ \, d\mu - \int_X f^- \, d\mu \tag{3.57}$$

Theorem 3.7 (Lebesgue's Dominated Convergence Theorem)**.** For a measure space $(X, \mathcal{A}, \mu)$, let $f_n : X \to \mathbb{R}$ be measurable function for each $n \in \mathbb{N}$. Suppose (f_n) converges point-wisely to $f : X \to \mathbb{R}$ almost everywhere w.r.t. measure μ[1]. If there exists $g \in \mathcal{L}^1(X, \mathcal{A}, \mu)$ such that $|f_n| \leq g$ for all $n \in \mathbb{N}$, then

1. $f_n \in \mathcal{L}^1(X, \mathcal{A}, \mu)$ for every $n \in \mathbb{N}$ and $f \in \mathcal{L}^1(X, \mathcal{A}, \mu)$,
2. we may exchange the limit and integral.

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu \tag{3.58}$$

Proof. Since $|f_n| \leq g$, for every $n \in \mathbb{N}$

$$\int_X |f_n| \, d\mu \leq \int_X g \, d\mu \leq \int_X |g| \, d\mu < \infty \tag{3.59}$$

Because f is the point-wise limit of f_n almost everywhere, $|f| \leq g$ almost everywhere as well.

[1]When we say a property holds almost everywhere w.r.t. measure μ, it means the set on which this property does not hold has measure zero under μ.

Therefore,

$$\int_X |f| \ d\mu \leq \int_X g \ d\mu < \infty \tag{3.60}$$

Hence $f_n, f \in \mathcal{L}^1$ for all n.

To prove the second conclusion, we are going to show

$$\lim_{n\to\infty} \int_X |f_n - f| \ d\mu = 0 \tag{3.61}$$

By triangle inequality, the following inequality holds almost everywhere for every $n \in \mathbb{N}$:

$$|f_n - f| \leq |f_n| - |f| \leq 2g \tag{3.62}$$

as a result,

$$\int_X |f_n - f| \ d\mu \leq \int_X |f_n| \ d\mu + \int_X |f| \ d\mu \tag{3.63}$$

Define

$$h_n := 2g - |f_n - f| \geq 0 \tag{3.64}$$

and h_n is measurable as well. By Fatou's lemma,

$$\int_X \liminf_{n\to\infty} h_n \ d\mu \leq \liminf_{n\to\infty} \int_X h_n \ d\mu \tag{3.65}$$

$$\implies \int_X 2g \ d\mu - \int_X \underbrace{\liminf_{n\to\infty} |f_n - f|}_{=0} \ d\mu \leq \int_X 2g \ d\mu - \limsup_{n\to\infty} \int_X |f_n - f| \ d\mu \tag{3.66}$$

$$\implies \int_X 2g \ d\mu \leq \int_X 2g \ d\mu - \limsup_{n\to\infty} \int_X |f_n - f| \ d\mu \tag{3.67}$$

$$\implies \liminf_{n\to\infty} \int_X |f_n - f| \ d\mu \leq \limsup_{n\to\infty} \int_X |f_n - f| \ d\mu \leq 0 \tag{3.68}$$

$$\implies \lim_{n\to\infty} \int_X |f_n - f| \ d\mu \leq 0 \tag{3.69}$$

Since $|f_n - f| \geq 0$, the limit must be non-negative as well, hence

$$\lim_{n\to\infty} \int_X |f_n - f| \ d\mu = 0 \tag{3.70}$$

Moreover,

$$0 \leq \left| \int_X f_n \; d\mu - \int_X f \; d\mu \right| \tag{3.71}$$

$$= \left| \int_X f_n - f \; d\mu \right| \tag{3.72}$$

$$\leq \int_X |f_n - f| \; d\mu \to 0 \text{ as } n \to \infty \tag{3.73}$$

By squeeze theorem,

$$\lim_{n\to\infty} \left| \int_X f_n \; d\mu - \int_X f \; d\mu \right| = 0 \tag{3.74}$$

$$\implies \lim_{n\to\infty} \int_X f_n \; d\mu = \int_X f \; d\mu \tag{3.75}$$

■

4 Constructing Measures

Definition 4.1. For an arbitrary set X and its power set $\mathcal{P}(X)$, $\mathcal{A} \subseteq \mathcal{P}(X)$ is said to be a **semi-ring** of sets if it satisfies

1. $\varnothing \in \mathcal{A}$,
2. $A \cap B \in \mathcal{A}$ for every $A, B \in \mathcal{A}$,
3. for every $A, B \in \mathcal{A}$, there exists finitely many pairwise disjoint sets $S_1, S_2, \cdots, S_n \in \mathcal{A}$ such that $A \backslash B = \bigcup_{i=1}^n S_i$.

Definition 4.2. Let $\mathcal{A}$ be a semi-ring, then a mapping $\mu : \mathcal{A} \to [0, \infty]$ is a **pre-measure** if it satisfies

1. $\mu(\varnothing) = 0$,
2. and the σ-additivity, for any disjoint sequence $(A_i)_{i\in\mathbb{N}}$ in $\mathcal{A}$ such that $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{4.1}$$

The main difference between a measure and a pre-measure is that a measure must be defined on a σ-algebra.

Theorem 4.1 (Caratheodory's Extension Theorem)**.** For a set X, a semi-ring $\mathcal{A} \subseteq \mathcal{P}(X)$, and a pre-measure $\mu : \mathcal{A} \to [0, \infty]$,

1. μ has an extension: a measure

$$\tilde{\mu} : \sigma(\mathcal{A}) \to [0, \infty] \tag{4.2}$$

where $\sigma(\mathcal{A})$ is the σ-algebra generated by $\mathcal{A}$, such that

$$\mu(A) = \tilde{\mu}(A) \quad \forall A \in \mathcal{A} \tag{4.3}$$

2. If there exists (S_j) such that every $S_j \in \mathcal{A}$, $\bigcup_{j\in\mathbb{N}} S_i\ X$, and $\mu(S_j) < \infty$, then the extension $\tilde{\mu}$ is unique.

Proof. ■

Definition 4.3. Let $F : \mathbb{R} \to \mathbb{R}$ be an monotonically non-decreasing function, then we may construct a pre-measure μ_F on semi-ring $\mathcal{A} := \{[a, b) \mid a, b \in \mathbb{R}, a \leq b\}$ such that

$$\begin{aligned} \tilde{mu}_F([a, b)) &= F(b^-) - F(a^-) &(4.4)\\ &= \lim_{x\to b^-} F(x) - \lim_{x\to a^-} F(x) \quad (\dagger) &(4.5) \end{aligned}$$

By the Caratheodory's extension theorem, there exists a unique measure μ_F on $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ satisfying $(\dagger)$. Then μ_F is essentially the measure constructed by F on measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Example 4.1. Let $F(x) := x$, then $\mu_F([a, b)) = b - a$ is the Lebesgue measure.

Example 4.2. Let $F(x) := 1\ \forall x \in \mathbb{R}$, then $\mu_F([a, b)) = 0$ is the zero measure.

Example 4.3. Let $F(x) = \mathbb{1}\{x \geq 0\}$, then for every $\varepsilon_1, \varepsilon_2 > 0$, $\mu_F([-\varepsilon_1, \varepsilon_2)) = 1$ and for any other half-open interval I such that $0 \notin I$, $\mu_F(I) = 0$. In this case, μ_F is the Dirac measure δ_0.

Example 4.4. Let $F : \mathbb{R} \to \mathbb{R}$ be any non-decreasing and continuously differentiable function, so that $F' : \mathbb{R} \to [0, \infty)$. Because $\mu_F([a, b))$ satisfies

$$\begin{aligned} \mu_F([a, b)) &= F(b^-) - F(a^-) &(4.6)\\ &= F(b) - F(a) &(4.7)\\ &= \int_a^b F'(x)\ dx &(4.8) \end{aligned}$$

where dx is the normal Lebesgue measure and $F'(x)$ is called the density function. Then for every $A \in \mathcal{B}(\mathbb{R})$, the measure

$$\mu_F(A) := \int_A F'(x)\ dx \tag{4.9}$$

Notation 4.1. For now, let's consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where λ is the Lebesgue measure defined as $\lambda([a, b)) := b - a$.

Definition 4.4. Let λ and μ be two measures on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (for our consideration here, λ is the Lebesgue measure), then μ is **absolutely continuous** (w.r.t. the Lebesgue measure) if

$$\forall A \in \mathcal{B}(\mathbb{R}),\ \lambda(A) = 0 \implies \mu(A) = 0 \tag{4.10}$$

denoted as $\mu \ll \lambda$.

Definition 4.5. A measure μ is **singular** w.r.t. λ if there exists $N \in \mathcal{B}(\mathbb{R})$ such that

$$\lambda(N) = 0 \wedge \mu(N^c) = 0 \tag{4.11}$$

denoted as $\mu \perp \lambda$.

Definition 4.6. A measure μ on $(X, \mathcal{A})$ is said to be σ**-finite** if there exists a sequence of (E_n) satisfying

$$X = \bigcup_{n=1}^{\infty} E_n \tag{4.12}$$

and $\mu(E_n) < \infty$ for all $n \in \mathbb{N}$.

Example 4.5. The Lebesgue measure is σ-finite: $\mathbb{R} = \bigcup_{k\in\mathbb{Z}}[k, k+1)$ and each $\lambda([k, k+1)) = 1 < \infty$.

Theorem 4.2 (Lebesgue Decomposition)**.** Let $\mu : \mathbb{R} \to [0, \infty)$ be a σ-finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, there exists a unique decomposition $\mu_{ac}, \mu_s : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ such that

$$\mu = \mu_{ac} + \mu_s \tag{4.13}$$

$$\mu_{ac} \ll \lambda \tag{4.14}$$

$$\mu_s \perp \lambda \tag{4.15}$$

Theorem 4.3 (Radon-Nikodym)**.** Let μ be σ-finite measure on measurable space $(X, \mathcal{A})$ such that μ is absolutely continuous (w.r.t. the Lebesgue measure λ). Then there is a (λ-)measurable map $h : \mathbb{R} \to [0, \infty)$ (the density function) satisfying

$$\mu(A) = \int_A h \, d\lambda \quad \forall A \in \mathcal{B}(\mathbb{R}) \tag{4.16}$$

The measurable map h is defined as the **Radon–Nikodym derivative**, and is often denoted as $\frac{d\mu}{d\lambda}$.

5 Image Measure and Change of Variables

Definition 5.1. Let $h : (X, \mathcal{A}) \to (Y, \mathcal{C})$ be a measureable function, let μ be a measure on $(X, \mathcal{A})$. The **image measure** (pushforward measure) of h and μ, denoted as $h_*\mu$, is a measure on $(Y, \mathcal{C})$

defined as following:

$$\forall c \in \mathcal{C},\ h_*\mu(c) := \mu(h^{-1}(c)) \tag{5.1}$$

Because h is measurable, $h^{-1}(c) \in \mathcal{A}$ all the time and the above notion of image measure is well-defined.

Example 5.1. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and $X : \Omega \to \mathbb{R}$ is a random variable. Then the probability distribution P of X on $\mathbb{R}$ is precisely the image measure of μ:

$$P(X = x) := \mu(X^{-1}(\{x\})) \tag{5.2}$$

Theorem 5.1 (Change of Variables)**.** Consider a measure space $(X, \mathcal{A}, \mu)$, measurable space $(Y, \mathcal{C})$, and a measurable function $h : X \to Y$. Let $h_{*\mu}$ denote the image measure on $(Y, \mathcal{C})$. Moreover, suppose there is a integrable function $g : Y \to \mathbb{R}$, then

$$\int_Y g\ d(h_*\mu) = \int_X g \circ h\ d\mu \tag{5.3}$$

Proof. (i) The change of variable formula holds for characteristic functions.

Let $c \in \mathcal{C}$ and $\mathcal{X}_c$ be the characteristic function, then

$$\int_Y \mathcal{X}_c(y)\ d(h_*\mu) = h_*\mu(c) = \mu(h^{-1}(c)) \tag{5.4}$$

$$\int_X \mathcal{X}_c(h(x))\ d\mu = \mu(\{x \in X : h(x) \in c\}) = \mu(h^{-1}(c)) \tag{5.5}$$

(ii) The change of variable formula holds for simple functions.

Let $g = \sum_{i=1}^n \lambda_i \mathcal{X}_{c_i}$ be a simple function on $(Y, \mathcal{C})$, by the linearity of integrals,

$$\int_Y \sum_{i=1}^n \lambda_i \mathcal{X}_{c_i}\ d(h_*\mu) = \sum_{i=1}^n \int_Y \lambda_i \mathcal{X}_{c_i}\ d(h_*\mu) \tag{5.6}$$

$$= \sum_{i=1}^n \lambda_i \mu(h^{-1}(c_i)) \tag{5.7}$$

and

$$\int_X \sum_{i=1}^n \lambda_i \mathcal{X}_{c_i}(h(x))\ d\mu = \sum_{i=1}^n \int_X \lambda_i \mathcal{X}_{c_i}(h(x))\ d\mu \tag{5.8}$$

$$= \sum_{i=1}^n \lambda_i \mu(h^{-1}(c_i)) \tag{5.9}$$

(iii) The change of variable formula holds for non-negative measurable functions.

Let $g : Y \to [0, \infty)$ be a measurable function, then by the definition of Lebesgue integral:

$$\int_Y g\ d(h_*\mu) \equiv \sup\{\int_Y \tilde{s}\ d(h_*\mu) \mid \tilde{s} \text{ is simple and } \tilde{s}(y) \leq g(y)\ \forall y \in Y\} \tag{5.10}$$

$$= \sup\{\int_Y \tilde{s}\ d(h_*\mu) \mid \tilde{s} \text{ is simple and } \tilde{s}(h(x)) \leq g(h(x))\ \forall x \in X\} \tag{5.11}$$

Note that $s := \tilde{s} \circ h$ is simple as well.

$$(5.11) = \sup\{\int_X \tilde{s}(h(x))\ d\mu \mid \tilde{s} \text{ is simple and } \tilde{s}(h(x)) \leq g(h(x))\ \forall x \in X\} \tag{5.12}$$

$$= \sup\{\int_X s\ d\mu \mid s : X \to \mathbb{R} \text{ is simple and } s \leq g \circ h\} \tag{5.13}$$

$$\equiv \int_X g \circ h\ d\mu \tag{5.14}$$

(iv) The general case. An arbitrary measurable function g may be written as the difference of two non-negative measurable function: $g = g^+ - g^-$. Applying the linearity of integrals and the result for non-negative functions leads to the general result. ■