Probability theory

ezz118

February 12, 2025

Contents

Chapter 1: Preliminary Math

1.1 Real number system

1.1.1 Basics of real number

Definition: Lower bound and upper bound
Let S be a non-empty subset of $\mathbb{R}$. Define $a \in \mathbb{R}$ is a lower bound of S if

$$\forall x \in S,\ a \leq x$$

Similarly, define $a \in \mathbb{R}$ is an upper bound of S if

$$\forall x \in S,\ a \geq x$$

If a set has lower or upper bound, we say the set is bounded from below or above.

Definition: Infimum and supremum of a set
Obviously, lower bound or upper bound of a set is not unique. Now define $x \in \mathbb{R}$ to be the infimum of S if

$$\text{for all lower bound } a \text{ of } S,\ x \geq a$$

denoted as

$$\inf S = x$$

Similarly, define $x \in \mathbb{R}$ to be the supremum of S if

$$\text{for all upper bound } a \text{ of } S, \; x \leq a$$

denoted as

$$\sup S = x$$

Since the upper bound is always greater or equal to the lower bound, we have $\inf S \leq \sup S$.
When $S = \emptyset$, we define $\inf \emptyset = +\infty$ and $\sup \emptyset = -\infty$.
We can regard a sequence $a_1, a_2, \cdots$, or denoted as $\{a_n\}_{n=1}^{\infty}$, to be a set $\{a_1, a_2, \cdots\}$. Then we can write $\sup a_n$ to be the supremum of the sequence. Similar for $\inf a_n$.

The existence of supremum and infimum is given by the following completeness axiom.

The Completeness Axiom
Least upper bound property: $\forall S \subseteq \mathbb{R}$ which is bounded from above, then its supremum must exist.
This property is an expression of the completeness axiom, which intuitively states that there is no gap on the real line. More expression of this axiom can be found on Wiki.
Such property do not hold for rational number set. We can construct a subset of $\mathbb{Q}$ which do not have a smallest upper bound in $\mathbb{Q}$, for example, $\{x \in \mathbb{Q} : x^2 < 2\}$.

For the set S which is not bounded either form above or below, we extent the definition of supremum and infimum to infinity, which is

$$\sup S = \infty$$

or

$$\inf S = -\infty$$

In the following discussion, infinity is often included with the real number set $\mathbb{R}$. So we write $\bar{\mathbb{R}}$ or $[-\infty, \infty]$ to represent the extended real number set $\mathbb{R} \cup \{-\infty, \infty\}$.
Then, we extended the completeness axiom such that every subset of $\bar{\mathbb{R}}$ has a supremum. The comparison between elements (i.e. order) is specified as follows:

- For $a, b \in \bar{\mathbb{R}}$, define $a \leq b$ if either 1) $a = -\infty$, 2) $b = \infty$ or 3) $a, b \in \mathbb{R}$ and $a \leq b$
- Define $a = b$ if $a \leq b$ and $b \leq a$. Define $a < b$ if $a \leq b$ but $b \leq a$ doesn't hold. Similar definition for $\geq$, $>$ and $\neq$.

This makes $(\bar{\mathbb{R}}, \leq)$ to be a totally ordered set.
We further use the convention that $0 \times \infty = 0$ in the context of measure theory and probability

theory.

In 1.3.2, we will further discuss $\bar{\mathbb{R}}$ in the context of Topology.

Proposition about supremum and infimum

Let $S \subset \mathbb{R}$ be a set. If $\inf S > -\infty$, for any $\epsilon > 0$, $\exists x \in S$ such that

$$\inf S > x - \epsilon$$

Otherwise, $\inf S + \epsilon \leq x \ \forall x \in S$, then $\inf S + \epsilon$ is a larger lower bound of S.

If $\inf S = -\infty$, then for any $M \in \mathbb{R}$, $\exists x \in S$ such that $x < M$. Otherwise M is a lower bound.

Similar proposition holds for $\sup S$.

Corollary: There exists a sequence $\{a_n \in S\}_{n=1}^{\infty}$ that $\lim_{n\to\infty} a_n = \inf S$. Same for $\sup S$.

It follows from the proposition and the definition of limit below.

1.1.2 Sequence of real number

Definition: Limit of a sequence

Let $\{a_n\}_{n=1}^{\infty}$, be a sequence of real numbers, when there exists $L \in \mathbb{R}$ such that

$$\forall \epsilon > 0,\ \exists n_0 \in \mathbb{N} \text{ such that } \forall n > n_0, \left|a_n - L\right| < \epsilon$$

We say the sequence converges to L, or the limit of the sequence exists, which is L. Denoted by:

$$\lim_{n\to\infty} a_n = L$$

This is similar to the limit of a function. And the limit can be infinity:

$$\lim_{n\to\infty} a_n = \infty \iff \forall M > 0,\ \exists n_0 \in \mathbb{N} \text{ such that } \forall n > n_0, a_n > M$$

Similar definition for negative infinity. Then the range of L is extended to $\bar{\mathbb{R}}$. We say a limit exists or not in the context of $\bar{\mathbb{R}}$.

Propositions about limit

1. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers whose limit exists. Then $\lim_{n\to\infty} a_n = l_0$ if and only if for any of its sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$, $\lim_{i\to\infty} a_{n_i} = l_0$.
2. Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be two sequences of real numbers whose limits exist. If $\forall i \in \mathbb{N}^*$, $a_i \leq b_i$, then $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.

3. (Sandwich theorem) Let $\{a_n\}_{n=1}^{\infty}$, $\{a_n\}_{n=1}^{\infty}$ and $\{c_n\}_{n=1}^{\infty}$ be three sequences of real numbers. If $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = l_0 \in [-\infty, \infty]$ and $\forall i \in \mathbb{N}^*$, $a_i \le b_i \le c_i$, then $\lim_{n\to\infty} b_n = l_0$.

They can be proved by definition. The $\Leftarrow$ direction of 1 and 2 can be proved by contradiction.

Proposition about monotonic sequence:
Any monotonic increasing and bounded (from above) sequence of real number $\{a_n\}_{n=1}^{\infty}$ converges to $\sup a_n$ (which must exist by the completeness axiom). Similarly, any monotonic decreasing and bounded (from below) sequence of real number $\{a_n\}_{n=1}^{\infty}$ converges to $\inf a_n$.
We can prove it by definition.

Proposition: limit of average
Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers such that $\lim_{n\to\infty} a_n = L \in \bar{\mathbb{R}}$, then define the average to be $b_n := \frac{1}{n}\sum_{i=1}^{n} a_n \ \forall n \in \mathbb{N}^*$. We have $\lim_{n\to\infty} b_n = L$.
Proof:
We discuss the case when $L \in \mathbb{R}$. The infinity case is similar. For any $\epsilon > 0$, there exist $N_1 \in \mathbb{N}^*$ such that

$$\begin{aligned}
&\forall n \ge N_1,\ |a_n - L| < \frac{\epsilon}{2} \\
&\qquad\qquad \Rightarrow\ |b_n - L| \le \frac{1}{n}\left(\sum_{i=1}^{N_1} |a_i - L| + \sum_{i=N_1+1}^{n} |a_i - L|\right) \le \frac{M}{n} + \frac{\epsilon(n-N_1)}{2n}
\end{aligned}$$

where $M := \sum_{i=1}^{N_1} |a_i - L| \in \mathbb{R}$. We can take N_2 such that $\frac{M}{n} + \frac{\epsilon(n-N_1)}{2n} < \epsilon \ \forall n \ge N_2$, which proves the result. The expression of N_2 is

$$N_2 := N_1 + \left\lceil \frac{2M}{\epsilon} \right\rceil + 1$$

The following theorems connect sequence limit with function limit.
Theorem: Sequential criteria for the limit of function
Let $f(x)$ be a function on domain $D \subset \mathbb{R}$. $\lim_{x\to c} f(x) = L$ if and only if for any sequence $\{a_n\}_{n=1}^{\infty}$ that converges to c, $\lim_{n\to\infty} f(a_n) = L$.
Note that the definition of function limit use the following ϵ-δ language:

$$\forall \epsilon > 0,\ \exists d > 0 \text{ such that } \forall x \in D \text{ and } \big|x - c\big| < d,\ \text{ we have } \big|f(x) - f(c)\big| < \epsilon$$

Proof:
The $\rightarrow$ direction comes from the definition. The $\leftarrow$ direction can be proved by contradiction.

Assume for any sequence $\{a_n\}_{n=1}^{\infty}$ that converges to c, $\lim_{n\to\infty} f(a_n) = L$. But the limit of $f(x)$ to c does not exist or does not equal to L, which means that

$$\exists \epsilon > 0,\ \forall d > 0,\ \text{we can always find } x \in D \text{ such that } \left|x - c\right| < d \text{ and } \left|f(x) - f(c)\right| \geq \epsilon$$

Fix such ϵ as ϵ_0, define a sequence of d that converges to 0, say, $\{\frac{1}{n}\}_{n=1}^{\infty}$. For every d_n, pick the x_n such that $\left|x_n - c\right| < d_n$ and $\left|f(x_n) - f(c)\right| \geq \epsilon_0$. Then we form the sequence $\{x_n\}_{n=1}^{\infty}$. We can check that $\{x_n\}_{n=1}^{\infty}$ converges to c but $\{f(x_n)\}_{n=1}^{\infty}$ does not converge to L. This shows the contradiction.
This theorem can be generated into one-sided limit or the case when c or L is infinity. For example:

$$\lim_{x\to c^+} f(x) = L \iff \lim_{n\to c^+} f(a_n) = L\ \forall \text{ monotonic decreasing } \{a_n\}_{n=1}^{\infty} \text{ such } \lim_{n\to\infty} a_n = c$$
$$\lim_{x\to -\infty} f(x) = \infty \iff \lim_{n\to -\infty} f(a_n) = \infty\ \forall \{a_n\}_{n=1}^{\infty} \text{ such } \lim_{n\to\infty} a_n = -\infty$$

Theorem: continuous mapping of limit
Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a continuous function and $\{a_n \in \mathbb{R}\}_{n=1}^{\infty}$ whose limit $L := \lim_{n\to\infty} a_n$ exists in $\mathbb{R}$. Then

$$\lim_{n\to\infty} f(a_n) = f(L) \text{ also exists}$$

Proof:
The continuity gives that for any $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(L)| < \epsilon\ \forall L - \delta < x < L + \delta$. Because $\lim_{n\to\infty} a_n = L$, there exists $N \in \mathbb{N}^*$ such that $\forall n > N$, $L - \delta < a_n < L + \delta$ and thus $|f(a_n) - f(L)| < \epsilon$. Therefore, $\lim_{n\to\infty} f(a_n) = f(L)$.
In 1.3.2, we will show that when L is infinite and f is a continuous $\bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ function, the result also hold.

We than state three important theorems about continuous function.

Intermediate value theorem
Let $f : [a, b] \to \mathbb{R}$ be a continuous function. For any $c \in [\min\{f(a), f(b)\}, \max\{f(a), f(b)\}]$, there exists $x \in [a, b]$ such that $f(x) = c$.
Proof:
The case $f(a) = f(b)$ is trivial. We then assume $f(a) < f(b)$, otherwise it is similar. If $c = f(a)$ or $c = f(b)$, a or b is as required. Fix $c \in (f(a), f(b))$, we claim that $x := \sup\{t \in [a, b] : f(t) < c\}$ is the solution. Since $f(a) < c$, the set $A := \{t \in [a, b] : f(t) < c\}$ is not empty. We first have $f(x) \leq c$ by the definition of the supremum. It suffices to prove $f(x) \geq c$.
Observe that $x < b$ since there exists a neighborhood of b with the function value greater than c. If

otherwise $f(x) < c$, there exists $\delta > 0$ such that $f(x+\delta) < c$ by the continuity. Then $(x+\delta) \in A$ and $\sup\{A\} \geq x + \delta$. The contradiction arises.

Extreme value theorem

If the function $f : [a,b] \to \mathbb{R}$ is continuous. Then it has both maximum and minimum.

With the fact that continuous function maps compact set to compact (see 1.3.2), this theorem is immediate. We offer another easier proof:

Take $A := \{f(t) : a \leq t \leq b\}$ and $M := \sup\{A\}$. We claim that M is the maximum, i.e. $\exists x \in [a,b],\ f(x) = M$. Let $\{a_i \in A\}_{i=1}^n$ be a sequence that $\lim_{i\to\infty} a_i = M$. For each $i \in \mathbb{N}^*$, there exists $t_i \in [a,b]$ such that $a_i = f(t_i)$. Let $\{t_{i_j}\}_{j=1}^\infty$ be a converging subsequence of $\{t_i\}_{i=1}^\infty$. We have $x := \lim_{j\to\infty} t_{i_j} \in [a,b]$ and $\lim_{j\to\infty} f(t_{i_j}) = M$. By continuity, $f(x) = M$. Likewise, the minimum is $\inf\{A\}$.

Corollary: Mean value theorem.

If the function $f : [a,b] \to \mathbb{R}$ $(b \neq a)$ is differentiable on (a,b), i.e., the limit

$$f'(x) := \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists, then there exists $x \in (a,b)$ such that

$$f'(x) = \frac{f(b) - f(a)}{b - a}$$

Proof:

We begin with a simple case when $f(a) = f(b)$, in which we prove $\exists x,\ f'(x) = 0$, known as the Rolle's theorem.

If $f(x) = f(a) = f(b)\ \forall x \in [a,b]$ is constant, the derivative are all zeros. By the extreme value theorem, the maximum and the minimum exists. Suppose the maximum $M > f(a)$ and it is attained by $x \in [a,b]$. Otherwise if $M = f(a)$, the minimum $m < f(a)$ unless f is constant, and the proof is similar.

First $x \in (a,b)$. We claim that $f'(x) = 0$. (It is known as the Fermat's theorem, that local extrema have derivative 0.) Suppose otherwise $f'(x) > 0$, $\exists h > 0$ such that $\frac{f(x+h)-f(x)}{h} > 0$, which gives $f(x+h) > f(x) = M$. Similar contradiction gives $f'(x) < 0$ does not hold.

More generally if $d = \frac{f(b)-f(a)}{b-a} \neq 0$. Let $g(x) = f(x) - d(x-a)\ \forall x \in [a,b]$. g satisfy all the assumption for the Rolle's theorem. Thus $\exists x \in [a,b]$, $0 = g'(x) = f'(x) - d$.

1.1.3 Limit supremum and limit infimum

The limit of a sequence may not exist. So we may have the following definition of limit point, which always exists.

Definition: Limit point of a sequence

For a sequence of real number $\{a_n\}_{n=1}^{\infty}$, if there exists a sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$ such that $\lim_{i\to\infty} a_{n_i} = L \in [\infty, \infty]$, we say L is a limit point of this sequence.

The set of all possible L is denoted as $\text{LIM}\{x_n\}$.

To prove the existence of LIM, it suffices to show that there exists a monotonic sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$. Then its limit exists, though probably goes to infinity. To find sub-sequence, we introduce a concept called *peak*. $\forall k \in \mathbb{N}^*$, we define a_k to be a *peak* if $a_k = \sup_{i\geq k} a_i$. There are two cases, both of which exists a monotonic sub-sequence.

1. There are infinite many *peaks* $a_{k_1}, a_{k_2}, \cdots$ Then $\{a_{k_i}\}_{i=1}^{\infty}$ is monotonic decreasing.
2. There is no *peak* or there are finitely many *peaks*. Then $\exists n_0$ such that there is no *peak* when $n \geq n_0$. Take $n_1 > n_0$, since a_{n_1} is not a peak, there exist a larger a_{n_2} with $n_2 > n_1$. Repeat the process for finding $a_{n_3} > a_{n_2}, a_{n_4} > a_{n_3}, \cdots$ Inductively, we can find a monotonic increasing sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$.

This definition helps to define the limit supremum and limit infimum of a sequence.

Definition: Limit supremum and limit infimum of a real number sequence

Given a sequence of real number $\{a_n\}_{n=1}^{\infty}$, we define its limit supremum and limit infimum respectively:

$$\limsup_{n\to\infty} a_n := \sup \text{LIM}\{x_n\}$$
$$\liminf_{n\to\infty} a_n := \inf \text{LIM}\{x_n\}$$

Limit supremum and limit infimum must exists in $\bar{\mathbb{R}}$.

Propositions of Limit supremum and limit infimum

Given a sequence of real number $\{a_n\}_{n=1}^{\infty}$, let $L = \liminf_{n\to\infty} a_n$ and $R = \limsup_{n\to\infty} a_n$.

1. $L \leq R$
2. When R is finite, $\forall \epsilon > 0$, the interval $(R + \epsilon, \infty)$ contains at most finitely many a_i. Similar result holds for $(-\infty, L - \epsilon)$.

3. When R is finite, $\forall \epsilon > 0$, the interval $(R-\epsilon, \infty)$ contains infinitely many a_i. Similar result holds for $(-\infty, L+\epsilon)$.

4. There exist a sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$ of $\{a_n\}_{n=1}^{\infty}$ such that $\lim_{i\to\infty} a_{n_i} = R$. Similar result holds for L.

5. $L = R$ if and only if $\{a_n\}_{n=1}^{\infty}$ converges. And $\lim_{n\to\infty} a_n = L$.

6. $\lim_{n\to\infty} a_n = L$ if any only if all the sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$ converges to L. To prove the convergence of a sequence, we can loosen the condition into all its convergent sub-sequence converges to a unique L.

1 immediately follows from the definition.

We prove 2 by contradiction. Assume there infinite many elements larger than $R+\epsilon$. Let them form a new sequence. We have shown that there exists a monotonic sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$ of the new sequence whose limit exist. Since $a_{n_i} > R+\epsilon$, we have $l_0 := \lim_{n\to\infty} a_{n_i} \in (R+\epsilon, \infty]$. Then because $\{a_{n_i}\}_{i=1}^{\infty}$ is also the sub-sequence of $\{a_n\}_{n=1}^{\infty}$, $l_0 \in \mathrm{LIM}\{x_n\}$ and $\sup \mathrm{LIM}\{x_n\} \geq l_0 > R$, which shows the contradiction.

3 is also proven by contradiction. Assume the negation holds. Then, any converging sequence $\{a_{n_i}\}_{i=1}^{\infty}$ with its limit l_0. There exist $i_0 \in \mathbb{N}^*$ such that $a_{n_i} \leq L - \epsilon \ \forall i \geq i_0$. Thus $l_0 \leq R - \epsilon$ and $\sup \mathrm{LIM}\{x_n\} \leq l_0 < R$, which shows the contradiction.

From 2 and 3, we have that $\forall \epsilon > 0$, $(R-\epsilon, R+\epsilon)$ contain infinitely many elements. Then, $\forall n_1 \in \mathbb{N}^*$, $\exists n_2 > n_1$ such that $a_{n_2} \in (R-\epsilon, R+\epsilon)$. Let $n_1 = 1$. Inductively, known n_i, we find $n_{i+1} > n_i$ such that $a_{n_{i+1}} \in (R - \frac{1}{i+1}, R + \frac{1}{i+1})$. Then the sub-sequence we have constructed has the limit R. (Details omitted)

5 is a further result of proposition 1 about limit.

6 follows from 5.

Proposition: Equivalent definitions of limit supremum and limit infimum of a sequence

Given a sequence of real number $\{a_n\}_{n=1}^{\infty}$, we can define its limit supremum and limit infimum in this way.

$$\limsup_{n\to\infty} := \inf_{k\in\mathbb{N}} \left(\sup_{m\geq k} a_m \right) = \lim_{k\to\infty} \left(\sup_{m\geq k} a_m \right)$$

$$\liminf_{n\to\infty} := \sup_{k\in\mathbb{N}} \left(\inf_{m\geq k} a_m \right) = \lim_{k\to\infty} \left(\inf_{m\geq k} a_m \right)$$

The equations of both lines hold because of the monotonicity of inf and sup.

Proof of equivalence: we only show $\lim_{n\to\infty} \left(\sup_{m\geq k} a_m \right) = \limsup_{n\to\infty}$, where $\limsup_{n\to\infty}$ uses the old definition. The proof of another is similar.

When $\limsup_{n\to\infty} = \infty$, it is not hard to show that $\sup_{m\geq k} a_m = \infty$ and thus, $\lim_{n\to\infty} \left(\sup_{m\geq k}\right) = \infty$. The following shows the case when $\limsup_{n\to\infty} < \infty$.

We have shown there exists a sub-sequence $\{a_{n_i}\}_{i=1}^{\infty}$ of $\{a_n\}_{n=1}^{\infty}$ whose limit is $\limsup_{n\to\infty} a_n$. Let $b_k := \sup_{m\geq k} a_m$. Then $b_{n_i} = \sup_{m\geq n_i} a_m \geq a_{n_i}$ $\forall i \in \mathbb{N}^*$, and

$$\lim_{k\to\infty} b_k = \lim_{i\to\infty} b_{n_i} \geq \lim_{i\to\infty} a_{n_i} = \limsup_{n\to\infty} a_n$$

Further we prove $\lim_{k\to\infty} b_k \leq \limsup_{n\to\infty} a_n$ by contradiction. Suppose the negation holds, $\exists \epsilon > 0$ such that $\inf b_m = \lim_{k\to\infty} b_k > 2\epsilon + \limsup_{n\to\infty} a_n$. By proposition 2 of limit supremum, their are at most finitely many elements in $\{a_n\}_{n=1}^{\infty}$ greater than $\epsilon + \limsup_{n\to\infty}$ and so as for $\inf b_m - \epsilon$. Then, $\exists n_0 \in \mathbb{N}^*$ such that $\forall n \geq n_0,\ a_n \leq \inf b_m - \epsilon$, which gives $a_n \leq b_n - \epsilon$. This contradicts to the definition of b_n.

Proposition: convergent sequence is equivalent to Cauchy sequence

Given a sequence of real number $\{a_n\}_{n=1}^{\infty}$, define it to be a Cauchy sequence if

$$\lim_{n\to\infty}\left(\sup_{i,j\geq n} |a_i - a_j|\right) = 0$$

Then $\{a_n\}_{n=1}^{\infty}$ converges to a real number L if and only if it is a Cauchy sequence.

We first prove the "$\Rightarrow$" direction. Since $\lim_{n\to\infty} a_n = L$, for any $\delta > 0$, there exist $N \in \mathbb{N}^*$ such that $|a_n - L| < \frac{\delta}{2}$ $\forall n \geq N$. It gives that

$$\forall i, j \geq N,\ |a_i - a_j| \leq |a_i - L| + |a_i - L| < \delta$$

and thus $\sup_{i,j\geq n} |a_i - a_j| < \delta$ $\forall n \geq N$. Then it is a Cauchy sequence.

We first prove the "$\Leftarrow$" direction. The definition of Cauchy sequence gives that for any $\delta > 0$, there exists $N \in \mathbb{N}^*$ such that $\forall n, i \geq N$, $|a_n - a_i| < \delta$ and then

$$\begin{cases} \sup_{i\geq n} a_i < a_n + \delta \\ \inf_{i\geq n} a_i > a_n - \delta \end{cases} \Rightarrow \sup_{i\geq n} a_i - \delta < a_n < \inf_{i\geq n} a_i + \delta$$

Since $\lim_{n\to\infty}(\sup_{i\geq n} a_i - \delta) = \limsup_{i\to\infty} a_i - \delta$ exists and so as the inf case, we have

$$\limsup_{i\to\infty} a_i - \delta \leq \liminf_{i\to\infty} a_i + \delta$$

Because $\limsup_{i\to\infty} a_i \geq \liminf_{i\to\infty} a_i$, we have $\liminf_{i\to\infty} a_i = \limsup_{i\to\infty} a_i$ by taking δ arbitrarily small. Then $\{a_i\}_{i=1}^{\infty}$ converges. We can further prove that it converges to finite value by contradiction.

1.2 Set theory

1.2.1 Power set, index set and countable set

Definition: Power set
For a set S, define the power set $\mathcal{P}(S)$ as the set of all the subsets of S.

Definition: Index set.
If every entry of the set S has a unique label i, all the label constitute the index set I. Then each entry of S can be expressed as $S_i, i \in I$, and $S = \{S_i : i \in I\}$. S_i can be considered as an injective map from I and S.

Definition: Countable set.
A set S is a countable set if the natural number set $\mathbb{N}$ (though often exclude 0) is an index set of S. In other words, all entries of a countable map can be injectively mapped to a natural number.

In the measure theory, we normally focus on countable sets. To construct more countable set, we can make use of these propositions.
Propositions about countable set (proof not given)

1. Any subset of a countable set is countable.
2. Any set is countable if and only if its index set is countable.
3. The Cartesian product of finitely many countable sets is countable.

Examples of countable sets:

1. Any finite set
2. $\mathbb{N}$ and $\mathbb{Z}$
3. $\mathbb{N}^2 = \{(x, y) : x, y \in \mathbb{N}\}$
4. $\mathbb{Z}^n$ and $\mathbb{Q}^n$
5. The set of ration number $\mathbb{Q}$

By definition, $\mathbb{N}$ is obviously countable, then we can map any integer x to $2|x| - \mathbf{1}_{x<0}$ injectively and show that $\mathbb{Z}$ is also countable.
3 and 4 is given by proposition 3. We can also show 3 by mapping (x, y) to the natural number $y + \sum_{i=1}^{x} i = \frac{x(x+1)}{2} + y$ and proving the mapping is injective (i.e. one-to-one).
$\mathbb{Q} = \{\frac{a}{b} : a, b \in \mathbb{Z}^* \text{ and } a \text{ is prime to } b\} \cup \{0\}$, we can construct its index set to be a subset of $\mathbb{Z}^2$, and 4 follows.
Note that $\mathbb{R}$ is uncountable.

The following proposition make use of the countability of ration number, which will be used in later chapter.

Proposition

Any open set (the definition of open set is in 1.3.1) A of $\mathbb{R}^n$ can be written as an union of at most countable open balls in $\mathbb{R}^n$, i.e.

$$A = \bigcup_{i=1}^{\infty} B(x_i, r_i)$$

We need to prove the set $\{B(x_i, r_i) : i \in I\}$ is countable. To prove it, let $A \subset \mathbb{R}^n$ be a open set. Then by definition, for any $x \in A$, we can take a $\epsilon_x > 0$ such that $B(x, 2\epsilon_x) \subset A$. Since $\mathbb{Q}$ is dense, we further let ϵ_x to be rational by making it smaller, and we can take a point $q_x \in \mathbb{Q}^n$ such that $d(x, q_x) < \epsilon_x$. Then, we have

$$\forall x \in A,\ x \in B(q_x, \epsilon_x) \ \Rightarrow\ A \subset \bigcup_{x \in A} B(q_x, \epsilon_x)$$

$$\forall x \in A,\ B(q_x, \epsilon_x) \subset B(x, 2\epsilon_x) \subset A \Rightarrow \bigcup_{x \in A} B(q_x, \epsilon_x) \subset A$$

thus,

$$A = \bigcup_{x \in A} B(q_x, \epsilon_x) = \bigcup_{(q_i, \epsilon_i) \in I} B(q_i, \epsilon_i), \text{ where } I := \{(q_x, \epsilon_x) : x \in A\}$$

which is a countable union because $(q_x, \epsilon_x) \in \mathbb{Q}^{n+1}$, and then I must be a countable set.

1.2.2 Closure

Definition: Closure

Let $\mathcal{C} \subset \mathcal{P}(S)$, we say $\mathcal{C}$ is closed under some certain operation if all the possible results of the operation belong to $\mathcal{C}$. Here are some examples of operation.

1. Complement: A^c, where $A \in \mathcal{C}$.

2. Arbitrary union: $\bigcup_{i \in I} A_i$, where I be an index set of $\{A_i : i \in I\}$, which satisfied $\forall i \in I,\ A_i \in \mathcal{C}$.

3. Finite union: $\bigcup_{i=1}^{n} A_i$, where $\{A_i\}_{i=1}^{n}$ is a finite sequence such $\forall i \in \{1, 2, \cdots, n\},\ A_i \in \mathcal{C}$. Such sequence will be denoted as $\{A_i \in \mathcal{C}\}_{i=1}^{n}$ in the following

4. Similarly, we have arbitrary intersection, countable intersection and finite intersection.

Here is an example of closure.

Example: Definition of Algebra

A non-empty subset $\mathcal{A} \subset \mathcal{P}(S)$ is called an algebra (aka. field) if it is closed under complement, finite union and finite intersection.

1.2.3 Limit of sets

Similar to comparing two real numbers, we can compare two sets with the partial order $\subset$.

Then the monotonicity of a sequence of sets can be defined.

Definition: Monotonic sequence of sets

A sequence of sets $\{A_n\}_{n=1}^{\infty}$ is monotonic increasing if

$$A_1 \subset A_2 \subset A_3 \subset \cdots \quad \text{i.e. } \forall i < j,\ A_i \subset A_j$$

The sequence is monotonic decreasing if

$$A_1 \supset A_2 \supset A_3 \supset \cdots \quad \text{i.e. } \forall i < j,\ A_i \supset A_j$$

We can also define sup and inf for a sequence of sets, although these notations are seldom used.

Definition: Supremum and infimum of a sequence of sets

Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets, define

$$\inf A_n := \bigcap_{n=1}^{\infty} A_n$$

$$\sup A_n := \bigcup_{n=1}^{\infty} A_n$$

Definition: Limit supremum and limit infimum of a sequence of sets.

Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets, define

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \inf_{k\geq n} A_k := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \sup_{k\geq n} A_k := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

The following proposition use the notation of indicator function.

Definition: indicator function

Let S be a set and $A \subset S$, define the indicator of A, $\mathbf{1}_A : S \mapsto \{0,1\}$ as

$$\mathbf{1}_A(x) := \begin{cases} 1, & x \in A \\ 0, & \text{otherwise} \end{cases}$$

Proposition about Limit supremum and limit infimum

1. $\liminf_{n\to\infty} A_n = \left\{x : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(x) < \infty\right\}$
2. $\limsup_{n\to\infty} A_n = \{x : \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(x) = \infty\}$
3. The first two statements immediately gives $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$

Proof of 1:

If $x \in \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$, then $\exists n_0 \in \mathbb{Z}^*$ such that $x \in \bigcap_{k=n_0}^{\infty} A_k$, then

$$\forall k \geq n_0,\ x \in A_n \Rightarrow \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(x) = \sum_{n=1}^{n_0-1} \mathbf{1}_{A_n^c}(x) + 0 \leq n_0 - 1 < \infty$$

Which gives $\liminf_{n\to\infty} A_n \subset \left\{x : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(x) < \infty\right\}$

Conversely, we can show $\liminf_{n\to\infty} A_n \supset \left\{x : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(x) < \infty\right\}$:

$$\forall x \in \left\{x : \sum_{n=1}^{\infty} \mathbf{1}_{A_n^c}(x) < \infty\right\},\ \exists L \text{ such that } \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{1}_{A_n^c}(x) = L$$

$$\Rightarrow \text{Let } \epsilon = 0.5,\ \exists n_0 \text{ such that } \forall n > n_0, \left|\sum_{i=1}^{n} \mathbf{1}_{A_n^c}(x) - L\right| < \epsilon$$

Since $\mathbf{1}_{A_n^c}(x) \in \{0,1\}$, we have

$$
\begin{aligned}
&\forall n > n_0, \sum_{i=1}^{n} \mathbf{1}_{A_n^c}(x) = \sum_{i=1}^{n_0} \mathbf{1}_{A_n^c}(x) + \sum_{i=n_0+1}^{n} \mathbf{1}_{A_n^c}(x) = L \\
\Rightarrow &\sum_{i=n_0+1}^{n} \mathbf{1}_{A_n^c}(x) = 0 \\
\Rightarrow &\forall n > n_0,\ x \in A_n \\
\Rightarrow &x \in \bigcap_{k=n_0}^{\infty} A_k \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \liminf_{n\to\infty} A_n
\end{aligned}
$$

Proof of 2 is similar.

Then we can define the limit of a sequence of sets:

Definition: Limit of a sequence of sets

If $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = A$, then the limit of this sequence exists, which is A, denoted as

$$
\lim_{n\to\infty} A_n = A
$$

Proposition about limit of monotonic sequence of sets

If $\{A_n\}_{n=1}^{\infty}$ is a monotonic increasing sequence of sets, then

$$
\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n
$$

If $\{A_n\}_{n=1}^{\infty}$ is a monotonic decreasing sequence of sets, then

$$
\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n
$$

Their proofs are similar, here we only work on the first one. Our goal is to show:

$$
\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n
$$

We first have

$$
\bigcap_{k=n}^{\infty} A_k = A_n \Rightarrow \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n
$$

Since $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$, it suffices to prove $\limsup_{n\to\infty} A_n \subset \liminf_{n\to\infty} A_n$, which can be shown by

$$
\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \subset \bigcup_{k=1}^{\infty} A_k \liminf_{n\to\infty} A_n
$$

These propositions follows that:

$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty}\left(\inf_{k\geq n} A_k\right)$$

$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty}\left(\sup_{k\geq n} A_k\right)$$

1.3 Metric space and topology space

1.3.1 Metric space

The goal of this section is to introduce the concept of continuity in the extended real number set $\bar{\mathbb{R}}$. It involves some basics of Topology. We start with discussing metric space, which is a special case of topology space.

Definition: metric space
Let S be a set and $d : S \times S \mapsto [0, \infty)$ be a function of distance. Then (S, d) is a metric space with metric d if it satisfies:

1. $\forall\ a, b \in S,\ d(a, b) \geq 0$
2. $\forall\ a, b \in S,\ a = b$ if and only if $d(a, b) = 0$
3. $\forall\ a, b \in S,\ d(a, b) = d(b, a)$
4. $\forall\ a, b, c \in S,\ d(a, b) \leq d(a, c) + d(c, b)$

Example of metric space

1. Euclidean space of dimension 1: the real number set with Euclidean distance: $(\mathbb{R}, d)$ where $d(x, y) = |x - y|$
2. Euclidean space of dimension n: real vector with Euclidean distance: $(\mathbb{R}^n, d)$ where $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$
3. Real vector of infinite dimension: $(\mathbb{R}^\infty, d)$, aka real sequence $(\{a_n\}_{n=1}^\infty, d)$, where

$$d(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{\infty} \frac{2^{-k} \sum_{n=1}^{k} |x_i - y_j|}{1 + \sum_{n=1}^{k} |x_i - y_j|}$$

Proof?

Definition: Open set and close set of a metric space
Given a metric space (S, d). A subset $A \subset S$ is defined to be an open set if $\forall x \in A,\ \exists \epsilon > 0$ such that

$$\forall\ y \in S \text{ with } d(x, y) < \epsilon, \text{ we have } y \in A$$

A subset $A \subset S$ is close if $A^c = S \setminus A$ is open.

Definition: Open ball
Given a metric space (S, d). Let $x_0 \in S,\ r \in (0, \infty)$, Define

$$B(x_0, r) := \{x \in S : d(x, x_0) < r\}$$

as an open ball at point x_0 with radius r.
We can show that any open ball is an open set. Because for an arbitrary $x_1 \in S$ with $r_1 = d(x_0, x_1)$, we have $\epsilon = r - r_1$ that satisfies the condition for open set.
The open interval of $\mathbb{R}$, (l, r), where $-\infty < l < r < \infty$ is an open ball at $\frac{l+r}{2}$ with radius $\frac{r-l}{2}$.

Definition: Limit in a metric space
Given a metric space (X, d_X), a sequence $\{a_n \in X\}_{n=1}^{\infty}$ converges to $L \in X$, written $\lim_{n\to\infty} a_n = L$, if

$$\forall \epsilon > 0,\ \exists n_0 \text{ such that } d_X(a_n, L) < \epsilon\ \forall n \geq n_0$$

or equivalently,

$$\lim_{n\to\infty} d_X(a_n, L) = 0$$

Definition: Continuous function
Given two metric spaces (X, d_X) and (Y, d_Y) and a function $f : X \mapsto Y$. The function f is continuous at point $a \in X$ if $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$\forall x \in X \text{ such } d_X(a, x) < \delta,\ \text{ we have } d_Y(f(a), f(x)) < \epsilon$$

or equivalently, $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$\forall x, y \in X \text{ such } d_X(a, x) < \delta, d_X(a, y) < \delta,\ \text{ we have } d_Y(f(x), f(y)) < \epsilon$$

If the function is continuous at all points $x \in X$, the function is continuous.

It is easy to verify the continuous mapping theorem of metric space.
Proposition: Continuous mapping of limit

Given two metric spaces (X, d_X) and (Y, d_Y) and a continuous function $f : X \mapsto Y$. If the sequence $\{a_n \in X\}_{n=1}^{\infty}$ converges to $L \in X$, then

$$\lim_{n\to\infty} f(a_n) = f(L)$$

The following is a stronger type of continuous.

Definition: Uniform continuity

Given two metric spaces (X, d_X) and (Y, d_Y) and a function $f : X \mapsto Y$. The function f is uniform continuous if $\forall\epsilon > 0$, $\exists\delta > 0$ such that

$$\forall a, b \in X \text{ such } d_X(a, b) < \delta, \text{ we have } d_Y(f(a), f(b)) < \epsilon$$

The following theorem about open set and continuous function will be used in later chapters.

Theorem: about continuity and open set

Given two metric spaces (X, d_X) and (Y, d_Y) and a continuous function $f : X \mapsto Y$. Then if $A \subset Y$ is an open set, the preimage of A, defined by

$$f^{-1}(A) = \{x \in X : f(x) \in A\}$$

is also an open set.

The inverse also holds. If $f^{-1}(A)$ is open for any open set A in Y, then f is continuous. Proof:

For any $a \in f^{-1}(A)$, since A is open and $f(a) \in A$, there exists $\epsilon > 0$ such that $\forall y \in Y$ satisfying $d_Y(y, f(a)) < \epsilon$, we have $y \in A$.

Further, since f is continuous, there exists $\delta > 0$ such that $\forall x \in X$ satisfying $d_X(x, a) < \delta$, we have $d_Y(f(x), f(a)) < \epsilon$ and thus, $f(x) \in A$. Then $x \in f^{-1}(A)$ follows, which gives $f^{-1}(A)$ is an open set.

The proof of the inverse is similar.

1.3.2 The topology space of extended real number set

Here are some basic concepts in Topology.

Definition: about Topology

Let S be a set and $\mathcal{P}(S)$ be its power set.

1. Topology $\mathcal{T}$ is a subset of $\mathcal{P}(S)$ that satisfies the following:

(a) $\emptyset, S \in \mathcal{T}$

(b) Close under arbitrary union: Let T be an index set, $\{A_t \in \mathcal{T}\}_{t\in T}$ gives $\bigcup_{t\in T} A_t \in \mathcal{T}$

(c) Close under finite intersection: $\{A_i \in \mathcal{T}\}_{i=1}^n$ gives $\bigcap_{i=1}^n A_i \in \mathcal{T}$

2. Open sets are defined as the elements of $\mathcal{T}$. $\mathcal{T}$ is the set of all open sets.

3. A subset $A \subset S$ is close if $A^c = S \setminus A$ is open.

4. S together with its $\mathcal{T}$ constitutes the topology space $(S, \mathcal{T})$

Example: metric space is a kind of topology space
We have defined open set of metric space. Take the topology to be the set of all open sets. We can check that it satisfies all the three requirements.

Proposition: Equivalent definition of open sets in $\mathbb{R}$
Let $\mathcal{T}'$ be the set of open sets $\mathbb{R}$ defined in 1.3.1 as a metric space, which is a topology. We state a new definition of it and claim that they are equivalent.

$$\mathcal{T}_{\mathbb{R}} := \left\{ \bigcup_{i\in I} A(a_i, b_i) : a_i, b_i \in \mathbb{R}\ \forall i \in I, I \text{ is a index set} \right\}$$

where $A(a,b) := \{a < x < b : x \in \bar{\mathbb{R}}\}$ is an interval. Proof:
We first prove $\mathcal{T}_{\mathbb{R}} \subset \mathcal{T}'$. Since any interval $A(a,b)$ is open by the old definition, $A(a,b) \in \mathcal{T}'\ \forall a, b \in \mathbb{R}$. As a topology, $\mathcal{T}'$ is closed under arbitrary union, thus $\mathcal{T}_{\mathbb{R}} \subset \mathcal{T}'$.
Next, we show $\mathcal{T}' \subset \mathcal{T}_{\mathbb{R}}$. For any $S \in \mathcal{T}'$, we have $\forall x \in S$, $\exists \epsilon_x$ such that $(x - \epsilon_x, x + \epsilon_x) \subset S$. Therefore,

$$S = \bigcup_{x\in S} A(x - \epsilon_x, x + \epsilon_x)$$

which shows $S \in \mathcal{T}_{\mathbb{R}}$.

Definition: Compact set
Given a topology space $(S, \mathcal{T})$, a subset $A \in \mathcal{P}(S)$ is defined to be compact if for any collection of open set that covers A, i.e.,

$$A \subset \bigcup_{i\in I} B_i, \text{ where } I \text{ is a index set, and } B_i \in \mathcal{T}\ \forall i \in I$$

there exists a finite sub-collection $\{B_i\}_{i\in J}$, where $J \subset I$ is a finite set, such that

$$A \subset \bigcup_{i\in J} B_i$$

Property: Compact set in Euclidean space

In Euclidean space $\mathbb{R}^k$, where $k \in \mathbb{N}^*$, a set $A \subset \mathbb{R}^k$ is compact if and only if it is closed and bounded.

Definition: Topology space of the extended real number set

In 1.1.1, we have specified the order of elements in $\bar{\mathbb{R}}$. We define the topology of $\bar{\mathbb{R}}$ based on it, known as the order topology. For any $a, b \in \bar{\mathbb{R}}$, let $A^-(a) := \{x < a : x \in \bar{\mathbb{R}}\}$, $A(a,b) := \{a < x < b : x \in \bar{\mathbb{R}}\}$ and $A^+(b) := \{x > a : x \in \bar{\mathbb{R}}\}$ be three types of set. Then the topology is the set of all these sets and their arbitrary union:

$$\mathcal{T} := \left\{ \bigcup_{i \in I} A^-(a_i) \cup \bigcup_{j \in J} A(b_j, c_j) \cup \bigcup_{k \in K} A^+(d_k) : a_i, b_j, c_j, d_k \in \bar{\mathbb{R}} \ \forall i, j, k \text{ in their index sets } I, J, K \right\}$$

Then $\mathcal{T}$ contains $\emptyset$ and $\bar{\mathbb{R}}$ obviously. And it is closed under arbitrary union from the definition. We can show that it is also closed under finite intersection because the finite intersection of A^-, A, A^+ is a set of one of these three types.

Thus we define the topology space $(\bar{\mathbb{R}}, \mathcal{T})$ properly.

Proposition: comparing the topology between $\mathbb{R}$ and $\bar{\mathbb{R}}$

Given $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ and $(\bar{\mathbb{R}}, \mathcal{T})$ defined as above, we have

$$\mathcal{T} = \left\{ S \cup A^-(a) \cup A^+(b) : S \in \mathcal{T}_{\mathbb{R}},\ a, b \in \bar{\mathbb{R}} \right\}$$

It follows directly from the definition.

Definition: Convergence

Let $\{A_n \in S\}_{n=1}^{\infty}$ be a sequence. A_n converges to point L if for any open set $B \in \mathcal{T}$ such that $L \in B$, there exists $N \in \mathbb{N}^*$ such that $A_n \in B \ \forall n > N$.

We can check that the old definition of limit of real sequence in 1.1.2, either finite (with $\mathcal{T}_{\mathbb{R}}$) or infinite (with the above $\mathcal{T}$) case, is equivalent to this new definition.

Definition: Continuity

Let $f : X \mapsto Y$ be a map between two topology spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$. f is defined to be continuous if for any open set $B \in \mathcal{T}_Y$, the preimage $f^{-1}(B) \in \mathcal{T}_X$

We have also shown in 1.3.1 that the continuity of function between metric spaces matches the new definition in Topology.

Proposition: About continuous $\bar{\mathbb{R}}$ function

We know that the definition of continuous real value function is equivalent to the new definition

in Topology, given the Topology $\mathcal{T}' = \mathcal{T}_{\mathbb{R}}$. The following shows how can we determine a function $f : \bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ is continuous.
If

1. $\forall a \in \mathbb{R}$, $\lim_{x\to a} f(x) = f(a)$
2. $\lim_{x\to +\infty} f(x) = f(+\infty)$
3. $\lim_{x\to -\infty} f(x) = f(-\infty)$

then f is continuous from the sense of Topology, i.e. the preimage of any open set is open. Note that all the three limits can be infinite.
Proof:
Let $\mathcal{T}$ be the topology of $\bar{\mathbb{R}}$. For any $Y \in \mathcal{T}$, we have shown that we can express Y as $S_y \cup A^-(a_y) \cup A^+(b_y)$, where $S_y \in \mathcal{T}_{\mathbb{R}},\ a_y, b_y \in \bar{\mathbb{R}}$. Then we have

$$f^{-1}(Y) = f^{-1}(S_y) \cup f^{-1}(A^-(a_y)) \cup f^{-1}(A^+(b_y))$$

To prove that $f^-(Y)$ is open, it suffices to show that each of $f^{-1}(S_y), f^{-1}(A^-(a_y)), f^{-1}(A^+(b_y))$ can also be expressed as $S \cup A^-(a) \cup A^+(b)$ for some $S \in \mathcal{T}_{\mathbb{R}},\ a, b \in \bar{\mathbb{R}}$. Then $f^{-1}(Y)$, as the union of three open sets, is also open.
We first prove $f^{-1}(S_y)$. If it does not contain $+\infty$ or $-\infty$, it becomes the case of real value function. We have shown that $f^{-1}(S_y) \in \mathcal{T}_{\mathbb{R}}$. If $+\infty \in f^{-1}(S_y)$, from condition 2, we can always find a large enough M such that $f(x) \in S_y\ \forall x > M$. Therefore, $A^+(M) \in f^{-1}(S_y)$. Similarly, if $-\infty \in f^{-1}(S_y)$, we have a small enough m and $A^-(m)$. Then

$$f^{-1}(S_y) = f^{-1}(S_y) \setminus \{+\infty, -\infty\} \cup A^+(M) \cup A^-(m)$$

where $f^{-1}(S_y) \setminus \{+\infty, -\infty\} \in \mathcal{T}_{\mathbb{R}}$ and $M, m \in \bar{\mathbb{R}}$.
Next, we prove $f^{-1}(A^-(a_y))$. We first have $f^{-1}(A^-(a_y)) \setminus \{+\infty, -\infty\} \in \mathcal{T}_{\mathbb{R}}$. Because for any $c \in f^{-1}(A^-(a_y))$, there exist $\epsilon > 0$ such that $f(x) \in S_y\ \forall c - \epsilon < x < x + \epsilon$, even when $f(c) = -\infty$. Therefore, $(c-\epsilon, c+\epsilon) \in f^{-1}(A^-(a_y))$, and $f^{-1}(A^-(a_y))$ is the union of these intervals, thus belongs to $\mathcal{T}_{\mathbb{R}}$. For the case when $f^{-1}(A^-(a_y))$ contains $+\infty$ or $-\infty$, similar to above, we can always find $A^+(M)$ and $A^-(m)$.
The part for $f^{-1}(A^+(b_y))$ is almost the same. Details omitted.

Theorem: continuous mapping of convergence
Given two topology spaces $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ and a continuous function $f : X \mapsto Y$, if a sequence

$\{A_n \in X\}_{n=1}^{\infty}$ converges to $L \in X$, then $\{f(A_n)\}_{n=1}^{\infty}$ converges to $f(L)$.

As a special case, let $f : \bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ be a continuous function and $\{a_n \in \mathbb{R}\}_{n=1}^{\infty}$ whose limit $L := \lim_{n\to\infty} a_n$ exists in $\bar{\mathbb{R}}$. Then

$$\lim_{n\to\infty} f(a_n) = f(L) \text{ also exists}$$

Proof:

Since f is continuous, for any $Y \in \mathcal{T}_Y$ with $f(L) \in Y$, $f^{-1}(Y) \in \mathcal{T}_X$. Then there exists $N \in \mathbb{N}^*$ such that $\forall n > N$, $a_n \in f^{-1}(Y)$ and thus $f(a_n) \in Y$. It gives that $\{f(A_n)\}_{n=1}^{\infty}$ converges to $f(L)$.

Theorem: Image of a continuous function from a compact set is compact

Given two topology spaces $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ and a continuous function $f : X \mapsto Y$, for any compact set $K \in \mathcal{P}(X)$, the image $F(K) := \{F(x) : x \in K\}$ is also a compact set.

Proof:

Take any open cover of $F(K)$, i.e.

$$\{A_i \in \mathcal{T}_Y\}_{i\in I} \text{ such that } F(K) \subset \bigcup_{i\in I} A_i$$

Since f is continuous, we have $f^{-1}(A_i) \in \mathcal{T}_X$. Thus,

$$K \subset f^{-1}[f(K)] \subset f^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f^{-1}(A_i)$$

to be an open cover of K. Since K is compact, there exists a finite set $J \subset I$ such that $K \subset \bigcup_{j\in J} f^{-1}(A_j)$, which gives

$$F\left[\bigcup_{j\in J} f^{-1}(A_j)\right] = \bigcup_{j\in J} F\left[f^{-1}(A_j)\right] = \bigcup_{j\in J} A_j \supset F(K)$$

Therefore, $F(K)$ is also compact.

Corollary: inverse of continuous function between compact Euclidean space is continuous

If a function $f : X \mapsto Y$ is bijective and continuous and $X, Y \subset \mathbb{R}^k$ ($k \in \mathbb{N}^*$) is two compact sets, its inverse $f^{-1} : Y \mapsto X$ is continuous.

Further, it suffices to let f be an injective and continuous real-vector-valued function on X and let $Y = f(X)$ be the image, which must be compact.

Proof:

In the context of Euclidean space, compact set X is closed and bounded. For any open set $A \subset X$, $A \cup (\mathbb{R}^k \setminus X)$ is open and $X \setminus A = \mathbb{R}^k \setminus [A \cup (\mathbb{R}^k \setminus X)]$ is closed. It is further compact given the boundedness of X. Since f is bijection and so as f^{-1}, $(f^{-1})^{-1}(A) = Y \setminus [(f^{-1})^{-1}(X \setminus A)] = Y \setminus f(X \setminus A)$.

By our last theorem, $f(X \setminus A)$ is compact and therefore $Y \setminus f(X \setminus A)$ is open. We have proved the continuity of f^{-1} since its preimage of any open set is open.

Chapter 2: Measure space and Probability space

2.1 Measure and Sigma algebra

2.1.1 Measure

The basic idea of measure comes from geometrical "volume", such as length in $\mathbb{R}$, area in $\mathbb{R}^2$, volume in $\mathbb{R}^3$ and even higher dimensions. In the physical world, the measure of a geometry should satisfy several requirements: 1) non-negativity; 2) Additivity of non-overlapping geometries; 3) Irrelevance of the position of the geometry.
This measure satisfying all these requirements is called the Lebesgue measure, which will be discussed later. More general cases of measure include other magnitudes, as well as probability. It is a function mapping from the subsets of a set to some positive real number or positive infinity, possessing the properties of non-negativity and countable additivity. The formal definition is given:

Definition: Measure
Let S be a set. We define a measure μ of some subsets $A \subset S$ which satisfies

1. $\mu(A) \in [0, \infty]$
2. $\mu(\emptyset) = 0$
3. Let $\{A_n \in \mathcal{S}\}_{n=1}^{\infty}$ be a sequence of mutually disjoint subsets of S,

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

It is easy to show that if $A \subset B$, $\mu(A) \leq \mu(B)$, given that the measure of A, B and $B \setminus A$ is well-defined.

However, some measures, such as the Lebesgue measure, cannot be defined on all subsets of S, which means these sets are not measurable. A counter-example can be constructed based on the Axiom of Choice.

Example: non-Lebesgue-measurable sets
Let $S = [0,1)$. By the basic significance of the Lebesgue measure, we have

$$\forall a, b \in \mathbb{R},\ \mu\left([a,b)\right) = b - a$$

It should further possess the proposition of irrelevance of the position,

$$\forall a, b, x \in \mathbb{R},\ \mu([a,b)) = \mu([a+x, b+x))$$

More generally,

$$\forall x \in \mathbb{R},\ \mu(A) \equiv \mu\left(\{a + x : a \in A\}\right)$$

Now the construction begins. $\forall x \in [0,1]$, let $S_x := \{x + q : q \in S \cap \mathbb{Q}\}$. Then S can be partitioned into an uncountable sequence of mutually disjoint subsets $\{S_i : i \in I\}$, where $I \subset [0,1]$ is the index set. This means the following holds:

1. $\bigcup_{i \in I} S_i = S$.
2. $\forall i, j \in I$ and $i \neq j$, $S_i \cap S_j = \emptyset$

Next, we choose an arbitrary value of every S_i to form a set
$A := \{\text{an arbitrary entry } a_i \in S_i : i \in I\}$. This is possible under the Axiom of Choice. We argue that A is not measurable.
Since ration number is countable, let $\{q_j\}_{j=1}^{\infty}$ be the enumeration of $S \cap \mathbb{Q}$. We have

$$S = [0,1) \subset \bigcup_{j=1}^{\infty} \{a + q_j : a \in A\} \subset [-1, 2)$$

The first subsetting is given by the fact that $\forall x \in S$, $\exists i \in I$ such $x \in S_i$ and thus, $\exists q_j = x - a_i$, such $x \in \{a + q_j : a \in A\}$. The second subsetting is trivial. Then,

$$\mu\left(\bigcup_{j=1}^{\infty} \{a + q_j : a \in A\}\right) = \mu\left(\bigcup_{j=1}^{\infty} A\right) = \sum_{i=0}^{\infty} \mu(A)$$

$$1 = \mu([0,1)) \leq \sum_{i=0}^{\infty} \mu(A) \leq \mu[-1, 2) = 3$$

Since $\sum_{i=0}^{\infty} \mu(A)$ is bounded by 3, $\mu(A) = 0$ and then $\sum_{i=0}^{\infty} \mu(A)$=0. The contradiction appears. Therefore, the set A is not measurable. That is to say that the domain of such basic measure μ is not $\mathcal{P}(S)$.

However, given a measure, it is sufficient to only analyse the measurable subset. This comes to the discussion of the proper domain of the measure, i.e. sigma algebra.

2.1.2 Sigma algebra and measure space

Definition: Sigma algebra
A non-empty subset $\mathcal{B} \subset \mathcal{P}(S)$ is called an σ-algebra (aka. σ-field) of S if it is closed under complement, countable union and countable intersection.

From the definition, we can derive that $\emptyset, S \in \mathcal{B}$, because for any $A \in \mathcal{B}$, $A^c \in \mathcal{B}$ and so as $A \cup A^c = \emptyset$ and $A \cap A^c = S$. In fact, a basic set of conditions of $\mathcal{B} \subset \mathcal{P}(S)$ being a σ-algebra is

1. $S \in \mathcal{B}$;
2. If $A \in \mathcal{B}$, $A^c \in \mathcal{B}$;
3. If $A_i \in \mathcal{B}$ $\forall i \in \mathbb{N}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Then the countable union holds automatically:

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c$$

Then we can define the measurable space and measure space.
Definition: Measurable space
A set S with its σ-algebra $\mathcal{B} \subset \mathcal{P}(S)$ is a measurable space $(S, \mathcal{B})$.

Definition: Measure space
If a measure μ is defined on a measurable space $(S, \mathcal{B})$, the triple $(S, \mathcal{B}, \mu)$ is a measure space.
Unless otherwise stated, we say a subset $A \subset S$ is measurable if $A \in \mathcal{B}$, because its measure $\mu(A)$ is properly defined.

Properties of measure and measure space $(S, \mathcal{B}, \mu)$

1. Sub-additivity: Let $\{A_n\}_{n=1}^{\infty}$ be a sequence set that $A_n \in \mathcal{B}\ \forall n \in \mathbb{N}^*$, we have
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$
2. Continuity from below: If the $\{A_n\}_{n=1}^{\infty}$ is monotonic increasing, we have
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$
3. Continuity from above: If the $\{A_n\}_{n=1}^{\infty}$ is monotonic decreasing, we have
$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

To prove 1 and 2, we can construct mutually disjoined set sequences by $B_i = A_i \backslash (A_1 \cup A_2 \cup \cdots \cup A_{i-1})$ and apply the countable additive rule. Noted that in monotonic increasing case, $B_i = A_i \backslash A_{i-1}\ \forall i > 1$. Details omitted.
To prove 3, construct the sequence $B_i = A_1 \setminus A_i$, which is monotonic increasing, then
$$\bigcup_{n=1}^{\infty} B_n = A_1 \setminus \left(\bigcap_{n=1}^{\infty} A_n\right)$$
Using property 2,
$$\begin{aligned}\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1) - \mu\left(\bigcup_{n=1}^{\infty} B_n\right) &= \mu(A_1) - \lim_{n\to\infty} \mu(B_n) \\ &= \mu(A_1) - \lim_{n\to\infty} \mu(A_1) - \mu(A_n) \\ &= \lim_{n\to\infty} \mu(A_n)\end{aligned}$$
Thus prove the result.

Examples of measure space

1. Zeros measure $(\Omega, \mathcal{P}(\Omega), \mu)$, where $\mu(A) = 0$ for all $A \in \mathcal{P}(\Omega)$.
2. Dirac measure $(\Omega, \mathcal{B}, \mathbf{1}_A)$, where $A \subset \Omega$ and $\mathbf{1}_A$ is the indicator function.
3. Counting measure $(\Omega, \mathcal{P}(\Omega), c)$, where Ω is a countable set, and $c(A)$ is the number of elements in set A, or ∞ if A is an infinite set.

2.1.3 Generating sigma algebra

Given a measure, we have a further problem of finding its σ-algebra. For some measures such as the measures we mentioned above, we can directly choose $\mathcal{P}(\Omega)$. But for other measures such as the Lebesgue measure, we have shown that $\mathcal{P}(\mathbb{R})$ cannot be its σ-algebra.

Here is a more general approach to constructing a σ-algebra. We first choose some subsets on which we can easily define the measure on it, for example, for the Lebesgue measure, we choose all the interval $(a, b]$ whose measure is $b - a$. Then, we find an appropriate σ-algebra containing the subsets, so that we can define the measure on this σ-algebra. In other words, this σ-algebra should not include the non-measurable subset. Therefore, we should try to find the σ-algebra as small as possible. The following shows that we can define the minimal σ-algebra containing the subsets.

Definition: σ-algebra generated by a subset

Let $\mathcal{A} \in \mathcal{P}(S)$. Denote $\sigma(\mathcal{A})$ be the σ-algebra generated by $\mathcal{A}$ which satisfy

1. $\mathcal{A} \subset \sigma(\mathcal{A})$
2. If $\mathcal{B} \subset \mathcal{P}(S)$ is another σ-algebra that contains $\mathcal{A}$, $\sigma(\mathcal{A}) \subset \mathcal{B}$

This is to define a minimal σ-algebra containing $\mathcal{A}$. Since $\mathcal{P}(S)$ is a σ-algebra, the minimal always exists. It is also unique because the intersection of two σ-algebras is a σ-algebra (proved by definition), then the intersection of all σ-algebras containing $\mathcal{A}$ is the unique minimal.

However, is the measure well-defined on all elements in $\sigma(\mathcal{A}) \setminus \mathcal{A}$? We will discuss some of its sufficient conditions and show how we can extend the "measure" on $\mathcal{A}$ to a measure on $\sigma(\mathcal{A})$ in 2.4.

An important construction of σ-algebra is the Borel σ-algebra that generates $\mathcal{B}$ from the open sets of a Topology.

Definition: Borel σ-algebra

Given a topology space $(S, \mathcal{T})$, the Borel σ-algebra is the σ-algebra generated by its topology (a set of all open sets), denoted by

$$\mathcal{B}(S) := \sigma(\mathcal{T})$$

The followings are examples of different topology space.

Definition: Borel σ-algebra of $\mathbb{R}$

Let $S := \mathbb{R}$ and $\mathcal{A} := \{\text{all open sets of } \mathbb{R}\}$. Define

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{A})$$

to be the Borel σ-algebra of $\mathbb{R}$.

In fact, such σ-algebra exists and the construction result is the same with the following $\mathcal{A}$.

- $\mathcal{A} := \{(a, b] : -\infty \leq a \leq b < \infty\}$
- $\mathcal{A} := \{[a, b] : -\infty < a \leq b < \infty\}$
- $\mathcal{A} := \{(-\infty, b] : -\infty < b < \infty\}$

They generate the same result because they can be transformed to each other using some countable operations.
For example, we have proved in 1.2.1 that any open set of $\mathbb{R}$ (can be generated to $\mathbb{R}^n$) is a countable union of open intervals (or open balls in $\mathbb{R}^n$). We also have the fact that

$$[a, b] = \bigcap_{n=1}^{\infty} [a, b + \frac{1}{n}) \text{ or } [\text{a, b}) = \bigcup_{\text{n=1}}^{\infty} [\text{a, b} - \frac{1}{\text{n}}]$$

Definition: Borel σ-algebra of $[-\infty, \infty]$
We extend $\mathbb{R}$ to be $\bar{\mathbb{R}}$. From its topology space $(\bar{\mathbb{R}}, \mathcal{T})$ stated in 1.3.2, define the Borel σ-algebra of $\bar{\mathbb{R}}$ to be $\sigma(\mathcal{T})$.
Equivalently,

$$\mathcal{B}(\bar{\mathbb{R}}) = \{A \cup B : A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{P}(\{-\infty, \infty\})\}$$
$$\mathcal{B}(\bar{\mathbb{R}}) = \sigma\left(\{[-\infty, a] : -\infty \leq a \leq \infty\}\right)$$

Definition: Product σ-algebra
Given two measurable spaces $(S_1, \mathcal{B}_1)$ and $(S_2, \mathcal{B}_2)$, we define the product σ-algebra on the Cartesian product set $S_1 \times S_2$ (which is $\{(a, b) : a \in S_1, b \in S_2\}$), by

$$\mathcal{B} = \sigma\left(\{A \times B : A \in \mathcal{B}_1, B \in \mathcal{B}_2\}\right)$$

Definition: Borel subset of $\mathbb{R}^n$
$\mathbb{R}^n$ is also a Topology space with its topology $\mathcal{T}$. Apart from the normal definition, the following

shows a equivalent but more concrete approach to define $\mathcal{B}(\mathbb{R}^n)$ as a product σ-algebra.
Let $S = \mathbb{R}^n$, define the σ-algebra $\mathcal{B}(\mathbb{R}^n)$ by $\sigma(\{\text{all open sets of } \mathbb{R}^n\})$. We can make equivalent definition using product σ-algebra:

$$\mathcal{B}(\mathbb{R}^n) = \sigma\left(\{X_1 \times X_2 \times \cdots \times X_n : X_i \in \mathcal{B}(\mathbb{R})\ \forall i \in \mathbb{Z} \cap [1, n]\}\right)$$

$\mathcal{B}(\mathbb{R}^n)$ can also be generated by $\{[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] : -\infty < a_i \leq b_i < \infty\ \forall i \in \mathbb{Z} \cap [1, n]\}$.
(Proof not given)
Using the same logic, we also have $\mathcal{B}(\bar{\mathbb{R}}^n)$ as a product σ-algebra. Details omitted.

Example of the usage of Borel subset
We will prove in 2.4 that the Lebesgue measure is well-defined on $\mathcal{B}(\mathbb{R})$. So its measure space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where $\lambda(A)$ means the length of $A \in \mathcal{B}(\mathbb{R})$, for example: $\lambda((a, b)) = b - a$, $\lambda([a, b) \cup (c, d)) = b - a + d - c$ (Given that $-\infty < a < b < c < d < \infty$). The rigorous definition of λ is in 2.4.

Sometimes we only need to generate the σ-algebra within a certain interval $[L, R]$. For example, $\mathcal{B} := \sigma\left(\{(a, b) : L \leq a \leq b \leq R\}\right)$. We now show $\mathcal{B} = \{A \cap [L, R] : A \in \mathcal{B}(\mathbb{R})\}$. This is given from the following proposition, which states a more general case.

Proposition:
Let a measurable space $(S, \mathcal{B})$ and $S_0 \subset S$. Let $\mathcal{C} \in \mathcal{P}(S)$ such $\sigma(\mathcal{C})$ exist. Let $\mathcal{C}_0 := \{A \cap S_0 : A \in \mathcal{C}\}$. Then

$$\sigma(\mathcal{C}_0) = \{A \cap S_0 : A \in \sigma(\mathcal{C})\}$$

Proof:
We prove this proposition in three steps.

1. $\mathcal{B}_0 := \{A \cap S_0 : A \in \mathcal{B}\}$ is a σ-algebra of S_0. Putting $\mathcal{B} = \sigma(\mathcal{C})$, $\{A \cap S_0 : A \in \sigma(\mathcal{C}_0)\}$ is also a σ-algebra
2. $\mathcal{G} := \{A \subset S : A \cap S_0 \in \sigma(\mathcal{C}_0)\}$ is a σ-algebra of $\mathcal{B}$.
3. Then it follows $\sigma(\mathcal{C}_0) = \{A \cap S_0 : A \in \sigma(\mathcal{C})\}$

1 and 2 can be verified by the definition of σ-algebra.
Since $\mathcal{C} \subset \sigma(\mathcal{C}) \Rightarrow \mathcal{C}_0 \subset \{A \cap S_0 : A \in \sigma(\mathcal{C})\}$, and $\{A \cap S_0 : A \in \sigma(\mathcal{C})\}$ is a σ-algebra, it contains the minimal $\sigma(\mathcal{C}_0)$. To prove 3, we only need to show $\{A \cap S_0 : A \in \sigma(\mathcal{C})\} \subset \sigma(\mathcal{C}_0)$.
Further, $\mathcal{C} \subset \mathcal{G}$ because $\forall A \in \mathcal{C}$, $A \cap S_0 \in \mathcal{C}_0 \subset \sigma(\mathcal{C}_0)$. Since $\mathcal{G}$ is a σ-algebra by 2, $\sigma(\mathcal{C}) \subset \mathcal{G}$. It

means $\forall A \in \sigma(\mathcal{C})$, $A \in \mathcal{G}$ and thus, $A \cap S_0 \in \sigma(\mathcal{C}_0)$. This gives the desired result that $\{A \cap S_0 : A \in \sigma(\mathcal{C})\} \subset \sigma(\mathcal{C}_0)$.

2.2 Probability space

2.2.1 Probability space and distribution function

We can define probability space as a special case of measure space.
Definition: probability space
Define a triple $(\Omega, \mathcal{F}, P)$ to be the probability space, where Ω is the set of all possible outcomes, $\mathcal{F}$ is the σ-algebra of Ω, and P is a probability measure, which is a function $P : \mathcal{F} \mapsto \{A \cap [0, 1] : A \in \mathcal{B}(\mathbb{R})\}$ satisfying the following requirements, knows as Kolmogorov's axioms,

1. $\forall A \in \mathcal{F},\ P(A) \geq 0$
2. $P(\Omega) = 1$
3. Other requirements as a measure, i.e. $P(\emptyset) = 0$ and countable additivity

Every element in $\mathcal{F}$ is called an event.
For convenience, we may not write the event A as a set in Ω, but rather the description of this event. For example, $P(X > Y)$ and $P(\{X > Y\})$ both represent $P(\{\omega \in \Omega : X > Y\})$.

Normally, it is unnatural to to obtain the probability for some events $E \in \mathcal{F}$ directly, but we usually begin with obtaining the probability for some basic events, then we derive the probability for other events. This process is known as the construction of probability space, there are two cases:

1. Discrete case: $\Omega = \{E_1, E_2, \cdots\}$ is countable. We first assign $p_i > 0,\ \forall i \in \mathbb{Z}^*$ to be the probability of event E_i, which satisfies, e.g. $\sum_{i=0}^{\infty} P(E_i) = 1$. Then,

   $$\forall A \in \mathcal{F} = \mathcal{P}(\Omega), \text{ let } P(A) := \sum_{E_i \in A} P(E_i)$$

   It can be proved by the definition that such $(\Omega, \mathcal{F}, P)$ is a probability space.

2. Continuous case: Ω is uncountable. We normally consider the case that $\Omega = \mathbb{R}$. We use $\{(-\infty, x) : x \in \mathbb{R}\}$ as the set of basic events. The probability of these events is defined by the probability distribution function as follows. In other cases, we can map Ω to $\mathbb{R}$, which is the content of random variable and will be discussed in the next chapter.

Definition: Probability distribution function
Given a probability space $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P)$, define

$$F(x) := P([-\infty, x]) : \bar{\mathbb{R}} \mapsto [0, 1]$$

be the probability distribution function, or *pdf.* of this probability space.
Usually, it is enough to consider $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ as the probability space, in which case $F := P((-\infty, x])$ is defined on $\mathbb{R}$.

Proposition: Properties of *pdf*
Consider the more general F defined on $\bar{\mathbb{R}}$,

1. F is monotone non-decreasing, i.e. $\forall a, b \in \bar{\mathbb{R}}$ and $a \leq b,\ F(a) \leq F(b)$
2. F is right continuous on $\bar{\mathbb{R}} \setminus \{\infty\}$, i.e. $\forall a \in [-\infty, \infty),\ \lim_{x \to a^+} F(x) = F(a)$
3. $F(-\infty) \geq 0, F(+\infty) = 1$
4. The set $C := \{x \in \mathbb{R} : f \text{ is discontinuous at } x\}$ is at most countable

Property 1 comes from the fact that $[-\infty, a] \subset [-\infty, b]$.
Property 2 uses the sequential criteria of function limit. For all monotonic decreasing sequences $\{a_n \in \mathbb{R}\}_{n=1}^{\infty}$ converging to $a \in [-\infty, \infty)$, we have

$$\lim_{n \to \infty} F(a_n) = \lim_{n \to \infty} P([-\infty, a_n])$$

Since $\{[-\infty, a_n]\}_{n=0}^{\infty}$, is monotonic decreasing and converges to $[-\infty, a]$, we apply the proposition of continuity from above,

$$\lim_{n \to \infty} P((-\infty, a_n]) = P\left(\bigcap_{i=1}^{\infty} (-\infty, a_n]\right) = P((-\infty, a]) = F(a)$$

And it proves the statement.
Property 3 directly comes from the definition of probability space.

To prove 4, first, the discontinuity and monotonicity gives that $\forall a \in C,\ \lim_{x \mapsto a^-} F(x) < F(a)$. We can construct an injective function $h : C \mapsto \mathbb{Q}$ as follows

$$\forall a \in C,\ h(a) = \text{an arbitrary } y \in \{q \in \mathbb{Q} : \lim_{x \mapsto a^-} F(x) < q < F(a)\}$$

Then $h(a) > \lim_{x \to a^-} F(x) = \sup\{F(x) : x < a\}$. Therefore, $h(a) > h(b)$ whenever $a > b$, thus h is injective. Since h maps C to a subset of a countable set ($\mathbb{Q}$) injectively, C is also countable.

2.2.2 Construction of probability space

Conversely, we can define that a function $F : \bar{\mathbb{R}} \mapsto [0, 1]$ satisfying the above properties 1,2 and 3 (then 4 holds) is a *pdf*. The following part of this chapter shows that every such *pdf* corresponds to a unique probability space, and we can construct a probability space with a *pdf*, i.e. the following statement:

1. Uniqueness of the construction: Given two probability spaces $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_1)$ and $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_2)$ with their *pdf* F_1, F_2 respectively, if $\forall x \in \bar{\mathbb{R}},\ F_1(x) = F_2(x)$, then $\forall A \in \mathcal{B}(\bar{\mathbb{R}}),\ P_1(A) = P_2(A)$, which means the two probability spaces are the same.
2. Feasibility of the construction: Given a *pdf* $F(x) = P([-\infty, x])$, we can always construct a probability space $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), \tilde{P})$ such that $\forall x \in \bar{\mathbb{R}},\ \tilde{P}([-\infty, x]) = F(x)$.

Their proofs require more definitions and theorems.
Definition: π-system
Given a set S, a nonempty set of its subsets $\mathcal{C} \subset \mathcal{P}(S)$ is a π-system of S if it is closed under finite intersection:

- $\forall A, B \in \mathcal{C},\ A \cap B \in \mathcal{C}$.

It's easy to check that $\{[-\infty, x] : x \in \bar{\mathbb{R}}\}$ is a π-system of $\bar{\mathbb{R}}$.

Definition: λ-system (aka. Dynkin system)
Given a set S, a nonempty set of its subsets $\mathcal{L} \subset \mathcal{P}(S)$ is a λ-system of S if:

1. $S \in \mathcal{L}$

2. $\forall A \in \mathcal{L}$, $A^c = S \setminus A \in \mathcal{L}$

3. If there is a sequence of mutually disjoint sets $\{A_i \in \mathcal{L}\}_{i=1}^{\infty}$, and $A_i \cap A_j = \emptyset \ \forall i, j$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$.

Here are some property of a λ-system:

4. If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A = B \cap A^c = (B^c \cup A)^c \in \mathcal{L}$

5. If $\{B_i\}_{i=0}^{\infty}$ (set $B_0 = \emptyset$) is monotonic increasing sequence that $B_i \in \mathcal{L}$, let $A_i := B_i \setminus B_{i-1} \in \mathcal{L}$, then $\{A_i\}_{i=1}^{\infty}$ are mutually disjoint. Therefore,
$$\bigcup_{i=0}^{\infty} B_i = \bigcup_{i=0}^{\infty} A_i \in \mathcal{L}$$
.

In fact, we can define λ-system by the conditions 1, 4 and 5. We can prove 2 and 3 from the new definition. Details omitted.

Example of λ-system:
Let $S = \{a, b, c, d\}$, then $\mathcal{L} = \{\emptyset, \{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}, S\}$ is a λ-system of S but not a σ-algebra or π-system.

The following proposition shows the relationship between π-system, λ-system and σ-algebra.
Proposition:
If $\mathcal{B}$ is a π-system and a λ-system at the same time, then $\mathcal{B}$ is a σ-algebra.
We can proof this proposition by checking the conditions of σ-algebra. Actually, we only need to check the third condition (i.e. countable additivity).
Let $\{B_i\}_{i=1}^{\infty}$ be a sequence such $B_i \in \mathcal{B}$ for all possible i, then

1. $\forall n \in \mathbb{N}^*, \bigcap_{i=1}^{n} B_i \in \mathcal{B}$ by the definition of π-system

2. $\forall n \in \mathbb{N}^*,\ A_n := \bigcup_{i=1}^{n} B_i = \left(\bigcap_{i=1}^{n} B_i^c\right)^c \in \mathcal{B}$ by 1 and the definition of λ-system

3. $\{A_i\}_{i=1}^{\infty}$ is monotonic increasing, by 2 and the property 5 of λ-system, $\bigcup_{i=0}^{\infty} A_i \in \mathcal{L}$. Then $\bigcup_{i=0}^{\infty} B_i = \bigcup_{i=0}^{\infty} A_i \in \mathcal{L}$, which gives the countable additivity.

Dynkin's Theorem

Given a set S, $\mathcal{C}$ is one of its π-system and $\mathcal{L}$ is one of its λ-system, such that $\mathcal{C} \subset \mathcal{L}$ then

$$\sigma(\mathcal{C}) \subset \mathcal{L}$$

To prove this theorem, we first construct a minimal λ-system $\mathcal{L}(\mathcal{C})$ that contains $\mathcal{C}$, i.e.

- $\mathcal{C} \subset \mathcal{L}(\mathcal{C})$
- And for every λ-system $\mathcal{L}'$ that contain $\mathcal{C}$, we have $\mathcal{L}(\mathcal{C}) \subset \mathcal{L}'$. This gives that $\mathcal{L}(\mathcal{C}) \subset \mathcal{L}$.

We first claim that $\mathcal{L}(\mathcal{C})$ is a σ-algebra containing $\mathcal{C}$, then $\sigma(\mathcal{C}) \subset \mathcal{L}(\mathcal{C}) \subset \mathcal{L}$ holds.
By the above proposition, to prove this claim, it suffices to show that $\mathcal{L}(\mathcal{C})$ is a π-system.

Proof of Dynkin's Theorem
For any fix $A \in \mathcal{L}(\mathcal{C})$, we first define:

$$\mathcal{G}_A := \{B \in \mathcal{P}(S) : A \cap B \in \mathcal{L}(\mathcal{C})\}$$

We can prove that $\mathcal{G}_A$ is a λ-system by checking the condition:

1. $A \cap S \in \mathcal{L}(\mathcal{C}) \to S \in \mathcal{G}_A$
2. $\forall B$ such $A \cap B \in \mathcal{L}(\mathcal{C})$, $A \cap B^c = [A^c \cup (A \cap B)]^c \in \mathcal{L}(\mathcal{C}) \Rightarrow$ condition 2 of λ-system.
3. For any sequences of mutually disjoint subsets $\{B_i\}_{i=1}^{\infty}$ such $A \cap B_i \in \mathcal{L}(\mathcal{C})$ for all possible i, $\{A \cap B_i\}_{i=1}^{\infty}$ are also mutually disjoint, then $A \cap \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A \cap B_i \in \mathcal{L}(\mathcal{C}) \Rightarrow$ condition 3 of λ-system.

Next we claim that when $A \in \mathcal{C}$ (satisfying the definition $A \in \mathcal{L}(\mathcal{C})$), we have $\mathcal{C} \subset \mathcal{G}_A$:

$$\mathcal{C} \text{ is a } \pi\text{-system } \Rightarrow \forall B \in \mathcal{C},\ A \cap B \in \mathcal{C} \subset \mathcal{L}(\mathcal{C})$$

and then by the definition of $\mathcal{L}(\mathcal{C})$, we have $\mathcal{L}(\mathcal{C}) \subset \mathcal{G}_A$ under the condition of $A \in \mathcal{C}$, which is equivalent to:

$$\text{If } A \in \mathcal{C} \text{ and } B \in \mathcal{L}(\mathcal{C}), \text{ then } A \cap B \in \mathcal{L}(\mathcal{C})$$

and

$$\text{If } B \in \mathcal{C} \text{ and } A \in \mathcal{L}(\mathcal{C}), \text{ then } A \cap B \in \mathcal{L}(\mathcal{C})$$

Therefore, $\mathcal{C} \subset \mathcal{G}_A$ and thus, $\mathcal{L}(\mathcal{C}) \subset \mathcal{G}_A$ also holds without the condition $A \in \mathcal{C}$. This immediately gives

$$\text{If } A \in \mathcal{L}(\mathcal{C}) \text{ and } B \in \mathcal{L}(\mathcal{C}), \text{ then } A \cap B \in \mathcal{L}(\mathcal{C})$$

which finally proves that $\mathcal{L}(\mathcal{C})$ is a π-system.

Next we are able to prove the Uniqueness of the construction.

Proof: Uniqueness of the construction

Given two probability spaces $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_1)$ and $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_2)$ with their *pdf* F_1, F_2 respectively. Let

$$\mathcal{L} := \{A \in \mathcal{B}(\bar{\mathbb{R}}) : P_1(A) = P_2(A)\}$$
$$\mathcal{C} := \{[-\infty, x] : x \in \bar{\mathbb{R}}\}$$

We can check by definition that $\mathcal{L}$ is a λ-system, and $\mathcal{C}$ is a π-system. Since $F_1(x) = F_2(x) \Rightarrow \forall A \in \mathcal{C},\ P_1(A) = P_2(A) \Rightarrow \mathcal{C} \subset \mathcal{L}$, by Dynkin's Theorem, $\mathcal{B}(\bar{\mathbb{R}}) = \sigma(\mathcal{C}) \subset \mathcal{L}$ and hence $P_1(A) = P_2(A)\ \forall A \in \mathcal{B}(\bar{\mathbb{R}})$.

Now given the *pdf*, we start to construct the probability space. First, we can obtain the probability for the following set of interval

$$\mathcal{S} := \{[a, b] : a, b \in \bar{\mathbb{R}}, a \leq b\} \cup \emptyset$$

whose probability is

$$\tilde{P}\left([a, b]\right) = \begin{cases} F(b) - \lim_{x \to a^-} F(x) & , \quad -\infty < a \leq b \leq \infty \\ F(b) & , \quad a = -\infty, b \geq a \end{cases}$$

and $\tilde{P}(\emptyset) = 0$. We will show that there exists a unique extension measure P defined on $\sigma(\mathcal{S}) = \mathcal{B}(\bar{\mathbb{R}})$ that

$$\forall A \in \mathcal{S},\ P(A) = \tilde{P}(A)$$

Then, $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P)$ is the probability space we want.

We will later show that $\tilde{P}$ is a premeasure on the semiring $\mathcal{S}$, and the above is a special case of the Carathéodory Extension Theorem.

2.3 Carathéodory extension theorem

2.3.1 The statement of Carathéodory extension theorem

Definition: semiring of set
Let S be a set, $\mathcal{S} \subset \mathcal{P}(S)$ is a semiring of S if

1. $\emptyset \in \mathcal{S}$
2. $\forall A, B \in \mathcal{S},\ A \cap B \in \mathcal{S}$, which means $\mathcal{S}$ is a π-system.
3. $\forall A, B \in \mathcal{S}$, there exists a finite sequence of mutually disjoint sets $\{C_i \in \mathcal{S}\}_{i=1}^n$ such that $A \setminus B = \bigcup_{i=1}^n C_i$

We can check that $\mathcal{S} := \{[a, b] : a, b \in \bar{\mathbb{R}}, a \leq b\} \cup \emptyset$ is a semiring.

Definition: premeasure
Let S be a set and $\mathcal{S}$ is a semiring of S, the function $\tilde{\mu} : \mathcal{S} \mapsto [0, \infty]$ is a premeasure if

1. $\tilde{\mu}(\emptyset) = 0$
2. For any sequence of mutually disjoint sets $\{A_n \in \mathcal{S}\}_{n=1}^{\infty}$,

$$\text{if } \bigcup_{i=1}^{\infty} A_i \in \mathcal{S}, \text{ then } \tilde{\mu}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \tilde{\mu}(A_i)$$

Premeasure can be regarded as a measure defined on a semiring. We can also define a premeasure on other structures such as algebra and ring (see below) in the same way.
We can check that some properties of measure also hold for premeasure, such as countable subadditivity. We can check that $\tilde{P}$ is a premeasure on the semiring.

Definition: σ-finite
A premeasure measure $\tilde{\mu}$ defined on $\mathcal{S}$, is σ-finite if there exists a countable sequence $\{A_i \in \mathcal{S}\}_{i=1}^{\infty}$ such that

$$\tilde{\mu}(A_i) < \infty \ \forall i \in \mathbb{N}^* \text{ and } \bigcup_{i=1}^{\infty} A_i = S$$

We can check that $\tilde{P}$ is σ-finite, since

$$\mathbb{R} = \bigcup_{i \in \mathbb{Z}} (i, i+1] \text{ and } \tilde{P}((i, i+1]) < 1$$

This concept also applies to measure.

Carathéodory extension theorem

Given S, any premeasure $\tilde{\mu}$ defined on a semiring $\mathcal{S} \subset \mathcal{P}(S)$ can be extended to the measure μ on $\sigma(\mathcal{S})$ such that

$$\forall A \in \mathcal{S},\ \mu(A) = \tilde{\mu}(A)$$

Further, if $\tilde{\mu}$ is σ-finite, the extended measure μ is unique.
This theorem states the feasibility of constructing a probability space from a *pdf*.

The following part of this chapter proves this theorem. We first extend the premeasure to ring, and then to σ-algebra.

2.3.2 The first extension

Definition: Ring of Let S be a set, $\mathcal{R} \subset \mathcal{P}(S)$ is a ring of S if

1. $\emptyset \in \mathcal{R}$
2. $\forall A, B \in \mathcal{R},\ A \cup B \in \mathcal{R}$.
3. $\forall A, B \in \mathcal{R},\ A \setminus B \in \mathcal{R}$

The condition of a ring is stronger than a semiring. We can check that a ring is also a semiring.
We can also check by definition that the set of all finite intersections of a semiring $\mathcal{S}$

$$\mathcal{R}(\mathcal{S}) := \left\{ \bigcup_{i=1}^{n} A_i : A_i \in \mathcal{S} \text{ and are mutually disjoint} \right\}$$

is a ring of S.

Lemma: Extension to ring

Given a set S, any premeasure $\tilde{\mu}$ defined on a semiring $\mathcal{S} \subset \mathcal{P}(S)$ can be uniquely extended to another premeasure $\bar{\mu}$ on $\mathcal{R}(\mathcal{S})$ by

$$\forall A \in \mathcal{R}(\mathcal{S}),\ \bar{\mu}(A) := \sum_{i=1}^{n} \tilde{\mu}(B_i) \text{ where } \bigcup_{i=1}^{n} B_i = A, B_i \in \mathcal{S} \text{ and are mutually disjoint}$$

To avoid ambiguity, we first need to verify that if a set has two representation $A = \bigcup_{i=1}^{n} B_i = \bigcup_{j=1}^{m} C_j$, where $B_i \in \mathcal{S},\ C_j \in \mathcal{S}$ for all possible i, j, whether

$$\sum_{i=1}^{n} \tilde{\mu}(B_i) = \sum_{j=1}^{m} \tilde{\mu}(B_j)$$

This is given by

$$\sum_{i=1}^{n} \tilde{\mu}(B_i) = \sum_{i=1}^{n} \tilde{\mu}(B_i \cap A) = \sum_{i=1}^{n} \tilde{\mu}\left(\bigcup_{j=1}^{m} B_i \cap C_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \tilde{\mu}\left(B_i \cap C_j\right)$$

$$\text{and similarly, } \sum_{j=1}^{m} \tilde{\mu}(C_j) = \sum_{j=1}^{m}\sum_{i=1}^{n} \tilde{\mu}\left(C_j \cap B_i\right)$$

To prove that $\bar{\mu}$ is also a premeasure, we can check the condition of countable additivity.
Let $\{A_i \in \mathcal{R}(\mathcal{S})\}_{i=1}^{\infty}$ be the sequence of mutually disjoint sets such $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{R}(\mathcal{S})$. Suppose A is the intersection of mutually disjoint $\{B_k \in \mathcal{R}\}_{k=1}^{n}$ and A_i is the intersection of mutually disjoint $\{C_{i,j} \in \mathcal{R}\}_{j=1}^{n_i}$ for all possible i, then

$$\bar{\mu}(A) = \sum_{k=1}^{n} \tilde{\mu}(B_k) = \sum_{k=1}^{n} \tilde{\mu}\left(B_k \cap A\right) = \sum_{k=1}^{n} \tilde{\mu}\left(\bigcup_{i=1}^{\infty}\bigcup_{j=1}^{n_i} B_k \cap C_{i,j}\right)$$

By the countable additivity of $\tilde{\mu}$,

$$\begin{aligned}\sum_{k=1}^{n} \tilde{\mu}\left(\bigcup_{i=1}^{\infty}\bigcup_{j=1}^{n_i} B_k \cap C_{i,j}\right) = \sum_{k=1}^{n}\sum_{i=1}^{\infty}\sum_{j=1}^{n_i} \tilde{\mu}(B_k \cap C_{i,j}) &= \sum_{i=1}^{\infty}\sum_{j=1}^{n_i} \tilde{\mu}\left(\bigcup_{k=1}^{n} B_k \cap C_{i,j}\right) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{n_i} \tilde{\mu}(C_{i,j}) = \sum_{i=1}^{\infty} \bar{\mu}(A_i)\end{aligned}$$

which gives the desired result $\bar{\mu}(A) = \sum_{i=1}^{\infty} \bar{\mu}(A_i)$

Proposition: preservation of σ-finite
If $\tilde{\mu}$ is σ-finite on the semiring $\mathcal{R}$, $\bar{\mu}$ is also σ-finite on the ring $\mathcal{R}(\mathcal{S})$.
Since $\mathcal{S} \subset \mathcal{R}(\mathcal{S})$ and $\bar{\mu}(A) = \tilde{\mu}(A)\ \forall A \in \mathcal{R}$, S can still be the intersection of countable subset with finite premeasure in $\mathcal{S}$.

2.3.3 The second extension

Next, we need to extend the premeasure on a ring to a measure. It utilizes the concept of outer measure.

Definition: Outer measure

Let S be a set. $\mathcal{A} \subset \mathcal{P}(S)$ is a set of some subsets that $\emptyset \in \mathcal{A}$ on which a function $\bar{\mu} : \mathcal{A} \mapsto [0, \infty]$ is defined. We define $\mu^* : \mathcal{P}(S) \mapsto [0, \infty]$ as the outer measure associated to $\mathcal{A}$:

$$\mu^*(A) := \inf \left\{ \sum_{i=1}^{\infty} \bar{\mu}(C_i) : A \subset \bigcup_{i=1}^{\infty} C_i \text{ and } C_i \in \mathcal{A}\ \forall i \in \mathbb{N}^* \right\}$$

Noted that outer measure is not a measure

Properties of outer measure

1. $\mu^*(\emptyset) = 0$
2. If $A, B \in \mathcal{P}(S)$ and $A \subset B$, $\mu^*(A) \leq \mu^*(B)$
3. Countable sub-additivity: $\forall \{A_i \in \mathcal{P}(S)\}_{n=1}^{\infty}$, we have
$$\mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$$

1 follows from the definition.

2 follows from the fact that when $A \subset B$,

$$\left\{ C := \bigcup_{i=1}^{\infty} C_i : A \subset C \text{ and } C_i \in \mathcal{A} \right\} \supset \left\{ C := \bigcup_{i=1}^{\infty} C_i : B \subset C \text{ and } C_i \in \mathcal{A} \right\}$$
$$\Rightarrow \left\{ \sum_{i=1}^{\infty} \bar{\mu}(C_i) : A \subset \bigcup_{i=1}^{\infty} C_i \text{ and } C_i \in \mathcal{A}\ \forall i \in \mathbb{N}^* \right\} \supset \left\{ \sum_{i=1}^{\infty} \bar{\mu}(C_i) : B \subset \bigcup_{i=1}^{\infty} C_i \text{ and } C_i \in \mathcal{A}\ \forall i \in \mathbb{N}^* \right\}$$

To prove 3, we first not that when the left-hand side goes to infinity, the inequality always holds. We only need to consider the finite case. We take an arbitrary small $\epsilon > 0$ and construct a sequence $\{\epsilon_n\}_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} \epsilon_n = \epsilon$. For example, let $\epsilon_n = \frac{\epsilon}{2^n}$. Then, given by the property of infimum, we have

$$\forall n \in \mathbb{N}^*,\ \exists \{C_{n,i} \in \mathcal{A}\}_{i=1}^{\infty} \text{ such that } A_n \subset \bigcup_{i=1}^{\infty} C_{n,i} \text{ and } \sum_{i=1}^{\infty} \bar{\mu}(C_{n,i}) - \epsilon_n < \mu^*(A_n) \leq \sum_{i=1}^{\infty} \bar{\mu}(C_{n,i})$$
$$\Rightarrow \bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n=1}^{\infty} \bigcup_{i=1}^{\infty} C_{n,i}$$
$$\Rightarrow \mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \bar{\mu}(C_{n,i}) < \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon_n = \epsilon + \sum_{n=1}^{\infty} \mu^*(A_n)$$

Take $\epsilon \to 0$, the result follows.

Definition: μ^*-measurable set

A subset $A \in \mathcal{P}(S)$ is μ^*-measurable if

$$\forall B \in \mathcal{P}(S),\ \mu^*(B) = \mu^*(A \cap B) + \mu^*(B \setminus A)$$

Lemma: Carathéodory lemma Let S be a set and μ^* is a outer measure associated to $\mathcal{A} \subset \mathcal{P}(S)$, then

1. $\mathcal{B} := \{A \subset \mathcal{P}(S) : A \text{ is } \mu^*\text{-measurable}\}$ is a σ-algebra
2. $\mu : \mathcal{B} \mapsto [0, \infty]$, defined by $\mu(A) = \mu^*(A)\ \forall A \in \mathcal{B}$, is a measure.

To prove 1, we check the conditions of σ-algebra. The first two conditions arrives easily, we then prove the closure of countable intersection.

First, prove the closure of finite intersection. Let A, B to be μ^*-measurable sets, we show that $A \cup B$ is also μ^*-measurable, which means

$$\forall X \in \mathcal{P}(S),\ \mu^*(X) = \mu^*((A \cup B) \cap X) + \mu^*(X \setminus (A \cup B))$$

Given by the Countable sub-additivity property, it suffices to prove

$$\forall X \in \mathcal{P}(S),\ \mu^*(X) \geq \mu^*((A \cup B) \cap X) + \mu^*(X \setminus (A \cup B))$$

Since A, B are μ^*-measurable,

$$\begin{aligned}\forall X \in \mathcal{P}(S),\ \mu^*(X) = \mu^*(A \cap X) + \mu^*(X \setminus A) &= \mu^*(A \cap X) + \mu^*((X \setminus A) \cap B) + \mu^*((X \setminus A) \setminus B) \\ &= \mu^*(A \cap X) + \mu^*(X \cap A^c \cap B) + \mu^*(X \setminus (A \cup B)) \\ &\geq \mu^*((A \cap X) \cup (X \cap A^c \cap B)) + \mu^*(X \setminus (A \cup B)) \\ &= \mu^*((A \cup B) \cap X) + \mu^*(X \setminus (A \cup B))\end{aligned}$$

Next, we show $\forall A, B \in \mathcal{B}$ that $A \cap B = \emptyset$, we have $A \cup B \in \mathcal{B}$ by above and

$$\forall X \in \mathcal{P}(S),\ \mu^*(X \cap (A \cup B)) = \mu^*(X \cap (A \cup B) \cap A) + \mu^*(X \cap (A \cup B) \cap A^c) = \mu^*(X \cap A) + \mu^*(X \cap B)$$

Finally, given a arbitrary sequence $\{B_n \in \mathcal{B}\}_{n=1}^{\infty}$, let $B_0 = \emptyset$ and $\forall n \in \mathbb{N}^*\ A_n = B_n \setminus \bigcup_{i=0}^{n-1} B_i$. Then $A_n \in \mathcal{B}$ because we have obtained that $\mathcal{B}$ is closed under complement and finite intersection, and the sequence $\{A_n\}_{n=1}^{\infty}$ is mutually disjoint. Then to prove $\bigcup_{i=1}^{\infty} B_n = \bigcup_{i=1}^{\infty} A_n \in \mathcal{B}$, it suffices to show

$$\forall X \in \mathcal{P}(S),\ \mu^*(X) \geq \mu^*\left(X \cap \bigcup_{n=1}^{\infty} A_n\right) + \mu^*\left(X \setminus \bigcup_{n=1}^{\infty} A_n\right)$$

By induction of the above result $\mu^*(X \cap (A \cup B)) = \mu^*(X \cap A) + \mu^*(X \cap B)$, we obtain

$$\forall n \in \mathbb{N}^*,\ \mu^*\left(X \cap \bigcup_{i=1}^{n} A_n\right) = \sum_{i=1}^{n} \mu^*(X \cap A_i)$$

Therefore,

$$\begin{aligned}\forall n \in \mathbb{N}^* \text{ and } X \in \mathcal{P}(S),\ \mu^*(X) &= \sum_{i=1}^{n} \mu^*(X \cap A_i) + \mu^*\left(X \setminus \bigcup_{i=1}^{n} A_i\right) \\ &\geq \sum_{i=1}^{n} \mu^*(X \cap A_i) + \mu^*\left(X \setminus \bigcup_{i=1}^{\infty} A_i\right)\end{aligned}$$

The inequality is given by $(X \setminus \bigcup_{i=1}^{\infty} A_i) \subset (X \setminus \bigcup_{i=1}^{n} A_i)$, then take $n \to \infty$, we have

$$\forall X \in \mathcal{P}(S),\ \mu^*(X) \geq \sum_{i=1}^{\infty} \mu^*(X \cap A_i) + \mu^*\left(X \setminus \bigcup_{n=1}^{\infty} A_n\right) \geq \mu^*\left(X \cap \bigcup_{n=1}^{\infty} A_n\right) + \mu^*\left(X \setminus \bigcup_{n=1}^{\infty} A_n\right)$$

which proves the lemma 1.

To prove 2, we check $\mu(\emptyset) = 0$ (trivial) and the countable additivitiy. First, we can check that the finite additivity is satisfied.

$$\forall A, B \in \mathcal{B} \text{ and } A \cap B = \emptyset,\ A \cup B \in \mathcal{P}(S),\ \mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \setminus A) = \mu^*(A) + \mu^*(B)$$

The properties 2 and 3 of outer measure give that

$$\sum_{n=1}^{\infty} \mu^*(A_n) \geq \mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \mu^*\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu^*(A_i)$$

where $\{A_n \in \mathcal{B}\}_{n=1}^{\infty}$ are mutually disjoint, and we have proved that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$. Since $\lim_{n \to \infty} \sum_{i=1}^{n} \mu^*(A_i) = \sum_{n=1}^{\infty} \mu^*(A_n)$, by the Sandwich Theorem,

$$\sum_{n=1}^{\infty} \mu^*(A_n) = \mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \Rightarrow \sum_{n=1}^{\infty} \mu(A_n) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right)$$

The following lemma ends the proof of the Carathéodory Extension Theorem.

Lemma: Extension to σ-algebra

Given a set S, any premeasure $\bar{\mu}$ defined on a ring $\mathcal{R} \subset \mathcal{P}(S)$ can be extended to the measure μ on a σ-algebra $\mathcal{B}$ containing $\mathcal{R}$ (which must contain $\mathcal{B}(\mathcal{R})$, the minimal σ-algebra) by

$$\forall A \in \mathcal{B},\ \mu(A) := \mu^*(A) = \inf\left\{\sum_{i=1}^{\infty} \bar{\mu}(C_i) : A \subset \bigcup_{i=1}^{\infty} C_i \text{ and } C_i \in \mathcal{R}\ \forall i \in \mathbb{N}^*\right\}$$

where $\mu^*(A)$ is the outer measure associated to $\mathcal{R}$ and $\mathcal{B} = \{A \subset \mathcal{P}(S) : A \text{ is } \mu^*\text{-measurable}\}$.
The extension is unique if $\bar{\mu}$ is σ-finite.
The Carathéodory lemma has shown that $\mathcal{B}$ is a σ-algebra, and that μ is a measure. It suffices to prove these remaining statements:

1. $\forall A \in \mathcal{R},\ \mu(A) = \bar{\mu}(A)$
2. $\mathcal{R} \subset \mathcal{B}$, then we immediately have $\sigma(\mathcal{R}) \subset \mathcal{B}$
3. The uniqueness of the extension when $\bar{\mu}$ is σ-finite

To prove 1, we first have $\bar{\mu}(A) \in \{\sum_{i=1}^{\infty} \bar{\mu}(C_i) : A \subset \bigcup_{i=1}^{\infty} C_i \text{ and } C_i \in \mathcal{R}\ \forall i \in \mathbb{N}^*\}$, which gives $\mu(A) \leq \bar{\mu}(A)$. Hence, it suffices to prove $\forall A \in \mathcal{R},\ \mu(A) \geq \bar{\mu}(A)$.
$\forall \{C_i \in \mathcal{R}\}_{i=1}^{\infty}$ such $A \subset \bigcup_{i=1}^{\infty} C_i$, letting $C_0 = \emptyset$, then,

$$\sum_{i=1}^{\infty} \bar{\mu}(C_i) \geq \sum_{i=1}^{\infty} \bar{\mu}(C_i \cap A) \geq \sum_{i=1}^{\infty} \bar{\mu}\left(A \cap C_i \setminus \bigcup_{j=0}^{i-1} C_j\right) = \bar{\mu}\left(A \cap \bigcup_{i=1}^{\infty} C_i \setminus \bigcup_{j=0}^{i-1} C_j\right) = \bar{\mu}(A)$$

The first equal sign is given by the countable additivity of the premeasure. It follows that $\mu(A) = \inf\{\sum_{i=1}^{\infty} \bar{\mu}(C_i)\} \geq \bar{\mu}(A)$.
To prove 2, we show that $\forall A \in \mathcal{R}$, A is μ^*-measurable. It suffices to show that $\forall X \in \mathcal{P}(S),\ \mu^*(X) \geq \mu^*(X \cap A) + \mu^*(X \setminus A)$.
$\forall \{C_i \in \mathcal{R}\}_{i=1}^{\infty}$ such $A \subset \bigcup_{i=1}^{\infty} C_i$, we have

$$\sum_{i=1}^{\infty} \bar{\mu}(C_i) = \sum_{i=1}^{\infty} \bar{\mu}(C_i \cap A) + \mu(C_i \setminus A) \geq \bar{\mu}\left(A \cap \bigcup_{i=1}^{\infty} C_i\right) + \bar{\mu}\left(A^c \cap \bigcup_{i=1}^{\infty} C_i\right) \geq \bar{\mu}(A \cap X) + \bar{\mu}(A \setminus X)$$

The first "$\geq$" sign is given by the countable sub-additivity of a premeasure. It follows that $\mu^*(X) = \inf\{\sum_{i=1}^{\infty} \bar{\mu}(C_i)\} \geq \mu^*(X \cap A) + \mu^*(X \setminus A)$.
Finally, we prove the extension is unique if $\bar{\mu}$ is σ-additive
Suppose a measure μ' is another extension of $\bar{\mu}$, by 1 we have $\forall A \in \mathcal{R},\ \mu'(A) = \bar{\mu}(A) = \mu(A)$. We check for any $X \in \mathcal{B}$, whether $\mu'(X) = \mu(X)$. $\forall \{C_i \in \mathcal{R}\}_{i=1}^{\infty}$ such $X \subset \bigcup_{i=1}^{\infty} C_i$, we have

$$\mu'(X) \leq \mu'\left(\bigcup_{i=1}^{\infty} C_i\right) \leq \sum_{i=1}^{\infty} \mu'(C_i) = \sum_{i=1}^{\infty} \mu(C_i)$$

The second "$\leq$" sign is given by the countable sub-additivity of a premeasure. It follows that $\forall X \in \mathcal{B},\ \mu(X) = \inf\{\sum_{i=1}^{\infty} \bar{\mu}(C_i)\} \geq \mu'(X)$.

Next, we prove the other direction. Since $\bar{\mu}$ is σ-additive, we can construct a countable sequence $\{A_i \in \mathcal{R}\}_{i=1}^{\infty}$ such that

$$\tilde{\mu}(A_i) < \infty \ \forall i \in \mathbb{N}^* \text{ and } \bigcup_{i=1}^{\infty} A_i = S$$

Let $A_0 = \emptyset$ and $B_i = A_i \setminus \bigcup_{j=0}^{i-1} A_j$ for all possible i. The sequence of mutually disjoint sets $\{B_i\}_{i=1}^{\infty}$ also satisfies the above condition and $B_i \in \mathcal{R} \ \forall i \in \mathbb{N}^*$. We then have $B_i \setminus X \in \mathcal{B}$ and $X = X \cap \bigcup_{i=1}^{\infty} B_i \ \forall X \in \mathcal{B}$. By previous result, we also have

$$\forall X \in \mathcal{R},\ \mu(X \cap B_i) = \mu(B_i) - \mu(B_i \setminus A) = \mu'(B_i) - \mu(B_i \setminus X) \leq \mu'(B_i) - \mu'(B_i \setminus X) = \mu'(X \cap B_i)$$

Then by countable additivity of a measure,

$$\forall X \in \mathcal{R},\ \mu(X) = \mu\left(X \cap \bigcup_{i=1}^{\infty} B_i\right) \leq \mu'\left(X \cap \bigcup_{i=1}^{\infty} B_i\right) = \mu'(X)$$

which proves the other direction.

2.3.4 Applying Carathéodory extension theorem

Example of using Carathéodory theorem

Given two measure spaces $(S_1, \mathcal{B}_1, \mu_1)$, $(S_2, \mathcal{B}_2, \mu_2)$ and the product σ-algebra $\mathcal{B}$ of $\mathcal{B}_1$ and $\mathcal{B}_2$, given by

$$\mathcal{B} = \sigma(\mathcal{S}) \text{ where } \mathcal{S} := \{A \times B : A \in \mathcal{B}_1, B \in \mathcal{B}_2\}$$

we define

$$\forall A \in \mathcal{B}_1, B \in \mathcal{B}_2,\ \tilde{\mu}(A \times B) := \mu_1(A)\mu_2(B)$$

We can check that $\mathcal{S}$ is a semiring and $\tilde{\mu}$ is a premeasure on it (the proof of σ-additivity is given in 4.2 using integral).

Given by Carathéodory theorem, we can further claim that if both μ_1 and μ_2 are σ-finite, then there exists a unique measure μ extended from $\tilde{\mu}$ on $\mathcal{B}$. The measure is called product measure.

It remains to prove that $\tilde{\mu}$ is σ-finite. Since both μ_1 and μ_2 are σ-finite, for $i \in \{1, 2\}$, $S_i = \bigcup_{n=1}^{\infty} A_{i,n}$ where $A_{i,j} \in \mathcal{B}_i$ and $\mu_i(A_{i,j}) < \infty \ \forall j \in \mathbb{N}^*$. Then we have $\forall j, k \in \mathbb{N}^*$, $A_{1,j} \times A_{2,k} \in \mathcal{S}$ and $\tilde{\mu}(A_{1,j} \times A_{2,k}) < \infty$. Since $S_1 \times S_2 = \bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} A_{1,j} \times A_{2,k}$, the result follows.

Based on the Carathéodory theorem, we can make some more rigorous definitions of measure.

Definition: Lebesgue measure on $\mathbb{R}$

Let $\tilde{\lambda}$ be a premeasure on a semiring $\mathcal{R} := \{(a, b] : -\infty < a \leq b < \infty\} \cup \emptyset$ defined by $\tilde{\lambda}((a, b]) := b - a$.

Since $\mathcal{R}$ is σ-finite, we define the measure λ on $\sigma(\mathcal{R}) = \mathcal{B}(\mathbb{R})$, which is the unique extension of $\tilde{\lambda}$, to be the Lebesgue measure, which is $\forall A \in \mathcal{B}(\mathbb{R})$,

$$\lambda(A) := \inf\left\{\sum_{i=1}^{\infty}\sum_{j=1}^{n_i}\tilde{\lambda}(C_{i,j}) : A \subset \bigcup_{i=1}^{\infty}\bigcup_{j=1}^{n_i} C_{i,j} \text{ and } \forall i \in \mathbb{N}^*, 1 \leq j \leq n_i,\ C_{i,j} \in \mathcal{R}\right\}$$

We can also take $\mathcal{R} := \{(a,b) : -\infty < a \leq b < \infty\} \cup \emptyset$ or other similar semiring. The results are the same. Note that $\bigcup_{j=1}^{n_i} C_{i,j}$ constitute a ring for any $i \in \mathbb{N}^*$, given $n_i \in \mathbb{N}^*$ and $C_{i,j} \in \mathcal{R}\ \forall j \in \{1, 2, \cdots, n_i\}$.
Remark: Lebesgue measure is defined on more than $\mathcal{B}(\mathbb{R})$ but all the λ^*-measurable sets,

$$\{A \in \mathcal{P}(\mathbb{R}) :\ \forall B \in \mathcal{P}(\mathbb{R}),\ \lambda^*(B) = \lambda^*(A \cap B) + \lambda^*(B \setminus A)\}$$

where λ^* is the corresponding outer measure. In literature, a set A being Lebesgue measurable generally refers to λ^*-measurable, but in our context, it usually means the narrower case that $A \in \mathcal{B}$ unless further specification.

Definition: Lebesgue-Stieltjes measure
Let $F : (L,R) \mapsto \mathbb{R}$, where $L, R \in \bar{\mathbb{R}}$ and $L < R$, be a non-decreasing function, and $\tilde{\mu}((a,b]) = \lim_{x\to b^+} F(x) - \lim_{x\to a^+} F(x)$ be a premeasure on $\mathcal{R} := \{(a,b] : L < a \leq b < R\}$. The Lebesgue-Stieltjes measure μ generated by F is the unique Carathéodory extension of $\tilde{\mu}$ from $\mathcal{R}$ to $\sigma(\mathcal{R}) = \mathcal{B}((L,R))$.
The probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the Lebesgue-Stieltjes measure generated by its *pdf.* $F : \mathbb{R} \mapsto [0,1]$. One needs to restrict

1. F to be right-continuous
2. F to be non-decreasing
3. $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$

to satisfy the Kolmogorov's axioms.

We can also define the Lebesgue measure in higher dimensions using the concept of product measure, which means area or volume.
Definition: Lebesgue measure on $\mathbb{R}^n$
Let λ_0 be the Lebesgue measure on $\mathbb{R}$. Define the premeasure $\tilde{\lambda}(A_1 \times A_2 \times \cdots \times A_n) = \prod_{i=1}^n \lambda_0(A_i)$ on a semiring $\mathcal{R} := \{A_1 \times A_2 \times \cdots \times A_n\}$, where $A_i \in \mathcal{B}(\mathbb{R})\ \forall i \in \mathbb{Z} \cap [1,n]$. The Lebesgue measure on $\mathbb{R}^n$ is λ, the unique Carathéodory extension of $\tilde{\lambda}$. Detail omitted.

Chapter 3: Measurable function and random variable

3.1 Measurable function

3.1.1 Measurable function

Definition: Inverse map

Let $(S_1, \mathcal{B}_1)$ and $(S_2, \mathcal{B}_2)$ be two measurable space and $f : S_1 \mapsto S_2$ be a function. We define the inverse map $f^{-1} : \mathcal{P}(S_2) \mapsto \mathcal{P}(S_1)$ by:

$$\forall\ Y \in \mathcal{P}(S_2),\ f^{-1}(Y) := \{x \in S_1 : f(x) \in Y\}$$

which is the preimage of Y.

Proposition:

1. $f^{-1}(S_2) = S_1$
2. $f^{-1}(S_2 \setminus A) = S_1 \setminus f^{-1}(A)$
3. $f^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f^{-1}(A_i)$
4. $\{f^{-1}(Y) : Y \in \mathcal{B}_2\}$ is a σ-algebra.
5. If $\mathcal{C} \in \mathcal{B}_2$, then $\{f^{-1}(Y) : Y \in \sigma(\mathcal{C})\} = \sigma\left(\{f^{-1}(Y) : Y \in \mathcal{C}\}\right)$

The Proof of 1-3 is trivial and 4 can be proved by 1-3 and checking the conditions. Here is the proof of 4:

Denote the LHS as $\mathcal{F}_{\sigma(\mathcal{C})}$ and the RHS as $\sigma(\mathcal{F}_{\mathcal{C}})$. By 1, $\mathcal{F}_{\sigma(\mathcal{C})}$ is a σ-algebra, then it suffices to show $\mathcal{F}_{\sigma(\mathcal{C})} \subset \sigma(\mathcal{F}_{\mathcal{C}})$.

Let $\mathcal{G} := \{A \in \mathcal{P}(S_2) : f^{-1}(A) \in \sigma(\mathcal{F}_{\mathcal{C}})\}$. We can prove by checking the condition that $\mathcal{G}$ is a σ-algebra containing $\mathcal{C}$. Therefore, $\sigma(\mathcal{C}) \subset \mathcal{G}$. It gives that

$$\forall A \in \sigma(\mathcal{C}),\ A \in \mathcal{G} \Rightarrow f^{-1}(A) \in \sigma(\mathcal{F}_{\mathcal{C}})$$

and thus $\mathcal{F}_{\sigma(\mathcal{C})} \subset \sigma(\mathcal{F}_{\mathcal{C}})$.

Definition: Measurable function

Let $(S_1, \mathcal{B}_1)$ and $(S_2, \mathcal{B}_2)$ be two measurable space and $f : S_1 \mapsto S_2$ be a function. f is measurable if $\forall\ Y \in \mathcal{B}_2,\ f^{-1}(Y) \in \mathcal{B}_1$. We denote

$$f : (S_1, \mathcal{B}_1) \mapsto (S_2, \mathcal{B}_2)$$

Propositions: Conditions for measurable function

1. If $\mathcal{B}_2 = \sigma(\mathcal{C})$, then f is measurable if $\forall Y \in \mathcal{C},\ f^{-1}(Y) \in \mathcal{B}_1$
2. When $(S_1, \mathcal{T}_1)$ and $(S_2, \mathcal{T}_2)$ are two topology spaces and for $i \in \{1, 2\}$, let $\mathcal{B}_i = \mathcal{B}(S_i) = \sigma(\mathcal{T}_i)$ be the Borel σ-algebra. Then if $f : S_1 \mapsto S_2$ is continuous, it is measurable.

Follow from the proposition of inverse set, 1 is given by $\{f^{-1}(Y) : Y \in \mathcal{C}\} \subset \mathcal{B}_1 \Rightarrow \{f^{-1}(Y) : Y \in \sigma(\mathcal{C})\} = \sigma\left(\{f^{-1}(Y) : Y \in \mathcal{C}\}\right) \subset \mathcal{B}_1$.
The proof of 2 follows from the definition of continuous function. $\forall A \in \mathcal{T}_2,\ f^{-1}(A) \in \mathcal{T}_1 \subset \mathcal{B}_1$. By 1, the result follows.

Proposition: about composition of measurable maps
Given three measurable space and two measurable maps: $f_1 : (S_1, \mathcal{B}_1) \mapsto (S_2, \mathcal{B}_2)$ and $f_2 : (S_2, \mathcal{B}_2) \mapsto (S_3, \mathcal{B}_3)$. Then the composition of the two maps, $f_2 \circ f_1 : (S_1, \mathcal{B}_1) \mapsto (S_3, \mathcal{B}_3)$, defined by

$$\forall x \in S_1,\ f_2 \circ f_1(x) = f_2[f_1(x)]$$

is measurable.
To prove this, we first check that $\forall A \in \mathcal{P}(S_3),\ f_2 \circ f_1^{-1}(A) = f_1^{-1}\left[f_2^{-1}(A)\right]$. Then $\forall A \in \mathcal{B}_3$, $f_2^{-1}(A) \in \mathcal{B}_2$ and thus $f_1^{-1}\left[f_2^{-1}(A)\right] \in \mathcal{B}_1$. It gives the result.

Definition: Push-forward measure
Let $(S_1, \mathcal{B}_1, \mu_1)$ be a measurable space and a measurable function $f : (S_1, \mathcal{B}_1) \mapsto (S_2, \mathcal{B}_2)$ maps it to another measureable space. We defined the push-forward measure (aka. image measure) on $(S_2, \mathcal{B}_2)$ by

$$\mu_2(A) = \mu_1 \circ f^{-1}(A) = \mu_1\left[f^{-1}(A)\right]\ \ \forall A \in \mathcal{B}_2$$

The definition is valid since $f^{-1}(A) \in \mathcal{B}_1$ from the definition of measurable function.

3.1.2 Surely and almost surely statement

In the remaining context, we usually treat a measurable function f as a variable. However, the image $f(x)$ is not fixed. The following definition is for disambiguation of statement involving f, such as $f = g$ or $f > g$. By default, a statement is a surely statement.

Definition: Surely and almost surely statement

Given the measure space $\{S, \mathcal{B}, \mu\}$. A statement involving the measurable function f is a surely statement if it holds for all $\{f(x) : x \in S\}$. For example, $f > g$ means $\forall x \in S, f(x) > g(x)$.

The statement is almost surely (aka. almost everywhere) if it holds for all $\{f(x) : x \in A\}$ where $A \in \mathcal{B}$ and $\mu(S \setminus A) = 0$. For example, $f > g$ almost surely, denoted by $f \overset{\text{a.s.}}{>} g$, means $\exists A \in \mathcal{B}$ such that $\mu(S \setminus A) = 0$ and $\forall x \in A, f(x) > g(x)$.

We then discuss the statement involving the limit of a sequence of measurable functions $\{f_n\}_{n=1}^{\infty}$. The surely statement that $\lim_{n\to\infty} f_n$ exists means $\forall x \in S, \lim_{n\to\infty} f_n(x)$ exists. In fact, $L := \lim_{n\to\infty} f_n$ is also a random variable (proof given in 3.2). Then it comes to the following concept.

Definition: Surely convergence and almost surely convergence

$\{f_n\}_{n=1}^{\infty}$ converges to f surely (aka. point wise) if $L = f$ surely, i.e., $\forall x \in S, \lim_{n\to\infty} f_n(x) = f(x)$.

$\{f_n\}_{n=1}^{\infty}$ converges to f almost surely (aka. almost everywhere) if $L \overset{\text{a.s.}}{=} f$, i.e., $\forall x \in A, \lim_{n\to\infty} f_n(x) = f(x)$, where $A \in \mathcal{B}$ and $\mu(S \setminus A) = 0$. It is denoted by

$$f_n \xrightarrow{\text{a.s.}} f \text{ or } \lim_{n\to\infty} f_n \overset{\text{a.s.}}{=} f$$

3.2 Random variable

Definition: Random variable

Given a probability space $(\Omega, \mathcal{F}, P)$, a measurable function $X : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ is defined to be a random variable.

If $X : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we say the variable X is finite.

Then we define the push-forward measure $P_X(A) = P[X^{-1}(A)]$ $\forall A \in \mathcal{B}(\mathbb{R})$ and its *pdf* F_X. We call P_X is the distribution of X.

Now $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_X)$ is a new probability space. We have shown in Chapter 2 that F_X and this probability space has one-to-one correspondence

We can also regard a scalar $c \in \bar{\mathbb{R}}$ as a random variable, in which case $P_X(A) = 1$ $\forall A \in \mathcal{B}(\bar{\mathbb{R}})$ such that $c \in A$, and $P_X(A) = 0$ whenever $c \notin A$.

Definition: Identical in distribution

Given a probability space $(\Omega, \mathcal{F}, P)$, two random variables X, Y are identical in distribution if

$(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_X)$ and $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P_Y)$ are identical, or equivalently, $F_X = F_Y$. It is denoted as

$$X \overset{d.}{=} Y$$

Definition: Discrete random variable

Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable $X : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ is discrete if its range, i.e. $\mathcal{R} := \{X(\omega) : \omega \in \Omega\}$, is a countable set and $P(X = x) > 0 \ \forall x \in \mathcal{R}$.

Proposition: Condition for random variable

A function $X : \Omega \mapsto \bar{\mathbb{R}}$ is a random variable if and only if $\forall a \in \bar{\mathbb{R}},\ X^{-1}([-\infty, a]) \in \mathcal{F}$.

This is given by condition 1 of measurable function.

Definition: Random vector

Given a probability space $(\Omega, \mathcal{F}, P)$, a measurable function $\mathbf{X} : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}^n, \mathcal{B}(\bar{\mathbb{R}}^n))$ is defined to be a random variable.

We say $\mathbf{X}$ is finite if it maps to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Definition: Random sequence

Given a probability space $(\Omega, \mathcal{F}, P)$, a measurable function $\mathbf{X} : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}^\infty, \mathcal{B}(\bar{\mathbb{R}}^\infty))$ is defined to be a random sequence.

We also say $\mathbf{X}$ is finite if it maps to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Propositions:

1. Let $\mathbf{X}$ be a random vector, if $f : (\bar{\mathbb{R}}^n, \mathcal{B}(\bar{\mathbb{R}}^n)) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ is a measurable function, then $f(\mathbf{X})$ is a random variable. To verify that f is measurable, we can show that f is continuous,
2. $\mathbf{X} = [X_1, X_2, \cdots, X_n]$ is a random vector if and only if X_i are random variables for all possible i.
3. Likewise, $\mathbf{X} = [X_1, X_2, \cdots]$ is a random sequence if and only if X_i are random variables for all $i \in \mathbb{N}^*$. Proof omitted.

Proof of 1:

The first statement is given by the proposition of composition (i.e. $f \circ \mathbf{X}$ is measurable if both f and $\mathbf{X}$ are measurable). The second statement is given by the condition 2 of measurable function.

Proof of 2:

Let $f_i(\mathbf{X}) = X_i$ be n measurable function. The "$\Rightarrow$" direction is given by 1. Then we have that $\forall A \in \mathcal{B}(\bar{\mathbb{R}})$, $X_i(A) := \mathbf{X}^{-1}[f_i^{-1}(A)] \in \mathcal{F}$ for all possible i.

Next we prove the "$\Leftarrow$" direction. Since $\mathcal{B}(\mathbb{R}^n) = \sigma\{\text{all open rectangles}\}$, by the condition 1 of measurable function, it suffices to shows that $\forall A \in \{\text{all open rectangles}\}$, $\mathbf{X}^{-1}(A) \in \mathcal{F}$. We can express $A = I_1 \times I_2 \times \cdots \times I_n$ such that $I_1, I_2, \cdots, I_n$ are the intervals of the sides of A. Then,

$$\forall A,\ \mathbf{X}^{-1}(A) = \bigcap_{i=1}^{n} X_i^{-1}(I_i) \text{ where } X_i^{-1}(I_i) \in \mathcal{F}$$

Therefore, $\mathbf{X}^{-1}(A) \in \mathcal{F}$ and it ends the proof.

The proposition 1 shows that many statistics of a sample $X_1, X_2, \cdots, X_n$ are random variables. Such statistics include sample means, sample variance, sample product, sample extremum and more.

Propositions: Random variables and limits

Given a probability space $(\Omega, \mathcal{F}, P)$, let $X_1, X_2, \cdots$ be some random variables.

1. $\sup X_n$ and $\inf X_n$ are random variables.
2. $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are random variables.
3. If $\forall \omega \in \Omega$, $\lim_{n\to\infty} X_n(\omega)$ exists, then $\lim_{n\to\infty} X_n$ is a random variable.

To prove 1, it suffices to show that $\forall A \in \{[-\infty, x] : x \in \bar{\mathbb{R}}\}$, $(\sup X_n)^{-1}(A) \in \mathcal{F}$. We first have

$$\forall x \in \bar{\mathbb{R}},\ \{\omega \in \Omega : (\sup X_n)(\omega) \le x\} = \bigcap_{i=1}^{\infty} \{\omega \in \Omega : X_n(\omega) \le x\}$$

Since $\forall i \in \mathbb{N}^*$, $\{\omega \in \Omega : X_n(\omega) \le x\} \in \mathcal{F}$, we have $\{\omega \in \Omega : (\sup X_n)(\omega) \le x\} \in \mathcal{F}\ \forall x \in \mathbb{R}$, which gives the result. The proof of $\inf A_n$ is similar.

Then 2 immediately follows since $\limsup_{n\to\infty} X_n = \inf_{n\in\mathbb{N}^*} \left(\sup_{k\ge n} X_k\right)$ and similar for $\liminf X_n$.

Further, when the limit exists, $\lim_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$ is also a random variable.

These propositions also hold for general measurable functions that map measurable space to $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$.

The following definition will be used in later chapters.

Definition: σ-algebra generated by random variables

Given a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$, define the σ-algebra generated by this random variable, denoted by $\sigma(X)$, to be

$$\sigma(X) := \{X^{-1}(L) : L \in \mathcal{B}(\bar{\mathbb{R}})\}$$

We have proved in 3.1 that $\{X^{-1}(L) : L \in \mathcal{B}(\bar{\mathbb{R}})\}$ is indeed a σ-algebra.
When we have multiple random variables $\{X_i\}_{i\in T}$ with index set T, define $\sigma(\{X_i\}_{i\in T})$ to be the smallest σ-algebra containing all $\{\sigma(X_i)\}_{i\in T}$.

3.3 Independence

3.3.1 Independent events

Definition: Independence for two events
Let $(\Omega, \mathcal{F}, P)$ be a probability measure. Two events $A, B \in \mathcal{B}$ are independent if

$$P(A \cap B) = P(A) \times P(B)$$

Definition: Independence for multiple events
Let $(\Omega, \mathcal{F}, P)$ be a probability measure. The event $A_1, A_2, \cdots, A_n$ ($A_i \in \mathcal{F}\ \forall i \in \{1, 2, \cdots, n\}$) are independent if

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i) \text{ for all finite } I \subset \{1, 2, \cdots, n\}$$

Definition: Independence for finite number of subsets
Let $(\Omega, \mathcal{F}, P)$ be a probability measure. The subsets $\{\mathcal{C}_i \subset \mathcal{F},\ i \in \{1, 2, .., n\}\}$, are independent if for any choices of $A_1, A_2 \cdots, A_n$ that $A_i \in \mathcal{C}_i$, the events $A_1, A_2, \cdots, A_n$ are independent.

Definition: Independence for infinite number of subsets
Let $(\Omega, \mathcal{F}, P)$ be a probability measure. The subsets $\{\mathcal{C}_i \subset \mathcal{F},\ i \in T\}$, where T is an index set, are independent if for any finite $I \subset T$, $\{\mathcal{C}_i, i \in I\}$ are independent.

When $\mathcal{C}_i$ are σ-algebras for all possible i, they are independent σ-algebras. The following criterion helps prove the independence among σ-algebras.
Proposition: Criterion for independent σ-algebras
If the non-empty subsets $\{\mathcal{C}_i \subset \mathcal{F}, i \in T\}$, satisfying

1. $\mathcal{C}_i$ are π-system for $i \in T$
2. $\mathcal{C}_i$, where $i \in T$, are independent

then $\sigma(\mathcal{C}_i)$, where $i \in T$, are independent σ-algebras.

We show the proof for the case $T = \{1, 2\}$

Let $A_2 \in \mathcal{C}_2$ be an arbitrary event, and let $\mathcal{L} := \{A \in \mathcal{F} : P(A \cap A_2) = P(A) \times P(A_2)\}$. We claim that $\mathcal{L}$ is a λ-system containing $\mathcal{C}_1$. Then, by Dykin's Theorem, $\sigma(\mathcal{C}_1) \subset \mathcal{L}$. Thus, $\sigma(\mathcal{C}_1)$ and $\mathcal{C}_2$ are independent. Adjust the definition of $\mathcal{L}$ to be $\{A \in \mathcal{F} : P(A \cap A_1) = P(A) \times P(A_1)\}$ where $A_1 \in \sigma(\mathcal{C}_1)$ is an arbitrary event, we prove $\sigma(\mathcal{C}_1)$ and $\sigma(\mathcal{C}_2)$ are independent in a similar way.

To prove our claim, we need to check

1. $\Omega \in \mathcal{L}$, given by $P(\Omega \cap A_2) = P(A_2) = P(\Omega) \times P(A_2)$

2. If $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$, since

$$P(A^c \cap A_2) = P(A_2) - P(A \cap A_2) = P(A_2) - P(A)P(A_2) = P(A^c)P(A_2)$$

3. If $\{B_n \in \mathcal{L}\}_{n=1}^{\infty}$ are mutually disjoint, then $\bigcup_{n=1}^{\infty} B_i \in \mathcal{L}$, because

$$P\left(A_2 \cap \bigcup_{i=1}^{\infty} B_i\right) = P\left(\bigcup_{i=1}^{\infty} A_2 \cap B_i\right) = \sum_{i=1}^{\infty} P(B_i \cap A_2) = \sum_{i=1}^{\infty} P(B_i)P(A_2) = P(A_2)P\left(\bigcup_{i=1}^{\infty} B_i\right)$$

4. $\mathcal{C}_1 \subset \mathcal{L}$ because $\mathcal{C}_1$ and $\mathcal{C}_2$ are independent.

Adjust the definition of $\mathcal{L}$, we can prove the finite case when $T = \{1, 2, \cdots, n\}$ inductively, and generalize the criteria to arbitrary index set T.

3.3.2 Independent random variables

Definition: Independence for random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability measure and $\{X_i : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), i \in T\}$ be some random variables, where T is the index set. The random variables are independent if the generated σ-algebras $\{\sigma(X_i), i \in T\}$ are independent.

Proposition: about independent random variables

If random variables Y and X_i are independent for any $i \in T$, then $\sigma(Y)$ and $\sigma(\{X_i\}_{i \in T})$ are independent.

More generally, let $\{X_i\}_{i \in T_1}$ and $\{Y_j\}_{j \in T_2}$ to be two sets of random variables such that X_i and Y_j are independent for any choices of $i \in T_1$ and $j \in T_2$. We have that $\sigma(\{X_i\}_{i \in T_1})$ and $\sigma(\{Y_j\}_{j \in T_2})$

are independent.

Proof:

From the criterion for independent σ-algebras, it suffices to show that $\sigma(Y)$ and $\bigcup_{i\in T}\sigma(X_i)$ are independent. This is given by the definition of independence of $\sigma(Y)$ and $\sigma(X_i)$. The second proposition holds similarly.

The criterion for independence among random variables makes use of the following definition.

Definition: Joint probability distribution function

Let $(\Omega, \mathcal{F}, P)$ be a probability measure and $\{X_i, i \in T\}$ be some random variables, where T is the index set. For a finite subset $I \subset T$, we define the joint probability distribution function, or joint *pdf*, to be

$$F_I(\{x_i \in \bar{\mathbb{R}}, i \in I\}) = P(\{\forall i \in I,\ X_i \leq x_i\})$$

Index I by $\{1, 2, \cdots, n\}$, then $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ is a random vector. The joint *pdf.* can be represented as

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}([x_1, x_2, \cdots, x_n]^T) = P(\{\forall i \in I,\ X_i \leq x_i\}) : \bar{\mathbb{R}}^n \mapsto [0, 1]$$

Theorem: Factorization criterion for independence among random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability measure and $\{X_i, i \in T\}$ be some random variables, where T is the index set. These random variables are independent if and only if for any finite subset $I \subset T$,

$$\forall x_i \in \mathbb{R} \text{ for all } i \in I,\ F_I(\{x_i, i \in I\}) = \prod_{i\in I} F_{X_i}(x_i)$$

where F_{X_i} is the *pdf* of X_i and F_I is the joint *pdf.*

The "$\Leftarrow$" direction comes directly from the definition of independent variables. The "$\Rightarrow$" direction is given by the criterion for independent σ-algebras and the fact that for any $i \in T$, $\{\{X_i \leq x_i\} : x_i \in \mathbb{R}\}$ is a π-system that generates $\mathcal{B}(\mathbb{R})$.

The following are weaker criteria when $\{X_i, i \in T\}$ satisfy other conditions.

Proposition: More criteria for independence among random variables

1. When T is finite, $\{X_i, i \in T\}$ are independent if and only if

$$\forall x_i \in \mathbb{R} \text{ for all } i \in T,\ F_T(\{x_i, i \in T\}) = \prod_{i\in T} F_{X_i}(x_i)$$

2. When T is finite, and X_i is discrete with its range $\mathcal{R}_i$ for all $i \in T$, $\{X_i, i \in T\}$ if and only if

$$\forall x_i \in \mathcal{R}_i \text{ for all } i \in T,\ P(\{\forall i \in T, X_i = x_i\}) = \prod_{i \in T} P(\{X_i = x_i\})$$

To prove 1, it suffices to show that this criterion implies for any subset $I \in T$, the factorization condition holds. Detail omitted.

For example, in the case $T = \{1,2,3\}$ and $I = \{1,2\}$, we can show that

$$F_T(x_1, x_2, x_3) = F_{X_1}(x_1) F_{X_2}(x_2) F_{X_3}(x_3)\ \forall x_1, x_2, x_3 \in \mathbb{R}$$

implies

$$F_I(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2)\ \forall x_1, x_2 \in \mathbb{R}$$

by taking $x_3 \to \infty$. Then, $F_{X_3}(x_3) \to 1$ and $F_T(x_1, x_2, x_3) = P(X_1 \le x_1 \cap X_2 \le x_2 \cap X_3 \le x_3) \to P(X_1 \le x_1 \cap X_2 \le x_2) = F_T(x_1, x_2)$.

To prove 2, we can express F_{X_i} to be $\sum_{a \le x_i, a \in \mathcal{R}_\rangle} P(X_i = a)$ and similar for F_T. Then use the result of 1. Details omitted.

Chapter 4: Integration and expectation

4.1 The Lebesgue integral

4.1.1 Integral of simple function

Definition: non-negative simple function:

Let $(S, \mathcal{B}, \mu)$ be a measure space. A measurable function $\phi : S \mapsto \mathbb{R}$ is simple if its range $\{\phi(x) : x \in S\}$ is a finite set.

Such function can always be written as

$$\phi(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)\ \forall x \in S$$

where $a_i \in \mathbb{R}\ \forall i \in \{1, 2, \cdots, n\}$, $\{A_i \in \mathcal{B}\}_{i=1}^n$ are mutually disjoint and $\bigcup_{i=1}^n A_i = S$. This is called the standard representation of simple function. To obtain it, we can take $\{a_i\}_{i=1}^n$ be the range of ϕ and let $A_i = \phi^{-1}(\{a_i\})$.

We denote $\mathcal{E}$ to be the set of all simple functions on $(S, \mathcal{B}, \mu)$ and $\mathcal{E}^+$ to be the set of all non-negative simple functions that further satisfy $a_i \geq 0 \ \forall i \in \{1, 2, \cdots, n\}$.

Theorem: Measurability theorem

Let $(S, \mathcal{B}, \mu)$ be a measure space. A non-negative function $f : S \mapsto [0, \infty]$ is measurable if and only if there exists a monotonic increasing sequence of non-negative simple functions $\{\phi_n \in \mathcal{E}^+\}_{n=1}^\infty$ (i.e. $\phi_n \leq \phi_{n+1} \ \forall n \in \mathbb{N}^*$) such that $\lim_{n\to\infty} \phi_n = f$.

The "$\Leftarrow$" direction is given by the fact that $\lim_{n\to\infty} \phi_n = \sup \phi_n$ preserve measurability, as shown in 3.3.

The "$\Rightarrow$" direction is given by the following construction. For any $n \in \mathbb{N}^*$, we partition $[0, \infty]$ into $\{A_i \in S\}_{i=1}^{n2^n+1}$:

$$A_1 = f^{-1}\left(\left[0, \frac{1}{2^n}\right)\right), A_2 = f^{-1}\left(\left[\frac{1}{2^n}, \frac{2}{2^n}\right)\right), \cdots, A_{(n2^n)} = f^{-1}\left(\left[\frac{n2^n - 1}{2^n}, n\right)\right), A_{(n2^n+1)} = f^{-1}([n, \infty])$$

Then $\forall x \in S$, let

$$\begin{aligned}
\phi_n(x) :&= \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathbf{1}_{A_i}(x) + n\mathbf{1}_{A_{n2^n+1}} f(x) \\
&= \begin{cases} \sup\left\{\frac{i-1}{2^n} : i \in \{1, 2, \cdots, n2^n\} \text{ and } \frac{i-1}{2^n} \leq f(x)\right\} & , f(x) < n \\ n & , f(x) \geq n \end{cases} \\
&= \begin{cases} \frac{i-1}{2^n}, \text{ where } i = \lfloor 2^n f(x) + 1 \rfloor & , f(x) < n \\ n & , f(x) \geq n \end{cases}
\end{aligned}$$

and $\{\phi_n \in \mathcal{E}\}_{n=1}^\infty$ is the desired sequence. Note that when $f(x) < \infty$, $\lim_{n\to\infty} f(x) - \phi_n(x) \leq \lim_{n\to\infty} 2^{-n} = 0$ and when $f(x) = \infty$, $\lim_{n\to\infty} \phi_n(x) = \lim_{n\to\infty} n = \infty$.

Definition: Integral of simple function

Let $(S, \mathcal{B}, \mu)$ be a measure space and $\phi \in \mathcal{E}$ be a simple function $\phi(x) = \sum_{i=1}^n a_i \mathbf{1}_{A_i}(x) \ \forall x \in S$ for some $a_i \in [0, \infty) \ \forall i \in \{1, 2, \cdots, n\}$ and mutually disjoint $\{A_i \in S\}_{i=1}^n$ such that $\bigcup_{i=1}^n A_i = S$. Define

$$\int_S \phi \mathrm{d}\mu := \sum_{i=1}^n a_i \mu(A_i)$$

to be the integral of ϕ on S. We also write $\int_S \phi(x) \mathrm{d}\mu(x)$ for disambiguation.

Propositions about the integral of simple function

1. If $\phi \in \mathcal{E}$ and $\phi > 0$, then $\int_S \phi \mathrm{d}\mu > 0$

2. If $\phi_1, \phi_2 \in \mathcal{E}$, then $c_1\phi_1 + c_2\phi_2 \in \mathcal{E}$ $\forall c_1, c_2 \in \mathbb{R}$ and $\int_S (c_1\phi_1 + c_2\phi_2)\mathrm{d}\mu = c_1 \int_S \phi_1 \mathrm{d}\mu + c_2 \int_S \phi_2 \mathrm{d}\mu$

3. If $\phi_1, \phi_2 \in \mathcal{E}$ and $\phi_1 \leq \phi_2$, then $\int_S \phi_1 \mathrm{d}\mu \leq \int_S \phi_2 \mathrm{d}\mu$

For 1, we first have $\phi \geq 0 \Rightarrow a_i \geq 0$ $\forall i \in \{1, 2, \cdots, n\}$ in its standard representation. Then the result follows.
To prove 2, let $\phi_1 = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ and $\phi_2 = \sum_{i=1}^m b_i \mathbf{1}_{B_i}$ be their standard representations, then

$$c_1\phi_1 + c_2\phi_2 = \sum_{i=1}^n \sum_{j=1}^m (c_1 a_i + c_2 b_j)\mathbf{1}_{A_i \cap Bj} \ \Rightarrow \ c_1\phi_1 + c_2\phi_2 \in \mathcal{E}$$

and

$$\int_S (c_1\phi_1 + c_2\phi_2)\mathrm{d}\mu = c_1 \sum_{i=1}^n a_i \sum_{j=1}^m \mu(A_i \cap Bj) + c_2 \sum_{j=1}^m b_j \sum_{i=1}^n \mu(A_i \cap Bj) = c_1 \int_S \phi_1 \mathrm{d}\mu + c_2 \int_S \phi_2 \mathrm{d}\mu$$

3 is given by 1 and 2, since $\phi_2 - \phi_1 \geq 0 \Rightarrow \int_S \phi_2 \mathrm{d}\mu - \int_S \phi_1 \mathrm{d}\mu \geq 0$

4.1.2 Integral of non-negative measurable function

Definition: Lebesgue integral of non-negative measurable function
Let $(S, \mathcal{B}, \mu)$ be a measure space and $f : S \mapsto [0, \infty]$ be a non-negative measurable function. Define

$$\int_S f \mathrm{d}\mu := \sup \left\{ \int_S \phi \mathrm{d}\mu : \phi \in \mathcal{E}^+ \text{ and } 0 \leq \phi \leq f \right\}$$

We denote $\mathcal{L}^+$ to be the set of all non-negative measurable functions on $(S, \mathcal{B}, \mu)$. We will later show that it suffices to say $f \in \mathcal{L}^+$ if f is defined and non-negative almost everywhere in S.

Definition: Lebesgue integral of $\mathcal{L}^+$ function on A
Let $(S, \mathcal{B}, \mu)$ be a measure space, $A \in \mathcal{B}$ be a subset and $f \in \mathcal{L}^+$. Define

$$\int_A f \mathrm{d}\mu = \int_S f \mathbf{1}_A \mathrm{d}\mu$$

Then, for a simple function $\phi = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$,

$$\int_A \phi \mathrm{d}\mu = \int_S a_i \mathbf{1}_A \mathbf{1}_{A_i} \mathrm{d}\mu = \int_S a_i \mathbf{1}_{A \cap A_i} \mathrm{d}\mu = \sum_{i=1}^n a_i \mu(A \cap A_i)$$

Proposition: Equivalent definition of Lebesgue integral of $\mathcal{L}^+$ function
Let $\{\phi_n \in \mathcal{E}^+\}_{n=1}^\infty$ to be a monotonic increasing sequence of non-negative simple functions such that

$\lim_{n\to\infty}\phi_n = f$, where $f \in \mathcal{L}^+$. Then we can define the integral to be

$$\int_S f\mathrm{d}\mu := \sup\left\{\int_S \phi_n \mathrm{d}\mu : n \in \mathbb{N}^*\right\} = \lim_{n\to\infty}\int_S \phi_n \mathrm{d}\mu$$

We have shown that such $\{\phi_n\}_{n=1}^\infty$ exists. Since $\{\phi_n\}_{n=1}^\infty$ is monotonic increasing, the second equation holds.

To prove this equivalent is equivalent, we can show that for any such monotonic increasing sequence $\{\phi_n \in \mathcal{E}^+\}_{n=1}^\infty$ such that $\lim_{n\to\infty}\phi_n = f$,

$$\sup\left\{\int_S \phi_n \mathrm{d}\mu : n \in \mathbb{N}^*\right\} = \int_S f\mathrm{d}\mu$$

where $\int_S f\mathrm{d}\mu$ uses the old definition. This is given by the monotone convergence theorem as follows.

Propositions about Lebesgue integral of $\mathcal{L}^+$ function

1. If $f, g \in \mathcal{L}^+$ and $f \leq g$, then $\int_S f\mathrm{d}\mu \leq \int_S g\mathrm{d}\mu$
2. (Linearity) If $f, g \in \mathcal{L}^+$, then $c_1 f + c_2 g \in \mathcal{L}^+ \ \forall c_1, c_2 \in [0, \infty)$ and $\int_S (c_1 f + c_2 g)\mathrm{d}\mu = c_1\int_S f\mathrm{d}\mu + c_2\int_S g\mathrm{d}\mu$
3. If $f \in \mathcal{L}^+$, $\int_S f\mathrm{d}\mu = 0$ if and only if $f \overset{\text{a.s.}}{=} 0$
4. If $f \in \mathcal{L}^+$ and there is a subset $A \in \mathcal{B}$ such that $\mu(S \setminus A) = 0$, then $f_S f\mathrm{d}\mu = f_A f\mathrm{d}\mu$
5. If $f, g \in \mathcal{L}^+$, then if $f \overset{\text{a.s.}}{=} g$, we have $\int_S f\mathrm{d}\mu = \int_S g\mathrm{d}\mu$
6. (Monotone convergence theorem) Let $\{f_n \in \mathcal{L}^+\}_{n=1}^\infty$ be a monotonic increasing sequence of non-negative measurable functions. If $\lim_{n\to\infty} f_n = f$, then $f \in \mathcal{L}^+$ and
$$\int_S f\mathrm{d}\mu = \int_S \lim_{n\to\infty} f_n \mathrm{d}\mu = \lim_{n\to\infty}\int_S f_n \mathrm{d}\mu$$
7. Continue from 4, if we loosen the conditions into almost surely monotonic increasing, i.e. $f_n \overset{a.s.}{\leq} f_{n+1} \ \forall n \in \mathbb{N}^*$, and almost surely convergence, i.e. $f_n \xrightarrow{\text{a.s.}} f$, by further ensuring $f \in \mathcal{L}^+$, the result still holds.

Proof:

1 follows from the fact that $\{\phi \in \mathcal{E}^+ \text{ and } 0 \leq \phi \leq f\} \subset \{\phi \in \mathcal{E}^+ \text{ and } 0 \leq \phi \leq g\}$ if $f \leq g$.

We have verified the new definition of Lebesgue integral is equivalent. Using this definition, we can prove 2 because the limit operation and the integral of a simple function preserve linearity.

To prove 3, we first prove that it holds for all non-negative simple functions. Then, we have $\phi \overset{\text{a.s.}}{=} 0\ \forall \phi \in \{\phi \in \mathcal{E}^+ \text{ and } \phi \leq f\}$ if and only if $f \overset{\text{a.s.}}{=} 0$. It gives the result.
4 follows because $\int_S f \mathrm{d}\mu = \int_S (f\mathbf{1}_A + f\mathbf{1}_{S\setminus A}) \mathrm{d}\mu = \int_A f \mathrm{d}\mu$ since $f\mathbf{1}_{S\setminus A} \overset{\text{a.s.}}{=} 0$.
Suppose $f(x) = g(x)$ holds for all $x \in A \subset \mathcal{B}$ such that $\mu(S \setminus A) = 0$. 5 is given by

$$\int_S f \mathrm{d}\mu = \int_A f \mathrm{d}\mu = \int_A g \mathrm{d}\mu = \int_S g \mathrm{d}\mu$$

To prove 6, we first note that f is non-negative and measurable from the limit, thus $f \in \mathcal{L}^+$. Meanwhile, $\forall n \in \mathbb{N}^*,\ f_n \leq f$ by the monotonicity, thus $\int_S f_n \mathrm{d}\mu \leq \int_S f \mathrm{d}\mu$. Then we will show $\int_S f_n \mathrm{d}\mu \geq \int_S f \mathrm{d}\mu$.
Let $\phi \in \mathcal{E}^+$ be any non-negative simple function such that $0 \leq \phi \leq f$. By the definition of the integral, it suffices to prove $\int_S f_n \mathrm{d}\mu \geq \int_S \phi \mathrm{d}\mu$ for all possible ϕ. For any $\epsilon > 0$, define

$$B_i := \{x \in S : f_i(x) \geq (1-\epsilon)\phi(x)\}\ \forall i \in \mathbb{N}^*$$

We have $\forall i \in \mathbb{N}^*$, $B_i \subset B_{i+1}$ since $f_i \leq f_{i+1}$, and we now show $\bigcup_{i=1}^\infty B_i = S$.
$\bigcup_{i=1}^\infty B_i \subset S$ is trivial. To prove $\bigcup_{i=1}^\infty B_i \supset S$, we first note that for all $x \in S$,

$$\forall \epsilon' > 0,\ \exists i_0 \in \mathbb{N}^* \text{ such that } \forall i \geq i_0,\ f(x) - f_i(x) < \epsilon' \Rightarrow f(x) + \epsilon' > \phi(x) \geq (1-\epsilon)\phi(x)$$

Take $\epsilon' \to 0$, $f_i(x) \geq (1-\epsilon)\phi(x)$ and thus $x \in B_i\ \forall i \geq i_0$.
Next, we have

$$\int_S f_i \mathrm{d}\mu \geq \int_S f_i \mathbf{1}_{B_i} \mathrm{d}\mu = \int_{B_i} f_i \mathrm{d}\mu \geq \int_{B_i} (1-\epsilon)\phi \mathrm{d}\mu = \sum_{j=1}^n (1-\epsilon) a_j \mu(B_i \cap A_j)\ \forall i \in \mathbb{N}^*$$

where $\sum_{j=1}^n a_j \mathbf{1}_{A_j} = \phi$ is the standard representation. By continuity from below,

$$\lim_{i\to\infty} \sum_{j=1}^n (1-\epsilon) a_j \mu(B_i \cap A_j) = \sum_{j=1}^n (1-\epsilon) a_j \mu\left(A_j \cap \bigcup_{i=1}^\infty B_i\right) = \sum_{j=1}^n (1-\epsilon) a_j \mu(A_j) = (1-\epsilon)\int_S \phi \mathrm{d}\mu$$

Take $\epsilon \to 0$, it gives

$$\lim_{i\to\infty} \int_S f_i \mathrm{d}\mu \geq \int_S \phi \mathrm{d}\mu\ \forall \phi \in \mathcal{E}^+ \text{ such that } 0 \leq \phi \leq f$$

Then the result follows.
To prove 7, we take S_n for each $n \in \mathbb{N}^*$ such that

$$\mu(S \setminus S_n) = 0 \text{ and } \forall a \in S_n,\ f_n(a) \leq f_{a+1}(a)$$

and S_0 such that

$$\mu(S_0 \setminus S_n) \text{ and } \forall a \in S_n,\ \lim_{n\to\infty} f_n(a) = f(a)$$

Let $S' = \bigcup_{n=0}^{\infty} S_n$, we have $S' \in \mathcal{B}$ and $\mu(S \setminus S') = 0$ and the sequence $\{f_n \mathbf{1}_{S'}\}_{n=1}^{\infty}$ satisfy the monotonic convergence condition, which gives

$$\int_{S'} f \mathrm{d}\mu = \int_S f \mathbf{1}_{S'} \mathrm{d}\mu = \int_S \lim_{n\to\infty} f_n \mathbf{1}_{S'} \mathrm{d}\mu = \lim_{n\to\infty} \int_S f_n \mathbf{1}_{S'} \mathrm{d}\mu = \lim_{n\to\infty} \int_{S'} f_n \mathrm{d}\mu$$

Then the result follows by property 5.

Property 5 gives insight that the integral is irrelevant to the value of the function in a set $A' \in \mathcal{B}$ such that $\mu(A') = 0$. Thus it is allowed that $f(x) < 0$ or $f(x)$ is not defined for some $x \in A'$. When f is defined and non-negative almost everywhere on S, we can write $f \in \mathcal{L}^+$ and compute $\int_S f \mathrm{d}\mu$, in which case we regard f as $f\mathbf{1}_{S'}$, where $S' \in \mathcal{B}$ is the set that f is defined.

The following is a lemma to prove the dominated convergence theorem.

Lemma: Fatou's lemma

Given $\{f_n \in \mathcal{L}^+\}_{n=1}^{\infty}$, then

$$\int_S \liminf_{n\to\infty} f_n \mathrm{d}\mu \le \liminf_{n\to\infty} \int_S f_n \mathrm{d}\mu$$

Proof:

We first show that $\liminf_{n\to\infty} f_n \in \mathcal{L}^+$. $\liminf_{n\to\infty} f_n \ge 0$ is trivial. And since both limit and infimum preserve measurability, $\liminf_{n\to\infty} f_n = \lim_{n\to\infty} (\inf_{k\ge n} f_k)$ is a measurable function. Then,

$$\int_S \liminf_{n\to\infty} f_n \mathrm{d}\mu = \int_S \lim_{n\to\infty} \left(\inf_{k\ge n} f_k\right) \mathrm{d}\mu = \lim_{n\to\infty} \int_S \inf_{k\ge n} f_k \mathrm{d}\mu \le \liminf_{n\to\infty} \int_S f_n \mathrm{d}\mu$$

The second equation is given by the monotone convergence theorem, since $\{\inf_{k\ge n} f_k\}_{n=1}^{\infty}$ is monotonic increasing.

4.1.3 Integral of integrable function

Definition: Lebesgue integral of integrable function

Given a measure space $(S, \mathcal{B}, \mu)$ and a measurable function $f : S \mapsto \bar{\mathbb{R}}$. Let $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0) = \max(-f, 0)$. Then both f^+ and f^- are non-negative and measurable. If $\int_S |f| \mathrm{d}\mu < \infty$, then f is integrable. Its integral on S is defined by

$$\int_S f \mathrm{d}\mu := \int_S f^+ \mathrm{d}\mu - \int_S f^- \mathrm{d}\mu$$

Note that $\int_S f^+ \mathrm{d}\mu, \int_S f^- \mathrm{d}\mu < \int_S |f| \mathrm{d}\mu < \infty$, so the definition has no ambiguity,
Further, for $A \in \mathcal{B}$, we define the integral on A by

$$\int_A f \mathrm{d}\mu = \int_S f \mathbf{1}_A \mathrm{d}\mu$$

The set of integrable functions is denoted as

$$\mathcal{L}^1 : \left\{ f : S \mapsto [-\infty, \infty] : f \text{ is measurable and } \int_S |f| \mathrm{d}\mu < \infty \right\} =$$

The set is dependent on the specific measure space. We may use $\mathcal{L}^1(\mu)$ to denote that the measurable function is on $(S, \mathcal{B}, \mu)$.
Same to $\mathcal{L}^+$ case, we may say $f \in \mathcal{L}^1$ and compute $\int_S f \mathrm{d}\mu$ even when f is defined almost everywhere in S, in which case we regard f as $f \mathbf{1}_{S'}$, where $S' \in \mathcal{B}$ is the set that f is defined.
We may also consider a specific subset $A \subset S$ only, in which case we write $f|_A \in \mathcal{L}^1$ which means $f \mathbf{1}_A \in \mathcal{L}^1$.

Propositions about Lebesgue integral of $\mathcal{L}^1$ function

1. If $f \in \mathcal{L}^1$, $|\int_S f \mathrm{d}\mu| \leq \int_S |f| \mathrm{d}\mu$
2. If $f, g \in \mathcal{L}^1$ and $f \leq h \leq g$, then $h \in \mathcal{L}^1$
3. If $f \in \mathcal{L}^1$ and $A \in \mathcal{B}$ such that $\mu(S \setminus A) = 0$, then $\int_S f \mathrm{d}\mu = \int_A f \mathrm{d}\mu = \int_S f \mathbf{1}_A \mathrm{d}\mu$
4. Continue from 3, $f \mathbf{1}_A \in \mathcal{L}^1$ if and only if $f \in \mathcal{L}^1$
5. If $f \in \mathcal{L}^1$, then for $c \in \mathbb{R}$, $cf \in \mathcal{L}^1$ and $\int_S cf \mathrm{d}\mu = c \int_S f \mathrm{d}\mu$
6. (Linearity) If $f, g \in \mathcal{L}^1$, then $f + g \in \mathcal{L}^1$ and $\int_S (f + g) \mathrm{d}\mu = \int_S f \mathrm{d}\mu + \int_S g \mathrm{d}\mu$
7. (Monotone convergence theorem) If $\{f_n \in \mathcal{L}^1\}_{n=1}^\infty$ is monotonic increasing and converges (either surely or almost surely) to $f \in \mathcal{L}^1$, then $\lim_{n \to \infty} \int_S f_n \mathrm{d}\mu = \int_S f \mathrm{d}\mu$.

They can be proved by definition. Details omitted.

Dominated convergence theorem
Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions that converge to a measurable function almost surely, i.e. $f_n \xrightarrow{\text{a.s.}} f$. If there exists a non-negative measurable function $g \in \mathcal{L}^+$ such that $\forall n \in \mathbb{N}^*$, $|f_n| \overset{\text{a.s.}}{\leq} g$, then $f \in \mathcal{L}^1$ and

$$\int_S f \mathrm{d}\mu = \lim_{n \to \infty} \int_S f_n \mathrm{d}\mu$$

which means we can interchange the integration and limit.

Proof:

We begin with proving these two lemmas

1. If $\forall n \in \mathbb{N}^*$, $f_n \geq g$, then
$$\int_S \liminf_{n\to\infty} f_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_S f_n \mathrm{d}\mu$$
This is given by the Fatou's Lemma by substituting f_n with $f_n - g \in \mathcal{L}^+$:
$$\int_S \liminf_{n\to\infty} f_n \mathrm{d}\mu - \int_S g\mathrm{d}\mu = \int_S \liminf_{n\to\infty}(f_n - g)\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_S (f_n - g)\mathrm{d}\mu = \liminf_{n\to\infty} \int_S f_n \mathrm{d}\mu - \int_S g\mathrm{d}\mu$$

2. If $\forall n \in \mathbb{N}^*$, $f_n \leq g$, then
$$\int_S \limsup_{n\to\infty} f_n \mathrm{d}\mu \geq \limsup_{n\to\infty} \int_S f_n \mathrm{d}\mu$$
This is because $-f_n \geq -g$, thus by 1, we have
$$\int_S \limsup_{n\to\infty} f_n \mathrm{d}\mu = -\int_S \liminf_{n\to\infty} -f_n \mathrm{d}\mu \geq -\liminf_{n\to\infty} \int_S -f_n \mathrm{d}\mu = \limsup_{n\to\infty} \int_S f_n \mathrm{d}\mu$$

Then, since $|f_n| \overset{\text{a.s.}}{\leq} g$ and $f_n \xrightarrow{\text{a.s.}} f$, suppose we have $\{A_n \in \mathcal{B}\}_{n=1}^{\infty}$ and $A_0 \in \mathcal{B}$ such that for any $n \in \mathbb{N}^*$, $|f_n(x)| \leq g(x)\ \forall x \in A_n$, $\lim_{n\to\infty} f_n(x) = f(x)\ \forall x \in A_0$ and $\mu(S \setminus A_n) = 0\ \forall n \in \mathbb{N}$. Let $A = \bigcup_{n=0}^{\infty} A_n$, we have $A \in \mathcal{B}$, $\mu(S \setminus A) = 0$ and the statements $|f_n \mathbf{1}_A| \leq g\mathbf{1}_A$ and $\lim_{n\to\infty} f_n \mathbf{1}_A = f\mathbf{1}_A$ hold surely. Therefore, $-g\mathbf{1}_A \leq f_n\mathbf{1}_A, f\mathbf{1}_A \leq g\mathbf{1}_A\ \forall n \in \mathbb{N}^*$ and thus, $f_n\mathbf{1}_A, f_n, f\mathbf{1}_A, f, g\mathbf{1}_A \in \mathcal{L}^1$ since $g \in \mathcal{L}^1$. We have
$$\int_S f\mathrm{d}\mu = \int_A f\mathrm{d}\mu = \int_S f\mathbf{1}_A \mathrm{d}\mu = \int_S \lim_{n\to\infty}(f_n\mathbf{1}_A)\mathrm{d}\mu$$
and
$$\int_S \liminf_{n\to\infty}(f_n\mathbf{1}_A)\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_S f_n\mathbf{1}_A \mathrm{d}\mu \leq \limsup_{n\to\infty} \int_S f_n\mathbf{1}_A \mathrm{d}\mu \leq \int_S \limsup_{n\to\infty}(f_n\mathbf{1}_A)d$$
by the lemmas. Then because the limit exists,
$$\int_S \lim_{n\to\infty}(f_n\mathbf{1}_A)\mathrm{d}\mu = \int_S \liminf_{n\to\infty}(f_n\mathbf{1}_A)\mathrm{d}\mu = \int_S \limsup_{n\to\infty}(f_n\mathbf{1}_A)\mathrm{d}\mu$$
Hence all the inequalities are equalities, which gives $\lim_{n\to\infty} \int_S f_n\mathbf{1}_A \mathrm{d}\mu$ exists and equals to $\int_S f\mathrm{d}\mu$. The result follows:
$$\int_S f\mathrm{d}\mu = \liminf_{n\to\infty} \int_S f_n\mathbf{1}_A \mathrm{d}\mu = \lim_{n\to\infty} \int_S f_n\mathbf{1}_A \mathrm{d}\mu$$

4.1.4 Integral with transformation

The transformation theorem

Given two measure space $(S_1, \mathcal{B}_1, \mu_1)$, $(S_2, \mathcal{B}_2, \mu_2)$, the measurable function $t : (S_1, \mathcal{B}_1) \mapsto (S_2, \mathcal{B}_2)$ between them, and $\mu_2 := \mu_1 \circ t^{-1}$ being the push-forward measure. The Lebesgue Integral of measurable function $f : S_2 \mapsto [0, \infty] \in \mathcal{L}^+$ can be transformed in this way:

$$\int_{S_2} f \mathrm{d}\mu_2 = \int_{S_1} (f \circ t) \mathrm{d}\mu_1$$

Further, if $f \in \mathcal{L}^1(\mu_2)$, then $f \circ t \in \mathcal{L}^1(\mu_1)$, and vice versa. Then

$$\int_{S_2} f \mathrm{d}\mu_2 = \int_{S_1} (f \circ t) \mathrm{d}\mu_1$$

also holds.

Proof:

We first noted that $f \circ t$ is measurable by the proposition in 3.1. We begin with the case that f is a simple function, then an $\mathcal{L}^+$ and later an $\mathcal{L}^1$ function.

When $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i} \in \mathcal{E}^+$ be a non-negative simple function, where $A_i \in \mathcal{B}_2$ and $a_i \geq 0$ for all possible i. We have

$$f \circ t = \sum_{i=1}^n a_i (\mathbf{1}_{A_i} \circ t) = \sum_{i=1}^n a_i \mathbf{1}_{t^{-1}(A_i)}$$

then,

$$\int_{S_1} (f \circ t) \mathrm{d}\mu_1 = \int_{S_1} \left[\sum_{i=1}^n a_i \mathbf{1}_{t^{-1}(A_i)} \right] \mathrm{d}\mu_1 = \sum_{i=1}^n a_i \mu_1 [t^{-1}(A_i)] = \sum_{i=1}^n a_i \mu_2(A_i) = \int_{S_2} f \mathrm{d}\mu_2$$

When $f \in \mathcal{L}^+$, by the measurablility theorem, there exists a monotonic increasing sequence of non-negative simple functions $\{\phi_n \in \mathcal{E}^+\}_{n=1}^\infty$ such that $\lim_{n\to\infty} \phi_n = f$. Then, $\lim_{n\to\infty} \phi_n \circ t = f \circ t$. And by the monotone convergence theorem

$$\int_{S_1} (f \circ t) \mathrm{d}\mu_1 = \lim_{n\to\infty} \int_{S_1} (\phi_n \circ t) \mathrm{d}\mu_1 = \lim_{n\to\infty} \int_{S_2} \phi_n \mathrm{d}\mu_2 = \int_{S_2} f \mathrm{d}\mu_2$$

The second equation is given by the simple function case.

When $f \in \mathcal{L}^1$, let $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$, then $f^+, f^- \in \mathcal{L}^+$ and

$$\int_{S_2} f \mathrm{d}\mu_2 = \int_{S_2} f^+ \mathrm{d}\mu_2 - \int_{S_2} f^- \mathrm{d}\mu_2 = \int_{S_1} (f^+ \circ t) \mathrm{d}\mu_1 - \int_{S_1} (f^- \circ t) \mathrm{d}\mu_1 = \int_{S_1} (f \circ t) \mathrm{d}\mu_1$$

We have $f \circ t \in \mathcal{L}^1(\mu_1)$ because $\int_{S_1} |f \circ t| \mathrm{d}\mu_1 = \int_{S_1} (f^+ \circ t) \mathrm{d}\mu_1 + \int_{S_1} (f^- \circ t) \mathrm{d}\mu_1 = \int_{S_2} f^+ \mathrm{d}\mu_2 + \int_{S_2} f^- \mathrm{d}\mu_2 < \infty$ The other direction is similar.

4.2 Integral on product space

4.2.1 Product measure space

The aim of 4.2 is to derive the Fubini Theorem which helps to calculate multiple Lebesgue integral. We begin with some properties of product space.

Definition: Cartesian product of set
Given two set S_1 and S_2, define the Cartesian product of these two set to be

$$S_1 \times S_2 := \{(a, b) : a \in S_1, b \in S_2\}$$

Let $A \subset S_1 \times S_2$, We further define the following notation for simplicity.

$$\begin{aligned} &\forall a \in S_1,\ S_2^A(a) = \{b : (a, b) \in A\} \\ &\forall b \in S_2,\ S_1^A(b) = \{a : (a, b) \in A\} \end{aligned}$$

Proposition

1. $S_2^{A^c}(a) = \left[S_2^A(a)\right]^c$ (i.e. $S_2 \setminus S_2^A(a)$)
2. If $A_i \subset S_1 \times S_2$ for a index set $i \in I$. Then,

$$S_2^{\bigcup_{i\in I} A_i}(a) = \bigcup_{i\in I} S_2^{A_i}(a),\ S_2^{\bigcap_{i\in I} A_i}(a) = \bigcap_{i\in I} S_2^{A_i}(a)$$

Then can be proved by simple set operation.

Definition: Product measure
Given two measure spaces $(S_1, \mathcal{B}_1, \mu_1)$, $(S_2, \mathcal{B}_2, \mu_2)$, from 2.3.4, we have discussed we can define the product measure space on $S_1 \times S_2$, i.e. $(S_1 \times S_2, \mathcal{B}, \mu)$ by

$$\mathcal{B} = \sigma(\mathcal{S}) \text{ where } \mathcal{S} := \{A \times B : A \in \mathcal{B}_1, B \in \mathcal{B}_2\}$$

And μ is the unique Caratheodory extension of the following premeasure $\tilde{\mu}$

$$\tilde{\mu}(A \times B) = \mu_1(A)\mu_2(B)\ \forall A \in \mathcal{B}_1, B \in \mathcal{B}_2,$$

To complete the story, we still need to prove the σ-additivity of $\tilde{\mu}$. Let $\{A_n \in \mathcal{S}\}_n^\infty$ to be mutually disjoint and $\bigcup_{n=1}^\infty A_n \in \mathcal{S}$. Assume $A_n = A_{n,1} \times A_{n,2}$ $\forall n \in \mathbb{N}^*$ and $B := \bigcup_{n=1}^\infty A_n = B_1 \times B_2$, where $A_{n,1}, B_1 \in \mathcal{B}_1$ and $A_{n,2}, B_2 \in \mathcal{B}_2$. Then, for $n \in \mathbb{N}^*$,

$$\begin{aligned}\tilde{\mu}(A_n) = \mu_1(A_{n,1})\mu_2(A_{n,2}) = \int_{S_1} \mathbf{1}_{A_{n,1}}(a)\mu_2(A_{n,2})\mathrm{d}\mu_1 &= \int_{S_1} \mathbf{1}_{A_{n,1}}(a) \int_{S_2} \mathbf{1}_{A_{n,2}}(b)\mathrm{d}\mu_2\mathrm{d}\mu_1 \\ &= \int_{S_1}\int_{S_2} \mathbf{1}_{A_{n,1}\times A_{n,2}}(a,b)\mathrm{d}\mu_2(a,*)\mathrm{d}\mu_1\end{aligned}$$

Similarly, $\tilde{\mu}(B) = \int_{S_1}\int_{S_2} \mathbf{1}_B(a,b)\mathrm{d}\mu_2\mathrm{d}\mu_1$. Further,

$$\begin{aligned}\tilde{\mu}(B) &= \int_{S_1}\int_{S_2} \mathbf{1}_{\bigcup_{n=1}^\infty A_n}(a,b)\mathrm{d}\mu_2\mathrm{d}\mu_1 \\ &= \int_{S_1}\int_{S_2}\sum_{n=1}^\infty \mathbf{1}_{A_n}(a,b)\mathrm{d}\mu_2\mathrm{d}\mu_1 \\ &= \sum_{n=1}^\infty \int_{S_1}\int_{S_2} \mathbf{1}_{A_n}(a,b)\mathrm{d}\mu_2\mathrm{d}\mu_1 \quad \text{(by using the monotone convergence theorem twice)} \\ &= \sum_{n=1}^\infty \tilde{\mu}(A_n)\end{aligned}$$

If $f : (S_1 \times S_2, \mathcal{B}) \mapsto (S_3, \mathcal{B}_3)$ is a measurable function, then we define

$$\forall a \in S_1,\ f_a : S_2 \to S_3, \text{ given by } f_a(b) = f(a,b)\ \forall b \in S_2$$
$$\forall b \in S_2,\ f_b : S_1 \to S_3, \text{ given by } f_b(a) = f(a,b)\ \forall a \in S_1$$

Proposition

1. If $A \in \mathcal{B}$, then $\forall a \in S_1$, $S_2^A(a) \in \mathcal{B}_2$. Similar result holds for $S_1^A(b)$
2. f_a is measurable, so as f_b

Proof:
To prove 1, we construct $\mathcal{G}_a := \{A \subset S_1 \times S_2 : S_2^A(a) \in \mathcal{B}_2\}$. We first show $\mathcal{S} := \{B_1 \times B_2 : B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\} \subset \mathcal{G}_a$. Then, we prove $\mathcal{G}_a$ is a λ-system. The result follows by applying Dynkin's theorem
For any $A \in \mathcal{S}$, suppose $A = B_1 \times B_2$, where $B_1 \in \mathcal{B}_1$ and $B_2 \in \mathcal{B}_2$, then

$$\forall a \in S_1, S_2^A(a) = \begin{cases} B_2, & a \in B_1 \\ \emptyset, & a \notin B_1 \end{cases} \Rightarrow S_2^A(a) \in \mathcal{B}_2$$

We check the conditions for a λ-system:

1. $S_1 \times S_2 \in \mathcal{S} \subset \mathcal{G}_a$

2. If $A \in \mathcal{G}_a$, then $S_2^A(a) \in \mathcal{B}_2$ and by the proposition $S_2^{A^c}(a) = \left[S_2^A(a)\right]^c \in \mathcal{B}_2$. Therefore, $A^c \in \mathcal{G}_a$

3. If $\{A_i \in \mathcal{G}_a\}_{i=1}^{\infty}$ and A_i are disjoint for $i \in \mathbb{N}^*$, we have

$$\forall i \in \mathbb{N}^*,\ S_2^{A_i}(a) \in \mathcal{B} \Rightarrow S_2^{\bigcup_{i\in\mathbb{N}^*} A_i}(a) = \bigcup_{i\in\mathbb{N}^*} S_2^{A_i}(a) \in \mathcal{B}$$

which gives $\bigcup_{i\in\mathbb{N}^*} A_i \in \mathcal{G}_a$ It is easy to prove that $\mathcal{S}$ is a π-system. By Dynkin's theorem, we have $\mathcal{B} = \sigma(\mathcal{S}) \subset \mathcal{G}_a$. Therefore, $\forall A \in \mathcal{B} \subset \mathcal{G}_a,\ S_2^A(a) \in \mathcal{B}_2$.
2 follows from 1. Since f is measurable, which gives $\forall Y \in \mathcal{B}_3,\ f^{-1}(Y) = \{(a,b) \subset S_1 \times S_2 : f(a,b) \in Y\} \in \mathcal{B}$. Then, $\forall a \subset S_1,\ f_a^{-1}(Y) = \{b \subset S_2 : f(a,b) \in Y\} = S_2^{f^{-1}(Y)}(a)$. By 1, we have $S_2^{f^{-1}(Y)}(a) \in \mathcal{B}_2$, and thus, $f_a^{-1}(Y) \in \mathcal{B}_2$.

4.2.2 Fubini's Theorem

Proposition: A special case of Fubini's Theorem

1. Given two finite measure spaces $(S_1, \mathcal{B}_1, \mu_1)$, $(S_2, \mathcal{B}_2, \mu_2)$ such that $\mu_1(S_1) < \infty$ and $\mu_2(S_2) < \infty$, let their product measurable space to be $(S_1 \times S_2, \mathcal{B}, \mu)$. Then $\mu(S_1 \times S_2) = \mu_1(S_1) \times \mu_2(S_2) < \infty$. Let $A \in \mathcal{B}$, then $\mu_2 \circ S_2^A : S_1 \mapsto [0, \infty]$ and $\mu_1 \circ S_1^A : S_2 \mapsto [0, \infty]$ are measurable. Further,
$$\mu(A) = \int_{S_1} (\mu_2 \circ S_2^A) \mathrm{d}\mu_1 = \int_{S_2} (\mu_1 \circ S_1^A) \mathrm{d}\mu_2$$

2. We can loosen the condition of the measure to σ-finite. Suppose $(S_1, \mathcal{B}_1, \mu_1)$, $(S_2, \mathcal{B}_2, \mu_2)$ are σ-finite, then we have shown in 2.4 that $(S_1 \times S_2, \mathcal{B}, \mu)$ is also σ-finite. The upper three statements still holds

The proof of 1 is similar to the previous one using Dynkin's theorem. For simplicity, we only show the half of the prove, i.e., the one regarding $\mu_2 \circ S_2^A$. The other is similar. Let

$$\mathcal{G} = \left\{A \subset S_1 \times S_2 : \mu_2 \circ S_2^A \text{ is measurable and } \mu(A) = \int_{S_1} (\mu_2 \circ S_2^A) \mathrm{d}\mu_1\right\}$$

First, $\mathcal{S} := \{B_1 \times B_2 : B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\} \subset \mathcal{G}$, because for any $A = B_1 \times B_2 \in \mathcal{S}$, where $B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2$, we have $\forall a \in S_1,\ \mu_2 \circ S_2^A(a) = \mu_2(B_2)\mathbf{1}_{B_1}(a)$. It is easy to check that it is measurable, and

$$\int_{S_1} (\mu_2 \circ S_2^A) \mathrm{d}\mu_1 = \int_{S_1} \mu_2(B_2) \mathbf{1}_{B_1}(a) \mathrm{d}\mu_1 = \mu_1(B_1)\mu_2(B_2) = \mu(A)$$

Then it suffices to show that $\mathcal{G}$ is a λ-system.

1. $S_1 \times S_2 \in \mathcal{G}$ since $S_1 \times S_2 \in \mathcal{S}$

2. If $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$. Because if $\mu_2 \circ S_2^A$ is measurable, then

$$\mu_2 \circ S_2^{A^c} = \mu_2 \left[(S_2^A)^c \right] = \mu_2(S_2) - \mu_2 \circ S_2^A$$

is also measurable since it is a continuous transformation of $\mu_2 \circ S_2^A$. And from $\mu(A) = \int_{S_1} (\mu_2 \circ S_2^A) \mathrm{d}\mu_1$, we have

$$\int_{S_1} (\mu_2 \circ S_2^{A^c}) \mathrm{d}\mu_1 = \int_{S_1} \mu_2(S_2) \mathrm{d}\mu_1 - \int_{S_1} (\mu_2 \circ S_2^A) \mathrm{d}\mu_1 = \mu_2(S_2)\mu_1(S_1) - \mu(A) = \mu(A^c)$$

3. If $\{A_i \in \mathcal{G}\}_{i=1}^{\infty}$ and A_i are disjoint for $i \in \mathbb{N}^*$, then $\forall a \in S_1,\ S_2^{A_i}(a) \cap S_2^{A_j}(a) = \emptyset$ for $i \neq j,\ i,j \in \mathbb{N}^*$ (prove by contradiction). We have $\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}$ because if $\mu_2 \circ S_2^{A_i}$ are all measurable for $i \in \mathbb{N}^*$, then

$$\mu_2 \circ S_2^{\bigcup_{i \in \mathbb{N}^*} A_i} = \mu_2 \left(\bigcup_{i \in \mathbb{N}^*} S_2^{A_i} \right) = \sum_{i=1}^{\infty} \mu_2 \left(S_2^{A_i} \right) = \lim_{i \to \infty} \sum_{i=1}^{\infty} \mu_2 \circ S_2^{A_i}$$

are measurable since both the sum and the limit of real-value measurable function and measurable. Further, since $f_n := \sum_{i=1}^{\infty} \mu_2 \circ S_2^{A_i}$ is monotonic increasing and non-negative, we have

$$\begin{aligned}
&\int_{S_1} (\mu_2 \circ S_2^{A_i}) \mathrm{d}\mu_1 = \mu(A_i) \ \forall i \in \mathbb{N}^* \\
\Rightarrow &\int_{S_1} \mu_2 \circ S_2^{\bigcup_{i \in \mathbb{N}^*} A_i} \mathrm{d}\mu_1 = \int_{S_1} \lim_{n \to \infty} f_n \mathrm{d}\mu_1 = \lim_{n \to \infty} \int_{S_1} f_n \mathrm{d}\mu_1 = \lim_{n \to \infty} \sum_{i=1}^{\infty} \int_{S_1} (\mu_2 \circ S_2^{A_i}) \mathrm{d}\mu_1 \\
&\qquad = \lim_{n \to \infty} \sum_{i=1}^{\infty} \mu(A_i) = \mu(A)
\end{aligned}$$

by monotone convergence theorem and the linearity property of Lebesgue intergral on $\mathcal{L}^+$.

Next, we proof (half of) 2 from 1. Since S_2 is σ-finite, we assume $S_2 = \bigcup_{i=1}^{\infty} B_i$ where $\{B_i \in \mathcal{B}_2\}_{i=1}^{\infty}$ are mutually disjoint and $\mu_2(B_i) < \infty \ \forall i \in \mathbb{N}^*$. Using the proposition about σ-algebra of subset in 2.2, let $(B_i, \mathcal{B}_{2,i}, \mu_{2,i})$, where $\mathcal{B}_{2,i} := \{A \cap B_i : A \in \mathcal{B}_2\}$ and $\mu_{2,i}(A) := \mu_2(A) \ \forall A \in \mathcal{B}_{2,i}$, be a countable sequence of measure spaces for $i \in \mathbb{N}^*$. For all $i \in \mathbb{N}^*$, $\mu_{2,i}(B_i) < \infty$, the statements in 1 hold for all the product measure space $(S_1 \times B_i, \mathcal{B}'_i, \mu'_i)$ $\forall i \in \mathbb{N}^*$, where $\mathcal{B}'_i = \{A \cap S_1 \times B_i : A \in \mathcal{B}\}$ and $\mu'_i(A) = \mu(A)$ $\forall A \in \mathcal{B}'_i$. Thus, $\forall A \in \mathcal{B}$, we have

$$\forall i \in \mathbb{N}^*,\ A \cap S_1 \times B_i \in \mathcal{B}'_i \Rightarrow \mu_{2,i} \circ S_2^{A \cap S_1 \times B_i} = \mu_{2,i}\left(S_2^A \cap B_i\right) = \mu_2\left(S_2^A \cap B_i\right)$$

are measurable, so as

$$\mu_2 \circ S_2^A = \mu_2\left[S_2^A \cap \bigcup_{i=1}^{\infty} B_n\right] = \lim_{n\to\infty} \sum_{i=1}^{n} \mu_2\left(S_2^A \cap B_n\right)$$

since measurablility is preserved within limit and addition operation. Further, we have

$$\forall i \in \mathbb{N}^*, A \in \mathcal{B},\ \mu(A \cap B_i) = \mu'_i(A \cap B_i) = \int_{S_1}\left(\mu_{2,i} \circ S_2^{A\cap S_1 \times B_i}\right)\mathrm{d}\mu_1 = \int_{S_1} \mu_2\left(S_2^A \cap B_i\right)\mathrm{d}\mu_1$$

$$\Rightarrow \int_{S_1}(\mu_2 \circ S_2^A)\mathrm{d}\mu_1 = \int_{S_1} \lim_{n\to\infty} \sum_{i=1}^{n} \mu_2\left(S_2^A \cap B_n\right)\mathrm{d}\mu_1$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \int_{S_1} \mu_2\left(S_2^A \cap B_i\right)\mathrm{d}\mu_1 = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A \cap B_i) = \mu\left(\bigcup_{i=1}^{\infty} A \cap B_i\right) = \mu(A)$$

which proof the entire proposition.

Fubini's Theorem

Given two σ-finite measure spaces $(S_1, \mathcal{B}_1, \mu_1)$, $(S_2, \mathcal{B}_2, \mu_2)$ and their product measure space $(S_1 \times S_2, \mathcal{B}, \mu)$, let $f \in \mathcal{L}^+ : S_1 \times S_2 \mapsto [0, \infty]$ be a function, then

1. For any fixed $a \in S_1$, $f_a : S_2 \mapsto [0, \infty]$ is in $\mathcal{L}^+$. Similar statement holds for $f_b : S_1 \mapsto [0, \infty]$ $\forall b \in S_2$

2. Let $g : S_1 \mapsto \bar{\mathbb{R}}$ be a function defined by $g(a) := \int_{S_2} f_a \mathrm{d}\mu_2$ is in $\mathcal{L}^+$. Similar result holds for $h(b) := \int_{S_1} f_b \mathrm{d}\mu_1$

3. $$\int_{S_1 \times S_2} f\mathrm{d}\mu = \int_{S_1} g\mathrm{d}\mu_1 = \int_{S_2} h\mathrm{d}\mu_2$$

4. If $f \in \mathcal{L}^1(\mu) : S_1 \times S_2 \mapsto \bar{\mathbb{R}}$, a statement similar to 1 hold almost surely with regard to S_1. That is $\exists B_1 \in \mathcal{B}_1$, where $\mu_1(S_1 \setminus B_1) = 0$, such that $f_a : S_2 \mapsto \bar{\mathbb{R}}$ is in $\mathcal{L}^1(\mu_2)$ for any fixed $a \in B_1$. Similar statement holds for f_b for a $B_2 \in \mathcal{B}_2$. Define the g on B_1 and h on B_2 correspondingly, we still have $g \in \mathcal{L}^1(\mu_1)$, $h \in \mathcal{L}^1(\mu_2)$ and 3 holds.

Proof:

When $f = \mathbf{1}_A$, where $A \in \mathcal{B}$, 1 holds from the definition and $g = \mu_2 \circ S_2^A$ (similar for h). 2 and 3 holds by the above theorem.

When $f \in \mathcal{E}^+$ and let $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$, where $a_i \geq 0, A_i \in \mathcal{B}\ \forall i \in \{1, 2, \cdots, n\}$, 1-3 still hold because measurability is preserved under finite addition and Lebesgue integral has linearity property.

When $f \in \mathcal{L}^+$, there exists a monotonic increasing sequence $\{\phi_n \in \mathcal{E}^+\}_{n=1}^\infty$ such that $\lim_{n\to\infty} \phi_n = f$. Then for any fixed $a \in S_1$, we have $\{\phi_n(a, *)\}_{n=1}^\infty$ monotonic increasing and $\lim_{n\to\infty} \phi_n(a, *) = f(a, *) = f_a$. Thus f_a is measurable and obviously non-negative (i.e. $f_a \in \mathcal{L}^+$). By the monotone convergence theorem,

$$g(a) = \int_{S_2} f(a, b) \mathrm{d}\mu_2 = \int_{S_2} \lim_{n\to\infty} \phi(a, b) \mathrm{d}\mu_2 = \lim_{n\to\infty} \int_{S_2} \phi(a, b) \mathrm{d}\mu_2$$

is a limit of some $\mathcal{L}^+$ functions, therefore in $\mathcal{L}^+$ and $\mathcal{L}^1$. f_b is similar. To verify 3, we start from what we have proved

$$\forall n,\ \phi \in \mathcal{E}^+ \Rightarrow \int_{S_1 \times S_2} \phi_n(a, b) \mathrm{d}\mu = \int_{S_1} \int_{S_2} \phi_n(a, b) \mathrm{d}\mu_1 \mathrm{d}\mu_2$$

Then we have

$$\begin{aligned}
LHS &= \int_{S_1 \times S_2} f \mathrm{d}\mu = \int_{S_1 \times S_2} \lim_{n\to\infty} \phi_n \mathrm{d}\mu = \lim_{n\to\infty} \int_{S_1 \times S_2} \phi_n \mathrm{d}\mu \\
RHS &= \int_{S_1} \left[\lim_{n\to\infty} \int_{S_2} \phi_n(a, b) \mathrm{d}\mu_2 \right] \mathrm{d}\mu_1 \\
&= \lim_{n\to\infty} \int_{S_1} \int_{S_2} \phi_n(a, b) \mathrm{d}\mu_2 \mathrm{d}\mu_1 \text{ (since the sequence } \left\{ \int_{S_2} \phi_n(a, b) \mathrm{d}\mu_2 \right\}_{n=1}^\infty \text{ is monotonic increasing)} \\
&= \lim_{n\to\infty} \int_{S_1 \times S_2} \phi_n \mathrm{d}\mu = LHS
\end{aligned}$$

by using monotone convergence theorem two times.

Finally, we prove 4. When $f \in \mathcal{L}^1$, we first have f_a, f_b measurable since they are continuous transformations. Let $f = f^+ - f^-$, where $f^+, f^- \in \mathcal{L}^1 \cap \mathcal{L}^+$. Thus 1-3 hold for both f^+ (with corresponding g^+, h^+) and f^- (with g^-, h^-) and further,

$$\int_{S_1 \times S_2} f^+ \mathrm{d}\mu = \int_{S_1} g^+ \mathrm{d}\mu_1 < \infty, \quad \int_{S_1 \times S_2} f^- \mathrm{d}\mu = \int_{S_1} g^- \mathrm{d}\mu_1 < \infty$$

Therefore, $\int_{S_1 \times S_2} f \mathrm{d}\mu = \int_{S_1 \times S_2} f^+ \mathrm{d}\mu - \int_{S_1 \times S_2} f^- \mathrm{d}\mu = \int_{S_1} g^+ \mathrm{d}\mu_1 - \int_{S_1} g^- \mathrm{d}\mu_1$ and since for any function $f_0 \in \mathcal{L}^+$, $\int_{S_1} f_0 \mathrm{d}\mu_2 < \infty$ gives $f_0 \overset{\text{a.s.}}{<} \infty$ (proved by expressing f_0 as limit of monotonic increasing sequence of simple function), we have $g^+, g^- \overset{\text{a.s.}}{<} \infty$. Suppose $g^+(a), g^-(a) < \infty$ on $a \in B_1$, where $B_1 \in \mathcal{B}_1$ and $\mu_1(S_1 \setminus C_1) = 0$, we have $g^+ \mathbf{1}_{B_1}, g^- \mathbf{1}_{B_1} < \infty$, which gives

$$\forall a \in S_1,\ g^+(a) \mathbf{1}_{B_1}(a) = \int_{S_2} \mathbf{1}_{B_1}(a) f_a^+ \mathrm{d}\mu_2 < \infty \text{ and similar for} g^-$$

Therefore, for any fixed $a \in B_1$, $\int_{S_2} f_a^+ \mathrm{d}\mu_2 < \infty$, so as $\int_{S_2} f_a^- \mathrm{d}\mu_2$ and $\int_{S_2} |f_a| \mathrm{d}\mu_2$, and thus,

$$f_a \in \mathcal{L}^1(\mu_2) \Rightarrow \int_{S_2} f_a \mathrm{d}\mu_2 = \int_{S_2} f_a^+ \mathrm{d}\mu_2 - \int_{S_2} f_a^- \mathrm{d}\mu_2 = g^+(a) - g^-(a)$$

Further,

$$\begin{aligned} &\forall a \in S_1,\ g(a) := \int_{S_2} \mathbf{1}_{B_1}(a) \times f_a \mathrm{d}\mu_2 = \left[g^+(a) - g^-(a)\right] \mathbf{1}_{B_1}(a) \in \mathcal{L}^1(\mu_1) \\ \Rightarrow &\int_{S_1} g(a) \mathrm{d}\mu_1 = \int_{S_1} \left[g^+(a) - g^-(a)\right] \mathbf{1}_{B_1}(a) \mathrm{d}\mu_1 = \int_{S_1} g^+ \mathrm{d}\mu_1 - \int_{S_1} g^- \mathrm{d}\mu_1 = \int_{S_1 \times S_2} f \mathrm{d}\mu \end{aligned}$$

The proof regarding f_b and h is similar.

Theorem: Extended Fubini's theorem

Fubini's theorem also applies to another measure on the product set defined in the following way. Given two measurable space $(S_1, \mathcal{B}_1)$, $(S_2, \mathcal{B}_2)$ and the σ-finite measure on $(S_1, \mathcal{B}_1)$ to be μ_1, now for every $a \in S_1$, we have a measure $\mu_2(a, *)$ on $(S_2, \mathcal{B}_2)$, which satisfies the following:

1. $\forall a \in S_1$, $\mu_2(a, *)$ is a σ-finite measure on $(S_2, \mathcal{B}_2)$
2. $\exists \{A_{2,i} \in \mathcal{B}_2\}_{i=1}^{\infty}$ such that $\bigcup_{i=1}^{\infty} A_{2,i} = S_2$, $\mu_2(a, A_{2,i}) < \infty\ \forall i \in \mathbb{N}^*, a \in S_1$ and $A_{2,i} \cap A_{2,j} = \emptyset\ \forall i, j \in \mathbb{N}^*$ and $i \neq j$
3. $\forall A_2 \in \mathcal{B}_2$, $\mu_2(*, A_2) : S_1 \mapsto [0, \infty]$ is a measurable function.

We call the structure $\mu_2 : S_1 \times \mathcal{B}_2 \mapsto [0, \infty]$ satisfying the above a measurable kernel.

Next we define the measure on $(S_1 \times S_2, \mathcal{B} := \sigma(\mathcal{S}))$, where $\mathcal{S} := \{A_1 \times A_2 : A_1 \in \mathcal{B}_1, A_2 \in \mathcal{B}_2\}$. Let

$$\forall A = A_1 \times A_2 \in \mathcal{S},\ \tilde{\mu}(A_1 \times A_2) = \int_{A_1} \mu_2(a, A_2) d\mu_1(a)$$

be a premeasure on the semiring $\mathcal{S}$. The measure μ on $(S_1 \times S_2, \mathcal{B})$ is the Carathéodory extension of $\tilde{\mu}$.

Let $f : S_1 \times S_2 \mapsto \bar{\mathbb{R}}$ be measurable and either in $\mathcal{L}^+$ or $\mathcal{L}^1$. Then the previous statements 1-4 hold for $(S_1 \times S_2, \mathcal{B}, \mu)$, after substituting the fixed $\mu_2(*)$ by $\mu_2(a, *)$ and omitting the part about h.

Proof:

We first prove that the measure μ is well-defined by verifying

1. $\mathcal{S}$ is a semiring. (By definition, trivial)
2. $\tilde{\mu}$ is a premeasure. (Similar to the product measure case)

3. $\tilde{\mu}$ is σ-finite. We first have that if $A = A_1 \times A_2 \in \mathcal{S}$, where $A_1 \in \mathcal{B}_1$, $\mu_1(A_1) < \infty$, $A_2 \in \mathcal{B}_2$, and $\mu_2(a, A_2) < \infty\ \forall a \in A_1$, then $\tilde{\mu}(A) < \infty$. Then from condition 2 of μ_2 and that μ_1 is σ-finite, we can decompose S_1 and S_2 into countable unions of subsets with finite measure, i.e. $\{A_{1,i}\}_{i=1}^{\infty}$, $\{A_{2,j}\}_{j=1}^{\infty}$ respectively. Thus $S_1 \times S_2$ is the countable union of $A_{1,i} \times A_{2,j}$ with finite measure.

Then, by the Carathéodory theorem, μ is the unique extension of $\tilde{\mu}$ on $(S_1 \times S_2, \mathcal{B})$.
The following proof is very similar to the proof of Fubini's theorem. It starts from the case when f is an indicator function, then a simple function, a $\mathcal{L}^+$ function and finally a $\mathcal{L}^1$ function. Details omitted.

4.3 Expectation

4.3.1 Expectation and moment

Definition: Mathematical expectation
Given a probability measure $(\Omega, \mathcal{F}, P)$ and a random variable $X : (\Omega, \mathcal{B}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$, let $X^+ := \max(X, 0)$ and $X^- = \min(-X, 0)$. If either $\int_\Omega X^+ dP$ or $\int_\Omega X^+ dP$ is finite, the (mathematical) expectation of X exists, defined to be

$$\mathbb{E}(X) := \int_\Omega X^+ dP - \int_\Omega X^- dP = \begin{cases} \int_\Omega X dP & , \quad X \in \mathcal{L}^1 \\ \infty & , \quad \int_\Omega X^+ dP = \infty \text{ and } \int_\Omega X^- dP < \infty \\ -\infty & , \quad \int_\Omega X^- dP = \infty \text{ and } \int_\Omega X^+ dP < \infty \end{cases}$$

If $\int_\Omega X^+ dP = \int_\Omega X^- dP = \infty$, the expectation of X does not exist.
Note that the condition for $\mathbb{E}(X)$ to exist is wider than $X \in \mathcal{L}^1$.
If $X \in \mathcal{L}^1$, we can always transform $\mathbb{E}(X)$ it into an integral on $\bar{\mathbb{R}}$. We have $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P \circ X^{-1})$ as the push-forward measure space. Define $g : \bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ by $g(x) = x$. Then,

$$\mathbb{E}(X) = \int_\Omega X \mathrm{d}P = \int_\Omega g(X) \mathrm{d}P = \int_{\bar{\mathbb{R}}} g \mathrm{d}P \circ X^{-1} = \int_{\bar{\mathbb{R}}} x \mathrm{d}P \circ X^{-1}(x)$$

Proposition: Properties of expectation
Let X, Y be two random variables whose expectations exist,

1. $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$

2. If $X \geq 0$, then $X \in \mathcal{L}^+$ and $\mathbb{E}(X) = \int_\Omega X dP$

3. If $X \leq Y$ (or $X \overset{\text{a.s.}}{\leq} Y$), then $\mathbb{E}(X) \leq \mathbb{E}(Y)$

4. $\mathbb{E}(cX) = c\mathbb{E}(X)$ exists, where $c \in \mathbb{R}$.

5. If $\mathbb{E}(X) > -\infty$ (i.e. either $Y \in \mathcal{L}^1$ or $\mathbb{E}(Y) = \infty$) and $\mathbb{E}(X) = \infty$, then $\mathbb{E}(X+Y) = \infty$ exists. Likewise, $\mathbb{E}(X+Y) = -\infty$ if $\mathbb{E}(Y) < \infty$ and $\mathbb{E}(X) = -\infty$.

6. (Linearity) If $X, Y \in \mathcal{L}^1$, so as $aX + bY$ and $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ exists, where $a, b \in \mathbb{R}$.

7. (Monotone convergence) Let $\{X_n \in \mathcal{L}^1\}_{n=1}^\infty$ be monotonic increasing, and $\lim_{n\to\infty} X_n = X$, then $\mathbb{E}(X)$ exist and $\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n)$. It applies to the case of monotonic decreasing.

8. We can loosen the surely convergence in 3 into $X_n \xrightarrow{\text{a.s.}} X$, further we ensure X is measurable, then $\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n)$.

1-6 follows from the properties of Lebesgue integral.
To prove 7 (monotonic increasing case), we first have X measurable from the limit. Further, we have $\{X_n + X_1^-\}_{n=1}^\infty$ monotone increasing and

$$\forall n \in \mathbb{N}^*,\ X_n + X_1^- \geq X_n + X_n^- = X_n^+ \geq 0 \Rightarrow X_n + X_1 \in \mathcal{L}^+$$

By 1 and monotone convergence theorem, $\mathbb{E}(X + X_1^-)$ exist and $\mathbb{E}(X + X_1^-) = \lim_{n\to\infty} \int_\Omega (X_n + X_1^-) dP = \lim_{n\to\infty} \mathbb{E}(X_n + X_1^-)$. Since $X_1^- < \infty$, then by 4, $\mathbb{E}(X_n + X_1^-) = \mathbb{E}(X_n) + \mathbb{E}(X_1^-)\ \forall n \in \mathbb{N}^+$ and $\mathbb{E}(X + X_1^-) = \mathbb{E}(X) + \mathbb{E}(X_1^-)$. $\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n)$ follows. Proof of 8 is similar.

Proposition: expectation and *pdf*.
Given the probability space $(\Omega, \mathcal{F}, P)$, if a random variable X is non-negative almost surely, then $\mathbb{E}(X)$ exist and

$$\mathbb{E}(X) = \int_{[0,\infty)} 1 - F(x) \mathrm{d}\lambda$$

where $F(x) := P(X \leq x)$ is the *pdf* and λ is the Lebesgue measure on $\mathbb{R}$.
Proof:
If $P(X = +\infty) > 0$, then $\mathbb{E}(X) = +\infty$. We also have $\forall x \in \mathbb{R}^+,\ 1 - F(x) \geq P(X = +\infty) > 0$, and therefore $\int_{[0,\infty)} 1 - F(x)\mathrm{d}\lambda = +\infty$.
When $0 \overset{a.s.}{\leq} X \overset{a.s.}{<} +\infty$, we make use of Fubini's theorem. Note that $\int_\mathbb{R} \mathbf{1}_{(0,x)} \mathrm{d}\lambda = \lambda[(0,x)] = x$. We

have

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{\bar{\mathbb{R}}} x \mathrm{d}P \circ X^{-1}(x) \\
&= \int_{\mathbb{R}^+} x \mathrm{d}P \circ X^{-1}(x) \\
&= \int_{\mathbb{R}^+} \int_{\mathbb{R}} \mathbf{1}_{(0,x)}(y) \mathrm{d}\lambda(y) \mathrm{d}P \circ X^{-1}(x) \\
&= \int_{\mathbb{R}} \mathbf{1}_{(0,\infty)}(y) \int_{\mathbb{R}^+} \mathbf{1}_{(y,\infty)}(x) \mathrm{d}P \circ X^{-1}(x) \mathrm{d}\lambda(y) \text{ by Fubini's theorem} \\
&= \int_{\mathbb{R}^+} P \circ X^{-1}[(y,\infty)] \mathrm{d}\lambda(y) \\
&= \int_{\mathbb{R}^+} P(X > y) \mathrm{d}\lambda(y) \\
&= \int_{\mathbb{R}^+} 1 - F(y) \mathrm{d}\lambda(y)
\end{aligned}
$$

Here is an important inequality about expectation.

Theorem: Markov's inequality

Given the probability space $(\Omega, \mathcal{F}, P)$, if a random variable X is non-negative almost surely, then

$$
\forall c > 0,\ P(X \geq c) \leq \frac{\mathbb{E}(X)}{c}
$$

Proof:

We only need to prove the case when $0 \leq \mathbb{E}(X) < \infty$. Then,

$$
\begin{aligned}
\forall c > 0,\ \mathbb{E}(X) &= \mathbb{E}\left[X\mathbf{1}_{(0,c)}(X) + X\mathbf{1}_{[c,\infty]}(X)\right] \\
&\geq \mathbb{E}\left[0 \times \mathbf{1}_{(0,c)}(X) + c \times \mathbf{1}_{[c,\infty]}(X)\right] \\
&= c \times P \circ X^{-1}([c,\infty]) \text{ (the Lebesgue integral of simple function)} \\
&= cP(X \geq c)
\end{aligned}
$$

With similar proof, this inequality also applies to other measure spaces $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), \mu)$. If $f \in \mathcal{L}^1$ and $f \overset{a.s.}{\geq} 0$, then

$$
\forall c > 0,\ \mu\left(\{x : f(x) \geq c\}\right) \leq \frac{1}{c} \int_{\bar{\mathbb{R}}} f \mathrm{d}\mu
$$

Definition: Moment

Given a probability measure $(\Omega, \mathcal{F}, P)$ and a random variable X, for $k \in \mathbb{N}$, the k-th moment is defined as $\mathbb{E}(X^k)$ (if exists), denoted as

$$
m_k(X) := \int_{\Omega} X^k \mathrm{d}P
$$

Note that $m_0(X) = \int_\Omega 1dP = P(\Omega) = 1$ must exist.
We are usually interested in the case when $X^k \in \mathcal{L}^1$ (or equivalently, write $X \in \mathcal{L}^k$), so that $m_k(X)$ exists and is finite. The following nice properties also hold:

1. $X \in \mathcal{L}^k$ gives $P(\{X^k = +\infty\}) = P(\{X^k = -\infty\}) = 0$, but the inverse does not hold.
2. $\mathcal{L}^k$ is a vector space, which means if $X, Y \in \mathcal{L}^k$, then both $X + Y$ and cX (for $c \in \mathbb{R}$) belongs to $\mathcal{L}^k$
3. If $X \in \mathcal{L}^n$ where $n \in \mathbb{N}^*$, then $X \in \mathcal{L}^k$ $\forall k \in \{1, 2, \cdots, n-1\}$

If $m_k(X)$ exists and is finite for any $k \in \mathbb{N}^*$, we write $X \in \mathcal{L}^\infty$

Definition: Moment generating function
Given a probability measure $(\Omega, \mathcal{F}, P)$ and a random variable X, its moment generating function is defined to be

$$M_X(t) := \mathbb{E}[\exp(tX)]$$

for $t \in \mathbb{R}$ such that $\mathbb{E}[\exp(tX)]$ exists. In fact, $\exp(x) = \mathrm{e}^x$ $\forall x \in \mathbb{R}$, but the rigorous definition is

$$\exp(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!} \ \forall x \in \mathbb{R}$$

More details of exp will be stated in 7.3.1.
Moment generating function can generate moments. If M_X is defined on at least a small open ball containing 0, say $B(0, r)$, and $M_X(t) < \infty$ $\forall t \in B(0, r)$, then $X \in \mathcal{L}^\infty$ and

$$\forall k \in \mathbb{N}, \ m_k(X) := M_X^{(k)}(0) \text{ (the } k \text{ times derivative at point 0)}$$

Proof:
From the rigorous definition of exp, for any $k \in \mathbb{N}$, we have

$$\forall t \in B(0, r), \ \frac{|t|^k}{k!} \mathbb{E}\left(|X|^k\right) = \mathbb{E}\left(\frac{|tX|^k}{k!}\right) \leq \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{|tX|^n}{n!}\right) = \mathbb{E}(\exp|tX|) \leq \mathbb{E}[\exp(tX) + \exp(-tX)]$$

Since both $\mathbb{E}[\exp(tX)]$ and $\mathbb{E}[\exp(-tX)]$ exist and are finite, $|X|^k \in \mathcal{L}^1$ $\forall k \in \mathbb{N}$. Then,

$$\begin{aligned}\forall t \in B(0, r), \ M_k(X) &= \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}\left(X^n\right) \text{ by dominated convergence theorem}\end{aligned}$$

The dominant is given by

$$\forall t \in B(0,r),\ \left|\sum_{n=0}^{\infty}\frac{(tX)^n}{n!}\right| \le \sum_{n=0}^{\infty}\frac{|tX|^n}{n!} = \exp(|tX|) < \exp(tX) + \exp(-tX) \in \mathcal{L}^1$$

Differentiate M_X once, we have

$$\begin{aligned}
\forall t \in B(0,r),\ M_X^{(1)}(t) &= \lim_{h\to 0}\frac{M_X(t) - M_X(t)}{h}\\
&= \lim_{h\to 0}\mathbb{E}\left\{\frac{\exp(tX)}{h}[\exp(hX) - 1]\right\}\\
&= \lim_{h\to 0}\mathbb{E}\left[\exp(tX)\frac{1}{h}\sum_{n=1}^{\infty}\frac{(hX)^n}{n!}\right]\\
&= \lim_{h\to 0}\mathbb{E}\left[\exp(tX)X\sum_{n=0}^{\infty}\frac{(hX)^n}{(n+1)!}\right]
\end{aligned}$$

We next make use of the dominated convergence theorem again. First pick $\delta \in (0, r - |t|)$. for any sequence $\{a_k\}_{k=1}^{\infty}$ such that $\lim_{k\to 0} a_k = 0$, take its subsequence $\{a_{i+K}\}_{i=1}^{\infty}$ such that $|a_{i+K}| < r - |t| - \delta\ \forall i \in \mathbb{N}^*$. Then we have

$$\begin{aligned}
\forall i \in \mathbb{N}^*,\ \left|\exp(tX)X\sum_{n=0}^{\infty}\frac{(a_{i+K}X)^n}{(n+1)!}\right| &\le \exp(|tX|)|X|\sum_{n=0}^{\infty}\frac{|a_{i+K}X|^n}{n!}\\
&\le \exp(|tX|) \times \frac{\exp(\delta|X|)}{\delta} \times \exp(|a_{i+K}X|)\\
&= \exp\left[(|t| + \delta + |a_{i+K}|)|X|\right]\\
&\le \exp\left[(|t| + \delta + |a_{i+K}|)X\right] + \exp\left[-(|t| + \delta + |a_{i+K}|)X\right] \in \mathcal{L}^1\\
&\quad (\text{since } |t| + \delta + |a_{i+K}| < r)
\end{aligned}$$

By dominated convergence theorem,

$$\begin{aligned}
\lim_{k\to\infty}\mathbb{E}\left[\exp(tX)X\sum_{n=0}^{\infty}\frac{(a_kX)^n}{(n+1)!}\right] &= \lim_{i\to\infty}\mathbb{E}\left[\exp(tX)X\sum_{n=0}^{\infty}\frac{(a_{i+K}X)^n}{(n+1)!}\right]\\
&= \mathbb{E}\left[\exp(tX)X\lim_{i\to\infty}\sum_{n=0}^{\infty}\frac{(a_{i+K}X)^n}{(n+1)!}\right] = \mathbb{E}[X\exp(tX)]
\end{aligned}$$

holds for any $\{a_k\}_{k=1}^{\infty}$ with limit 0. Thus, $M_X^{(1)}(t) = \mathbb{E}[X\exp(tX)]$ exists. By induction, we can do further differentiation and obtain $M_X^{(k)}(t) = \mathbb{E}[X^k\exp(tX)]\ \forall t \in B(0,r)$ for any $k \in \mathbb{N}^*$. Plug in $t = 0$, we have $m_k(X) := M_X^{(k)}(0)\ \forall k \in \mathbb{N}$.

In fact, when the moment generating function is defined on $\mathbb{R}$ and is finite, it uniquely determines its distribution. Details are shown in Chapter 7.

The following properties follow directly from the definition.

Proposition of moment generating function

Suppose random variables X has moment generating functions M_X defined on $R \subset \mathbb{R}$, and $c \in \mathbb{R}^*$ is a non-zero constant.

1. Let $Z_1 := cX$ be a random variable. Then M_{Z_1} is defined on $R' := \{\frac{t}{c} : t \in R\}$ and $M_{Z_1}(t) = M_X(tc)\ \forall t \in R'$
2. Let $Z_2 := X+c$ be a random variable. Then M_{Z_2} is defined on R and $M_{Z_2}(t) = M_X(t)\exp(t)\ \forall t \in R$

4.3.2 Variance and covariance

Definition: Variance

Given a probability measure $(\Omega, \mathcal{F}, P)$ and a random variable $X \in \mathcal{L}^1$, define its variance to be

$$\mathrm{var}(X) := \mathbb{E}\left([X - \mathbb{E}(X)]^2\right)$$

To make the definition valid, we can check that $[X - \mathbb{E}(X)]^2 \in \mathcal{L}^+$ given $X \in \mathcal{L}^1$.

Propositions: Properties of variance

1. $\mathrm{var}(X) = \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right) = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = m_2(X) - m_1(X)^2$
2. $\mathrm{var}(X) \geq 0$, which gives $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$
3. $\mathrm{var}(X)$ is finite if and only if $X \in \mathcal{L}^2$
4. $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$, where $a, b \in \mathbb{R}$ are constants
5. $\mathrm{var}(X) = 0$ if and only if X equals to a constant in $\mathbb{R}$ almost surely.

We further define

$$\sigma_X := \sqrt{\mathrm{var}(X)}$$

to be the standard deviation.

Analogous to variance, we define the covariance between two variables.

Definition: Covariance

Given a probability measure $(\Omega, \mathcal{F}, P)$ and two random variables $X, Y \in \mathcal{L}^2$, define its covariance to be

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left([X-\mathbb{E}(X)][Y-\mathbb{E}(Y)]\right)$$

We first show $[X-\mathbb{E}(X)][Y-\mathbb{E}(Y)] \in \mathcal{L}^1$ so that this definition is valid. First, we have $[X - \mathbb{E}(X)], [Y-\mathbb{E}(Y)] \in \mathcal{L}^2$. It suffices to prove that whenever $X', Y' \in \mathcal{L}^2$, $X'Y' \in \mathcal{L}^1$.
From 3.2, we have that $[X', Y']^T$ is a random vector and further, $X'Y'$ is measurable because the mapping $[X', Y']^T \mapsto X'Y'$ is continuous. Next, $\int_\Omega |X'Y'|\mathrm{d}P \leq \int_\Omega X'^2 + Y'^2 \mathrm{d}P < \infty$. The result follows.

Propositions: Properties of covariance

1. Similar to variance, $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
2. $\mathrm{Cov}(X,X) = \mathrm{var}(X)$
3. $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{Cov}(X,Y)$
4. Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be some random variables of $\mathcal{L}^2$ on $(\Omega, \mathcal{F}, P)$, $a_1, \cdots, a_n$ and $b_1, \cdots, b_m$ be some real constant, we have
$$\mathrm{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j)$$

Theorem: Cauchy-Schwarz Inequality (integral case)
Without introducing the complete version (which include the discussion of inner product space), we show some special cases of this inequality in the context of Lebesgue integral

1. Given a general measure space $(S, \mathcal{B}, \mu)$ and two function $f, g \in \mathcal{L}^+(\mu)$, we have $\left(\int_S fg\mathrm{d}\mu\right)^2 \leq \int_S f^2 \mathrm{d}\mu \times \int_S g^2 \mathrm{d}\mu$
2. Given a probability measure $(\Omega, \mathcal{F}, P)$ and two random variables X, Y such that $\mathbb{E}(XY)$ exists, we have $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$
3. Continued from 2, if $X, Y \in \mathcal{L}^2$, then $\mathrm{Cov}(X,Y)^2 \leq \mathrm{var}(X)\mathrm{var}(Y)$

Proof:
Considering the product measure μ^2 of two $(S, \mathcal{B}, \mu)$, the function $[f(x)g(y) - f(y)g(x)]^2 : \Omega^2 \mapsto$

$[0, +\infty]$ is of $\mathcal{L}^+(\mu^2)$. Thus we can use Fubini's theorem,

$$\begin{aligned}
0 \leq & \int_{\Omega^2} [f(x)g(y) - f(y)g(x)]^2 \mathrm{d}\mu^2(x, y) \\
= & \int_\Omega \int_\Omega f^2(x)g^2(y) - 2f(x)f(y)g(x)g(y) + f^2(y)g^2(x) \mathrm{d}\mu(y)\mathrm{d}\mu(x) \\
= & \left[\int_\Omega f^2(x)\mathrm{d}\mu(x)\right]\left[\int_\Omega g^2(y)\mathrm{d}\mu(y)\right] - 2\left[\int_\Omega f(x)g(x)\mathrm{d}\mu(x)\right]\left[\int_\Omega f(y)g(y)\mathrm{d}\mu(y)\right] \\
& \qquad + \left[\int_\Omega f^2(y)\mathrm{d}\mu(y)\right]\left[\int_\Omega g^2(x)\mathrm{d}\mu(x)\right] \\
= & 2\left[\int_\Omega f^2 \mathrm{d}\mu\right]\left[\int_\Omega g^2 \mathrm{d}\mu\right] - 2\left[\int_\Omega fg\mathrm{d}\mu\right]^2
\end{aligned}$$

which gives

$$\left[\int_\Omega fg\mathrm{d}\mu\right]^2 \leq \left[\int_\Omega g^2\mathrm{d}\mu\right]\left[\int_\Omega f^2\mathrm{d}\mu\right]$$

Then 2 follows,

$$\mathbb{E}(XY)^2 = |\mathbb{E}(XY)|^2 \leq \mathbb{E}(|XY|) \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

By substituting $X := X' - \mathbb{E}(X')$ and $Y := Y' - \mathbb{E}(Y')$, we can show 3.

Theorem: Zero correlation between independent variable
Given a probability measure $(\Omega, \mathcal{F}, P)$ and two random variables $X, Y \in \mathcal{L}^2$, if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$, and as corollaries, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.
Proof:
Record the definition of independent variables in 3.3: X and Y are independent if $\forall A \in \sigma(X), B \in \sigma(Y)$, $P(A \cap B) = P(A)P(B)$, where $\sigma(X) = \{X^{-1}(L) : L \in \mathcal{B}(\bar{\mathbb{R}})\}$ and so as $\sigma(Y)$.
We then construct a product measure and use Fubini's theorem to prove this theorem.
Let $\mathbf{X} := [X, Y]^T : (\Omega, \mathcal{F}) \mapsto (\bar{\mathbb{R}}^2, \mathcal{B}(\bar{\mathbb{R}}^2))$ be a measurable vector. We claim that $(\bar{\mathbb{R}}^2, \mathcal{B}(\bar{\mathbb{R}}^2), \mu)$, where $\mu := P \circ \mathbf{X}^{-1}$, is the product measure space of $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P \circ X^{-1})$ and $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}), P \circ Y^{-1})$.
From the independence, we have

$$\begin{aligned}
&\forall L_A, L_B \in \bar{\mathbb{R}},\ X^{-1}(L_A) \in \sigma(X) \text{ and } Y^{-1}(L_B) \in \sigma(Y) \\
&\qquad \Rightarrow \mu(L_A \times L_B) = P \circ \mathbf{X}^{-1}(L_A \times L_B) = P[X^{-1}(L_A) \cap Y^{-1}(L_B)] \\
&\qquad\qquad = P[X^{-1}(L_A)]P[Y^{-1}(L_Y)]
\end{aligned}$$

Therefore,

$$\text{if } A = L_A \times L_B,\ \mu(A) = P \circ X^{-1}(L_A) \times P \circ X^{-1}(L_B)$$

holds for all $A \in \mathcal{R} := \{A_1 \times A_2 : A_1, A_2 \in \bar{\mathbb{R}}\}$. By Carathéodory theorem, the extension from pre-measure on $\mathcal{R}$ to a measure on $\sigma(\mathcal{R}) = \mathcal{B}(\bar{\mathbb{R}}^2)$ is unique, thus μ is the product measure.
Let g be a measurable function on $(\bar{\mathbb{R}}^2, \mathcal{B}(\bar{\mathbb{R}}^2))$ defined by $g(a,b) = ab$. We first proceed with the case that $X, Y \in \mathcal{L}^+$, then

$$\mathbb{E}(XY) = \int_\Omega XY \mathrm{d}P = \int_{\bar{\mathbb{R}}^2} g \mathrm{d}\mu$$

by the transformation theorem.
Finally, by Fubini's theorem

$$\begin{aligned}
\int_{\bar{\mathbb{R}}^2} g\mathrm{d}\mu &= \int_{\bar{\mathbb{R}}} \left[\int_{\bar{\mathbb{R}}} g(a,b) \mathrm{d}P \circ Y^{-1}(b) \right] \mathrm{d}P \circ X^{-1}(a) \\
&= \int_{\bar{\mathbb{R}}} \left[a \times \int_{\bar{\mathbb{R}}} b \mathrm{d}P \circ Y^{-1}(b) \right] \mathrm{d}P \circ X^{-1}(a) \\
&= \int_{\bar{\mathbb{R}}} a \left(\int_\Omega Y \mathrm{d}P \right) \mathrm{d}P \circ X^{-1}(a) \text{ (by the transformation theorem)} \\
&= \int_{\bar{\mathbb{R}}} a \mathbb{E}(Y) \mathrm{d}P \circ X^{-1}(a) = \mathbb{E}(Y) \int_\Omega X \mathrm{d}P = \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

Then, for general X, Y whose expectation exists, we split them into $X = X^+ - X^+$ such that $X^+, X^- \in \mathcal{L}^+$ (similar for Y). We have

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}\left[(X^+ - X^-)(Y^+ - Y^-) \right] \\
&= \mathbb{E}(X^+Y^+) + \mathbb{E}(X^-Y^-) - \mathbb{E}(X^+Y^-) - \mathbb{E}(X^-Y^+) \\
&= \mathbb{E}(X^+)\mathbb{E}(Y^+) + \mathbb{E}(X^-)\mathbb{E}(Y^-) - \mathbb{E}(X^+)\mathbb{E}(Y^-) - \mathbb{E}(X^-)\mathbb{E}(Y^+) \\
&= \left[\mathbb{E}(X^+) - \mathbb{E}(X^-) \right] \left[\mathbb{E}(Y^+) - \mathbb{E}(Y^-) \right] \\
&= \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

The following property of moment generating function follows from $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ given the independence of X and Y.
Corollary: Moment generating function of summation of random variables
Suppose two independent random variables, X and Y, have moment generating functions M_X and M_Y defined on $R_1, R_2 \subset \mathbb{R}$ respectively. Then $Z := X + Y$ has its moment generating function defined on $R_1 \cap R_2$ (if it is not empty) and

$$M_Z(t) = M_X(t) M_Y(t) \ \forall t \in R_1 \cap R_2$$

The following theorem is useful in proving convergence.
Theorem: Chebyshev's inequality

Given a probability measure $(\Omega, \mathcal{F}, P)$ and a random variable $X \in \mathcal{L}^2$, we have finite $\mathbb{E}(X)$ and $\text{var}(X)$. If $\text{var}(X) \neq 0$, then

$$\forall k > 0,\ P(|X - \mathbb{E}(X)| \geq k\sigma_X) \leq \frac{1}{k^2} \text{ and equivalently, } P(|X - \mathbb{E}(X)| \geq k) \leq \frac{\sigma_X^2}{k^2}$$

where $\sigma_X = \sqrt{\text{var}(X)}$ is the standard deviation.
Proof:

$$\begin{aligned} \forall k > 0,\ P(|X - \mathbb{E}(X)| \geq k\sigma_X) &= P([X - \mathbb{E}(X)]^2 \geq k^2 \text{var}(X)) \\ &\leq \frac{\mathbb{E}([X - \mathbb{E}(X)]^2)}{k^2 \text{var}(X)} \text{ by Markov's inequality} \\ &= \frac{1}{k^2} \text{ by the definition of variance} \end{aligned}$$

Chapter 5: Differentiation and density

5.1 Derivative of real function

5.1.1 Lebesgue integral and Rienann integral

We have defined the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ in 2.3.4. We start with discussing its proposition.

Proposition: about Lebesgue measure
Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ be the Lebesgue measure space and $S \in \mathcal{B}(\mathbb{R})$, we have

1. If S is a countable set, then $\lambda(S) = 0$
2. If S is compact (i.e., close and bounded in this case), then $\lambda(S) < \infty$
3. $\lambda(S) = \inf\{\lambda(A) : A \supset S \text{ and } A \text{ is open}\}$
4. $\lambda(S) = \sup\{\lambda(A) : A \subset S \text{ and } A \text{ is compact}\}$
5. $\forall S \in \mathcal{B}(\mathbb{R})$, $\inf\{\lambda(A \setminus S) : A \supset S \text{ and } A \text{ is open}\} = 0$,
   $\inf\{\lambda(S \setminus A) : A \subset S \text{ and } A \text{ is compact (or closed)}\} = 0$

1 and 2 are easy and proofs are omitted. 3 follows from the definition by using $\mathcal{R} := \{a < x < b : a, b \in \mathbb{R}\}$ as the semiring. In 1.2.2, we have shown that any open set can be the countable union of sets in $\mathcal{R}$.
To prove 4, we first note that when S is bounded, we can take a compact set $B \supset S$. Then, use 3 to approach $\lambda(B \setminus S)$ and the result follows. Note that $B \setminus A$ is compact whenever B is compact and A is open.
When S is unbounded, let $S = \bigcup_{i \in \mathbb{Z}} S_i$ where

$$S_i = S \cap (i, i+1] \text{ is bounded}, \forall i \in \mathbb{Z}$$

Since 4 holds for all S_i, for any $\epsilon > 0$ and $i \in \mathbb{Z}$, there exists a compact set A_i such that

$$\lambda(S_i) - \frac{\epsilon}{2^i} \leq \lambda(A_i) \leq \lambda(S_i)$$

Then, for $n \in \mathbb{N}^*$, add up all the disjoint S_i with index set $\{i : |i| \leq n\}$,

$$\lambda(S \cup (-n, n+1]) - \epsilon \leq \lambda(B_n) \leq \lambda(S \cup (-n, n+1]) \ \forall \epsilon > 0$$

where $B_n := \bigcup_{-n \leq i \leq n} A_i$ is compact and thus $B_n \subset S \cup (-n, n+1]$ for all $n \in \mathbb{N}^*$. When n goes to infinity,

$$\lambda(S) - \epsilon \leq \sup_{n \in \mathbb{N}^*} \lambda(B_n) \leq \lambda(S)$$

and therefore $\lambda(S) = \sup_{n \in \mathbb{N}^*} \lambda(B_n)$ when taking ϵ towards 0.
For the cases that $\lambda(S) < \infty$, 5 follows from 3 and 4. For infinity cases, we can prove 5 from the fact that λ is σ-finite.
They can be generated to Lebesgue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, where $k \in \mathbb{N}^*$, with similar proofs.

Definition: Integral with regard to Lebesgue measure
Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ be the Lebesgue measure space and $f : \mathbb{R} \mapsto \mathbb{R}$ be a function. For an interval (a, b), where $a, b \in \bar{\mathbb{R}}$, $a < b$ and $f|_{(a,b)} \in \mathcal{L}^1$, we can compute the Lebesgue integral with regard to the Lebesgue measure, denoted as

$$\int_a^b f(x)\mathrm{d}x := \int_{(a,b)} f \mathrm{d}\lambda = \int_{\mathbb{R}} f \mathbf{1}_{(a,b)} \mathrm{d}\lambda$$

If $a = b$, $\int_a^b f(x)\mathrm{d}x = 0$ naturally. If $a > b$, define $\int_a^b f(x)\mathrm{d}x$ to be $-\int_b^a f(x)\mathrm{d}x$. If either a or b is infinity, it represents the corresponding open and unbounded interval.
Further, fixed an $a \in \mathbb{R}$, we define

$$F(x) := \int_a^x f(t)\mathrm{d}t$$

on where either $f|_{(a,x)} \in \mathcal{L}^1$ or $f|_{(x,a)} \in \mathcal{L}^1$, so that

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a)$$

We have some primary questions. Is this Lebesgue integral equal to Rienann integral? Is F continuous or differentiable? Is the derivative F' equal to f? We will solve the first question in this section. The remaining are stated by the fundamental theorem of calculus in the next section.

5.1.2 The fundamental theorem of calculus

Definition: derivative of function
Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a function, define

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) := f'(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

on where the limit exists (in $\mathbb{R}$) to be the derivative of f, then we call f is differentiable at point x. If f is differentiable at almost every point in $\mathbb{R}$ or $\mathbb{R}^k$ except for a set with Lebesgue measure 0, it is differentiable almost everywhere.

From basic calculus, we know that if f is differentiable at x (or everywhere, or almost everywhere), it is continuous at x (or everywhere, or almost everywhere, respectively).

The fundamental theorem of calculus
Let $f : \mathbb{R} \mapsto \mathbb{R}$ be an $\mathcal{L}^1$ function. Then given a constant $a \in [-\infty, \infty]$, $F(x) := \int_a^x f(t)\mathrm{d}t$, is differentiable almost everywhere on $\mathbb{R}$, and further, the derivative $F' \overset{\text{a.s.}}{=} f$.
If $f \notin \mathcal{L}^1$, but it is integrable locally, i.e. $\forall x \in \mathbb{R}$, $\exists r > 0$ such that $f|_{B(x,r)} \in \mathcal{L}^1$, fix $a \in \mathbb{R}$ and let $D \ni a$ be an open set such that $f|_D \in \mathcal{L}^1$. Then $F\mathbf{1}_D$ is differentiable almost everywhere and $F'\mathbf{1}_D \overset{\text{a.s.}}{=} f\mathbf{1}_D$.

The above are equivalent to

$$f(x) = \lim_{h \to 0} \frac{\int_a^{x+h} f(t)\mathrm{d}t - \int_a^x f(t)\mathrm{d}t}{h} = \lim_{h \to 0} \frac{\int_x^{x+h} f(t)\mathrm{d}t}{h} \text{ for almost every } x \in \mathbb{R}$$

because the integral will not be infinity.

More generally, if $f : \mathbb{R}^k \mapsto \mathbb{R}$, where $k \in \mathbb{N}^*$, is also Lebesgue integrable locally, we have a lemma

$$f(\mathbf{x}) = \lim_{r\to 0} \frac{\int_{B(\mathbf{x},r)} f(\mathbf{t})\mathrm{d}\lambda(\mathbf{t})}{\lambda\left[B(\mathbf{x},r)\right]} \text{ for almost every } \mathbf{x} \in \mathbb{R}^k$$

where $B(\mathbf{x}, r)$ denote the open ball at point $\mathbf{x}$ with radius r, and λ is the Lebesgue measure on $\mathbb{R}^k$.

The following part will prove the lemma and a stronger result

$$\lim_{r\to 0} \frac{1}{\lambda\left[B(\mathbf{x},r)\right]} \int_{B(\mathbf{x},r)} |f(\mathbf{t}) - f(\mathbf{x})|\mathrm{d}\mathbf{t} = 0 \text{ for almost every } \mathbf{x} \in \mathbb{R}^k$$

which gives the above because

$$\begin{aligned}
\lim_{h\to 0}\left|\frac{\int_x^{x+h} f(t)\mathrm{d}t}{h} - f(x)\right| &= \lim_{h\to 0}\frac{1}{h}\left|\int_x^{x+h} f(t)\mathrm{d}t - \int_x^{x+h} f(x)\mathrm{d}t\right| \\
&\leq \lim_{h\to 0}\frac{\int_x^{x+h} |f(t) - f(x)|\mathrm{d}t}{h} \\
&\leq \lim_{h\to 0}\frac{\int_{x-h}^{x+h} |f(t) - f(x)|\mathrm{d}t}{h} \\
&= \lim_{r\to 0} 2\frac{\int_{B(x,r)} |f(t) - f(x)|\mathrm{d}t}{\lambda\left[B(x,r)\right]}
\end{aligned}$$

We begin with an auxiliary definition.

Definition: The Hardy-Littlewood maximal function
Let $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \lambda)$ be the Lebesgue measure space and $f : \mathbb{R}^k \mapsto \mathbb{R}$ be an $\mathcal{L}^1$ function, where $k \in \mathbb{N}^*$. The Hardy-Littlewood maximal function of f is defined as

$$f^*(\mathbf{x}) = \sup\left\{\frac{1}{\lambda(B)}\int_B |f(\mathbf{t})|\mathrm{d}\mathbf{t} : B \ni \mathbf{x} \text{ and is an open ball}\right\} : \mathbb{R}^k \mapsto \bar{\mathbb{R}}$$

Then define the set

$$I(y) := f^{*-1}((y, \infty]) = \{\mathbf{x} \in \mathbb{R}^k : f^*(\mathbf{x}) > y\}\ \forall y \in \bar{\mathbb{R}}$$

They have the following properties.

1. $I(y)$ is open for any $y \in \bar{\mathbb{R}}$
2. f^* is measurable with regard to $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \lambda)$
3. $\forall y \in \mathbb{R}^+$,
$$\lambda\left[I(y)\right] \leq \frac{3^k}{y}\int_{\mathbb{R}^k} |f(\mathbf{x})|\mathrm{d}\mathbf{x}$$

4. $f^* \overset{a.s.}{<} +\infty$

Proofs:

First, $I(+\infty) = \emptyset$ and $I(-\infty) = \mathbb{R}^k$ are open. Fix an arbitrary $y \in \mathbb{R}$. When $I(y) \neq \emptyset$ (otherwise trivial), for any $\mathbf{x} \in I(y)$, there exists an open ball B such that $\mathbf{x} \in B$ and

$$y < \frac{1}{\lambda(B)} \int_B |f(\mathbf{t})| \mathrm{d}\mathbf{t} \leq f^*(\mathbf{x})$$

Then, $B \subset I(y)$ because $f^*(\mathbf{x}') > y$ for any $\mathbf{x}' \in B$. Therefore, $I(y)$ is open.

2 follows because $\forall A \in \mathcal{R} := \{(y, \infty] : y \in \bar{\mathbb{R}}\}$, we have $f^{*-1}(A)$ open and thus $f^{*-1}(A) \in \mathcal{B}(\mathbb{R}^k)$. Note the fact that $\mathcal{B}(\bar{\mathbb{R}}) = \sigma(\mathcal{R})$.

We need the following Lemma to prove 3. 4 follows from 3 because $\lim_{y\to\infty} \lambda[I(y)] = 0$ from the inequality.

Lemma: Let $\{B_i \in \mathbb{R}^k\}_{i=1}^n$ be n open balls. Then there exists a mutually disjoint sub-collection $\{B_{i_j}\}_{j=1}^m$ with index set $\{i_j \in \{1, 2, \cdots, n\}\}_{j=1}^m$ such that

$$\lambda\left(\bigcup_{i=1}^n B_i\right) \leq 3^k \sum_{j=1}^m \lambda(B_{i_j})$$

Proof:

We construct the sub-collection using recursion. First, pick B_{i_1} (say, $B(\mathbf{x}_1, r_1)$) with the largest radius. Next, remove all the open balls B_i if $B_i \subset B(\mathbf{x}_1, 3r_1)$. Thus, all the remaining open balls, having radius $r_i \leq r_1$, will be disjoint from B_{i_1}. We also have $3^k\lambda(B(\mathbf{x}_1, r_1)) = \lambda(B(\mathbf{x}_1, 3r_1))$ (proved by definition).

Repeat the two steps if there are still balls remaining, i.e. for the j-th recursion, we pick B_{i_j} with the largest radius in the remaining set. We then have

$$\lambda\left(\bigcup_{i=1}^n B_i\right) \leq \lambda\left(\bigcup_{j=1}^m B(\mathbf{x}_j, 3r_j)\right) \leq \sum_{j=1}^m \lambda[B(\mathbf{x}_j, 3r_j)] = \sum_{j=1}^m 3^k\lambda(B_{i_j})$$

Now we continue to prove property 3 of the Hardy-Littlewood maximal function. Fix an arbitrary $y \in \mathbb{R}^+$ and pick a compact set $A \subset I(y)$ (We only discuss the non-trival case that $I(y) \neq \emptyset$). As the proof of 1, for any $\mathbf{x} \in A$, there exists an open ball $B_\mathbf{x}$ such that $\mathbf{x} \in B_\mathbf{x}$ and $B_\mathbf{x} \subset I(y)$. From the definition of compact subset in Topology, we also have

$$A \subset \bigcup_{\mathbf{x}\in A} B_\mathbf{x} \ \Rightarrow \ A \subset \bigcup_{i=1}^n B_\mathbf{x}, \text{ where } \{\mathbf{x}_i : i \in \{1, 2, \cdots, n\}\} \subset A$$

Then, the above lemma implies

$$\lambda(A) \leq \lambda\left(\bigcup_{i=1}^{n} B_{\mathbf{x}_i}\right) \leq 3^k \sum_{j=1}^{m} \lambda(B_j)$$

where $\{B_j\}_{j=1}^m$ are mutually disjoint and $\{B_j\}_{j=1}^m \subset \{B_{\mathbf{x}_i}\}_{i=1}^n$. Further,

$$\lambda(A) \leq 3^k \sum_{j=1}^{m} \lambda(B_j) < 3^k \sum_{j=1}^{m} \frac{1}{y} \int_{B_j} |f(\mathbf{t})| \mathrm{d}\mathbf{t} \leq \frac{3^k}{y} \int_{\mathbb{R}^k} |f(\mathbf{t})| \mathrm{d}\mathbf{t}$$

The second inequality is given by

$$\frac{1}{\lambda(B_j)} \int_{B_j} |f(\mathbf{t})| \mathrm{d}\mathbf{t} > y \ \forall 1 \leq j \leq m$$

Thus, $\lambda(A) < \frac{3^k}{y} \int_{\mathbb{R}^k} |f(\mathbf{t})| \mathrm{d}\mathbf{t}$ holds for all compact set $A \in I(y)$. Since $\lambda[I(y)]$ is the supremum of $\lambda(A)$, the result holds.

Lemma: approaching integrable function with continuous function
Let $f : \mathbb{R}^k \mapsto \mathbb{R} \in \mathcal{L}^1$ with regard to the Lebesgue measure space. Then,

1. Given a subset $A \in \mathcal{B}(\mathbb{R}^k)$ that $\lambda(A) < \infty$, for any $\epsilon > 0$, there exists a simple function $\phi \in \mathcal{E} : A \mapsto \mathbb{R}$ such that $\int_A |f - \phi| \mathrm{d}\lambda < \epsilon$
2. Continued from 1, there exists a continuous function $g : \mathbb{R}^k \mapsto \mathbb{R}$ such that $\int_{\mathbb{R}^k} |\phi - g| \mathrm{d}\lambda < \epsilon$ and thus $\int_{\mathbb{R}^k} |f - g| \mathrm{d}\lambda < 2\epsilon$

Proof:
$f \in \mathcal{L}^1$ gives f to be measurable and bounded almost surely. We use the construction in the Measurability theorem in 4.1.1 to prove 2. After discarding the set on which f is unbounded and partitioning f into positive and negative parts, we can show that there exists a sequence of simple function $\{\phi_n \in \mathcal{E}\}_{n=1}^{\infty}$ that

$$\forall n \overset{a.s.}{>} |f|, \ |f - \phi_n| \overset{a.s.}{\leq} 2^{-n} \text{ and } |\phi_n| < |f|, f\phi_n \geq 0$$

Next, for any $\epsilon > 0$, there exists an open ball $B := B(\mathbf{0}, r)$ with large enough r such that $\int_{\mathbb{R}^k \setminus B} |f| \mathrm{d}\lambda < \frac{\epsilon}{2}$ and $\int_B |f| \mathrm{d}\lambda < \infty$ whenever $f \in \mathcal{L}^1$ (proved by contradiction). Then, we take a simple function $\phi \in \mathcal{E}$ such that

$$|f(\mathbf{x}) - \phi_n(\mathbf{x})| \overset{a.s.}{\leq} \frac{\epsilon}{2\lambda(B)}$$

It gives

$$\int_{\mathbb{R}^k} |f-\phi| \mathrm{d}\lambda = \int_B |f-\phi| \mathrm{d}\lambda + \int_{\mathbb{R}^k \setminus B} |f-\phi| \mathrm{d}\lambda < \int_B \frac{\epsilon}{2\lambda(B)} \mathrm{d}\lambda + \int_{\mathbb{R}^k \setminus B} |f| \mathrm{d}\lambda < \epsilon$$

Next we prove 2. Let $\phi = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$, where $a_i \in \mathbb{R}$ and $\{A_i \in \mathcal{B}(\mathbb{R}^k)\}_{i=1}^n$ are mutually disjoint. Since

$$\int_{\mathbb{R}^k} |\phi - g| \mathrm{d}\lambda = \int_{\mathbb{R}^k} \left| \sum_{i=1}^n a_i \left(\mathbf{1}_{A_i} - g_i\right) \right| \mathrm{d}\lambda \le \sum_{i=1}^n |a_n| \int_{\mathbb{R}^k} |\mathbf{1}_{A_i} - g_i| \mathrm{d}\lambda \text{ where } g = \sum_{i=1}^n b_n g_i$$

It suffices to prove that for any $\epsilon > 0$ (which can be taken to be $\frac{\epsilon'}{|a_i|}$), there exists a continuous function $h : \mathbb{R}^k \mapsto [0,1]$ such that $\int_{\mathbb{R}^k} |h - \mathbf{1}_A| \mathrm{d}\lambda < \epsilon$, where $A \in \mathcal{B}(\mathbb{R}^k)$. We further let $A \in \mathcal{B}(\mathbb{R}^k) \setminus \{\emptyset, \mathbb{R}^k\}$ since those two cases are trivial.

First, the property of Lebesgue measure gives that $\forall \epsilon > 0$, there exist a close set $C \subset A$ and a open set $D \supset A$ such that $C, D \in \mathcal{B}(\mathbb{R}^k)$ and

$$\lambda(D \setminus A) < \frac{\epsilon}{2}, \lambda(A \setminus C) < \frac{\epsilon}{2}$$

Then, C and D^c are both close and disjoint. There exists a continuous function g such that $\lambda(\{\mathbf{x} : h(\mathbf{x}) \neq \mathbf{1}_A\}) \le \lambda(D \setminus C) < \epsilon$ with the following construction

$$h(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in C \\ 0, & \mathbf{x} \in D^c \\ \frac{d(\mathbf{x},C)}{d(\mathbf{x},C)+d(\mathbf{x},D^c)}, & \text{otherwise} \end{cases}$$

$$\text{where } d(\mathbf{x}, K) := \inf\{d(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in K\} > 0 \ \forall \mathbf{x} \in K^c \text{ when } K \text{ is close}$$

The proof of h being continuous is omitted. In fact this construction is a special case of Urysohn's lemma in Topology. Then,

$$\int_{\mathbb{R}^k} |h - \mathbf{1}_A| \mathrm{d}\lambda = \int_{\{\mathbf{x}: h(\mathbf{x}) \neq \mathbf{1}_A\}} |h - \mathbf{1}_A| \mathrm{d}\lambda < \int_{\{\mathbf{x}: h(\mathbf{x}) \neq \mathbf{1}_A\}} 1 \mathrm{d}\lambda < \epsilon$$

which complete the proof.

Proof of lemma

$$f(\mathbf{x}) = \lim_{r \to 0} \frac{\int_{B(\mathbf{x},r)} f(\mathbf{t}) \mathrm{d}\mathbf{t}}{\lambda\left[B(\mathbf{x},r)\right]} \text{ for almost every } \mathbf{x} \in \mathbb{R}^k$$

given f integrable locally.

First, we verify that it holds for any almost surely continuous function f. For any $\epsilon > 0$ and almost every $\mathbf{x} \in \mathbb{R}^k$, there exists r_0 such that $f|_{B(\mathbf{x},r_0)} \in \mathcal{L}^1$ and $|f(\mathbf{t}) - f(\mathbf{x})| < \epsilon \ \forall \mathbf{t} \in B(\mathbf{x}, r_0)$. Then, for any $r \le r_0$, we have

$$\left| \frac{\int_{B(\mathbf{x},r)} f \mathrm{d}\lambda}{\lambda[B(\mathbf{x},r)]} - f(\mathbf{x}) \right| \le \frac{1}{\lambda[B(\mathbf{x},r)]} \int_{B(\mathbf{x},r)} |f(\mathbf{t}) - f(\mathbf{x})| \mathrm{d}\mathbf{t} < \frac{\int_{B(\mathbf{x},r)} \epsilon \mathrm{d}\mathbf{t}}{\lambda[B(\mathbf{x},r)]} = \epsilon$$

Thus the result follows. We move on to the general function case.

For any $\mathbf{x} \in \mathbb{R}^k$, take the open ball $B := B(\mathbf{x}, r)$ such that $f|_B \in \mathcal{L}^1$. We know from the above lemma that for any $\epsilon_1 > 0$, there exists a continuous function $g : \mathbb{R}^k \mapsto \mathbb{R} \in \mathcal{L}^1$ such that $\int_{\mathbb{R}^k} |f\mathbf{1}_B - g| \mathrm{d}\lambda < \epsilon_1$. Then,

$$g(\mathbf{x}) = \lim_{r \to 0} \frac{\int_{B(\mathbf{x},r)} g(\mathbf{t}) \mathrm{d}\mathbf{t}}{\lambda\left[B(\mathbf{x}, r)\right]} \text{ for almost every } \mathbf{x} \in \mathbb{R}^k$$

We also have (letting $h = |f\mathbf{1}_B - g|$)

$$\begin{aligned}
\left| \frac{\int_B f \mathrm{d}\lambda}{\lambda(B)} - f(\mathbf{x}) \right| &\leq \left| \frac{\int_B f\mathbf{1}_B \mathrm{d}\lambda}{\lambda(B)} - \frac{\int_B g \mathrm{d}\lambda}{\lambda(B)} \right| + \left| \frac{\int_B g \mathrm{d}\lambda}{\lambda(B)} - g(\mathbf{x}) \right| + \left| g(\mathbf{x}) - f(\mathbf{x})\mathbf{1}_B(\mathbf{x}) \right| \\
&\leq \frac{1}{\lambda(B)} \int_B |f\mathbf{1}_B - g| \mathrm{d}\lambda + \left| \frac{\int_B g \mathrm{d}\lambda}{\lambda(B)} - g(\mathbf{x}) \right| + \left| g(\mathbf{x}) - f(\mathbf{x})\mathbf{1}_B(\mathbf{x}) \right| \\
&\leq h^*(\mathbf{x}) + \left| \frac{\int_B g \mathrm{d}\lambda}{\lambda(B)} - g(\mathbf{x}) \right| + |h| \\
&\forall \mathbf{x} \in \mathbb{R}^k, \ (\mathbf{1}_B \text{ can be always added since we take } B \text{ according to } \mathbf{x})
\end{aligned}$$

where $h^*(\mathbf{x})$ is the Hardy-Littlewood maximal function of h. From its property 3, we have that for any $\delta > 0$,

$$\lambda\left(\left\{\mathbf{x} : h^*(\mathbf{x}) > \frac{\delta}{3}\right\}\right) \leq \frac{3^{k+1}}{\delta} \int_{\mathbb{R}^k} h \mathrm{d}\lambda < \frac{3^{k+1}\epsilon_1}{\delta}$$

We also have

$$\lambda\left(\left\{\mathbf{x} : h > \frac{\delta}{3}\right\}\right) \leq \frac{3}{\delta} \int_{\mathbb{R}^k} h \mathrm{d}\lambda < \frac{3\epsilon_1}{\delta}$$

by Markov's inequality. And there exists ϵ_2 such that whenever $r < \epsilon_2$,

$$\lambda\left(\left\{\mathbf{x} : \left| \frac{\int_{B(\mathbf{x},r)} g \mathrm{d}\lambda}{\lambda[B(\mathbf{x}, r)]} - g(\mathbf{x}) \right| > \frac{\delta}{3}\right\}\right) = 0$$

Therefore,

$$\lambda\left(\left\{\left| \frac{\int_{B(\mathbf{x},r)} f \mathrm{d}\lambda}{\lambda[B(\mathbf{x}, r)]} - f(\mathbf{x}) \right| > \delta\right\}\right) < \frac{3^k + 1}{\delta} \times \epsilon_1 \ \forall r < \epsilon_2$$

For any $\delta > 0$, we can take ϵ_1 small enough and find ϵ_2 such that

$$r < \epsilon_2 \ \Rightarrow \ \left| \frac{\int_{B(\mathbf{x},r)} f \mathrm{d}\lambda}{\lambda[B(\mathbf{x}, r)]} - f(\mathbf{x}) \right| \leq \delta \text{ for almost every } \mathbf{x} \in \mathbb{R}^k$$

which gives the result.

Proof of the fundamental theorem of calculus

It suffices to prove that if f is integrable locally,

$$\lim_{r \to 0} \frac{1}{\lambda\left[B(\mathbf{x}, r)\right]} \int_{B(\mathbf{x},r)} |f(\mathbf{t}) - f(\mathbf{x})| \mathrm{d}\mathbf{t} = 0 \text{ for almost every } \mathbf{x} \in \mathbb{R}^k$$

Let $\epsilon > 0$ and take $q_{\mathbf{x}} \in \mathbb{Q}$ such that $|f(\mathbf{x}) - q_{\mathbf{x}}| < \frac{\epsilon}{2}$ and $|f(\mathbf{x}) - q_{\mathbf{x}}|$ is integrable locally. We have

$$\frac{1}{\lambda\left[B(\mathbf{x},r)\right]}\int_{B(\mathbf{x},r)}|f(\mathbf{t}) - f(\mathbf{x})|\mathrm{d}\mathbf{t} \leq \frac{1}{\lambda\left[B(\mathbf{x},r)\right]}\int_{B(\mathbf{x},r)}|f(\mathbf{t}) - q_{\mathbf{x}})|\mathrm{d}\mathbf{t} + |f(\mathbf{x}) - q_{\mathbf{x}}|$$

The last lemma shows that there exists a set $A(q_{\mathbf{x}})$ such that $\lambda\left[A(q_{\mathbf{x}})^c\right] = 0$ and $\forall \mathbf{x} \in A(q_{\mathbf{x}})$,

$$|f(\mathbf{x}) - q_{\mathbf{x}}| = \lim_{r\to 0}\frac{1}{\lambda\left[B(\mathbf{x},r)\right]}\int_{B(\mathbf{x},r)}|f(\mathbf{t}) - q_{\mathbf{x}}|\mathrm{d}\mathbf{t}$$

Thus, when r is close enough to 0,

$$\frac{1}{\lambda\left[B(\mathbf{x},r)\right]}\int_{B(\mathbf{x},r)}|f(\mathbf{t}) - f(\mathbf{x})|\mathrm{d}\mathbf{t} \leq 2|f(\mathbf{x}) - q_{\mathbf{x}}| < \epsilon$$

Take $A \in \mathcal{B}(\mathbb{R}^k)$ such that

$$A^c = \bigcup_{\mathbf{x}\in\mathbb{R}^k} A(q_{\mathbf{x}})^c = \bigcup_{q\in\mathbb{Q}} A(q)^c \text{ with 0 measure}$$

We have

$$\lim_{r\to 0}\frac{1}{\lambda\left[B(\mathbf{x},r)\right]}\int_{B(\mathbf{x},r)}|f(\mathbf{t}) - f(\mathbf{x})|\mathrm{d}\mathbf{t} = 0 \ \forall \mathbf{x} \in A$$

5.1.3 Absolute continuous and differentiable

Definition: absolute continuity of function

Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a function. f is defined to be absolute continuous on a interval $I \subset \mathbb{R}$ if for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\sum_{i=1}^{n}|f(b_i) - f(a_i)| < \epsilon \text{ whenever } \sum_{i=1}^{n} b_i - a_i < \delta$$

where $\{(a_i, b_i) \subset I\}_{i=1}^n$ are n disjoint open intervals.

It is easy to check that absolute continuous implies uniform continuous by taking $n = 1$. It is a stronger property than uniform continuity and continuity.

Proposition: equivalent definition of absolute continuous function

If a function $f : I \mapsto \mathbb{R}$ is absolute continuous, for any $\epsilon > 0$, there exists $\delta > 0$ such that any countable sequence of disjoint open intervals $\{(a_i, b_i) \subset I\}_{i=1}^\infty$ which $\sum_{i=1}^\infty (b_i - a_i) < \delta$, we have $\sum_{i=1}^n |f(b_i) - f(a_i)| < \epsilon$.

This property replaces the finite sequence condition with countably infinite. It is obvious that functions with this property also satisfy the almost continuity condition. Therefore it is an equivalent definition.

Proof:

Take $\epsilon > 0$ and $\delta > 0$ such that any finite $\{(a_i', b_i') \subset I\}_{i=1}^n$ satisfies

$$\sum_{i=1}^n |f(b_i') - f(a_i')| < \frac{\epsilon}{2} \text{ if } \sum_{i=1}^n b_i' - a_i' < \delta$$

Then, for any countable $\{(a_i, b_i) \subset I\}_{i=1}^\infty$, such that $\sum_{i=1}^\infty (b_i - a_i) < \delta$, we have

$$\forall N \in \mathbb{N}^*, \sum_{i=1}^N b_i - a_i < \delta \Rightarrow \sum_{i=1}^N |f(b_i) - f(a_i)| < \frac{\epsilon}{2}$$

Take $N \to \infty$, we have

$$\sum_{i=1}^N |f(b_i) - f(a_i)| \leq \frac{\epsilon}{2} < \epsilon$$

giving the result.

Example:

Let $f : [a, b] \mapsto \mathbb{R}$ be an $\mathcal{L}^1$ function with regard to Lebesgue measure, then the function $F(x) := c + \int_a^x f(t)\mathrm{d}t : [a, b] \mapsto \mathbb{R}$, where $c \in \mathbb{R}$ is a constant, is absolute continuous on $\mathbb{R}$.

Proof:

Since

$$\sum_{i=1}^n |F(b_i) - F(a_i)| = \sum_{i=1}^n \left| \int_a^x f(t)\mathrm{d}t \right| \leq \sum_{i=1}^n \int_a^x |f(t)|\mathrm{d}t$$

It suffices to prove that for any $\epsilon > 0$, there exists $\delta > 0$ such that whenever $A \in \mathcal{B}([a, b])$ and $\lambda(A) < \delta$, $\int_A |f|\mathrm{d}\lambda < \epsilon$, where λ is the Lebesgue measure.

Since $|f| \in \mathcal{L}^+$, by the measurability theorem in 4.1.1, there exists a sequence of simple function $\{\phi_n \in \mathcal{E}^+\}_{n=1}^\infty$ such that $\lim_{n\to\infty} |f| - \phi_n = 0$ and $|f| \geq \phi_{n+1} \geq \phi_n \ \forall n \in \mathbb{N}^*$. Since $|f| \in \mathcal{L}^1$, by dominated convergence theorem,

$$\lim_{n\to\infty} \int_{[a,b]} f - \phi_n \mathrm{d}\lambda = \int_{[a,b]} \lim_{n\to\infty} f - \phi_n \mathrm{d}\lambda = 0$$

Therefore, for any $\epsilon > 0$, there exists a simple function ϕ choosing from the sequence that $\int_{[a,b]}(f - \phi)\mathrm{d}\lambda < \frac{\epsilon}{2}$. We can then easily choose an $\delta > 0$ such that $\int_A \phi \mathrm{d}\lambda < \frac{\epsilon}{2}$ whenever $A < \delta$. The results $\int_A |f|\mathrm{d}\lambda = \int_A \phi \mathrm{d}\lambda + \int_A (|f| - \phi)\mathrm{d}\lambda < \epsilon$ follows.

Theorem: Integrable derivative implies absolute continuity

Suppose a continuous function $f : [a, b] \mapsto \mathbb{R}$ is differentiable everywhere on (a, b), and the derivative

$|f'| < \infty$ and $f' \in \mathcal{L}(\lambda)^1$, where λ is the Lebesgue measure in $\mathbb{R}$, then f is absolute continuous on $[a, b]$, and

$$f(x) = f(a) + \int_a^x f'(t)\mathrm{d}t \ \forall x \in [a, b]$$

A interesting prove is based on the following definition and some lemmas.

Definition: Derived number

Let $f : I \mapsto \mathbb{R}$ be a function defined on an open interval $I = (a, b) \subset \mathbb{R}$. Fix a point $x_0 \in I$, γ is a derived number of f on x_0 if there exists a sequence $\{a_n \in \mathbb{R}\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} a_n = 0$ and $\gamma = \lim_{n\to\infty} \frac{f(x_0+a_n)-f(x_0)}{a_n} \in [-\infty, \infty]$. Denote $\mathrm{D}f(x_0)$ be the set of all derived number on x_0, which is the union of all limit point of $\{\frac{f(x_0+a_n)-f(x_0)}{a_n}\}_{n=1}^{\infty}$ for all sequence $\{a_n \in \mathbb{R}\}_{n=1}^{\infty}$.

If f is differentiable at x_0, then $\mathrm{D}f(x_0) = \{f'(x_0)\}$.

Lemma:

If $\forall x \in (a, b)$, $\gamma \geq 0 \ \forall \gamma \in \mathrm{D}f(x)$, then f is non-decreasing on (a, b). Proof:

Fix $\epsilon > 0$. Let $g(x) := f(x) + \epsilon x$ We prove that g is non-decreasing by contradiction. Suppose there exists x, y such that $a < x < y < b$ and $g(x) > g(y)$. Take $z := \frac{x+y}{2}$, then one of $g(x) - g(z)$ and $g(z) - g(y)$ is positive. Take $[\alpha_1, \beta_1]$ be either the interval $[x, z]$ or $[z, y]$ on which $g(\alpha_1) - g(\beta_1) > 0$. Repeat this process on $[\alpha_1, \beta_1]$ instead of $[x, y]$, we can obtain $[\alpha_2, \beta_2]$, then recursively $\{[\alpha_i, \beta_i]\}_{i=1}^{\infty}$ such that $g(\alpha_i) - g(\beta_i) > 0$, $\beta_i - \alpha_i > 0$ and $[\alpha_i, \beta_i] \subset [\alpha_j, \beta_j]$ for all $1 \leq j \leq i$.

Due to our construction, $\bigcap_{i=1}^{\infty} [\alpha_i, \beta_i]$ contains only a unique point, say x_0. For any $i \in \mathbb{N}^*$ one of $g(\alpha_i) - g(x_0)$ and $g(x_0) - g(\beta_i)$ is positive. Let $a_i = \beta_i - x_0$ if $g(x_0) - g(\beta_i) > 0$, and $a_i = \alpha_i - x_0$ otherwise. Then,

$$a_i \neq 0, \ \frac{g(x_0 + a_i) - g(x_0)}{a_i} < 0 \ \forall i \in \mathbb{N}^*$$

Observe that $\lim_{i\to\infty} a_i = 0$ since it is bounded by $\{\alpha_i - x_0\}_{i=1}^{\infty}$ and $\{\beta_i - x_0\}_{i=1}^{\infty}$. There exists a subsequence $\{a_{i_n}\}_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} \frac{g(x_0 + a_{i_n}) - g(x_0)}{a_{i_n}} \leq 0$$

which is impossible because for any $x \in (a, b)$, $\inf\{\mathrm{D}g(x)\} = \inf\{\gamma + \epsilon, \gamma \in \mathrm{D}f(x)\} \geq \epsilon$.

Therefore, we have shown that g is non-decreasing, i.e., $f(x)+\epsilon x \leq f(y)+\epsilon y$ whenever $a < x < y < b$. Taking $\epsilon \to 0$, $f(x) \leq f(y)$ and the result follows.

Lemma:

If for almost every $x \in (a, b)$ with respect to Lebesgue measure λ, $\gamma \geq 0 \ \forall \gamma \in \mathrm{D}f(x)$ and $-\infty$ does not appear as a derived number throughout (a, b), then f is also non-decreasing on (a, b).

Proof:

Let $A \in \mathbb{B}((a,b))$ such that $\lambda(A) = 0$ and $\forall x \in A,\ \exists \gamma \in \mathrm{D}f(x)$ and $\gamma < 0$. We will prove that there exists a function $h : (a,b) \mapsto \mathbb{R}$ such that h is continuous, non-decreasing, non-negative, bounded and $h'(x) = +\infty\ \forall x \in A$. Using this result, let

$$g(x) = f(x) + \epsilon h(x)\ \forall x \in (a,b) \text{ for some } \epsilon > 0$$

We have that $\inf\{\mathrm{D}g(x)\} \geq 0\ \forall x \in (a,b)$, because $\mathrm{D}g(x) = \{+\infty\}\ \forall x \in A$ and $\inf\{\mathrm{D}g(x)\} \geq \inf\{\mathrm{D}f(x)\} \geq 0\ \forall x \in (a,b) \setminus A$. By the previous lemma, g is non-decreasing. Taking $\epsilon \to 0$, so is f.
We now construct such h. Let $\{B_n\}_{n=1}^{\infty}$ be some open sets such that

$$\forall n \in \mathbb{N}^*, B_n \supset A \text{ and } \lambda(B_n) \leq \frac{1}{2_n}$$

For any $n \in \mathbb{N}^*$, let

$$h_n(x) := \lambda(B_n \cap [a,x])\ \forall x \in \mathbb{N}^*$$

be a function. Observe that it is non-decreasing and the image is within $[0, \frac{1}{2^n})$. We can show that h_n is also continuous on (a,b) using the continuity of measure. Then define

$$h = \sum_{n=1}^{\infty} h_n$$

which is also non-decreasing, non-negative and bounded.
h is also continuous. for any point $x \in (a,b)$, let $\{a_i \in (a,b)\}_{i=1}^{n}$ be a sequence with limit x. We first have $\lim_{i\to\infty} h_n(a_i) = h_n(x)\ \forall n \in \mathbb{N}^*$. Since $h_n(a_i)$ is bounded by $\frac{1}{2^n}$,

$$\lim_{i\to\infty} h(a_i) = \lim_{i\to\infty} \sum_{n=1}^{\infty} h_n(a_i) = \sum_{n=1}^{\infty} \lim_{i\to\infty} h_n(a_i) = h(x)$$

by dominated convergence theorem using counting measure.
It remains to prove that h has derivative $+\infty$ on every point in A. For any $x \in A$, $x \in G_n\ \forall n \in \mathbb{N}^*$. Note that $\{G_n\}_{n=1}^{\infty}$ are open, for any $N \in \mathbb{N}^*$, there exists $h > 0$ such that $x + h \in (a,b) \cap \bigcap_{n=1}^{N} G_n$. Thus, also due to h_n being non-decreasing,

$$\frac{h(x+h) - h(x)}{h} \geq \sum_{i=1}^{N} \frac{h_n(x+h) - h_n(x)}{h} = \sum_{i=1}^{N} \frac{\lambda((x, x+h])}{h} = N$$

We can conclude that $h'(x) = \lim_{h\to 0} \frac{h(x+h)-h(x)}{h} = \infty$.

We are ready to prove the theorem.
Proof of theorem: Integrable derivative implies absolute continuity:
'For any $n \in \mathbb{N}^*$, construct

$$g_n := \min\{n, f'\} \text{ and } G_n(x) := f(x) - \int_a^x g_n(t)\mathrm{d}t\ \forall x \in [a,b]$$

By linearity of derivative and the fundamental theorem of calculus, $G_n'(x) \overset{a.s.}{=} f'(x) - g_n \overset{a.s.}{\geq} 0$. Therefore, $\inf\{\mathrm{D}G_n(x)\} \geq 0$ for x almost everywhere in (a, b). Also, if $\lambda \in \mathrm{D}G_n(x)$ for some $x \in (a, b)$, we can prove that $\lambda \geq f'(x) - n > -\infty$. Using the above lemma, G_n is non-decreasing on (a, b). Since G_n is continuous on $[a, b]$, $G_n(b) \geq G_n(a)$ and thus

$$f(b) - f(a) = g(b) + \int_a^x g_n(t) - g(a) \geq \int_a^b g_n(t)\mathrm{d}t \ \forall n \in \mathbb{N}^*$$

Using dominated convergence theorem, $f(b) - f(a) \geq \int_a^b f'(t)\mathrm{d}t$. Replace f with $-f$, we can get $f(b) - f(a) \leq \int_a^b f'(t)\mathrm{d}t$. Thus, substituting b with any $x \in (a, b]$, we have

$$f(x) = f(a) + \int_a^x f'(t)\mathrm{d}t$$

which is absolute continuous by the example. The above holds obviously for $x = a$.

Further, we are interesting that under what condition can a function be differentiable (almost everywhere) and satisfy above.

Theorem: absolute continuous implies differentiable almost everywhere
If $f : [a, b] \mapsto \mathbb{R}^k$ is absolute continuous, where $a, b \in \mathbb{R}$, then

$$\forall x \in [a, b],\ f(x) - f(a) = \int_a^x f'(t)\mathrm{d}t$$

Observe that we have proved its converse.
We prove a special case that f is non-decreasing in 5.2.2., which is sufficient in our context since we mostly differentiate non-decreasing *pdf.* For a general absolute continuous f, one can prove that it can be written as the difference of two non-decreasing functions, $f_1 - f_2$, and the theorem then holds.

5.2 Derivative of measure

5.2.1 Radon–Nikodym theorem

The following definition extends the concept of absolute continuous and differentiable into an arbitrary measure.
Definition: Absolute continuous measure
Given two measures, μ and τ, on a measurable space $(S, \mathcal{B})$, we say τ is absolute continuous to with

respect to μ, written $\tau \ll \mu$ if

$$\forall A \in \mathcal{B} \text{ such that } \mu(A) = 0,\ \tau(A) = 0$$

Radon–Nikodym theorem Given two measures, μ and τ, on a measurable space $(S, \mathcal{B})$, if $\tau \ll \mu$ and τ is σ-finite, then there exists a function $f : S \to \mathbb{R}^+$ such that $f \in \mathcal{L}^+(\mu)$ and

$$\forall A \in \mathcal{B},\ \tau(A) = \int_A f \mathrm{d}\mu$$

Further, for any such choice of f, they are equal almost everywhere with regard to μ. f are the Radon–Nikodym derivative of τ with regard to μ, denoted as

$$f = \frac{\mathrm{d}\tau}{\mathrm{d}\mu}$$

We first verify that f is almost-surely-unique. Assume there exist f and g such that

$$\forall A \in \mathcal{B},\ \tau(A) = \int_A f \mathrm{d}\mu = \int_A g \mathrm{d}\mu \ \Rightarrow \ \int_A f - g \mathrm{d}\mu$$

Take $A_1 := \{x \in S : f(x) > g(x)\}$ and $A_2 := \{x \in S : f(x) < g(x)\}$. $\int_{A_1} f - g \mathrm{d}\mu = \int_{A_2} f - g \mathrm{d}\mu = 0$ gives $f - g \overset{a.s.}{=} 0$.

Its proof of existence requires to generate the concept of normal measure into signed measure and use the Hahn Decomposition. We state a simple corollary about Lebesgue measure before proceeding to its proof.

Corollary: Radon-Nikodym derivative and symmetric derivative

Let λ be the Lebesgue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and μ be a σ-finite measure on $(S, \mathcal{B}(S))$, where $k \in \mathbb{N}^*$ and $S \in \mathcal{B}(\mathbb{R}^k)$ is a subset. If $\mu \ll \lambda$ and the Radon-Nikodym derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is defined on an open set $S' \in \mathcal{B}(S)$, then for almost every $x \in S'$,

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x) = \lim_{r \to 0} \frac{1}{\lambda[B(x,r)]} \int_{B(x,r)} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t) \mathrm{d}\lambda = \lim_{r \to 0} \frac{\mu[B(x,r)]}{\lambda[B(x,r)]}, \text{ where } B(x,r) \text{ denote a open ball}$$

The first equation is given by $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \in \mathcal{L}^1$ and the fundamental theorem of calculus. The RHS is called the symmetric derivative of μ with respect to λ.

Conversely, given $\mu \ll \lambda$ and the symmetric derivative exists everywhere on a open set S', it is almost surely equal to any of its Radon-Nikodym derivative.

Definition: Signed measure

Let $(S, \mathcal{B})$ be a measurable space, a signed measure μ is a $\mathcal{B} \mapsto [-\infty, \infty]$ that satisfies

1. $\mu(\emptyset) = 0$

2. Either $\mu : \mathcal{B} \mapsto (-\infty, \infty]$ or $\mu : \mathcal{B} \mapsto [-\infty, \infty)$, i.e., the image only includes one of the infinity

3. For any sequence of disjoint subset $\{A_n \in \mathcal{B}\}_{n=1}^{\infty}$,

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

We also define a subset $A \in \mathcal{B}$ to be

1. a positive set, if $\forall B \subset A$ and $B \in \mathcal{B}$, $\mu(B) \geq 0$

2. a negative set, if $\forall B \subset A$ and $B \in \mathcal{B}$, $\mu(B) \leq 0$

3. a null set, if if $\forall B \subset A$ and $B \in \mathcal{B}$, $\mu(B) = 0$

Proposition: properties of signed measure

1. Continuity from below: If the $\{A_n\}_{n=1}^{\infty}$ is monotonic increasing, we have

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

2. Continuity from above: If the $\{A_n\}_{n=1}^{\infty}$ is monotonic decreasing, and $\mu(A_1)$ is finite, we have

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

3. Positive set is closed under countable union, i.e., if $\{A_n \in \mathcal{B}\}_{n=1}^{\infty}$ is a sequence of positive set, then $\bigcup_{n=1}^{\infty} A_n$ is also positive. The same holds for negative set.

Proof:
1 and 3 follow from the definition by constructing a mutually disjoint sequence $\{B_n := A_n \backslash B_{n-1}\}_{n=1}^{\infty}$ (setting $B_0 = \emptyset$). 3 also uses a simple fact that any measurable subset of a positive set is positive.
2 follows from 1 because

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1) - \mu\left(A_1 \setminus \bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1) - \mu\left(\bigcup_{n=1}^{\infty} A_1 \setminus A_n\right)$$

Then we can use the result of 1. Note that since $\mu(A_1)$ is finite, any of its measurable subset (e.g., $A_2, A_3, \cdots$) has finite measure, thus the equality holds.

Theorem: The Hahn Decomposition

Let μ be a signed measure on $(S, \mathcal{B})$. There exists a decomposition of $S = P \cup N$, where $P, N \in \mathcal{B}$ and $P \cap N = \emptyset$, such that P is a positive set and N is a negative set. (P, N) is called the Hahn Decomposition of signed measure space $(S, \mathcal{B}, N)$.

Proof:

We assume $\mu : \mathcal{B} \mapsto [-\infty, \infty)$, otherwise take $\mu := -\mu$ and the result is the same. Let

$$M := \sup \{\mu(A) : A \in \mathcal{B} \text{ is a positive set}\} \in [0, \infty]$$

Note that $\emptyset$ is positive thus $\{\mu(A) : A \in \mathcal{B} \text{ is a positive set}\}$ is non-empty. Since it is a subset of $\mathbb{R}$, there exists a sequence of positive set $\{P_i\}_{i=1}^{\infty}$ such that $\lim_{i\to\infty} \mu(P_i) = M$. Let $P := \bigcup_{i=1}^{\infty} P_i$, then P is positive (by property 3) and we can show $\mu(P) = M$ ($\mu(P) \leq M$ is obvious, while $\mu(P) \geq M$ because $\mu(P) \geq \mu(P_i)\ \forall i \in \mathbb{N}^*$ and $\lim_{i\to\infty} \mu(P_i) = M$). This also gives that $M < \infty$

We claim that $N := S \setminus P$ is a negative set. First, any nonempty subset $A \in \mathcal{B}$ of N is not positive, otherwise $P \cup A$ is positive and $\mu(P \cup A) = M + \mu(A) > M$.

We then prove by contradiction. If N is not negative, then there exists $A_1 \subset N$, $A_1 \in \mathcal{B}$ such that $\mu(A_1) > 0$. And there also exists $A_2 \subset A_1$, $A_2 \in \mathcal{B}$ that $\mu(A_2) > \mu(A_1)$, because A_1 is not positive and we can let $\mu(A_1 \setminus A_2) < 0$. Let

$$n_1 := \inf \left\{ N \in \mathbb{N}^* : \exists A_2 \subset A_1 \text{ and } A_2 \in \mathcal{B} \text{ such that } \frac{1}{N} < \mu(A_2) - \mu(A_1) \right\}$$

and pick one A_2 that $\frac{1}{n_1} < \mu(A_2) - \mu(A_1)$. Recursively, for $j \geq 2$, we always pick (n_j, A_{j+1}) such that $A_{j+1} \subset A_j$, $A_{j+1} \in \mathcal{B}$, $\frac{1}{n_j} < \mu(A_{j+1}) - \mu(A_j)$, and n_j is the smallest positive integer among all the other choices. Then we obtain a monotonic decreasing $\{A_j\}_{i=1}^{\infty}$ and $0 < \frac{1}{N_i} + \mu(A_j) < \mu(A_{j+1})\ \forall j \in \mathbb{N}^*$.

Let $A := \bigcap_{j=1}^{\infty} A_j$. Since $\mu(A_1) < \infty$, property 2 gives that

$$\mu(A) = \lim_{j\to\infty} \mu(A_j) \geq \mu(A_1) + \sum_{j=1}^{\infty} \frac{1}{n_j} > 0$$

A is non-empty because $\mu(A) > 0$. Then A is not positive because it is a subset of N. There exists a large enough $n \in \mathbb{N}^*$ and a measurable subset $B \subset A$ such that $\frac{1}{n} < \mu(B) - \mu(A)$.

Also, $\sum_{j=1}^{\infty} \frac{1}{n_j}$ converges and $\lim_{j\to\infty} n_j = +\infty$. Then we can pick $j_0 \in \mathbb{N}^*$ such that $n < n_{j_0}$. Because $B \subset A \subset A_{j_0}$, we have $\mu(B) > \mu(A) + \frac{1}{n} > \mu(A_{j_0}) + \frac{1}{n}$. This contradicts to the restriction that n_{j_0} is the minimum, i.e. we should pick (n, B) instead of (n_{j_0}, A_{j_0+1}). Note that $B \neq A_{j_0+1}$ because $\mu(B) > \mu(A_{j_0+1})$.

Proof of Radon–Nikodym theorem:

We first prove the case when τ and μ are finite. Construct the set of function

$$\mathcal{G} := \left\{ f \in \mathcal{L}^+(\mu) : \forall A \in \mathcal{B},\ \int_A f \mathrm{d}\mu \le \tau(A) \right\}$$

$\mathcal{G} \ne \emptyset$ because we can let f to be constant 0. We first show that $\mathcal{G}$ is closed under max operation. Suppose $f_1, f_2 \in \mathcal{G}$. Let $A_1 := \{x \in S : f_1(x) \ge f_2(x)\}$ $A_2 := \{x \in S : f_1(x) < f_2(x)\}$ such that $A = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$. We have $A_1, A_2 \in \mathcal{B}$ since f_1, f_2 and $f_1 - f_2$ are measurable. Then, for any $A \in \mathcal{B}$,

$$\int_A \max\{f_1, f_2\} \mathrm{d}\mu = \int_{A_1 \cap A} f_1 \mathrm{d}\mu + \int_{A_2 \cap A} f_2 \mathrm{d}\mu \le \tau(A_1) + \tau(A_2) = \tau(A)$$

which gives $\max(f_1, f_2) \in \mathcal{G}$. Further, let

$$\alpha := \sup_{f \in \mathcal{G}} \left\{ \int_S f \mathrm{d}\mu \right\}$$

Obviously, $\alpha \le \tau(S)$. We will show that there exists $f_0 \in \mathcal{G}$ such that $\int_S f_0 \mathrm{d}\mu = \alpha$. Let $\{f_n' \in \mathcal{G}\}_{n=1}^\infty$ such that $\lim_{n\to\infty} \int_S f_n' \mathrm{d}\mu = \alpha$. Then let $f_1 = f_1'$ and $f_n = \max\{f_n', f_{n-1}\}$ $\forall n \ge 2$. We have $\{f_n \in \mathcal{G}\}_{n=1}^\infty$ monotonic increasing and $\lim_{n\to\infty} \int_S f_n \mathrm{d}\mu = \lim_{n\to\infty} \int_S f_n' \mathrm{d}\mu = \alpha$. By monotone convergence theorem, we have

$$\tau(S) \ge \alpha = \lim_{n\to\infty} \int_S f_n \mathrm{d}\mu = \int_S \lim_{n\to\infty} f_n \mathrm{d}\mu$$

and $f_0 := \lim_{n\to\infty} f_n \in \mathcal{G}$ because

$$\forall A \in \mathcal{B},\ \int_A \lim_{n\to\infty} f_n \mathrm{d}\mu = \lim_{n\to\infty} \int_A f_n \mathrm{d}\mu \le \tau(A)$$

The following prove that $\tau(A) - \int_A f_0 \mathrm{d}\mu = 0$ $\forall A \in \mathcal{B}$. Let $v(A) := \tau(A) - \int_A f_0 \mathrm{d}\mu = 0$ be an unsigned measure on $(S, \mathcal{B})$, whose σ-additivity can be verified by dominated convergence theorem. We prove $v(X) = 0$ by contradiction. If $v(X) > 0$, since $\mu(X) < \infty$, there exists $\epsilon > 0$ such that $v(X) - \epsilon\mu(X) > 0$. Let $u := v - \epsilon\mu$ be a signed measure on $(S, \mathcal{B})$. And let $S = P \cup Q$ be the Hahn Decomposition. For any $A \in \mathcal{B}$, $u(A \cap P) = v(A \cap P) - \epsilon\mu(A \cap P) \ge 0$, then

$$\tau(A) = v(A) + \int_A f_0 \mathrm{d}\mu \ge v(A \cap P) + \int_A f_0 \mathrm{d}\mu \ge \epsilon\mu(A \cap P) + \int_A f_0 \mathrm{d}\mu = \int_A f_0 + \epsilon \mathbf{1}_P \mathrm{d}\mu$$

It shows that $(f_0 + \epsilon\mathbf{1}_P) \in \mathcal{G}$. Since $\alpha = \int_S f_0 \mathrm{d}\mu$ is the supremum of $\{\int_S f \mathrm{d}\mu : f \in \mathcal{G}\}$, $\int_S (f_0 + \epsilon\mathbf{1}_P) \mathrm{d}\mu \le \alpha$ and thus $\mu(S \cap P) = \mu(P) = 0$. $v(P) = \tau(P) = 0$ follows since $\tau \ll \mu$ and $v \le \tau$. However,

$$u(X) = u(P) + u(N) \le u(P) = v(P) - \epsilon\mu(P) = 0$$

which contradicts to $u(X) = v(X) - \epsilon\mu(X) > 0$.
When τ and μ are σ-finite, we decomposite S into a countable number of disjoint subsets $\{S_i\}_{i=1}^{\infty}$ such that v and u are both finite on each S_i, $i \in \mathbb{N}^*$. There exists $\{f_i\}_{i=1}^{\infty}$ such that each f_i is a Radon–Nikodym derivative of $\tau|S_i$. Then $f := \sum_{i=1}^{\infty} f_i \mathbf{1}_{S_i}$ is the Radon–Nikodym derivative of τ.

Proposition: about Radon–Nikodym derivative
Let u, v, μ be three σ- finite measures on $(S, \mathcal{B})$,

1. (Linearity) If $u \ll \mu$ and $v \ll \mu$, then for any $a, b \in \mathbb{R}^+$,
$$\frac{\mathrm{d}(au+bv)}{\mathrm{d}\mu} \overset{a.s.}{=} a\frac{\mathrm{d}u}{\mathrm{d}\mu} + b\frac{\mathrm{d}v}{\mathrm{d}\mu}$$
2. (Chain rule) If $u \ll v \ll \mu$, then
$$\frac{\mathrm{d}u}{\mathrm{d}\mu} \overset{a.s}{=} \frac{\mathrm{d}u}{\mathrm{d}v} \times \frac{\mathrm{d}v}{\mathrm{d}\mu}$$
3. (Change of variable) If $v \ll \mu$ and g is either a $\mathcal{L}^+(v)$ or $\mathcal{L}^1(v)$ function,
$$\int_S g \mathrm{d}v = \int_S g\frac{\mathrm{d}v}{\mathrm{d}\mu}\mathrm{d}\mu$$

Proof:
To prove 1, first we can show that $au + bv$ is a σ-finite measure and $au + bv \ll \mu$. For any $A \in \mathcal{B}$,
$$(au+bv)(A) = au(A) + bv(v) = a\int_A \frac{\mathrm{d}u}{\mathrm{d}\mu}\mathrm{d}\mu + b\int_A \frac{\mathrm{d}v}{\mathrm{d}\mu}\mathrm{d}\mu = \int_A a\frac{\mathrm{d}u}{\mathrm{d}\mu} + b\frac{\mathrm{d}v}{\mathrm{d}\mu}\mathrm{d}\mu$$
Since for any integrable functions f, g,
$$\int_A f\mathrm{d}\mu = \int_A g\mathrm{d}\mu \ \forall A \in \mathcal{B} \iff f \overset{a.s.}{=} g$$
the result follows.
To prove 2, it suffices to show the last equation of
$$\forall A \in \mathcal{B},\ \int_A \frac{\mathrm{d}u}{\mathrm{d}\mu}\mathrm{d}\mu = u(A) = \int_A \frac{\mathrm{d}u}{\mathrm{d}v}\mathrm{d}v = \int_A \frac{\mathrm{d}u}{\mathrm{d}v} \times \frac{\mathrm{d}v}{\mathrm{d}\mu}\mathrm{d}\mu$$
which is given by 3.
To prove 3, we first proceed with the case when g is an indicator function $\mathbf{1}_A$, where $A \in \mathcal{B}$, then
$$\int_R g\mathrm{d}\mu = \mu(A) = \int_A \frac{\mathrm{d}v}{\mathrm{d}\mu}\mathrm{d}\mu = \int_S g \times \frac{\mathrm{d}v}{\mathrm{d}\mu}\mathrm{d}\mu$$
Next, suppose g is a simple function. The above still holds because g is a linear combination of indicator functions. Then we can extend it to $\mathcal{L}^+$ function by the monotone convergence theorem. Finally, it holds for any function $g \in \mathcal{L}^1$, proved by decomposing $g = g^+ - g^-$.

Definition: Mutually singular

Let u, v be two measure on $(S, \mathcal{B})$, if there exists a decomposition $S = A \cup B$, where $A, B \in \mathcal{B}$, $A \cap B = \emptyset$ such that $u(A) = 0$ and $v(B) = 0$, then u and v are mutually singular, denoted as $u \perp v$.

Theorem: Radon-Lebegues decomposition

Given two σ-finite signed measures, μ and τ, on a measurable space $(S, \mathcal{B})$, there exists a unique decomposition into two signed measures u and v, $\tau = v + u$ such that $v \ll \mu$ and $u \perp v$.

Since in probability theory, we are only interested in normal unsigned measure. We restrict μ, τ to be unsigned, then u, v are also unsigned.

Proof:

First, we consider the case when μ and τ are finite. The same as the proof of the Radon-Nikodym theorem, we let

$$\mathcal{G} := \left\{ f \in \mathcal{L}^+(\mu) : \forall A \in \mathcal{B},\ \int_A f \mathrm{d}\mu \leq \tau(A) \right\}$$

and obtain $f_0 \in \mathcal{G}$ that

$$\int_S f_0 \mathrm{d}\mu = \alpha := \sup_{f \in \mathcal{G}} \left\{ \int_S f \mathrm{d}\mu \right\}$$

Then, $v(A) := \int_A f_0 \mathrm{d}\mu$ is a measure on $(S, \mathcal{B})$ that $v \ll \mu$. It remains to show that $u := (\tau - v) \perp \mu$. Note that both v and u are unsigned measures.

Suppose u and μ are not mutually singular. The contradiction arises with the following construction. For all $n \in \mathbb{N}^*$, consider the Hahn Decomposition $\{(P_n, N_n)\}_{n=1}^{\infty}$ of the signed measure $u - \frac{1}{n}\mu$ on $(S, \mathcal{B})$. We have $\{P_n\}_{n=1}^{\infty}$ monotonic increasing and $\{N_n\}_{n=1}^{\infty}$ monotonic decreasing. Let

$$P := \bigcup_{i=1}^{\infty} P_n \text{ and } N := S \setminus P = \bigcap_{i=1}^{\infty} N_n$$

For all $n \in \mathbb{N}^*$, $N \subset N_n$ and thus $u(N) - \frac{1}{n}\mu(N) \leq 0$. Since μ is finite, taking $n \to \infty$ yields $u(N) \leq 0$. We have $u(N) = 0$. By our assumption that $u \perp \mu$ not hold, $0 < \mu(P) = \lim_{n \to \infty} \mu(P_n)$. There exists $n_0 \in \mathbb{N}^*$ such that

$$\mu(P_{n_0}) > 0, \text{ and } u(P_{n_0}) - \frac{1}{n_0}\mu(P_{n_0}) \geq 0 \ \Rightarrow \tau(P_{n_0}) = \int_{P_{n_0}} f_0 \mathrm{d}\mu + u(P_{n_0}) \geq \int_{P_{n_0}} f_0 \mathrm{d}\mu + \frac{1}{n_0}\mu(P_{n_0})$$

Then similar to the proof of the Radon-Nikodym theorem, we can show that $(f_0 + \frac{1}{n_0}\mathbf{1}_{P_{n_0}}) \in \mathcal{G}$ and contradict to $\alpha = \int_S f_n \mathrm{d}\mu$ being the supremum.

It is also similar to the previous in the case of τ, μ being σ-finite, i.e, decomposite S into finite measure case and glue the resulting measures.

To show the uniqueness, suppose $\tau = v + u = v' + u'$ are two decompositions. First, $(u' - u) = (v - v') \ll \mu$ because $v \ll \mu, v' \ll \mu$. Similarly, $(u' - u) \perp \mu$. Then the result follows by a simple fact that whenever $a \ll b$ and $a \perp b$, $a = 0$ (where a, b are measures on the same measurable space).

5.2.2 Relationship of two type of absolute continuity

This section discusses the relationship between absolute continuous function and absolute continuous measure and use Radon-Nikodym theorem to prove the important theorem in 5.1.3.

Proposition: Suppose a σ-finite measure μ is absolute continuous to the Lebesgue measure λ on $(I \subset \mathbb{R}, \mathcal{B}(I))$. Let a be a fixed point in I such that $\mu([\min\{a,x\}, \max\{a,x\}]) < \infty \ \forall x \in I$. Then the function

$$F(x) := \operatorname{sign}(x-a)\mu([\min\{a,x\}, \max\{a,x\}]) = \int_a^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\mathrm{d}\lambda : I \to \mathbb{R}$$

is absolute continuous and $F' \overset{a.s.}{=} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$.

This is a direct result of the fundamental theorem of calculus. Note that for any closed interval $[L,R] \subset \mathbb{R}$, $\mu([L,R]) \le \lambda([L,R]) \sup_{x\in[l,R]}\{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x)\} < \infty$, thus a can be chosen to be a arbitrary finite number.

Theorem: Luzin N property of differentiable function

Let $F : I \mapsto \mathbb{R}$ be a differentiable function on an open interval $I \subset \mathbb{R}$ and λ is the Lebesgue measure on $\mathbb{R}$. Then F has the Luzin N property, which is

$$\forall B \in \mathcal{B}(I), \text{ if } \lambda(B) = 0, \text{ then } \lambda^*[F(B)] := \lambda^*(\{F(x) : x \in B\}) = 0$$

where λ^* is the outer measure defined in 2.3.3 to be

$$\forall A \in \mathcal{P}(\mathbb{R}),\ \lambda^*(A) = \inf\left\{\sum_{i=1}^{\infty}\sum_{j=1}^{n_i}(b_{i,j}-a_{i,j}) : A \subset \bigcup_{i=1}^{\infty}\bigcup_{j=1}^{n_i}(a_{i,j}, b_{i,j}), -\infty \le a_{i,j} \le b_{i,j} \le \infty \ \forall i,j\right\}$$

Proof:

For all $n, p \in \mathbb{N}^*$, let

$$A_{n,p} := \left\{x \in I : \frac{|f(x)-f(y)|}{|x-y|} \le n \ \forall y \in I \cap B\left(x, \frac{1}{p}\right)\right\}$$

Note that $A_{n,p} \in \mathcal{B}(\mathbb{R})$ because

$$A_{n,p} = \bigcap_{h \in \mathbb{Q}\cap B\left(0,\frac{1}{p}\right)} \left\{x \in I : \frac{|f(x)-f(x+h)|}{|h|} \le n\right\}$$

is a countable union of inverse maps from $[0,n]$ of measurable functions $\frac{|f(x)-f(x+h)|}{|h|}$.

From the definition of differentiable function, $I = \bigcup_{n\in\mathbb{N}^*}\bigcup_{p\in\mathbb{N}^*} A_{n,p}$ is a countable union. By countable sub-additivity of the outer measure, for any $B \in \mathcal{B}(I)$ such that $\lambda(B) = 0$,

$$\lambda^*[F(B)] = \lambda^*\left[F\left(\bigcup_{n\in\mathbb{N}^*}\bigcup_{p\in\mathbb{N}^*} A_{n,p}\cap B\right)\right] = \lambda^*\left[\bigcup_{n\in\mathbb{N}^*}\bigcup_{p\in\mathbb{N}^*} F\left(A_{n,p}\cap B\right)\right] \le \sum_{n=1}^{\infty}\sum_{p=1}^{\infty}\lambda^*[F(A_{n,p}\cap B)]$$

It suffices to prove that $\lambda^*[F(A_{n,p} \cap B)] = 0 \ \forall n, p \in \mathbb{N}^*$.

Fix n and p. Let $0 < \epsilon < \frac{1}{p}$. Since $\lambda(B \cap A_{n,p}) \leq \lambda(B) = 0$, there exists a sequence of open balls $\{B(x_i, r_i)\}_{i=1}^{\infty}$ such that $\forall i \in \mathbb{N}^*$, $x_i \in B \cap A_{n,p}$, $\sum_{i=1}^{\infty} 2r_i \leq \epsilon$ and $B \cap A_{n,p} \subset \bigcup_{i=1}^{\infty} B(x_i, r_i)$. Then, for any $x \in B \cap A_{n,p} \cap B(x_i, r_i)$, we have

$$|x - x_i| < r_i < \epsilon < \frac{1}{p} \ \Rightarrow \ |f(x) - f(x_i)| \leq n|x - x_i| < nr_i \ \Rightarrow \ f(x) \in B(f(x_i), nr_i)$$

Therefore, $F(B \cap A_{n,p} \cap B(x_i, r_i)) \subset B(f(x_i), nr_i)$ and

$$F(B \cap A_{n,p}) \subset F\left(\bigcup_{i=1}^{n} B \cap A_{n,p} \cap B(x_i, r_i)\right) = \bigcup_{i=1}^{\infty} F[B \cap A_{n,p} \cap B(x_i, r_i)] \subset \bigcup_{i=1}^{\infty} B(f(x_i), nr_i)$$

We have

$$\lambda^*[F(B \cap A_n, p)] \leq \lambda^*\left[\bigcup_{i=1}^{\infty} B(f(x_i), nr_i)\right] \leq \sum_{i=1}^{\infty} \lambda^*[B(f(x_i), nr_i)] = 2n \sum_{i=1}^{\infty} r_i < n\epsilon$$

Taking $\epsilon \to 0$ yields $\lambda^*[F(B \cap A_{n,p})] = 0$.

Absolute continuous function, thought differentiable almost everywhere, also satisty this property.

Theorem: Luzin N-property of absolute continuous function

Let F be an absolute continuous function on an interval $I \subset \mathbb{R}$. Then F has the Luzin-N property stated above.

Proof:

By the property of absolute continuous function, for any ϵ, there exists $\delta > 0$ such that

$$\forall \text{ disjoint } \{(a'_n, b'_n) \subset I\}_{n=1}^{\infty} \text{ that } \sum_{n=1}^{\infty} (b'_n - a'_n) < \delta, \ \sum_{n=1}^{\infty} |f(b'_n) - f(a'_n)| < \epsilon$$

Take I' be the largest open interval of I. We can omit at most 2 end points. Let $B \in \mathcal{B}(\mathbb{R})$ and $B \subset I'$ such that $\lambda(B) = 0$. Since $\lambda(B) = \inf\{\lambda(A) : A \supset B \text{ and } A \text{ is open}\}$, there exists an open set A such that $B \subset A \subset I'$ and $\lambda(A) < \delta$. We can further express A as a countable union of disjoint open interval $A = \bigcup_{i=1}^{\infty} (a_i, b_i)$.

For each $i \in \mathbb{N}^*$, Let c_i, d_i be the minimum and maximum points of F on $[a_i, b_i]$. The extremums exist because F is continuous. Then, since

$$\sum_{i=1}^{\infty} |d_i - c_i| \leq \sum_{i=1}^{\infty} b_i - a_i = \lambda(A) < \delta$$

we have

$$\lambda^*[F(B)] \leq \lambda^*[F(A)] = \lambda^*\left[\bigcup_{i=1}^{\infty} \lambda^*[F((a_i, b_i))]\right] \leq \sum_{i=1}^{\infty} F((a_i, b_i)) = \sum_{i=1}^{\infty} |f(d_i) - f(c_i)| < \epsilon$$

Taking $\epsilon \to 0$, the result follows.

Proof of absolute continuous implies differentiable almost everywhere
We are now ready to prove the important theorem in 5.1.3. that absolute continuous and non-decreasing function imply almost surely integrable differentiable.
We will prove the equivalency of the followings.

1. Non-descreasing $f$is absolute continuous
2. f satisfy the Luzin N-property
3. f is differentiable almost surely and $\forall x \in [a,b],\ f(x) - f(a) = \int_a^x f'(t)\mathrm{d}t$

We have proved that 1 implies 2, and the first example in 5.1.3 gives 3 implies 1. It suffices to prove 2 implies 3.
Instead of using f directly, we construct a strictly increasing (and thus injective) function $g(x) = f(x) + x : [a,b] \to \mathbb{R}$ such that it has an inverse $g^{-1} : g([a,b]) \mapsto [a,b]$. The following shows that the measure $\mu := \lambda \circ g$ is properly defined on $([a,b], \mathcal{B}([a,b]))$, and it is an absolute continuous measure with respect to λ.
First we prove that g is absolute continuous, given f is. It follows from that for any finite collection of intervals $\{(a_i, b_i) \in I\}_{i=1}^n$,

$$\sum_{i=1}^n |g(b_i) - g(a_i)| = \sum_{i=1}^n f(b_i) - f(a_i) + b_i - a_i = \sum_{i=1}^n |f(b_i) - f(a_i)| + \sum_{i=1}^n b_i - a_i$$

Then for any $\epsilon > 0$, taking δ' such that $\sum_{i=1}^n |f(b_i) - f(a_i)| < \frac{\epsilon}{2}$ and choosing $\delta = \min(\delta', \frac{\epsilon}{2})$, we have $\sum_{i=1}^n |g(b_i) - g(a_i)| < \epsilon$ whenever $\sum_{i=1}^n b_i - a_i < \delta$. Thus g is absolute continuous and satisfies 2.
Since g is continuous and strictly increasing on compact interval $[a,b]$, $g^{-1} : g([a,b]) \mapsto [a,b]$ is properly defined and continuous (see 1.3.2). Therefore, it is a measurable function on $(G, \mathcal{B}(G)) \mapsto ([a,b], \mathcal{B}([a,b]))$, where $G := g([a,b])$. For any $A \in \mathcal{B}([a,b])$, $(g^{-1})^{-1}(A) = g(A) \in \mathcal{B}(G)$, thus $\lambda[g(A)]$ is properly defined. Further, $\mu := \lambda \circ g$ is a finite measure on $([a,b], \mathcal{B}([a,b]))$, because the bijection g gives the σ-additivity and G is closed and bounded. The Luzin N property gives that whenever $\lambda(A) = 0$, $\lambda^*[g(A)] = \lambda[g(A)] = 0$, therefore we have $\mu \ll \lambda$, and the Radon-Nikodym derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is defined almost everywhere on $[a,b]$. We have $\forall x \in [a,b]$,

$$f(x) - f(a) = g(x) - x - g(a) + a = \lambda[g([a,x])] - (x-a) = \int_a^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t)\mathrm{d}t - \int_a^x 1\mathrm{d}t = \int_a^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t) - 1\mathrm{d}t$$

By the fundamental theorem of calculus, when $f(x) = f(a) + \int_a^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t) - 1\mathrm{d}t$, f is differentiable almost surely and $f' \overset{a.s.}{=} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} - 1$.
Remark: A complete proof should consider any measurable set such that the outer measure matched, instead of the Boral set.

Corollary: Let F be a non-decreasing function on $I = [L, R] \subset \mathbb{R}$ and λ is the Lebesgue measure on $\mathbb{R}$. Suppose μ is the Lebesgue-Stieltjes measure generated by nondecreasing F on $((L, R), \mathcal{B}((L, R)))$, then

1. F is absolute continuous if and only if $\mu \ll \lambda$
2. The Radon–Nikodym derivative is almost surely (with respect to λ) equal to the normal derivative,
$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \overset{a.s.}{=} F'$$
3. When I is an arbitrary interval with possibly infinite Lebesgue measure, and F is absolute continuous in every closed interval in $\mathbb{R}$, it still implies $\mu \ll \lambda$ and 2.

Proof:
The proof of the "$\Rightarrow$" direction is similar to the above one. Note that when F is continuous, the premeasure of the Lebesgue-Stieltjes measure can be $\tilde{\mu}((a, b)) = F(b) - F(a)$ on $\mathcal{R} := \{(a, b) : L \leq a \leq b \leq R\}$, which yield the same result to the definition in 2.3.4.
Now we prove the other direction. Since $\mu \ll \lambda$, we have
$$\forall L < x < y < R,\ F(y) - F(x) = \mu((x, y)) = \int_{(x,y)} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \mathrm{d}\lambda$$
Therefore, fix $x_0 \in (L, R)$, we have $F(x) = F(x_0) + \int_{x_0}^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t)\mathrm{d}t$ Since the function $\int_{x_0}^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t)\mathrm{d}t : I \to \mathbb{R}$ is absolute continuous (proved in 5.1.3), so as $F(x)$.
To prove 2, we show that the derivative F' satisfies
$$\forall A \in \mathcal{B}(I),\ \mu(A) = \int_A F' \mathrm{d}\lambda$$
By the theorem in 5.1.3, F' is defined almost everywhere and the above holds for any $A \in \mathcal{R}$. Let
$$\mathcal{L} := \left\{ A \subset I : \mu(A) = \int_A F' \mathrm{d}\lambda \right\}$$
We can check that $\mathcal{L}$ is a λ-system, and $\mathcal{R}$ is a π-system. Since $\mathcal{R} \subset \mathcal{L}$, Dynkin's Theorem gives the result $\mathcal{B}(I) = \sigma(\mathcal{R}) \subset \mathcal{L}$.

When I is an interval in $\mathbb{R}$ with possibly infinite Lebesgue measure, $\mu \ll \lambda$ still holds obviously. For any $a, b \in I$, partition $[a, b]$ into at most countable number of intervals $\{[a_n, a_{n+1}]\}_{n=0}^{\infty}$ ($a_0 = a, a_\infty = b$) with all finite Lebesgue measure. We have

$$\mu((a,b)) = F(b) - F(a) = \sum_{n=0}^{\infty} F(a_{n+1}) - F(a_n) = \sum_{n=0}^{\infty} \int_{a_n}^{a_{n+1}} F'(x)\mathrm{d}x = \int_a^b F'(x)\mathrm{d}x$$

by monotone convergence, which gives the base case for proving 2.

We look into the case for higher dimension. Instead of considering the measure defining on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, we look into a wider case that the measure τ is defined on a product set (with its σ-algebra), in which one of its dimension is $\mathbb{R}$ and the absolute continuity holds on this dimension. Lastly, we will show that if the absolute continuity holds on the remaining dimensions, it will then hold for the whole product set and the density is thereby defined. We can recursively use this approach to check wether $\tau \ll \lambda_k$, where λ_k is the Lebesgue measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and obtain the density $\frac{\mathrm{d}\tau}{\mathrm{d}\lambda_k}$.

Proposition: The relationship in multivariable case
Let τ be a finite measure, , μ be a measure and λ be the Lebesgue measure on $\mathbb{R}$, such that τ and $\lambda \times \mu$ are on a product measurable space $(\mathbb{R} \times S_2, \mathcal{B} := \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}_2))$. Assume for any $A_2 \in \mathcal{B}_2$, the measure $\tau_1^{A_2}(*) := \tau(* \times A_2) : \mathcal{B}(\mathbb{R}) \mapsto [0, \infty]$ is the Lebesgue-Stieltjes measure generated by nondecreasing F_{A_2}. Or equivalently, we start with the measure τ and $\tau_1^{A_2}$ and consider a cumulative function

$$F_{A_2}(x) := \tau^{A_2}((a, x]) \text{ when } x > a, \text{ and } F_{A_2}(x) : -\tau^{A_2}((x, a]) \text{ otherwise, where } a \in \mathbb{R} \text{ is arbitrary}$$

We have

1. Pick a π-system $\mathcal{C} \subset \mathcal{B}_2$ such that $\mathcal{B}_2 = \sigma(\mathcal{C})$. If $\tau_1^{A_2} \ll \lambda \ \forall A_2 \in \mathcal{C}$ and $\tau_1^{S_2} \ll \lambda$, then $\tau_1^{A_2} \ll \lambda \ \forall A_2 \in \mathcal{B}_2$, which is equivalent that F_{A_2} is absolute continuous for any $A_2 \in \mathcal{B}_2$. Then the two derivatives coincide almost surely,

$$\forall A_2 \in \mathcal{B}_2, \ \frac{\mathrm{d}\tau_1^{A_2}}{\mathrm{d}\lambda} \overset{a.s.}{=} F_{A_2}$$

   A simple way to verify the condition is to

2. Further from 1, if $(S_2, \mathcal{B}_2)$ is $(\bar{\mathbb{R}}^k, \mathcal{B}(\bar{\mathbb{R}}^k))$, $k \in \mathbb{N}^*$, the two derivatives almost surely equal to a measurable kernel, i.e.,

$$\frac{\mathrm{d}\tau_1^{A_2}}{\mathrm{d}\lambda} \overset{a.s.}{=} F'_{A_2} \overset{a.s.}{=} \tau'(*, A_2), \text{ where } \tau' : \mathbb{R} \times \mathcal{B}_2 \mapsto [0, \infty]$$

such that $\tau'(*, A_2)$ is a measurable function for any $A_2 \in \mathcal{B}_2$, and $\tau'(x, *)$ is a measure on $(S_2, \mathcal{B}_2)$.

3. From 2, we have

$$\forall A_1 \in \mathcal{B}(\mathbb{R}), A_2 \in \mathcal{B}_2,\ \tau(A_1 \times A_2) = \int_{A_1} \tau'(x, A_2) \mathrm{d}\lambda(x)$$

   Thus, τ is the measure constructed using the kernel in the exteneded Fubini theorem. Also by this theorem, we have that for any $f \in \mathcal{L}^+(\tau)$ or $f \in \mathcal{L}^1(\tau)$,

$$\int_{\mathbb{R}\times S_2} f(x, y) \mathrm{d}\tau(x, y) = \int_{\mathbb{R}} \int_{S_2} f(x, y) \mathrm{d}\tau'(x, y) \mathrm{d}\lambda(x)$$

   (The inner integral is with respect to y)

4. To complete the story, we suppose there exists a measure v on $(S_2, \mathcal{B}_2)$ such that for $x \in \mathbb{R}$ almost everywhere (with respect to λ), $\tau'(x, *) \ll v$. Then setting $\tau'(x, *) = 0$ for x that the absolute continuity doesn't hold, there exists the density $\frac{\mathrm{d}\tau'(x,y)}{\mathrm{d}v}$ such that

$$\forall A_1 \in \mathcal{B}(\mathbb{R}), A_2 \in \mathcal{B}_2,\ \tau(A_1 \times A_2) = \int_{A_1} \int_{A_2} \frac{\mathrm{d}\tau'(x, y)}{\mathrm{d}v} \mathrm{d}v(y) \mathrm{d}\lambda(x) = \int_{A_1 \times A_2} \frac{\mathrm{d}\tau'(x, y)}{\mathrm{d}v} \mathrm{d}\lambda \times v(x, y)$$

   We further have $\tau \ll \lambda \times v$ and $\frac{\mathrm{d}\tau'(x,y)}{\mathrm{d}v}$ is a density of τ with respect to $\lambda \times v$, denoted as $\frac{\mathrm{d}\tau}{\mathrm{d}\lambda\times v} : \mathbb{R} \times S_2 \to \mathbb{R}$. Conversely, $\tau \ll \lambda \times v$ also gives that $\tau'(x, *) \ll v$ for for $x \in \mathbb{R}$ almost everywhere.

One can replace $\mathbb{R}$ (where λ is defined) by any open interval of $\mathbb{R}$, or replace $(\bar{\mathbb{R}}^k, \mathcal{B}(\bar{\mathbb{R}}^k))$ by any of its sub-measurable space or any structure isomorphic to it, in which case we should choose the corresponding dense set in the proof.

Proof:

1 and 4 can be proved by Dynkin's Theorem. 3 follows from 2 and the extended Fubini theorem. Details omit for proving 1, 3, 4.

We prove 2 in the case of $k = 1$, others are similar by using its countable dense set.

For every $q \in \mathbb{Q}$, define

$$f(x, q) := \begin{cases} F'_{[-\infty,q]}(x) & , \quad F_{[-\infty,q]}(x) \text{ is defined and nonnegative} \\ 0 & , \quad \text{otherwise} \end{cases} : \mathbb{R} \times \mathbb{Q} \mapsto [0, \infty)$$

Because for any $p, r \in \mathbb{Q},\ p \leq r$,

$$\int_{A_1} F'_{[-\infty,q]} \mathrm{d}\lambda = \tau(A_1 \times [-\infty, q]) \leq \tau(A_1 \times [-\infty, r]) = \int_{A_1} F'_{[-\infty,r]} \mathrm{d}\lambda \ \Rightarrow\ F'_{[-\infty,q]} \overset{a.s.}{\leq} F'_{[-\infty,r]}$$

Let

$$N_1 := \bigcup_{q\in\mathbb{Q}} \{x \in \mathbb{R} : F'_{[-\infty,q]}(x) \text{ is not defined or negative}\}$$

$$N_2 := \bigcup_{q,r\in\mathbb{Q},q\le r} \{x \in \mathbb{R}\setminus N_1 : F'_{[-\infty,q]}(x) > F'_{[-\infty,r]}(x)\}$$

Since $\mathbb{Q}$ is countable, we have $N_1 \cup N_2 \in \mathcal{B}(\mathbb{R})$ and $\lambda(N_1\cup N_2) = 0$. Set $f(x,q) := 0\ \forall x \in N_1\cup N_2, q \in \mathbb{Q}$. We have $f(x,q)$ measurable $\forall q \in \mathbb{Q}$, $f(x,q) = F'_{[-\infty,q]}(x)$ for all $q \in \mathbb{Q}$ and x almost everywhere in $\mathbb{R}$, and $f(x,q) \le f(x,r)\ \forall x \in \mathbb{R}, q, r \in \mathbb{Q}, q \le r$.
Then, let

$$\forall x \in \mathbb{R}, y \in [-\infty,\infty),\ f(x,y) = \inf\{f(x,q) : q \in \mathbb{Q}, q \ge y\} \text{ and } f(x,\infty) = \sup\{f(x,q) : q \in \mathbb{Q}\}$$

to be a $\mathbb{R}\times\bar{\mathbb{R}} \mapsto [0,\infty]$ function such that for any $x \in \mathbb{R}$, $f(x,*)$ is nondecreasing and right-continuous. We can check that $0 \le f(*,+\infty) \overset{a.s.}{<} \infty$ since τ is finite. For $x \in \mathbb{R}$ such that $f(*,+\infty) = +\infty$, set $f(x,y) := 0\ \forall y \in \bar{\mathbb{R}}$
Define the premeasure

$$\tilde{\tau}_x([a,b]) = \begin{cases} f(x,b) - \lim_{t\to a^-} f(x,t) & , \quad -\infty < a \le b \le \infty \\ f(x,b) & , \quad a = -\infty, b \ge a \end{cases} \quad \forall (a,b] \in \mathcal{R} := \{[a,b] : -\infty \le b \le \infty\}$$

with $\tilde{\tau}_x(\emptyset) = 0$ on the semiring $\mathcal{R}\cup\emptyset$. $\tilde{\tau}_x$ is finite. We claim that $\tau'(x,*)$ is the Carathéodory extension of $\tilde{\tau}_x$ into $\sigma(\mathcal{R}\cup\emptyset) = \mathcal{B}(\bar{\mathbb{R}})$.
It remains to verify that

1. $\forall A_2 \in \mathcal{B}(\bar{\mathbb{R}}),\ \tau'(*,A_2) : \mathbb{R} \mapsto [0,\infty]$ is measurable with respect to λ
2. $\forall A_2 \in \mathcal{B}(\bar{\mathbb{R}}),\ \tau'(*,A_2) \overset{a.s.}{=} F'_{A_2}$ with respect to λ

Let $\mathcal{L} \subset \mathcal{B}(\bar{\mathbb{R}})$ be the set that the two conditions holds for any $A_2 \in \mathcal{L}$. We first claim that $\mathcal{C} := \{[-\infty,y] : y \in \bar{\mathbb{R}}\} \subset \mathcal{L}$. Condition 1 holds for all $[-\infty,y] \in \mathcal{C}$ because $\tau'(*,[-\infty,y]) = f(*,y)$ is either the infimum or the supremum of countable number of measurable functions $f(*,p)$, $p \in \mathbb{Q}$. To prove condition 2, since for any monotonic sequence $\{q_i \in \mathbb{Q}\}_{i=1}^{\infty}$ such that $\lim_{i\to\infty} q_i = a \in \bar{\mathbb{R}}$, we have

$$\begin{aligned}
\forall A_1 \in \mathcal{B}(\mathbb{R}),\ \int_{A_1} \lim_{i\to\infty} F'_{[-\infty,q_i]}\mathrm{d}\lambda &= \lim_{i\to\infty}\int_{A_1} F'_{[-\infty,q_i]}\mathrm{d}\lambda \\
&= \lim_{i\to\infty} \tau(A_1 \times [-\infty, q_i]) \\
&= \tau\left(\lim_{i\to\infty} A_1 \times [-\infty, q_i]\right) \\
&= \tau(A_1, [-\infty, a]) = \int_{A_1} F'_{[-\infty,a]}\mathrm{d}\lambda
\end{aligned}$$

Note that the first equation is given by dominated convergence theorem, which rely on our assumption that $\int_{A_1} F'_{[-\infty,q_1]}\mathrm{d}\lambda = \tau(A_1, [-\infty, q_1]) < \infty\ \forall A_1 \in \mathcal{B}(\mathbb{R})$. Then,

$$\begin{aligned}
\forall y \in [-\infty, \infty),\ f(*, y) &= \lim_{i\to\infty} f(*, q_i) \text{ where } \{q_i \in \mathbb{Q}\}_{i=1}^{\infty} \text{ is monotonic descreasing, and } \lim_{i\to\infty} q_i = y \\
&\overset{a.s.}{=} \lim_{i\to\infty} F'_{[-\infty,q_i]} \\
&\overset{a.s.}{=} F'_{[-\infty,y]}
\end{aligned}$$

When $y = \infty$, we have the same result with similar proof. Now we have shown $\mathcal{C} \subset \mathcal{L}$.
Note that $\mathcal{C}$ is a π-system, we use Dynkin's Theorem to complete the proof. It suffices to show that $\mathcal{L}$ is a λ-system.

1. $\bar{\mathbb{R}} \in \mathcal{L}$ since $\bar{\mathbb{R}} \in \mathcal{C}$
2. When $A_2 \in \mathcal{L}$, $\tau(*, A_2^c) \overset{a.s.}{=} \tau(*, \bar{\mathbb{R}}) - \tau(*, A_2)$ is measurable since $\tau(*, A_2)$ is measurable, and $\tau(*, A_2^c) \overset{a.s.}{=} F'_{\bar{\mathbb{R}}} - F'_{A_2} \overset{a.s.}{=} F'_{A_2^c}$ by linearity. Thus $A_2^c \in \mathcal{L}$.
3. Let $\{A_{2,j} \in \mathcal{L}\}_{j=1}^{\infty}$ be disjoint, we can show $\bigcup_{j=1}^{\infty} A_{2,j} \in \mathcal{L}$ because (the last step given by monotone convergence theorem)

$$\tau\left(*, \bigcup_{j=1}^{\infty} A_{2,j} \in \mathcal{L}\right) = \sum_{j=1}^{\infty} \tau\left(*, A_{2,j} \in \mathcal{L}\right) \overset{a.s.}{=} \sum_{j=1}^{\infty} F'_{A_{2,j}} \overset{a.s.}{=} F'_{\bigcup_{j=1}^{\infty} A_{2,j}}$$

Therefore, $\mathcal{L}$ is a λ-system. We have proved 2 for the finite measure τ case.

Corollary: for the σ-finite case
Follow from the theorem, when τ is σ-finite, $\tau_1^{A_2} \ll \lambda\ \forall A_2 \in \mathcal{C}$ (a π system) and $\tau_1^{S_2} \ll \lambda$ still guarantee $\tau_1^{A_2} \ll \lambda\ \forall A_2 \in \mathcal{B}_2$ and the existence of kernel τ' that satisfy 2, 3 and 4.
Proof:

When it is σ-finite, consider a partition $\mathbb{R}\times S_2 = \bigcup_{k=1}^{\infty} B_k$ such that $\forall k \in \mathbb{N}^*, B_k \in \mathcal{B}$ and $\tau(B_k) < \infty$. Let $\tau_{1,k}^{A_2} := \tau((* \times A_2) \cap B_k)$ be the finite measures. Since $\tau_1^{A_2} \ll \lambda \ \forall A_2 \in \mathcal{C}$ gives $\tau_{1,k}^{A_2} \ll \lambda \ \forall A_2 \in \mathcal{C}, k \in \mathbb{N}^*$ and $\tau_1^{S_2} \ll \lambda$ gives $\tau_{1,k}^{S_2} \ll \lambda \ \forall k \in \mathbb{N}^*$, We have shown that $\tau_{1,k}^{A_2} \ll \lambda \ \forall A_2 \in \mathcal{B}_2, k \in \mathbb{N}^*$, and thus $\tau_1^{A_2} = \sum_{k=1}^{\infty} \tau_{1,k}^{A_2} \ll \lambda \ \forall A_2 \in \mathcal{B}_2$

We define τ_k' be the corresponding measurable kernel of $\tau_{1,k}^{A_2}$ with reference to λ. By monotone convergence theorem, we have $\forall A_1 \in \mathcal{B}(\mathbb{R}), A_2 \in \mathcal{B}(\bar{\mathbb{R}})$,

$$\tau(A_1 \times A_2) = \tau^{A_2}(A_1) = \sum_{k=1}^{\infty} \tau_{1,k}^{A_2}(A_1) = \sum_{k=1}^{\infty} \int_{A_1} \tau_k'(x, A_2) \mathrm{d}\lambda(x) = \int_{A_1} \sum_{k=1}^{\infty} \tau_k'(x, A_2) \mathrm{d}\lambda$$

Let

$$\forall x \in \mathbb{R}, A_2 \in \mathcal{B}(\bar{\mathbb{R}}),\ \tau'(x, A_2) = \sum_{k=1}^{\infty} \tau_k'(x, A_2)$$

We can check that τ' is the kernel we need.

But given the absolute continuity, the almost surely condition still have trouble for finding the kernel τ'. The following consider a stronger case of integrable derivative, with directly gives τ'.

Example: derivative with reference to higher dimension Lebesgue measure

Let τ be a finite measure and λ_k be the Lebesgue measure, both on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ $(k \in \mathbb{N}^*)$. Define the cumulative function

$$F(x_1, \cdots, x_k) := \tau\left(\prod_{i=1}^{n}(-\infty, x_i]\right) : \mathbb{R}^k \to \mathbb{R}$$

If all permutations $\{a_i\}_{i=1}^k$ of $\{i\}_{i=1}^k$, $\frac{\partial F}{\partial x_{a_1} \cdots \partial x_{a_k}}$ exists and is of $\mathcal{L}^1(\lambda)$, then all of them are Radon–Nikodym derivative of τ with reference to λ_k.

To show the result, we use the Fubini theorem,

$$\begin{aligned}
\forall [x_{a_1}, \cdots, x_{a_k}]^T \in \mathbb{R}^k,\ \tau&\left(\prod_{i=1}^{n}(-\infty, x_{a_i}]\right) \\
&= F(x_{a_1}, \cdots, x_{a_1}) \\
&= \int_{-\infty}^{x_{a_1}} \cdots \int_{-\infty}^{x_{a_k}} \frac{\partial}{\partial x_{a_1} \cdots \partial x_{a_k}} F(t_{a_1}, \cdots, t_{a_1}) \mathrm{d}t_{a_k} \cdots \mathrm{d}t_{a_1} \\
&= \int_{(-\infty, x_{a_1}]} \int_{\prod_{i=2}^{n}(-\infty, x_{a_i}]} \frac{\partial F}{\partial x_{a_2} \cdots \partial x_{a_k}} \mathrm{d}\lambda_{k-1} \mathrm{d}\lambda_1
\end{aligned}$$

Assuming that $\frac{\partial F}{\partial x_{a_2} \cdots \partial x_{a_k}}$ is the density for the a measure $\tau'(x_{a_1}, *)$, we have

$$\tau^{\prod_{i=2}^{n}(-\infty, x_{a_i}]}((-\infty, x_{a_1}]) = \tau\left(\prod_{i=1}^{n}(-\infty, x_{a_i}]\right) = \int_{(-\infty, x_{a_1}]} \tau'\left(t_{a_1}, \prod_{i=2}^{n}(-\infty, x_{a_i}]\right) \mathrm{d}\lambda_1(t_{a_1})$$

Thus by the above proposition, $\tau^{A_2} \ll \lambda\ \forall A_2 \in \mathcal{B}(\mathbb{R}^{k-1})$, τ' is the kernel, and finally $\frac{\partial F}{\partial x_{a_1} \cdots \partial x_{a_k}}$ is the density of τ. To justify the assumption, we recursively use the proposition for each measure $\tau'(x_{a_1}, *)$ until the base case that

$$\tau'(x_{a_1}, \cdots, x_{a_n-1}, (-\infty, x_{a_n}]) = \int_{-\infty}^{x_{a_n}} \frac{\partial F}{\partial x_{a_1} \cdots \partial x_{a_k}} \mathrm{d}\lambda_1$$

which is a measure with density $\frac{\partial F}{\partial x_{a_1} \cdots \partial x_{a_k}}$ for every $[x_{a_1}, \cdots, x_{a_{k-1}}]^T \in \mathbb{R}^{k-1}$.
$\frac{\partial F}{\partial x_{a_1} \cdots \partial x_{a_k}}$ are equal almost surely for every permutation, if further, all of $\frac{\partial F}{\partial x_{a_1} \cdots \partial x_{a_k}}$ are continuous, they are equal surely by Clairaut's theorem.

5.2.3 The chain rule

Theorem: The chain rule of $\mathbb{R}$ case
If $t : (a, b) \mapsto \mathbb{R}$ is bijective, differentiable everywhere and $|t'| < \infty$, for $g \in \mathcal{L}^1$, we have $g \circ t(x) \times t'(x) \in \mathcal{L}^1$

$$\int_{t(a)}^{t(b)} g(y)\mathrm{d}y = \int_a^b g \circ t(x) \times t'(x)\mathrm{d}x$$

Note that t is invertible and either strictly increasing or decreasing.
Proof:
Suppose t is strictly increasing, otherwise both sides are negative and it is similar. Its inverse t^{-1} is defined. First let $g = \mathbf{1}_A$ where $A \in \mathcal{B}((t(a), t(b)))$. Note that $\lambda \circ t$ is a measure on $((a, b), \mathcal{B}((a, b))$ such that $\lambda \circ t \ll \lambda$ and $\frac{\mathrm{d}\lambda \circ}{\mathrm{d}\lambda} \overset{a.s.}{=} t'$ by the Luzin N property. We have

$$RHS = \int_{t^{-1}(A)} t'(y)\mathrm{d}\lambda(y) = \lambda \circ t[t^{-1}(A)] = \lambda(A) = LHS$$

Then we generalize the result into when g is simple function, $\mathcal{L}^1$ function and $\mathcal{L}^+$ function.

We extend the theorem to a more general case in $\mathbb{R}^n$. First we define a equivalent role of derivative in higher dimension.

Definition: Total derivative
Let $\mathbf{f} : A \mapsto \mathbb{R}^m$ be a function on a open set $A \in \mathbb{R}^n$ $(n, m \in \mathbb{N}^*)$. $\mathbf{f}$ is (totally) differentiable at point $\mathbf{x} \in A$ if the finite linear transformation $\mathbf{f}' : \mathbb{R}^n \mapsto \mathbb{R}^m$ (represented as a $\mathbb{R}^{m\times n}$ matrix) exists such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{f}'\mathbf{h}\|}{\|\mathbf{h}\|} = 0$$

The limit means it holds for any sequence in $\mathbb{R}^n$ towards $\mathbf{0}$. Or equivalently, write $\|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{f}'(\mathbf{x})\mathbf{h}\| \in o(\|\mathbf{h}\|)$ as $\|\mathbf{h}\| \to \mathbf{0}$, which means for any $\mathbf{x} \in A$, $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$\|\mathbf{f}(\mathbf{x}+\mathbf{h})-\mathbf{f}(\mathbf{x})-\mathbf{f}'(\mathbf{x})\mathbf{h}\|<\epsilon\|\mathbf{h}\|$ whenever $\mathbf{h}\leq B(\mathbf{0},\delta)$.
$\mathbf{f}'$ can be represented as a $m\times n$ matrix, called the total derivative of $\mathbf{f}$ at $\mathbf{x}$.
Note that at point $\mathbf{x}$, $\mathbf{f}$ is differentiable only if it is continuous. If otherwise $\lim_{i\to\infty}\|\mathbf{f}(\mathbf{x}+\mathbf{h}_i)-\mathbf{f}(\mathbf{x})\|=c>0$ for some sequence $\{\mathbf{h}_i\}_{i=1}^{\infty}$ with limit $\mathbf{0}$, the required limit is $+\infty$ regarding to $\{\mathbf{h}_i\}_{i=1}^{\infty}$ for any linear $\mathbf{f}'$.

Proposition: about total derivative
Write $\mathbf{f}=[f_1,\cdots,f_m]^T$. We can prove that at point $\mathbf{x}\in A$, $\mathbf{f}'$ exists if and only if all of f_i' exists, and

$$\mathbf{f}'=\begin{pmatrix} f_1' \\ \vdots \\ f_m' \end{pmatrix}$$

It suffices to check the function $f:A\mapsto\mathbb{R}$, where $A\in\mathbb{R}^n$ is open, for the following properties.

1. f' at each point $\mathbf{x}\in A$ is unique if exists.
2. If $n=1$, f' is the derivative of f at x defined in the one dimensional case.
3. (The chain rule of total derivative) Let $\mathbf{g}:B\mapsto A$, where B is a open set in $\mathbb{R}^m$, $m\in\mathbb{N}^*$. If $\mathbf{g}$ is differentiable at $\mathbf{x}\in B$ and f is differentiable at $\mathbf{g}(\mathbf{x})$, then $f\circ\mathbf{g}$ is differentiable at $\mathbf{x}$, and

$$(f\circ\mathbf{g})'=f'\times\mathbf{g}'$$

   Here $(f\circ\mathbf{g})'$ and $\mathbf{g}'$ is at $\mathbf{x}$, and f' is at $\mathbf{g}(\mathbf{x})$. $\times$ is matrix multiplication.

Proof:
Suppose two linear transformation f' and f'' satisfies the definition, we have

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|f''(\mathbf{h})-f'(\mathbf{h})\|}{\|\mathbf{h}\|}\leq\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|f(\mathbf{x}+\mathbf{h})-f(\mathbf{x})-f'(\mathbf{h})\|}{\|\mathbf{h}\|}+\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|-f(\mathbf{x}+\mathbf{h})+f(\mathbf{x})+f''(\mathbf{h})\|}{\|\mathbf{h}\|}=0$$

Since f', f'' are linear, $(f''-f')(\mathbf{h})=f''(\mathbf{h})-f'(\mathbf{h})\in o(\|\mathbf{h}\|)$. For any $\mathbf{h}\in\mathbb{R}^n$,

$$0=\lim_{\alpha\to 0}\frac{\|(f''-f')(\alpha\mathbf{h})\|}{|\alpha|\|\mathbf{h}\|}=\frac{\|(f''-f')(\mathbf{h})\|}{\|\mathbf{h}\|}$$

Thus $f''-f'=O$ (zero matrix).
2 follows from the definition. To prove 3, it suffices to show that for any $\{\mathbf{h}_i\in\mathbb{R}^m\}_{i=1}^{\infty}$ that $\lim_{i\to\infty}\mathbf{h}_i=\mathbf{0}$, $\|f\circ\mathbf{g}(\mathbf{x}+\mathbf{h})-f\circ\mathbf{g}(\mathbf{x})-(f'\times\mathbf{g}')(\mathbf{h}_i)\|\in o(\|\mathbf{h}_i\|)$.
Since $\|\mathbf{g}(\mathbf{x}+\mathbf{h}_i)-\mathbf{g}(\mathbf{x})-\mathbf{g}'(\mathbf{h}_i)\|\in o(\|\mathbf{h}_i\|)$. Taking any $\epsilon>0$, $\exists i_\epsilon$ such that when $i\geq i_\epsilon$,

$\mathbf{g}(\mathbf{x}+\mathbf{h}_i) = \mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{h}_i) + \epsilon\|\mathbf{h}_i\|\mathbf{e}_i$ for some $\mathbf{e}_i \in \mathbb{R}^m$ that $0 \leq \|\mathbf{e}_i\| \leq 1$, the choice is unique for each i. Further, by the derivative of f, and since $\{\epsilon\|\mathbf{h}_i\|\mathbf{e}_i\}_{i=1}^{\infty}$ has limit zero,

$$\begin{aligned} S_1 :&= \left\| f \circ \mathbf{g}(\mathbf{x}+\mathbf{h}_i) - f[\mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{h}_i)] - f'\left[\epsilon\|\mathbf{h}_i\|\mathbf{e}_i\right] \right\| \\ &= \left\| f\left[\mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{h}_i) + \epsilon\|\mathbf{h}_i\|\mathbf{e}_i\right] - f\left[\mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{h}_i)\right] - f'\left[\epsilon\|\mathbf{h}_i\|\mathbf{e}_i\right] \right\| \in o(\epsilon\|\mathbf{h}_i\|\|\mathbf{e}_i\|) \end{aligned}$$

Also, $\|f(\mathbf{g}(\mathbf{x})+\mathbf{h}') - f(\mathbf{g}(\mathbf{x})) - f'(\mathbf{h}')\| \in o(\|\mathbf{h}'\|)$ whenever $\mathbf{h}' \in \mathbb{R}^n$ and $\mathbf{h}' \to \mathbf{0}$. Put $\mathbf{h}' = \mathbf{g}'(\mathbf{h}_i)$ (note that $\{\mathbf{g}'(\mathbf{h}_i)\}_{i=1}^{\infty}$ have limit zero), we have

$$S_2 := \|f[\mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{h}_i)] - f[\mathbf{g}(\mathbf{x})] - f'[\mathbf{g}'(\mathbf{h}_i)]\| \in o(\|\mathbf{g}'(\mathbf{h}_i)\|)$$

The target is bounded by

$$\|f \circ \mathbf{g}(\mathbf{x}+\mathbf{h}_i) - f \circ \mathbf{g}(\mathbf{x}) - (f' \times \mathbf{g}')(\mathbf{h}_i)\| \leq \|f'\left[\epsilon\|\mathbf{h}_i\|\mathbf{e}_i\right]\| + S_1 + S_2$$

Since $\mathbf{g}'$ is linear, taking $c = \sup\{\|\mathbf{g}'(\mathbf{t})\| : \mathbf{t} \in \mathbb{R}^m, \|\mathbf{t}\| = 1\}$, we have $0 \leq c < \infty$ and $\|\mathbf{g}'(\mathbf{h})\| \leq c\|\mathbf{h}\|$ for any $\mathbf{h} \in \mathbb{R}^m$. For any $\epsilon' > 0$, it suffice to restrict $S_2 \leq \frac{\epsilon'}{c}\|\mathbf{g}'(\mathbf{h})\|$ to ensure $S_2 \leq \epsilon'\|\mathbf{h}\|$ when $c \neq 0$. When $c = 0$, $S_2 = 0 \in o(\|\mathbf{h}\|)$. It means that $o(\mathbf{g}'(\|\mathbf{h}\|))$ is no weaker than $o(\|\mathbf{h}\|)$. Similarly, $o(\epsilon\|\mathbf{h}\|\|\mathbf{e}_i\|)$ is no weaker than $o(\|\mathbf{h}\|)$ for any $\epsilon > 0$. For any ϵ', pick $\epsilon > 0$ such that $\epsilon f'(\mathbf{t}) \leq \frac{\epsilon'}{3}$ $\forall \mathbf{t} \in B(\mathbf{0}, 1)$, then choose the corresponding i_ϵ, i_2, i_3 to ensure $S_1, S_2 \leq \frac{\epsilon'}{3}\|\mathbf{h}_i\|$ when $i \geq \max\{i_\epsilon, i_2, i_3\}$. We have prove the result.

Theorem: Total derivative as Jacobian matrix
Let $\mathbf{f} := [f_1, \cdots, f_m]^T : A \mapsto \mathbb{R}^m$, where $A \subset \mathbb{R}^n$ is open. If at point $\mathbf{x} \in \mathbb{R}^n$ and its neighborhood B, the partial derivative of each f_i with respect to each dimension $1 \leq j \leq n$ exists and is continuous at $\mathbf{x}$, then $\mathbf{f}$ is differentiable at $\mathbf{x}$, and the derivative is the Jacobian matrix, $J_{\mathbf{f}}(\mathbf{x})$, a $m \times n$ matrix whose (i,j)-entry is

$$\frac{\partial}{\partial x_j} f_i(\mathbf{x}) := \lim_{h \to 0} \frac{f_i(\mathbf{x} + h\mathbf{e}_j) - f_i(\mathbf{x})}{h}, \text{ where } \mathbf{e}_j \text{ is the } j\text{-th standard basis on} \mathbb{R}^n$$

If the above holds for any $\mathbf{x} \in A$, $J_{\mathbf{f}} : A \mapsto \mathbb{R}^{m \times n}$ defines a function.
Proof:
As previously mentioned, it suffices to prove the case when $m = 1$, i.e., $\mathbf{f}$ degraded to f and $J_f = \left[\frac{\partial}{\partial x_1} f(\mathbf{x}), \cdots, \frac{\partial}{\partial x_n} f(\mathbf{x})\right]$. Due to the continuity, for any $\epsilon > 0$, there exists $\delta > 0$ such that $\|\frac{\partial}{\partial x_j} f(\mathbf{x}+\mathbf{h}) - \frac{\partial}{\partial x_j} f(\mathbf{x})\| \leq \epsilon$ whenever $\|\mathbf{h}\| \leq \delta$. Let $\mathbf{h} = [h_1, \cdots, h_n]^T$. Denote $g_i(\mathbf{h}) =$

$[h_1, \cdots, h_i, 0, \cdots, 0]^T \in \mathbb{R}^n$ be the first i entries for $0 \le i \le n$. We have

$$\begin{aligned}\|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - J_f\mathbf{h}\| &= \left\| \sum_{i=1}^n f(\mathbf{x}+g_i(\mathbf{h})) - f(\mathbf{x}+g_{i-1}(\mathbf{h})) - \frac{\partial}{\partial x_i} f(\mathbf{x})h_i \right\| \\ &\le \sum_{i=1}^n \left\| f(\mathbf{x}+g_i(\mathbf{h})) - f(\mathbf{x}+g_{i-1}(\mathbf{h})) - \frac{\partial}{\partial x_i} f(\mathbf{x})h_i \right\|\end{aligned}$$

By the mean value theorem, for any $1 \le i \le n$ there exists $\mathbf{c}_n \in \mathbb{R}^n$ on the closed segment between $(\mathbf{x}+g_i(\mathbf{h}))$ and $(\mathbf{x}+g_{i-1}(\mathbf{h}))$ (with length h_i and parallel to $\mathbf{e}_i$) such that $h_i \frac{\partial}{\partial x_i} f(\mathbf{c}_i) = f(\mathbf{x}+g_i(\mathbf{h})) - f(\mathbf{x}+g_{i-1}(\mathbf{h}))$. Note that $\|\mathbf{c}_i - \mathbf{x}\| \le \|h\| \le \delta$. Apply the continuity property,

$$\|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - J_f\mathbf{h}\| \le \sum_{i=1}^n \left\| h_i \frac{\partial}{\partial x_i} f(\mathbf{c}_i) - \frac{\partial}{\partial x_i} f(\mathbf{x})h_i \right\| \le \sum_{i=1}^n |h_i|\epsilon \le \epsilon\sqrt{n}\|\mathbf{h}\|$$

The last inequality is due to Cauchy inequality, i.e,

$$\forall \mathbf{h} \in \mathbb{R}^n,\ \sum_{i=1}^n |h_i| \times 1 \le \sqrt{\left(\sum_{i=1}^n h_i^2\right)}\sqrt{\left(\sum_{i=1}^n 1\right)} = \sqrt{n}\|\mathbf{h}\|$$

We therefore prove the theorem.

We are ready to establish the main theorem.

Theorem: The chain rule
Let $A \subset \mathbb{R}^n$ be open and $\mathbf{t} : A \to \mathbb{R}^m$ is injective, differentiable, and $J_\mathbf{t}$ exists everywhere in A and the λ is the Lebesgue measure on $(A, \mathcal{B}(A))$. We have

$$\forall f \in \mathcal{L}^+ \cup \mathcal{L}^1,\ \int_{\mathbf{t}(A)} f \mathrm{d}\lambda = \int_A f \circ \mathbf{t} \times |\det(J_\mathbf{t})| \mathrm{d}\lambda$$

In fact, we do not require the existence of the Jacobian, in which case we use the derivative $\mathbf{t}'$.

We need two Lemmas. The first one is based on Brouwer fixed point theorem, that is a continuous function mapping a closed ball (or more generally, a compact and convex set) in the Euclidean space to itself have a fix point, i.e.

$$f : \bar{B} \mapsto \bar{B} \ \Rightarrow\ \exists x \in A \text{ such that } f(x) = x, \text{ given } \bar{B} \in \mathbb{R}^n \text{ is a closed ball}$$

A interesting proof with elementary algebra using Sperner's lemma can be found in here.

Lemma: Let $B := B(\mathbf{0}, r) \subset \mathbb{R}^n$, $S := \{\mathbf{x} : \|\mathbf{x}\| = r\}$ be its boundary and $\bar{B}$ be the closed ball of B. If for some $0 < \epsilon < 1$, $\mathbf{f} : \bar{B} \mapsto \mathbb{R}^k$ satisfies $\|\mathbf{f}(\mathbf{x}) - \mathbf{x}\| < \epsilon r\ \forall \mathbf{x} \in S$, then $B(\mathbf{0}, (1-\epsilon)r) \subset \mathbf{f}(B)$

Proof:

We prove by contradiction in the case when $r = 1$. Other cases hold through scaling.

Suppose, otherwise, there exists $\mathbf{x}_0 \in B(\mathbf{0}, 1-\epsilon)$ such that $\mathbf{x}_0 \notin \mathbf{f}(B)$. Note that $\mathbf{x}_0 \notin \mathbf{f}(S)$ by definition, thus $\mathbf{x}_0 \notin \mathbf{f}(\bar{B})$. We can define the continuous function $\mathbf{g} : \bar{B} \mapsto \bar{B}$ by

$$\mathbf{g}(\mathbf{x}) := \frac{\mathbf{x}_0 - \mathbf{f}(\mathbf{x})}{\|\mathbf{x}_0 - \mathbf{f}(\mathbf{x})\|}$$

Obviously $\mathbf{g}(\bar{B}) = S$. We claim that g do not have fixed point, which contradict to the Brouwer's theorem. Because the fixed point only exists in S, but $\forall \mathbf{x} \in S$,

$$\mathbf{x} \cdot (\mathbf{x}_0 - \mathbf{f}(\mathbf{x})) = \mathbf{x} \cdot \mathbf{x}_0 + \mathbf{x} \cdot (\mathbf{x} - \mathbf{f}(\mathbf{x})) - \mathbf{x} \cdot \mathbf{x} \le \|\mathbf{x}\|\|\mathbf{x}_0\| + \|\mathbf{x}\|\epsilon - 1 = \|\mathbf{x}_0\| + \epsilon - 1 < 0$$

thus $\mathbf{x} \cdot \mathbf{g}(\mathbf{x}) < 0$ and $\mathbf{x} \neq \mathbf{g}(\mathbf{x})$.

Lemma

If $A \subset \mathbb{R}^n$ is a open set, $\mathbf{t} : A \mapsto \mathbb{R}^m$ is differentiable at $\mathbf{x} \in A$, the Jacobian $J_{\mathbf{t}}(\mathbf{x})$ exists, and is continuous at the neighborhood of $\mathbf{x}$, then

$$\lim_{r \to 0} \frac{\lambda[\mathbf{t}(B(\mathbf{x}, r))]}{\lambda[B(\mathbf{x}, r)]} = |\det [J_{\mathbf{t}}(\mathbf{x})]\,|$$

Proof:

First note that a open ball is a countable union of closed ball. $\mathbf{t}$, being continuous near $\mathbf{x}$, maps compact set to compact set. Therefore $\mathbf{t}(B(\mathbf{x}, r)) \in \mathcal{B}(\mathbb{R}^m)$ when r is small.

Denote $J := J_{\mathbf{t}}(\mathbf{x})$. If J is invertible, define $\mathbf{g} = J^{-1}\mathbf{t} : J \mapsto \mathbb{R}^m$. We have $J_{\mathbf{g}} = J^{-1}J = I$, and by a theorem in linear algebra (see the supplementary note), $\lambda[\mathbf{t}(B)] = \lambda[J(\mathbf{g}(B))] = |\det(J)|\lambda[\mathbf{g}(B)]\ \forall B \in \mathcal{B}(A)$. It suffices to prove that

$$\lim_{r \to 0} \frac{\lambda[\mathbf{g}(B(\mathbf{x}, r))]}{\lambda[B(\mathbf{x}, r)]} = 1$$

By the definition of derivative, for any $0 < \epsilon < 1$, $\exists \delta > 0$ such that $\|\mathbf{h}\| < \delta$ gives $\|\mathbf{g}(\mathbf{h}+\mathbf{x}) - \mathbf{g}(\mathbf{x}) - \mathbf{h}\| \le \epsilon\|\mathbf{h}\|$. By the previous lemma, letting $\mathbf{h}$ be the variable, $0 < r < \delta$, and $\mathbf{f}(\mathbf{h}) = \mathbf{g}(\mathbf{h}+\mathbf{x}) - \mathbf{g}(\mathbf{x}) : \bar{B}(\mathbf{0}, r)$, we have $B(\mathbf{0}, (1-\epsilon)r) \subset \mathbf{f}(B(\mathbf{0}, r))$. Also, $\|\mathbf{f}(\mathbf{h})\| \le (1+\epsilon)\|\mathbf{h}\| \le (1+\epsilon)r\ \forall \mathbf{h} \in B(\mathbf{0}, r)$. Therefore,

$$\forall 0 < r < \delta,\ B(\mathbf{0}, (1-\epsilon)r) \subset \mathbf{f}(B(\mathbf{0}, r)) \subset B(\mathbf{0}, (1+\epsilon)r), \text{ and then } (1-\epsilon)^n \le \frac{\lambda[\mathbf{f}(B(\mathbf{0}, r))]}{\lambda[B(\mathbf{0}, r)]} \le (1-\epsilon)^n$$

The result follows by taking $\epsilon \to 0$. Note that $\lambda[\mathbf{f}(B(\mathbf{0}, r))] = \lambda[\mathbf{g}(B(\mathbf{x}, r))]$.

If J is not invertible. $\det(J) = 0$ and $J(\mathbb{R}^n)$ is a subspace in $\mathbb{R}^m$ with strictly smaller dimension.

Thus $\lambda[J(\mathbb{R}^n)] = 0$. Construct the open set

$$Q_{c,r} = \{\mathbf{x} \in \mathbb{R}^m : \exists \mathbf{y} \in J(B(\mathbf{0}, r)),\ d(\mathbf{x}, \mathbf{y}) < cr\}\ \forall c, r > 0$$

We have $\forall r > 0, \epsilon > 0$, $\exists c$ such that $\lambda(Q_{c,r}) < \epsilon r^m$. Still, by the definition of derivative, for any $c > 0$, $\exists \delta > 0$ such that $\|\mathbf{h}\| < \delta$ gives $\|\mathbf{g}(\mathbf{h} + \mathbf{x}) - \mathbf{g}(\mathbf{x}) - J\mathbf{h}\| \leq c\|\mathbf{h}\|$. Then for any $0 < r < \delta$, $\mathbf{f}(B(\mathbf{0}, r)) \subset Q_{c,r}$ ($\mathbf{f}$ is as defined above). By choosing c according to ϵ, we have

$$\lambda[\mathbf{g}(B(\mathbf{x}, r))] = \lambda[\mathbf{f}(B(\mathbf{0}, r))] \leq \lambda(Q_{c,r}) \leq \epsilon r^m = \epsilon \frac{\lambda[B(\mathbf{0}, r]}{\lambda[B(\mathbf{0}, 1]}$$

The taget limit is therefore 0.

Proof of the chain rule

Note that the Luzin N property of differentiable function (see 5.2.2) holds for higher dimension case with similar proof. Since $\mathbf{t} : A \to \mathbf{t}(A)$ is invertible. Thus $\mathbf{t}(B) \in \mathcal{B}(\mathbb{R}^m)$ whenever $B \in \mathcal{B}(\mathbb{R}^n)$, and $\lambda \circ \mathbf{t}$ is a measure in $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ that $\lambda \circ \mathbf{t} \ll \lambda$. For $i \in \mathbb{N}^*$, define

$$A_i := \{\mathbf{x} \in A : \|\mathbf{t}(\mathbf{x})\| < i\}$$

to be open set and the measure

$$\mu_i(B) = \lambda[\mathbf{t}(B)]\ \forall B \in \mathcal{B}(A_i)$$

Thus μ_i is bouned, and $\mu_i \ll \lambda$. Then $\frac{\mathrm{d}\mu_i}{\mathrm{d}\lambda} \in \mathcal{L}'(\lambda)$ exists almost everywhere.
We claim that $\frac{\mathrm{d}\mu_i}{\mathrm{d}\lambda} \overset{a.s.}{=} |\mathrm{det} J_{\mathbf{t}}|$. Observe that for any $i \in \mathbb{N}^*$, fix $\mathbf{x} \in A_i$, $B(\mathbf{x}, r) \subset A_i$ for some small enough r, thus $\mu_i[B(\mathbf{x}, r)] = \lambda \circ \mathbf{t}[B(\mathbf{x}, r)]$. By the previous lemma,

$$\lim_{r \to 0} \frac{\mu_i[B(\mathbf{x}, r)]}{\lambda[B(\mathbf{x}, r)]} = \lim_{r \to 0} \frac{\lambda \circ \mathbf{t}[B(\mathbf{x}, r)]}{\lambda[B(\mathbf{x}, r)]} = |\mathrm{det} J_{\mathbf{t}}|$$

It thus can be a Radon-Nikodym derivative of μ_i with reference to λ. Therefore,

$$\forall i \in \mathbb{N}^*, B \in \mathcal{B}(A),\ \lambda \circ \mathbf{t}[B \cap A_i] = \mu_i(B \cap A_i) = \int_{A_i} \mathbf{1}_{B \cap A_i} |\mathrm{det} J_{\mathbf{t}}| \mathrm{d}\lambda = \int_A \mathbf{1}_{B \cap A_i} |\mathrm{det} J_{\mathbf{t}}| \mathrm{d}\lambda$$

Since for any $B \in \mathcal{B}(A)$, $B = \bigcup_{i=1}^{\infty} B \cap A_i$, by monotone convergence theorem, $\lambda \circ \mathbf{t}(B) = \int_B |\mathrm{det} J_{\mathbf{t}}| \mathrm{d}\lambda$. Now suppose $f = \mathbf{1}_{\mathbf{t}(B)} : \mathbb{R}^m \to \mathbb{R}$ (or equivalently choose $B' \in \mathbb{R}^m$ and let $B = \mathbf{t}^{-1}(B')$), we have

$$\int_{\mathbf{t}(A)} f \mathrm{d}\lambda = \lambda[\mathbf{t}(B)] = \int_A \mathbf{1}_B |\mathrm{det} J_{\mathbf{t}}| \mathrm{d}\lambda = \int_A f \circ \mathbf{t} |\mathrm{det} J_{\mathbf{t}}| \mathrm{d}\lambda$$

By linearity, the chain rule holds for every simple function. We can then use monotone convergence theorem to extend to $\mathcal{L}^+$ or $\mathcal{L}^1$ function.

5.2.4 Conditional expectation and conditional probability

Definition: Conditional expectation
Given a probability measure $(\Omega, \mathcal{B}, P)$ and a random variable $X \in \mathcal{L}^1$, let $\mathcal{G} \subset \mathcal{B}$ be a sub σ-algebra of the set $G \in \mathcal{B}$. The expectation conditional to $\mathcal{G}$ is defined to be the function (or a random variable)

$$\mathbb{E}(X|\mathcal{G})(\omega) : (G, \mathcal{G}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$$

whose of $\mathcal{L}^1$ with regard to $(G, \mathcal{G}, P)$ and

$$\forall A \in \mathcal{G},\ \mathbb{E}(X\mathbf{1}_A) = \mathbb{E}\left[\mathbb{E}(X|\mathcal{G})\mathbf{1}_A\right] = \int_A \mathbb{E}(X|\mathcal{G})(x)\mathrm{d}P(x)$$

We now show that this definition is valid. First, suppose $X \in \mathcal{L}^+ \cap \mathcal{L}^1$, we can verify that

$$\tau(A) := \mathbb{E}(X\mathbf{1}_A) = \int_A X\mathrm{d}P\ \forall A \in \mathcal{G}$$

is a finite measure on $(G, \mathcal{G})$, and we have $\tau \ll P$. Thus, Radon–Nikodym derivative $f = \frac{\mathrm{d}\tau}{\mathrm{d}P} \in \mathcal{L}^1(P)$ exists, and

$$\forall A \in \mathcal{G},\ \mathbb{E}(X\mathbf{1}_A) = \int_A f\mathrm{d}P$$

Then the expectation conditional is defined to be f.
Next, when $X \in \mathcal{L}^1$, we can decompose it into $X = X^+ - X^-$ such that $X^+, X^- \in \mathcal{L}^+$. Then the expectation conditional $\mathbb{E}(X|\mathcal{G})$ is defined to be the difference between the corresponding Radon–Nikodym derivative, $f^+ - f^-$.

Conditional expectation shares some common properties with expectation. **Proposition**: properties of conditional expectation

1. (Linearity) Let $X, Y \in \mathcal{L}^1$ be two random variables and $a, b \in \mathbb{R}$, then

$$\mathbb{E}(aX + bY|\mathcal{G}) \overset{a.s.}{=} a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$$

2. (Monotonicity) Let $X, Y \in \mathcal{L}^1$ be two random variables. If $X \overset{a.s.}{\leq} Y$, then $\mathbb{E}(X|\mathcal{G}) \overset{a.s.}{\leq} \mathbb{E}(Y|\mathcal{G})$

3. (Monotone convergence theorem) Let $\{X_n \in \mathcal{L}^1\}_{n=1}^{\infty}$ be an almost surely monotonic increasing sequence of random variables, i.e. $X_n \overset{a.s.}{\leq} X_{n+1}\ \forall n \in \mathbb{N}^*$, and $\lim_{n\to\infty} X_n \overset{a.s.}{=} X$, where X is also measurable, then

$$\mathbb{E}(X|\mathcal{G}) \overset{a.s.}{=} \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{G})$$

4. Let $X, Y \in \mathcal{L}^1$ be two random variables. If $\sigma(Y) \subset \mathcal{G}$ (i.e. Y is measurable with regard to $(G, \mathcal{G}, P)$, in which case $G = \Omega$), then $\mathbb{E}(XY|\mathcal{G}) \overset{a.s.}{=} Y\mathbb{E}(X|\mathcal{G})$

5. As a special case of 4, when $\sigma(Y) \in \mathcal{G}$, $\mathbb{E}(Y|\mathcal{G}) \overset{a.s.}{=} Y$

Proof:
Proof of 1 and 2 is trivial by the definition. To prove 3, we first note that $\{\mathbb{E}(X_n|\mathcal{G})\}_{n=1}^{\infty}$ is almost surely monotonic increasing. We then can apply the monotone increasing theorem of normal expectation

$$\begin{aligned}\forall A \in \mathcal{G},\ \mathbb{E}\left[\mathbf{1}_A \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{G})\right] &= \lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_A \mathbb{E}(X_n|\mathcal{G})\right] \\ &= \lim_{n\to\infty} \mathbb{E}(X_n \mathbf{1}_A) \\ &= \mathbb{E}(\mathbf{1}_A X) = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}(X|\mathcal{G})\right]\end{aligned}$$

thus $\mathbb{E}(X|\mathcal{G}) \overset{a.s.}{=} \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{G})$.
To prove 4, we first consider when Y is an indicator function, i.e. $Y = \mathbf{1}_B$ where $B \in \mathcal{G}$, then

$$\forall A \in \mathcal{G},\ \int_A \mathbb{E}(XY|\mathcal{G})\mathrm{d}P = \int_A X\mathbf{1}_B \mathrm{d}P = \int_{A\cap B} X\mathrm{d}P = \int_{A\cap B} \mathbb{E}(X|\mathcal{G})\mathrm{d}P = \int_A Y\mathbb{E}(X|\mathcal{G})\mathrm{d}P$$

thus $\mathbb{E}(XY|\mathcal{G}) \overset{a.s.}{=} Y\mathbb{E}(X|\mathcal{G})$ when $Y = \mathbf{1}_B$. It further holds for all simple functions by linearity. When $Y \in \mathcal{L}^1 \cap \mathcal{L}^+$, express Y as the limit of a monotonic increasing sequence of $\mathcal{E}^+$ function, the result is proved by the monotone convergence theorem (property 3). Finally, when $Y \in \mathcal{L}^1$, decomposite $Y = Y^+ - Y^-$ and the result follows.

Definition: Conditional probability
Given a probability measure $(\Omega, \mathcal{B}, P)$, let $\mathcal{G} \subset \mathcal{B}$ be a sub σ-algebra of the set $G \in \mathcal{B}$. The probability of event $A \in \mathcal{B}$ conditional to $\mathcal{G}$ is defined to be

$$P(A|\mathcal{G}) := \mathbb{E}(\mathbf{1}_A|\mathcal{G})$$

or equivalently, the $\mathcal{L}^1$ random variable $(G, \mathcal{G}) \mapsto (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ satisfying the following

$$\forall A \in \mathcal{B}, B \in \mathcal{G},\ P(A \cap B) = \int_B P(A|\mathcal{G})\mathrm{d}P$$

By monotonicity of conditional expectation, one can show that $P(A|\mathcal{G})(\omega_2) \in [0, 1]\ \forall A \in \mathcal{B}, \omega_2 \in G$.

Proposition: regular conditional distribution
Let $\mathbf{X} : (\Omega, \mathbb{B}) \mapsto (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $k \in \mathbb{N}^*$ be a random vector. Then, for any $B \in \mathcal{B}(\mathbb{R}^k)$ and almost

every $\omega_2 \in G$, $P(\mathbf{X} \in B|\mathcal{G})(\omega_2)$ equals to a measurable kernel $\tau'(\omega_2, B)$, i.e., $\tau' : G \times \mathcal{B}(\mathbb{R}^k) \mapsto [0,1]$, $\tau'(\omega_2, *)$ is a σ-finite measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)$ for any $\omega_2 \in G$, and $\tau'(*, B)$ is a measurable function on $(G, \mathcal{G})$ for any $B \in \mathcal{B}(\mathbb{R}^k)$. Such kernel is called the regular conditional distribution, denoted as $P \circ X^{-1}|_{\mathcal{G}}(B|\omega_2)$. We have for almost every $\omega_2 \in G$,

$$P \circ X^{-1}|_{\mathcal{G}}(\mathbb{R}^k|\omega_2) = P(\mathbf{X} \in \mathbb{R}^k|\mathcal{G})(\omega_2) = P(\Omega|\mathcal{G})(\omega_2) = 1$$

Thus $P \circ X^{-1}|_{\mathcal{G}}(*|\omega_2)$ is a probability measure.
Further, for any real function $g : \mathbb{R}^k \mapsto \mathbb{R}$ such that $g(X) \in \mathcal{L}^1$,

$$\mathbb{E}[g(X)|\mathcal{G}] \overset{a.s.}{=} \int_{\mathbb{R}^k} g(\mathbf{x}) \mathrm{d}\tau'(*, \mathbf{x}) = \int_{\mathbb{R}^k} g(\mathbf{x}) \mathrm{d}P \circ X^{-1}|_{\mathcal{G}}(\mathbf{x}|*)$$

The almost surely statement is with respect to $(G, \mathcal{G}, P)$.
Proof:
The proof of the equation almost surely to a kernel is very similar to a proof in a proposition in 5.2.2, where we utilised the dense set $\mathbb{Q}$ in $\mathbb{R}$. When $k \geq 1$, the claim is still valid since it is isomorphic to $\mathbb{R}$.
We then begin with $g = \mathbf{1}_B$, where $B \in \mathcal{B}(\mathbb{R}^k)$, then

$$LHS = \mathbb{E}[\mathbf{1}_{\mathbf{X}^{-1}(B)}|\mathcal{G}] = P[\mathbf{X} \in B|\mathcal{G}] \overset{a.s.}{=} \tau'(*, B) = RHS$$

Next, we can extend g into $\mathcal{L}^+$ and then $\mathcal{L}^1$ function by linearity and monotone convergence theorem of conditional expectation.

Definition: Conditional expectation, variance and probability to a random variable
Given a probability measure $(\Omega, \mathcal{B}, P)$, X and $\{X_t\}_{t \in T}$ are some random variables (where T is a index set). Let $\mathcal{G} := \sigma(\{X_t\}_{t \in T})$ be the generated σ-algebra. The conditional expectation and probability conditional to $\{X_t\}_{t \in T}$ are defined to be

$$\mathbb{E}(X|\{X_t\}_{t \in T}) := \mathbb{E}(X|\mathcal{G}),\ P(X|\{X_t\}_{t \in T}) = P(X|\mathcal{G})$$

We then define the conditional variance whenever $X \in \mathcal{L}^2$

$$\mathrm{var}(X|Y) := \mathbb{E}\left\{[X - \mathbb{E}(X|Y)]^2 \,\middle|\, Y\right\}$$

The definition is valid since $[X - \mathbb{E}(X|Y)]^2$ is a $\mathcal{L}^1$ random variable in $(\Omega, \mathcal{B}, P)$.
We will show the approach to obtain conditional expectations in 5.3.2 using probability density.

When T is finite and we express $\{X_t\}_{t \in T}$ as a vector $\mathbf{X}$, we can write $\mathbb{E}(X|\mathbf{X})$ and $P(X|\mathbf{X})$. We also use $\mathbb{E}(X|Y)$ or $P(X|X_1, X_2)$ to express the conditional expectation or probability of the specific random variables.

Proposition: properties about conditional expectation and variance with respect to random variable

1. For random variable X, Y and any function $f : \bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ such that $X, Xf(Y) \in \mathcal{L}^1$, we have $\mathbb{E}\left[Xf(Y)|Y\right] \overset{a.s.}{=} f(Y)\mathbb{E}(X|Y)$
2. There exists a function $g : \bar{\mathbb{R}} \mapsto \bar{\mathbb{R}}$ such that $\mathbb{E}(X|Y) = g(Y)$
3. (Law of total expectation) When $X \in \mathcal{L}^1$, $\mathbb{E}(X) = \int_\Omega \mathbb{E}(X|Y)\mathrm{d}P = \mathbb{E}\left[\mathbb{E}(X|Y)\right]$
4. (Law of total variance) When $X \in \mathcal{L}^2$, $\mathrm{var}(X) = \mathbb{E}\left[\mathrm{var}(X|Y)\right] + \mathrm{var}\left[\mathbb{E}(X|Y)\right]$

Proof:
1 is given by property 4 of conditional expectation.
2 follows from this lemma,

$$\text{If function } f \text{ is measurable on } (\Omega, \sigma(Y), P)\text{, then } \exists g \text{ such that } f = g(Y)$$

To prove this lemma, we first consider the case when f is an indicator function $\mathbf{1}_A$, where $A \in \sigma(Y)$, then

$$f = \mathbf{1}_A = \mathbf{1}_{\{Y(\omega):\omega\in\Omega\}}(Y)$$

By linearity, we can ensure the existence of g when f is a simple function. Next, when $f \in \mathcal{L}^+$, it is the point-wise limit of a monotonic increasing sequence of $\mathcal{E}^+$ simple functions, i.e.

$$f = \lim_{n\to\infty} \phi_n, \text{ where } \forall n \in \mathbb{N}^*,\ \phi_n \le \phi_{n+1} \text{ and } \phi_n \in \mathcal{E}^+ \ \Rightarrow\ \exists g_n \text{ such that } \phi_n = g_n(Y)$$

Let $g = \sup_{n\in\mathbb{N}^*} g_n$, we have $g(Y)$ measurable and $f = g(Y)$. For a general measurable function f, the result follows by considering the decomposition $f = f^+ - f^-$.
Proposition 2 holds because $E(X|Y)$ is a measurable function with respect to $(\Omega, \sigma(Y), P)$.
3 is given by the definition and $\Omega \in \sigma(Y)$
To prove 4, we observe

$$\begin{aligned}
\mathrm{var}(X|Y) =& \mathbb{E}\left\{\left[X^2 - 2X\mathbb{E}(X|Y) + \mathbb{E}(X|Y)^2\right]\big|Y\right\} \\
\overset{a.s.}{=}& \mathbb{E}(X^2|Y) - 2\mathbb{E}\left[X\mathbb{E}(X|Y)|Y\right] + \mathbb{E}\left[\mathbb{E}(X|Y)^2|Y\right] \text{ (by linearity)} \\
\overset{a.s.}{=}& \mathbb{E}(X^2|Y) - 2\mathbb{E}(X|Y)\mathbb{E}\left(X\right) + \mathbb{E}(X|Y)^2 \text{ (by 1 and 2)} \\
=& \mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2
\end{aligned}$$

Thus,

$$\mathbb{E}\left[\mathrm{var}(X|Y)\right] = \mathbb{E}\left\{\mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2\right\} = \mathbb{E}(X^2) - \mathbb{E}\left[\mathbb{E}(X|Y)^2\right]$$

and

$$\mathrm{var}\left[\mathbb{E}(X|Y)\right] = \mathbb{E}\left[\mathbb{E}(X|Y)^2\right] - \mathbb{E}\left[\mathbb{E}(X|Y)\right]^2 = \mathbb{E}\left[\mathbb{E}(X|Y)^2\right] - \mathbb{E}(X)^2$$

Therefore, $\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}\left[\mathrm{var}(X|Y)\right] + \mathrm{var}\left[\mathbb{E}(X|Y)\right]$.

5.3 Probability density

5.3.1 Probability density defined in different cases

Definition: Probability density of random variable with absolute continuous *pdf.*
Given a random variable X on a probability space $(\Omega, \mathcal{F}, P)$ that $|X| \overset{a.s.}{<} \infty$, if its *pdf.* F is absolute continuous on every closed interval in $\mathbb{R}$, we define $f = F'$ to be the probability density function of X.
Further, since $P \circ X^{-1}$ is the Lebesgue-Stieltjes measure generated by $F|_{\mathbb{R}}$, from 5.2.2, we have

$$\forall A \in \mathcal{B}(\mathbb{R}),\ P(X \in A) = P \circ X^{-1}(A) = \int_A f \mathrm{d}\lambda$$

We have shown that F is absolute continuous if and only if $P \circ X^{-1}|_{\mathbb{R}} \ll \lambda$. (we used the $P \circ X^{-1}|_{\mathbb{R}}(A) := P \circ X^{-1}(\mathbb{R} \cap A)$ to denote a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.)

Definition: Probability density of discrete random variable
Given a discrete random variable X on a probability space $(\Omega, \mathcal{F}, P)$ which only takes a countable set $R \in \mathcal{B}(\bar{\mathbb{R}})$ of values (then $\sigma(R) = \mathcal{P}(R)$, $P(X \in R) = 1$ and $P(X = x) > 0\ \forall x \in R$). We define

$$\forall x \in R,\ f(x) := P(X = x) = \sum_{t \in R} f(t) \mathbf{1}_{\{t\}}(\{x\})$$

to be the probability mass function of X. We can express the probability as an integral of f with regard to the counting measure c on $(R, \sigma(R))$.

$$\forall A \in \sigma(R),\ P(X \in A) = \sum_{x \in A} f(x) = \sum_{t \in A} f(t) c(\{t\}) = \int_A f \mathrm{d}c$$

because when A is finite, $f = \sum_{t \in A} f(t) \mathbf{1}_{\{t\}}$ is a simple function, when A is infinite but countable, the equation still holds by monotone convergence theorem. Then we can still regard f as probability density with regard to counting measure. We also have $P \circ X^{-1}|_R \ll c$

It leads to a more general definition of probability density that use the Radon-Nikodym derivative.
Definition: Probability density
Let X be a random variable on the probability measure space $(\Omega, \mathcal{F}, P)$. Obviously, the measure $P \circ X^{-1}$ is finite and thus σ-finite. If there exists a reference measure μ on $(R, \mathcal{B})$ such that $R \in \mathcal{B}(\bar{\mathbb{R}})$, $P \circ X^{-1}(\bar{\mathbb{R}} \setminus R) = 0$ and $P \circ X^{-1}|_R \ll \mu$, then we define the joint probability density function

$$f : (R, \mathcal{B}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

to be the Radon-Nikodym derivative of $P \circ X^{-1}$ with regard to μ.
Then, we have

$$\forall A \in \mathcal{B},\ P \circ X^{-1}(A) = \int_A f \mathrm{d}\mu$$

and therefore $f \overset{a.s.}{\geq} 0$. Also,

$$\mathbb{E}(X) = \int_R x f(x) \mathrm{d}\mu$$

It is clear that probability density f uniquely determines the *pdf.* of X. Because if two *pdf.* F_1, F_2 share the same density f, then

$$\forall x \in \bar{\mathbb{R}},\ F_1(x) = \int_{R \cap [-\infty, x]} f \mathrm{d}\mu = F_2(x)$$

Thus, we can recover the *pdf.* from any proper density on $(R, \mathcal{B}, \mu)$. For a function f to be a proper density, it suffices that

1. $f : (R, \mathcal{B}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable
2. $f \overset{a.s.}{\geq} 0$
3. $f \in \mathcal{L}^1(\mu)$ and $\int_R f \mathrm{d}\mu = 1$

However, even when a reference measure μ is chosen, a *pdf.* can still correspond to multiple almost surely equal density functions (with respect to μ).

Proposition: The limit of density functions is a density
Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of density functions with reference measure μ that converge point-wise to f, $\lim_{n\to\infty} f_n = f$. If the dominated convergence applies, i.e., $\exists g \in \mathcal{L}^1(\mu)$ such that $|f_n| \overset{a.s.}{\leq} g\ \forall n \in \mathbb{N}^*$, then f is a proper density with reference measure μ.
The proof is easy by verifying the three conditions. Further, in 7.2.1 we will show that this implies convergence in distribution.

The following is given by the properties of Radon-Nikodym derivatives

Proposition: calculating expectation from density

If f is the density (with the reference measure space $(R, \mathcal{B}, \mu)$) of a random variable X, then

$$\mathbb{E}(X) = \int_R x f(x) \mathrm{d}\mu(x) \text{ whenever } X \in \mathcal{L}^+$$

If $g : R \mapsto \bar{\mathbb{R}}$, then

$$\mathbb{E}[g(X)] = \int_R g(x) f(x) \mathrm{d}\mu(x) \text{ whenever } g(X) \in \mathcal{L}^+$$

Then by considering $g^+ := \max\{g, 0\}$ and $g^- := \max\{-g, 0\}$ separately, we have

$$\mathbb{E}[g(X)] = \int_R g^+(x) f(x) \mathrm{d}\mu(x) - \int_R g^-(x) f(x) \mathrm{d}\mu(x) \text{ whenever } \mathbb{E}[g(X)] \text{ exists}$$

Note that not all random variable has a density. In the above two cases (absolute continuous and discrete), μ is either the Lebesgue measure or a counting measure. In fact, it is enough to consider these two measures to cover most cases, since we have this theorem.

Theorem: Decomposition of probability measure

Let X be a random variable on $(\Omega, \mathcal{B}, P)$ and λ is the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then $P \circ X^{-1}$ can be expressed as the sum of three measures,

$$\forall A \in \mathcal{B}(\bar{\mathbb{R}}),\ P \circ X^{-1}(A) = u(A \cap \mathbb{R}) + v(A \cap R) + w(A) \text{ such that } u \ll \lambda, v \ll c, w|_{\mathbb{R}} \perp \mu$$

where u is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, v and c on $(R, \mathcal{P}(R))$, $R \subset \bar{\mathbb{R}}$ is countable and c is the counting measure.

The construction follows from the Radon-Lebegues decomposition theorem. Since $P \circ X^{-1}|_{\mathbb{R}}$ and λ are σ-finite, there exists a unique decomposition $P \circ X^{-1}|\mathbb{R} = u + v'$ such that $u \ll \lambda$ and $v' \perp \lambda$. To get v, consider all the discontinuous points of the *pdf.* F, which is at most countable. Let

$$R := \{x \in \bar{\mathbb{R}} : F \text{ is discontinuous on } x\} \text{ and let } R := R \cup \{-\infty\} \text{ if } F(-\infty) \neq 0$$

Let $v := P \circ X^{-1}|_R$ and c be the counting measure, both on $(R, \mathcal{P}(R))$. It is obvious that $v|_{\mathbb{R}} \perp \lambda$. The remaining is $w := P \circ X^{-1} - u|_{\mathbb{R}} - v|_R$, which is also a measure on $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ that $w|_{\mathbb{R}} \perp \lambda$. We can define the density of u and v and scale them to be probability density functions, then

$$\forall A \in \mathcal{B}(\bar{\mathbb{R}}),\ P \circ X^{-1}(A) = \alpha_1 \int_{A \cap \mathbb{R}} f_1 \mathrm{d}\lambda + \alpha_2 \int_{A \cap R} f_2 \mathrm{d}c + w(A)$$

where f_1, f_2 are two probability density functions of scaled u, v, and $\alpha_1 = u(\mathbb{R})$ and $\alpha_2 = v(R)$ is the scaling factor.

A normal push-forward measure of a random variable does not have the part w, except for some weird cases (e.g. those with the Cantor Function being the *pdf.*). In the real practice of probability and statistics, we will always avoid these *pdf.* and deal with only absolute continuous variables (only have u), discrete variables (only have v) and their mixed type.
To calculate the expectation of a mixed-type random variable, we have

$$\mathbb{E}[g(X)] = \alpha_1 \int_{\mathbb{R}} g \times f_1 \mathrm{d}\lambda + \alpha_2 \int_{R} g \times f_2 \mathrm{d}c$$

which is not hard to prove. It gives insight that it is enough to study the random variable with a density, while those who do not have (e.g. mixed-type) can be transformed to be the convex sum of what we know.

5.3.2 Probability density of multiple random variables

We define density for multiple random variables (or random vector) in a similar way.
Definition: Joint probability density
Let $(\Omega, \mathcal{F}, P)$ be a probability measure and $\{X_i\}_{i=1}^n$ be n ($n \in \mathbb{N}^*$) variables, then $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ is a random vector. Fix a set $R \in \mathcal{B}(\bar{\mathbb{R}}^n)$ such that $P(\mathbf{X} \in R) = 1$ and find a reference measure μ on $(R, \mathcal{B})$ such that $P \circ \mathbf{X}^{-1}|_R \ll \mu$. Then there exists a Radon-Nikodym derivative

$$f_{\mathbf{X}} : (R^n, \mathcal{B}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

of $P \circ \mathbf{X}^{-1}$ with regard to μ. The joint probability density is defined to be such derivative when μ is a product measure of n measures

$$(R_i, \mathcal{B}_i, \mu_i) \text{ where } R_i \in \mathcal{B}(\mathbb{R}) \text{ and } R = R_1 \times R_2 \times \cdots \times R_n$$

Denoted as $\mu = \prod_{i=1}^n \mu_i$.
The following properties of probability density still holds,

1. $$\forall A \in \mathcal{B}(\bar{\mathbb{R}}^n),\ P \circ \mathbf{X}^{-1}(A) = \int_A f \mathrm{d}\mu$$
2. For any function $g : \bar{\mathbb{R}}^n \mapsto \bar{\mathbb{R}}$,
$$\mathbb{E}[g(\mathbf{X})] = \int_R g(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x}$$
3. A joint probability density also uniquely determines the joint *pdf.*

Further by Fubini's theorem, we have

$$1 = \int_R f \mathrm{d}\mu = \int_{R_1} \int_{R_2} \cdots \int_{R_n} f \ \mathrm{d}\mu_1 \mathrm{d}\mu_2 \cdots \mathrm{d}\mu_n$$

and the order of $\{\mu_i\}_{i=1}^n$ do not affect the result.

In most practical scenarios in Statistics, the following construction of μ and R is enough. If some entry, say, X_1 only takes a countable set of values and all with positive probability (a.k.a. discrete), simply let R_1 be the set of these values and μ_1 be the corresponding counting measure, then τ' : $R_1 \times \mathcal{B}(\mathbb{R}^{k-1})$,

$$\tau'(x_1, A_2) = P(\mathbf{X} \in \{x_1\} \times A_2) \ \Rightarrow \ \tau(A_1 \times A_2) = \int_{A_1} \tau'(x_1, A_2) \mathrm{d}\mu_1 \ \forall A_1 \subset R_1$$

is a kernel defined in 5.2.2. Then we recursively consider the measure $\mu_{x_1}(*) := P(\{x_1\} \times *)$ for each x_1, taking out discrete entries, until there are no more discrete X_i in the remaining vector $\mathbf{X}' \in \mathbb{R}^k$. Further, if for all outcomes of discrete entries $\mathbf{x}_d \in \prod_{i=1}^{n-k} R_i$, the cumulative function $F'_{\mathbf{x}_d}$ of measure $\mu_{\mathbf{x}_d}(*) := P(\{\mathbf{x}_d\} \times *)$ satisfies some integrably differentiable condition, i.e., for all permutations $\{a_i\}_{i=1}^k$ of $\{i\}_{i=1}^k$, $\frac{\partial}{\partial x_{a_1} \cdots \partial x_{a_k}} F'_{\mathbf{x}_d}$ exists and is of $\mathcal{L}^1(\lambda)$, its density $f'_{\mathbf{x}_d}$ is well-defined in 5.2.2 with respect to product Lebesgue measure λ^k. Combing $f'_{\mathbf{x}_d}$ and the discrete case density, we obtain the density of $\mathbf{X}$,

$$f(\mathbf{x}) = P(\mathbf{X}_d = \mathbf{x}_d) \times f'_{\mathbf{x}_d}, \text{ where } \mathbf{x}_d, \ \mathbf{X}_d \text{ are the discrete entries}$$

with respect to the product of corresponding discrete measures and Lebesgue measures.

We restrict μ to be a product measure to ensure the existence of marginal probability density function.

Definition: Marginal probability density function

Let f be a joint density of a random vector $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ on $(\Omega, \mathcal{F}, P)$ with the reference measure $\mu = \prod_{i=1}^n \mu_i$, where μ_i is a measure on $(R_i \in \mathcal{B}(\mathbb{R}), \mathcal{B}_i)$ for all $i \in \{1, 2, \cdots, n\}$. Define

$$f_{X_i} := \int_{R_1} \cdots \int_{R_{i-1}} \int_{R_{i+1}} \cdots \int_{R_n} f \ \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{i-1} \mathrm{d}\mu_{i+1} \cdots \mathrm{d}\mu_n$$

to be the marginal probability density of X_i for $i \in \{1, 2, \cdots, n\}$.

f_{X_i} can be defined on a subset $R'_i \in \mathcal{B}_i$ where the integral exists. By Fubini's theorem, $\mu_i(R_i \backslash R'_i) = 0$. Set $f_{X_i}(x_i) = 0$ for all $x_i \in R_i \setminus R'_i$. Then

$$f_{X_i} \in \mathcal{L}^1(\mu_i) \text{ and } \int_{R_i} f_{X_i} \mathrm{d}\mu_i = 1$$

Also, from $P(\mathbf{X} \in R) = 1$, we have $P(X_i \in R_i) = 1$ and

$$\forall A \in \mathcal{B}_i,\ P(X_i \in A) = P(\mathbf{X} \in R_1 \times \cdots \times R_{i-1} \times A \times R_{i+1} \times \cdots \times R_n) = \int_A f_{X_i} \mathrm{d}\mu_i$$

which shows that f_{X_i} is a probability density of X_i with reference measure μ_i on $(R_i, \mathcal{B}_i)$.

Theorem: Density condition for independent variables
Let $(\Omega, \mathcal{F}, P)$ be a probability measure and $\{X_i\}_{i=1}^n$ be some random variables. If $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ has an density f with reference to $(\prod_{i=1}^n R_i, \mathcal{B}, \prod_{i=1}^n \mu_i)$ and

$$\forall [x_1, x_2, \cdots, x_n]^T \in R_1 \times R_2 \times \cdots \times R_n,\ f\left([x_1, x_2, \cdots, x_n]^T\right) = \prod_{i=1}^n f_{X_i}(x_i)$$

where f_{X_i} is the marginal density of X_i, then $\{X_i\}_{i=1}^n$ is mutually independent, and vice versa.
Proof:
We only show the proof of the "$\Rightarrow$" direction, while the other one is easy. It suffices to show that for any subset $I \subset \{1, 2, \cdots, n\}$, the *pdf.* satisfies

$$\forall x_i \in \mathbb{R} \text{ for all } i \in I,\ F_I(\{x_i, i \in I\}) = \prod_{i \in I} F_{X_i}(x_i)$$

where F_{X_i} is the *pdf* of X_i and F_I is the joint *pdf.*

Suppose J is the size of I and $I = \{i_1, i_2, \cdots, i_J\}$, then

$$
\begin{aligned}
\forall [x_{i_1}, x_{i_2}, \cdots, x_{i_J}]^T \in \mathbb{R}^J,\ F_I\left([x_{i_1}, x_{i_2}, \cdots, x_{i_J}]^T\right)
&= P\left(\bigcap_{j=1}^{n} \{X_{i_j} \le x_{i_j}\}\right) \\
&= P\left(\bigcap_{j=1}^{n} \{X_{i_j} \le x_{i_j}\} \cap \bigcap_{i \notin I} \{X_i \in \bar{\mathbb{R}}\}\right) \\
&= P\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) \quad \text{by setting the corresponding } A_i \\
&= P\left(\mathbf{X} \in A_1 \times A_2 \times \cdots \times A_n\right) \\
&= \int_{A_1 \times A_2 \times \cdots \times A_n} f \mathrm{d}\mu \\
&= \int_{A_1 \times A_2 \times \cdots \times A_n} \prod_{i=1}^{n} f_{X_i} \mathrm{d}\mu \quad \text{(the condition)} \\
&= \prod_{i=1}^{n} \int_{A_i} f_{X_i} \mathrm{d}\mu_i \ \text{(Fubini's theorem)} \\
&= \prod_{j=1}^{J} F_{X_{i_j}}(x_{i_j}) \times \prod_{i \notin I} 1 \\
&= \prod_{j=1}^{J} F_{X_{i_j}}(x_{i_j})
\end{aligned}
$$

Proposition: density and conditional expectation

Let X, Y be two random variables on $(\Omega, \mathcal{F}, P)$ that agree on a joint probability density f with regard to $(R_X \times R_Y, \mathcal{B}, \mu)$, and f_X, f_Y be the marginal density of Y with regard to $(R_X, \mathcal{B}_X, \mu_X)$ and $(R_Y, \mathcal{B}_Y, \mu_Y)$ respectively, then

$$
\mathbb{E}(X|Y) \overset{a.s.}{=} \int_{S_1} x \frac{f(x,Y)}{f_Y(Y)} \mathrm{d}\mu_X(x)
$$

The almost surely statement is with regard to $(\Omega, \sigma(Y), P)$. A similar statement holds for $\mathbb{E}(Y|X)$.

Proof:

It suffices to show that for all $A \in \sigma(Y)$,

$$
\int_A \mathbb{E}(X|Y) \mathrm{d}P = \int_A \int_{R_X} x \frac{f(x,Y)}{f_Y(Y)} \mathrm{d}\mu_X(x) \mathrm{d}P
$$

or equivalently,

$$
\mathbb{E}(X \mathbf{1}_A) = \int_A \int_{R_X} x \frac{f(x,Y)}{f_Y(Y)} \mathrm{d}\mu_X(x) \mathrm{d}P
$$

Let $A' = \{Y(\omega) : \omega \in A\}$. Then $A' \in \mathcal{B}(\bar{\mathbb{R}})$, $A = Y^{-1}(A')$ and $\mathbf{1}_A = \mathbf{1}_{A'}(Y)$ since $A \in \sigma(Y)$. Then,

$$\begin{aligned}
\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}\left[X\mathbf{1}_{A'}(Y)\right] &= \int_{R_X \times R_Y} x\mathbf{1}_{A'}(y)f(x,y)\mathrm{d}\mu(x,y) \\
&= \int_{R_Y} \mathbf{1}_{A'}(y) \int_{R_X} xf(x,y)\mathrm{d}\mu_X(x)\mathrm{d}\mu_Y(y) \text{ (Fubini's theorem)} \\
&= \int_{R_Y} f_Y(y) \times \left[\mathbf{1}_{A'}(Y) \int_{R_X} x\frac{f(x,y)}{f_Y(y)}\mathrm{d}\mu_X(x)\right] \mathrm{d}\mu_Y(y) \text{ (since } f_Y \overset{a.s.}{>} 0) \\
&= \int_{\Omega} \mathbf{1}_{A'}(Y) \left[\int_{R_X} x\frac{f(x,Y)}{f_Y(Y)}\mathrm{d}\mu_X(x)\right] \mathrm{d}P \;\; \text{(property of density)} \\
&= \int_A \int_{R_X} x\frac{f(x,Y)}{f_Y(Y)}\mathrm{d}\mu_X(x)\mathrm{d}P
\end{aligned}$$

which gives the result.

Here is an easy corollary about conditional density and independence.

$$X \text{ and } Y \text{ are independent} \iff f_{X|Y=y} = f_X \ \forall y \in R_Y \text{ such that } f_Y(y) > 0$$

Definition: Conditional density

Continued from the above proposition, define

$$f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)}$$

on where $f_Y(y) > 0$ to be the density of X conditional to Y. Then we write

$$\mathbb{E}(X|Y) \overset{a.s.}{=} \int_{S_1} xf_{X|Y}(x)\mathrm{d}\mu_X(x)$$

where $f_{X|Y}(x) = \frac{f(x,Y)}{f_Y(Y)}$.

Similarly, we can also define the density of a random vector $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ conditional to $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_m]^T$ by

$$f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) = \frac{f(\mathbf{x},\mathbf{y})}{\int_{R_{Y1}}\int_{R_{Y1}}\cdots\int_{R_{Ym}} f(\mathbf{x},[y_1,y_2,\cdots,y_m]^T)\mathrm{d}\mu_m(y_m)\mathrm{d}\mu_2(y_2)\mathrm{d}\mu_1(y_1)}$$

on where the denominator is larger than 0, in which $\mathbf{y} = [y_1, y_2, \cdots, y_m]^T$, and f is the joint probability density of $[X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_m]^T$ with the reference measure μ being the product measure of $\left\{\{(R_{Xi}, \mathcal{B}_{Xi}, \mu_{Xi})\}_{i=1}^n,\ \{(R_{Yj}, \mathcal{B}_{Yj}, \mu_{Yj})\}_{j=1}^m\right\}$

5.3.3 Density with change of variable

This section discusses the method to obtain the new probability density given the original random vector $\mathbf{X} \in \mathbb{R}^n$ (or random variable if $n = 1$) with known density $f_{\mathbf{X}}$ is transformed by a mapping

$\mathbf{t}: \mathbb{R}^n \mapsto \mathbb{R}^m$. When $\mathbf{X}$ a discrete random vector, $\mathbf{Y}$ is also discrete, and the density of $\mathbf{Y}$ can be obtained by summing up the density of the preimage.

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in \mathbf{X}(\Omega)} f_{\mathbf{X}}(\mathbf{x}) \times \mathbf{1}_{\mathbf{t}(\mathbf{x})=\mathbf{y}} \text{ for any possible } \mathbf{y}$$

When $\mathbf{X}$ is continuous. The (joint) *pdf.* of $\mathbf{Y} := \mathbf{t}(\mathbf{X})$ is first avaliable,

$$F_{\mathbf{Y}}(\mathbf{y}) = P\left(\bigcap_{i=1}^{m} \mathbf{Y}_i \leq \mathbf{y}_i\right) = P\left(\bigcap_{i=1}^{m} \mathbf{t}_i(\mathbf{X}) \leq \mathbf{y}_i\right) \text{ for any possible } \mathbf{y}$$

If the set of possible $\mathbf{y}$ is discrete, the density of $\mathbf{Y}$ with reference to the counting measure is available. If $F_{\mathbf{Y}}$ satisfies some differentiable conditions mentioned in 5.3.2 such that $\mathbf{Y}$ is a continuous random vector, the density can be obtained by differentiating $F_{\mathbf{Y}}$ in all its dimension.
For the case that both $\mathbf{X}$ and $\mathbf{Y}$ are continuous random vectors, under some conditions on $\mathbf{t}$, we can use the chain rule to get a simplified result.

Theorem: Change of variable
Let $\mathbf{X}$ be a n dimension continuous random vector on the probability space $(\Omega, \mathcal{F}, P)$, and $f_{\mathbf{X}}: \mathbb{R}^n \mapsto [0, \infty]$ is its density with reference to Lebesgue measure λ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n)$. Suppose $\mathbf{t}: \mathbb{R}^n \mapsto \mathbb{R}^m$ is an injective transformation whose Jacobian $J_{\mathbf{t}}$ exists and is almost everywhere invertible (with respect to Lebesgue measure). Then the density of $\mathbf{Y} := t(\mathbf{X})$ is given by

$$f_{\mathbf{Y}} \overset{a.s.}{=} f_{\mathbf{X}} \circ t^{-1} \times |\det[J_{\mathbf{t}^{-1}}]|$$

with reference to Lebesgue measure λ_m on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R})^m)$. It also has the property

$$f_{\mathbf{Y}}[\mathbf{t}(\mathbf{x})] = f_{\mathbf{X}}(\mathbf{x}) \times |\det[J_{\mathbf{t}}(\mathbf{x})]|^{-1} \text{ for } \mathbf{x} \text{ almost everywhere on } \mathbf{X}(\Omega)$$

Plug in $\mathbf{x} = \mathbf{t}^{-1}(\mathbf{y})$ and compare with the first equation, we can observe that $|\det[J_{\mathbf{t}^{-1}}(*)]| \overset{a.s.}{=} |\det(J_{\mathbf{t}}[\mathbf{t}^{-1}(*)])|^{-1}$. This is the reason why we should restrict on $\mathbf{t}$ to have non-zero $\det(J_{\mathbf{t}})$ almost everywhere. In the case of 0 appearing in the denominator of $\det[J_{\mathbf{t}^{-1}}(\mathbf{y})]$ for some $\mathbf{y} \in \mathbf{t} \circ \mathbf{X}(\Omega)$, the corresponding density is $+\infty$. The set of such points $\mathbf{y}$ must have Lebesgue measure 0.
Proof:
Note that t, t^{-1} are continuous. For any open set $B \subset \mathbf{Y}(\Omega)$, $\mathbf{t}^{-1}(Y)$ is a open set in $\mathbf{X}(\Omega)$. We use the chain rule in 5.2.3. with respect to $\mathbf{t}^{-1}$,

$$\int_{\mathbf{t}^{-1}(B)} f_{\mathbf{X}} \mathrm{d}\lambda_n = \int_B f_{\mathbf{X}} \circ \mathbf{t}^{-1}(\mathbf{y}) \times |\det[J_{\mathbf{t}^{-1}}(\mathbf{y})]| \mathrm{d}\lambda_m(\mathbf{y})$$

The above holds for all open sets $B \subset \mathbf{Y}(\Omega)$. Thus they also holds for any $B \in \mathcal{B}(\mathbf{Y}(\Omega))$. It follows that the part in the RHS integral is a density.
The second equation can be derived similarly, using the chain rule on $\mathbf{t}$.