Real Analysis and Machine Learning

Ruixin Guo

Department of Computer Science
Kent State University

Mar 7, 2025

Contents

Introduction

The development of mathematics is driven by problems. Some problems arise from the real world, especially from fields like physics and computer science. Others originate within mathematics itself, often motivated by the goals of simplification or greater rigor. The branch of mathematics that focuses on the first type of problem is known as applied mathematics, while the branch focusing on the second type is known as pure mathematics.

Modern calculus, primarily developed by Newton and Leibniz, was motivated by the need for mathematical tools to explain experimental findings in physics and astronomy during the 17th century. Early calculus lacked rigor and contained logical inconsistencies, making it unsatisfactory to the mathematical community. To address this, mathematicians sought rigorous definitions for limits and continuity, which led to the development of real analysis. The first precise definition of limits – the (ϵ, δ)-definition – was introduced in the early 19th century. Based on this definition, the well-known Riemann integral was developed.

Introduction

Early real analysis focused on calculus on the real line. Later, mathematicians realized that the real line was too limited to describe many objects in the real world, and they began to consider sets as a more general framework. Set theory was developed by Cantor in 1870s. Modern real analysis is built upon set theory and is grounded in the (ϵ, δ)-definition of limits over sets.

There are several important branches of modern real analysis, including topology, measure theory, and functional analysis. These areas focus on different questions but are often closely interconnected.

Topology studies the geometric properties of sets – such as openness and closedness – under the set-theoretic definition of limits. Once openness is defined, we can then define the compactness of a set, which plays a crucial role in analyzing the continuity of functions on sets.

Introduction

Measure Theory studies the Lebesgue integral and the conditions under which it can be defined. Mathematicians found the Riemann integral to be inflexible for handling functions on general sets, so they developed the Lebesgue integral as a replacement. One key component of the Lebesgue integral is the measure, a nonnegative number assigned to a set to represent its "size". For a given measure, there often exist sets on which it cannot be defined. Conversely, the collection of all sets on which the measure is defined forms a σ-algebra. To ensure that a measure is well-defined, one must first show the existence of its corresponding σ-algebra.

Functional Analysis studies linear mappings from a vector space to a real line. The elements of the vector space can be functions, and such linear mappings are called linear functionals. Under certain conditions, the Lebesgue integral can be a linear functional that maps a space of measurable functions to a real line.

Contents

Finite, Countable and Uncountable Sets

Definition 1.1: Let A and B be two sets. If there exists a bijection between A and B, we say A and B are equivalent, and write $A \sim B$.
The equivalence operation $\sim$ has the following properties:

- Reflexive: $A \sim A$
- Symmetric: If $A \sim B$, then $B \sim A$
- Transitive: If $A \sim B$ and $B \sim C$, then $A \sim C$.

Definition 1.2: For any integer n, let $J_n = \{1, 2, ..., n\}$, we say:

- A is finite if $A \sim J_n$ for some n. (The empty set is also considered to be finite.)
- A is infinite if A is not finite.
- A is countable if $A \sim \mathbb{N}^+$, where $\mathbb{N}^+$ is the set of all positive integers.
- A is uncountable if A is neither finite nor countable.

Theorem 1.3: The set of all integers are countable.

Proof: The elements of $\mathbb{N}^+$ and $\mathbb{N}$ can be paired by

$$\mathbb{N} : 0, 1, -1, 2, -2, ...$$
$$\mathbb{N}^+ : 1, 2, 3, 4, 5, ...$$

Then there is a bijective function $f : \mathbb{N}^+ \to \mathbb{N}$ that $f(k) = k/2$ if k is even and $f(k) = -(k-1)/2$ if k is odd. Thus $\mathbb{N}^+ \sim \mathbb{N}$, $\mathbb{N}$ is countable. $\square$

Theorem 1.4: Every infinite subset of a countable A is countable.

Proof: Arrange the elements of A in a sequence $\{x_i\}_{i=1}^{\infty}$. Suppose an infinite subset $E \subset A$ is obtained by taking the subsequence $\{x_{i_k}\}_{k=1}^{\infty}$ where $1 \leq i_1 < i_2 < i_3 ....$ We define $f(k) = x_{i_k}$, then f is a bijection between $\mathbb{N}^+$ and E. $\square$

Finite, Countable and Uncountable Sets

Theorem 1.5: Let $\{E_i\}_{i=1}^{\infty}$ be a sequence of countable sets, and define $S = \bigcup_{i=1}^{\infty} E_i$, then S is countable.

Proof: Suppose the elements in E_i is arranged in a sequence $\{x_{ik}\}_{k=1}^{\infty}$. We arrange all elements in $\{E_i\}_{i=1}^{\infty}$ into an infinite array:

$$
\begin{array}{ccccc}
x_{11} & x_{12} & x_{13} & x_{14} & \cdots \\
x_{21} & x_{22} & x_{23} & x_{24} & \cdots \\
x_{31} & x_{32} & x_{33} & x_{34} & \cdots \\
x_{41} & x_{42} & x_{43} & x_{44} & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots
\end{array}
$$

And we get

$$x_{11}, x_{21}, x_{12}, x_{31}, x_{22}, x_{13}, x_{41}, x_{32}, x_{23}, x_{14} \ldots \tag{1}$$

We can map the integer sequence 1, 2, 3, 4, ... to the above sequence. Let S be the set of the sequence Eq (1) which is formed by removing the redundant elements, then there must exist $T \subset \mathbb{N}^{+}$ such that $T \sim S$. Since $\mathbb{N}^{+}$ is countable, by Theorem 1.4, T is countable, thus $\mathbb{N}^{+} \sim T \sim S$, which implies S is countable. $\square$

Theorem 1.6: The set of all rational numbers is countable.

Proof: Since any rational number can be expressed in the form p/q where $p, q \in \mathbb{N}$ and $q \neq 0$, we define $E_q = \{p/q \,|\, p \in \mathbb{N}\}$ where $q \neq 0$, then E_q is a countable set. Since $\mathbb{Q} = \bigcup_{q \in \mathbb{N}/\{0\}} E_q$, $\mathbb{Q}$ is a countable union of countable sets. By Theorem 1.5, $\mathbb{Q}$ is countable. $\square$

Theorem 1.7: The set of all real numbers is uncountable.

Proof: First we show that $(0, 1) \sim \mathbb{R}$: $(-\frac{\pi}{2}, \frac{\pi}{2}) \sim \mathbb{R}$ because of the bijection $f(x) = \tan(x)$, and $(0, 1) \sim (-\frac{\pi}{2}, \frac{\pi}{2})$ because of the bijection $f(x) = \frac{\pi}{2}x$. Thus $\mathbb{R}$ is countable if and only if $(0, 1)$ is countable.

Let $E = \{s_1, s_2, s_3 ...\}$ be a subset of $(0, 1)$ where $s_i = 0.a_{i1}a_{i2}a_{i3}a_{i4}...$ for $i \in \mathbb{N}^+$. Each a_{ij} is a digit ranged in $\{0, 1, 2, ..., 9\}$. Clearly $\mathbb{N}^+ \sim E$ because of the bijection $f(i) = s_i$.

We take the sequence $a_{11}, a_{22}, a_{33}...$ and construct a sequence $b_1, b_2, b_3...$ in a way that for $k \in \mathbb{N}^+$, if $a_{kk} = 1$, then $b_k = 2$; if $a_{kk} \neq 1$, then $b_k = 1$. Therefore, the number $b = 0.b_1b_2b_3b_4...$ will have at least one digit that different from any s_i. Thus $b \notin E$. But note that $b \in (0,1)$, which implies $E \subset (0,1)$.

We have shown that any countable subset E of $(0,1)$ is a proper subset. If $(0,1)$ is countable, then $(0,1)$ is a proper subset of $(0,1)$, which is impossible. Thus $(0,1)$ is uncountable. $\square$

Definition 1.8: A set X, whose elements are called points, is said to be a metric space if with any two points $p, q \in X$, there is an associated real number $d(p, q)$, called the distance of p, q, satisfying that

- $d(p, q) > 0$ if $p \neq q$; $d(p, p) = 0$.
- $d(p, q) = d(q, p)$.
- $d(p, q) \leq d(p, r) + d(r, q)$ for any $r \in X$.

For example, the Euclidean space $\mathbb{R}^d$ is a metric space which defines $d(p, q) = \|p - q\|$ where $\|\cdot\|$ is the Euclidean norm.

Definition 1.9: Let A, B be two sets, $A \subset B$ means A is the subset of B, i.e., either A is a proper subset of B or $A = B$. Similarly, $A \supset B$ means B is the subset of A. $A = B$ if and only if $A \subset B$ and $B \subset A$ [1].

---

[1] It is common to use the notation $\subseteq$ for subset and $\subsetneq$ for proper subset. Most theorems in real analysis only describe the subset relationships, while the proper subset relationships are less discussed unless specified. This makes $\subseteq$ more often used, and some textbooks [1][11][12][17] use $\subset$ instead of $\subseteq$ for simplicity.

Definition 1.10: Let X be a metric space, all points and sets mentioned below are the elements and subsets of X.

- A neighborhood of a point p is the set $N_r(p) = \{q \mid d(p,q) < r\}$, where r is called the radius of $N_r(p)$.
- A point p is the limit point of the set E if for any $r > 0$, there exists q such that $q \in N_r(p) \cap E$ and $q \neq p$.
- If $p \in E$ and p is not a limit point of E, then p is called an isolated point of E.
- E is closed if every limit point of E is a point of E.
- A point p is an interior point of E if there is $r > 0$ such that $N_r(p) \subset E$.
- E is open if every point of E is an interior point of E.
- The complement of E, denoted as E^c, is the set of all points $p \in X$ such that $p \neq E$.
- E is perfect if E is closed and every point of E is a limit point of E, i.e., there is no isolated point in E.
- E is bounded if there is a real number M and a point $q \in X$ such that $d(p,q) < M$ for all $p \in E$.
- E is dense in X if every point in X is a point in E, or a limit point of E, or both. In other words, for any $p \in X$, we have either $p \in E$ or $p \notin E$. If $p \notin E$, then for any $r > 0$, there exists a $q \in E$ such that $d(p,q) < r$.

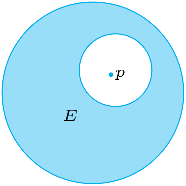
Theorem 1.11: If p is a limit point of a set E, then every neighborhood of p contains infinitely many points of E.

Proof: Let p be a limit point of E. Suppose there exists a neighborhood N of p such that $N \cap E$ contains finite many number of points $q_1, q_2, ..., q_n$. Let $r = \min_{1 \leq i \leq n}\{d(p, q_i)\}$, then for any $\epsilon < r$, $N_\epsilon(p) \cap E = \varnothing$, implying that p is not a limit point of E, which is contradict. $\square$

Theorem 1.12: A set E is open if and only if its complement is closed.

Proof: $\Longrightarrow$: Let x be a limit point of E^c, then any neighborhood of x contains a point of E^c. Since E is open, x is not a interior point of E, we must have $x \in E^c$. This holds for any x, which implies E^c is closed.
$\Longleftarrow$: Since E^c is closed, for any $x \in E$, $x \notin E^c$, thus x is not a limit point of E^c. This implies that there exists a neighborhood N of x such that $N \cap E^c = \varnothing$, hence $N \subset E$, which means x is an interior point of E. Thus E is open. $\square$

Examples:



- As shown in the right figure, E is a closed set and p is a isolated point. Then $E \cup \{p\}$ is closed but not perfect.
- Let $E \in X$ be a finite point set, then E is closed. This is because, by the definition, E is closed if for any $x \notin E$, x is not a limit point of E. By Theorem 1.11, the neighborhood of a limit point of E must contain infinite number of points. Since E is finite, it has no limit points, which holds for any $x \notin E$.
- The whole Euclidean space $\mathbb{R}^d$ is both open and closed, because every point $p \in \mathbb{R}^d$ is both a limit point and a interior point of $\mathbb{R}^d$. The empty set $\varnothing$ is both open and closed since it is the complement of $\mathbb{R}^d$. [2]

[2] https://en.wikipedia.org/wiki/Clopen_set

Definition 1.13: Let E be a set of the metric space X, and $\{G_\alpha\}$ be a (finite, countable, or uncountable) collection of open sets. If $E \subset \bigcup_\alpha G_\alpha$, then $\{G_\alpha\}$ is an open cover of E. E is compact if every open cover of E contains a finite subcover, i.e., for any $\{G_\alpha\}$, there exist finite number of indices $\alpha_1, \alpha_2, ...\alpha_n \in \alpha$ such that $E \subset \bigcup_{i=1}^n G_{\alpha_i}$.

Theorem 1.14 (Heine-Borel): A subset E in the Euclidean space $\mathbb{R}^d$ is compact if and only if it is closed and bounded.

Proof: We list three statements:
(1) E is closed and bounded.
(2) E is compact.
(3) Every infinite subset of E has a limit point in E.
And prove $(1) \iff (2)$ by showing $(1) \implies (2)$, $(2) \implies (3)$ and $(3) \implies (1)$.

$(1) \implies (2)$:
We call the set $I = \{x = (x_1, x_2, ..., x_d) | a_j \leq x_j \leq b_j, j = 1, 2, ..., d\}$ as a d-cell.
Then we have the following lemmas:

Lemma 1.15: Every d-cell is compact.

Proof: Put $\delta = \|a - b\|$ where $a = [a_1, ..., a_d]^T$ and $b = [b_1, ..., b_d]^T$, then for any $x, y \in I$, $\|x - y\| \leq \delta$.
We prove by contradiction. Suppose I is not compact, there exists an open cover $\{G_\alpha\}$ of I which contain no finite subcovers of I.
Take $c_j = (a_j + b_j)/2$. The intervals $[a_j, c_j], [c_j, b_j]$ partition I into 2^d d-cells whose union is I. Since I cannot be covered by finite open sets, there exists a partitioned d-cell, denoted as I_1, cannot be covered by open sets. Continue the partitioning process, we get
(a) $I \supset I_1 \supset I_2 \supset I_3 ....$
(b) I_n is not covered by any finite open sets $\{G_\alpha\}$.
(c) If $x, y \in I_n$, then $\|x - y\| \leq 2^{-n}\delta$.
For any n, I_n is not empty. let $x^* \in I_n$, we take $r > 2^{-n}\delta$ such that $I_n \subset N_r(x^*)$. This implies I_n can be covered by a open set for any n, which is contract to (b). □

Compact Sets

Lemma 1.16: Closed subsets of compact sets are compact.

Proof: Suppose $F \subset K \subset X$ [3], F is closed, K is compact. Let $\{V_\alpha\}$ be a open cover of F. By Theorem 1.12, F^c is open. Thus $K \subset X = \{V_\alpha\} \cup F^c$. Since K is compact, there exists finite subcovers in $\{V_\alpha\} \cup F^c$ of K, i.e., $K \subset (\bigcup_{i=1}^n V_{\alpha_i}) \cup F^c$ where n is finite. For any $x \in F$, we have $x \in K$ and $x \notin F^c$, thus $F \subset (\bigcup_{i=1}^n V_{\alpha_i}) - F^c \subset (\bigcup_{i=1}^n V_{\alpha_i})$, which shows F is compact. $\square$

Since E is bounded, there must exists a d-cell such that $E \subset I$. By Lemma 1.15, I is compact. Since E is a closed subset of a compact set I, by Lemma 1.16, E is compact.

$(2) \Longrightarrow (3)$:
This can be shown by the following Lemma.
Lemma 1.17: If F is an infinite subset of a compact set E, then F has a limit point in E.

Proof: We prove the contraposition: If E is compact and F has no limit point in E, then F is not an infinite subset of E.

[3] This lemma holds for any general metric space X, including $\mathbb{R}^d$.

Compact Sets

If F has no limit point in E, then for any $p \in E$, there exists $\epsilon > 0$ such that for $p \notin F$, $N_\epsilon(p) \cap F = \varnothing$ and for $p \in F$, $N_\epsilon(p) \cap F = \{p\}$. Thus $\{N_\epsilon(p) | p \in F\}$ is an open cover of F.
Since for any $p, q \in F$, $N_\epsilon(p) \cap N_\epsilon(q) = \varnothing$, if F is infinite, then the number of open sets in the cover $\{N_\epsilon(p) | p \in F\}$ must be infinite, thus F is not compact.
F is closed because for any $p \notin F$, p is not a limit point of F. Since $F \subset E$ and E is compact, by Lemma 1.16, F is compact, which is contradict.
Thus F cannot be infinite. $\square$

$(3) \implies (1)$:
We prove the contraposition: If E is not closed or not bounded, then there exists an infinite subset of E that has no limit point in E.
If E is not bounded, let $\|x_i\| > i$ for $i = 1, 2, 3...$ and take $S = \{x_1, x_2, x_3...\}$, then S is an infinite subset of E. For $p \in \mathbb{R}^d$ and $\epsilon > 0$, any $q \in N_\epsilon(p)$ satisfies $\|q - p\| < \epsilon \implies \|q\| < \|p\| + \epsilon$. Since the number of x_is that satisfy $\|x_i\| < \|p\| + \epsilon$ is finite, i.e., $S \cap \{x \in \mathbb{R}^d \mid \|x\| < \|p\| + \epsilon\}$ is finite, and $N_\epsilon(p) \subset \{x \in \mathbb{R}^d \mid \|x\| < \|p\| + \epsilon\}$, then $S \cap N_\epsilon(p)$ must be finite. If p is a limit point of S, by Theorem 1.11, $S \cap N_\epsilon(y)$ should be infinite for any ϵ. Thus any $p \in \mathbb{R}^d$ is not a limit point of S, which implies S has no limit point in E.

If E is not closed, then there exists a point $x \in \mathbb{R}^d$ which is a limit point of E but not a point in E. Let $x_0 \notin E$, $x_1, x_2, x_3... \in E$ where $\|x_n - x_0\| \leq \frac{1}{n}$, then $S = \{x_1, x_2, x_3...\}$ is an infinite subset of E. S has x_0 as a limit point, and S has no other limit point in $\mathbb{R}^d$. This is because, for any $y \in \mathbb{R}^d$ and $y \neq x_0$, let $\epsilon = \frac{1}{2}\|y - x_0\|$, then

$$\begin{aligned} \|x_n - y\| < \epsilon &\Longrightarrow \|x_0 - y\| - \|x_n - x_0\| \leq \frac{1}{2}\|y - x_0\| \\ &\Longrightarrow \frac{1}{2}\|y - x_0\| \leq \|x_n - x_0\| \leq \frac{1}{n} \Longrightarrow n \leq \frac{1}{\epsilon} \end{aligned}$$

This implies the set $S \cap N_\epsilon(y)$ is finite because n is finite. If y is a limit point of S, by Theorem 1.11, $S \cap N_\epsilon(y)$ should be infinite for any ϵ.
Therefore, any $y \in \mathbb{R}^d/\{x_0\}$ is not a limit point of S. Since $x_0 \notin E$, S has no limit point in E. $\square$

Continuity

Definition 1.18: Let A, B be two sets and $f : A \to B$ be a function. If $E \subset A$, denote $f(E) = \{f(x) : x \in E\}$, and we call $f(E)$ the image of E under f. If $E \subset B$, denote $f^{-1}(E) = \{x \in A : f(x) \in E\}$, and we call $f^{-1}(E)$ the inverse image of E under f.

Note that there can exist some subset M in B such that there does not exist element x in A satisfying $f(x) \in M$. Let $E \subset B$ and $M \subset E$ such that $f^{-1}(M) = \varnothing$, then $f^{-1}(E) = f^{-1}(E - M)$.

The inverse image f^{-1} satisfies the following properties [4]:

- For any $E, F \in B$, $f^{-1}(E^c) = f^{-1}(E)^c$, $f^{-1}(E - F) = f^{-1}(E) - f^{-1}(F)$.
- For any (possibly uncountable) family of subsets $\{E_\alpha\}$ of B, $f^{-1}(\bigcup_\alpha E_\alpha) = \bigcup_\alpha f^{-1}(E_\alpha)$, $f^{-1}(\bigcap_\alpha E_\alpha) = \bigcap_\alpha f^{-1}(E_\alpha)$.

Definition 1.19: Let X and Y be metric spaces, $E \subset X$, $p \in E$, $f : E \to Y$. f is said to be continuous at p if for any $\epsilon > 0$ there exists $\delta > 0$ such that $d_Y(f(p), f(x)) < \epsilon$ for any $x \in E$ satisfying $d_X(p, x) < \delta$.

[4] https://planetmath.org/theinverseimagecommuteswithsetoperations

Continuity

Theorem 1.20: A mapping from a metric space X to a metric space Y is continuous if and only if $f^{-1}(V)$ is open in X for every open set V in Y.

Proof: $\Longrightarrow$: Suppose f is a continuous function and V is an open set in Y. Let $p \in X$ be a point and $f(p) \in Y$ be its image. V is open implies that any point $f(p) \in V$ is an interior point of V. Thus, for any $f(p)$, there exists $\epsilon > 0$ such that $N_\epsilon(f(p)) \subset V$. f being continuous implies that for any $\epsilon > 0$, there exists $\delta > 0$ such that for any $q \in N_\delta(p) \subset X$ we have $f(q) \in N_\epsilon(f(p)) \subset V$. Thus for any $q \in N_\delta(p)$, we have $q \in f^{-1}(V)$, which implies $N_\delta(p) \subset f^{-1}(V)$. Since for any $p \in f^{-1}(V)$, we can always find $\delta > 0$ such that $N_\delta(p) \subset f^{-1}(V)$, hence p is always an interior point, which means $f^{-1}(V)$ is open.
$\Longleftarrow$: Suppose $f^{-1}(V) \in X$ is open for every open $V \in Y$. We fix $p \in X$ and $\epsilon > 0$, and take $V = N_\epsilon(f(p))$, then V is open, and by our assumption, $f^{-1}(V)$ is open. Since $f^{-1}(V)$ is open, p is an interior point of $f^{-1}(V)$, thus there exists $\delta > 0$ such that $N_\delta(p) \subset f^{-1}(V)$. Since any $q \in f^{-1}(V)$ gives $f(q) \in N_\epsilon(f(p))$, every q in $N_\delta(p)$ must give $f(q) \in N_\epsilon(f(p))$, which implies f is continuous.

$\square$

Corollary 1.21: A mapping from a metric space X to a metric space Y is continuous if and only if $f^{-1}(V)$ is closed in X for every closed set V in Y.

Proof: V is closed if and only if V^c is open (Theorem 1.12), thus $f^{-1}(V)$ is closed if and only if $f^{-1}(V^c) = f^{-1}(V)^c$ is open. □

Theorem 1.22: Suppose f is a continuous mapping of X to Y where X is compact, then $f(X)$ is compact.

Proof: Let $\{V_\alpha\}$ be an open cover of $f(X)$. Since f is continuous, by Theorem 1.20, each $f^{-1}(V_i)$ is open for $i \in \alpha$.
Since X is compact, there are finite indices $\alpha_1, \alpha_2, ..., \alpha_n$ such that $X \subset \bigcup_{i=1}^n f^{-1}(V_{\alpha_i})$. Using the fact $f(f^{-1}(E)) \subset E$ [5] and the property $f(\bigcup_{i=1}^n A_i) \subset \bigcup_{i=1}^n f(A_i)$ [6], we have

$$f(X) \subset f\left(\bigcup_{i=1}^n f^{-1}(V_{\alpha_i})\right) = \bigcup_{i=1}^n f\left(f^{-1}(V_{\alpha_i})\right) \subset \bigcup_{i=1}^n V_{\alpha_i}$$

which implies $f(X)$ is compact. □

[5] If $f: X \to E$ is non-surjective, then $f(f^{-1}(E)) \subset E$.

[6] https://math.stackexchange.com/questions/75247/functional-equality-of-unions-intersections

Definition 1.23: Let f be a mapping from a metric space X to a metric space Y. We say f is uniformly continuous on X if for every $\epsilon > 0$ there exists $\delta > 0$ such that $d_Y(f(p), f(q)) < \epsilon$ for any p, q satisfying $d_X(p, q) < \delta$.

The differences between continuity and uniform continuity include: (1) uniform continuity is a property of a function on a set, while continuity can be defined on either a single point or a set. (2) If f is continuous, for any $\epsilon > 0$ and for each $p \in X$, we can find $\delta > 0$ satisfying the property of Definition 1.19. However, the δ may differ for different p. If f is uniformly continuous on X, then for any $\epsilon > 0$, there exists one $\delta > 0$ that satisfies the property of Definition 1.19 for all $p \in X$.

Theorem 1.24: Let f be a continuous mapping from a compact metric space X to a metric space Y. Then f is uniformly continuous on X.

Proof: Since f is continuous on any $p \in X$, for any $\epsilon > 0$, we can assign each point $p \in X$ a positive number $\phi(p)$ such that $d_Y(f(p), f(q)) < \frac{\epsilon}{2}$ for any $q \in X$ satisfying $d_x(p, q) < \phi(p)$. Note that $\phi(p)$ may differ for different p.
Since X is compact, there exists an open cover $\{N_{\phi(r_i)/2}(r_i)\}_{i=1}^n$ of x where n is finite and $r_1, r_2, ..., r_n$ are points in X.

Continuity

We take $\delta = \frac{1}{2}\min\{\phi(r_1), \phi(r_2), ..., \phi(r_n)\}$, then δ is a positive number. For any $p \in X$, p must be in at least one open ball $N_{\phi(r_i)/2}(r_i)$. Suppose $p \in N_{\phi(r_m)/2}(r_m)$ where $1 \leq m \leq n$, then

$$d_X(p, r_m) < \frac{\phi(r_m)}{2} < \phi(r_m)$$

$$d_X(q, r_m) < d_X(p, q) + d_X(p, r_m) < \delta + \frac{\phi(r_m)}{2} < \phi(r_m)$$

This implies that for any q satisfies $d_X(p, q) < \delta$, $q \in N_{\phi(r_m)}(r_m)$. Of course, $p \in N_{\phi(r_m)}(r_m)$. Since any point $r \in N_{\phi(r_m)}(r_m)$ gives $d_Y(f(r), f(r_m)) < \epsilon/2$, we have

$$d_Y(f(p), f(q)) \leq d_Y(f(p), r_m) + d_Y(f(q), r_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Thus we find a positive δ that any p, q satisfying $d_X(p, q) < \delta$ give $d_Y(f(p), f(q)) < \epsilon$.

$\square$

Closure

Definition 1.25: Let X be a metric space. Let $E \subset X$ and E' be the set of all limit points of E in X, then $\overline{E} = E \cup E'$ is called the closure of E.

Theorem 1.26: E is dense in $\overline{E}$.

Proof: $x \in \overline{E}$ implies $x \in E$ or $x \in E'$. That is, any point x in E is a point in E, or a limit point of E, or both. $\square$

Theorem 1.27: $\overline{E}$ is closed.

Proof: For any $p \notin \overline{E}$, we have $p \notin E$ and $p \notin E'$, which means p is neither a point in E nor a limit point of E. Thus there exists a neighborhood $N_\epsilon(p)$ such that $N_\epsilon(p) \cap \overline{E} = \varnothing \implies N_\epsilon(p) \subset \overline{E}^c$. This implies that any $p \in \overline{E}^c$ is an interior point of $\overline{E}^c$, which means $\overline{E}^c$ is open. By Theorem 1.12, $\overline{E}$ is closed. $\square$

Corollary 1.28: Any limit point of $\overline{E}$ is in $\overline{E}$.

Proof: If there exists a limit point of $\overline{E}$ not in $\overline{E}$, then $\overline{E}$ is not closed, which is contradict. $\square$

Topological Space

We would like to find a general mathematical space that allows to define limit and continuity. In a metric space, continuity is defined by the open sets (Definition 1.19), where the open sets are defined under the axioms of metric (Definition 1.8). However, limit and continuity are all about the open sets, where the metrics are usually irrelevant [7]. Topological space is a general space defined by the axioms of open sets and without a metric involved.

Definition 1.29: Let τ be a family of subsets of a set X. τ is said to be a topology in X if it satisfies the three axioms:

- $\varnothing \in \tau$ and $X \in \tau$.
- If $\{V_\alpha\}$ is an arbitrary family of the members of τ (finite, countable or uncountable), then $\bigcup_\alpha V_\alpha \in \tau$.
- If $V_i \in \tau$ for $i = 1, 2, ..., n$, then $\bigcap_{i=1}^{n} V_i \in \tau$.

If τ is a topology in X, then X is called a topological space, denoted as (X, τ). Any member in τ is called an open set in X. A subset E of X is said to be closed if $E^c \in \tau$.
If X and Y are topological spaces and if f is a mapping from X to Y, then f is said to be continuous if for every open $V \subset Y$, $f^{-1}(V)$ is open (A general case of Theorem 1.20).

[7] https://math.stackexchange.com/questions/3521499/

Topological Space

It is easy to verify that open sets of a metric space satisfies the axioms of Definition 1.29. Uncountable unions of open sets is still open, but countable intersections of open sets is not necessary open. For example [8], in $\mathbb{R}$, let $\{(-\frac{1}{n}, \frac{1}{n})\}_{n=1}^{\infty}$ be a family of open sets, then $\bigcap_{i=1}^{\infty}(-\frac{1}{n}, \frac{1}{n}) = \{0\}$, where $\{0\}$ is closed.

The continuity can be defined only with open sets. For example [9], let $f : \mathbb{R} \to \mathbb{R}$, under $\epsilon - \delta$ definition, we say f is left continuous on $a \in \mathbb{R}$ if for any $\epsilon > 0$, there exists $\delta > 0$ such that $a - x < \delta \implies |f(a) - f(x)| < \epsilon$. Now, let τ_1 be generated from $\{(-\infty, a - \delta)\}_{\delta>0}$ and τ_2 be generated from $\{(f(a) - \epsilon, f(a) + \epsilon)\}_{\epsilon>0}$. It is easy to verify that both τ_1 and τ_2 are topologies. We can rewrite the definition of continuity as follows: Let $(\mathbb{R}, \tau_1), (\mathbb{R}, \tau_2)$ be two topological spaces and $f : (\mathbb{R}, \tau_1) \to (\mathbb{R}, \tau_2)$, f is continuous if for any $Y \in \tau_2$, there exists $X \in \tau_1$ such that $Y \subset f(X)$ (or $f^{-1}(Y) \subset X$).

The open sets can be defined without a metric. For example, a topology of $[0, 1]$ can be generated from $\{[0, a), (1 - a, 1], (a, b), \varnothing, [0, 1]\}_{0<a<b<1}$. Here $[0, a)$ and $(1 - a, 1]$ are considered open sets in the topological space, but their openness is not defined though a metric (obviously they are not open in Euclidean space).

[8] https://math.stackexchange.com/questions/3763170/

[9] https://math.stackexchange.com/questions/451142/

Definition 1.30: Let X be a topological space and $A \subset X$ be a set. Let A^o be the set of all interior points of A such that A^o is the largest open set contained in A. We call A^o the interior of A. Let $\partial A = \overline{A} - A^o$. We call ∂A the boundary of A.

∂A is closed, since $\partial A = \overline{A} \cap (A^o)^c$ where both $\overline{A}$ and $(A^o)^c$ are closed.

If A is an open set, then $A = A^o$, and $\partial A \cap A = \varnothing$.

Note that we may not have $\partial A = \overline{A^o} - A^o$, i.e., $\overline{A} = \overline{A^o}$ does not always hold. For example [10], let $A = \mathbb{Q} \cap (0, 1)$, then $\overline{A} = \mathbb{R}$ (Lemma C.5), $A^o = (0, 1)$, $\overline{A^o} = [0, 1]$.

Definition 1.31: If $x \in X$ (or $E \subset X$), a neighborhood of x (or E) is a set $A \subset X$ such that $x \in A^o$ (or $E \subset A^o$). Thus a set A is open if and only if it is a neighborhood of itself.

[10] https://math.stackexchange.com/questions/3760906/

Lemma 1.32: $(A^o)^c = \overline{A^c}$, and $(A^c)^o = (\overline{A})^c$.

Proof: We first prove $(A^o)^c = \overline{A^c}$. (1) Since $A^o \subset A$, $(A^o)^c \supset A^c$. Since $(A^o)^c$ is closed, any limit point of $(A^o)^c$ is in $(A^o)^c$, thus any limit point of A^c is in $(A^o)^c$. Hence, $(A^o)^c \supset \overline{A^c}$. (2) Since $\overline{A^c}$ is closed, $(\overline{A^c})^c$ is open. Thus for any $x \in (\overline{A^c})^c$, there exists a neighborhood $N_x \subset (\overline{A^c})^c$. This means $(N_x)^c \supset \overline{A^c} \supset A^c \implies N_x \subset A$, implying that x is an interior point of A. Since A^o is the set of all interior points of A, we have $(\overline{A^c})^c \subset A^o \iff \overline{A^c} \supset (A^o)^c$. Therefore, $\overline{A^c} = (A^o)^c$, and by replacing A with A^c we get $(A^c)^o = (\overline{A})^c$. □

For a topological space (X, τ), the set τ can be large or small. We would like to distinguish distinct points or disjoint subsets of X using their open neighborhoods. Remember that open neighborhoods and are supposed to be contained in τ. However, if τ is too small and does not contain enough open neighborhoods, some points or subsets would be indistinguishable. A more distinguishable topological space is easier for applying a metric on it [11]. We use Separation Axioms [12] to describe how distinguishable a topological space is.

[11] https://mathstrek.blog/2013/03/15/topology-separation-axioms/
[12] https://en.wikipedia.org/wiki/Separation_axiom

Topological Space

Definition 1.33: Let (X, τ) be a topological space.

- If for any points $x, y \in X$ with $x \neq y$, there exist disjoint open sets $U, V \in \tau$ such that $x \in U$ and $y \in V$, then we say (X, τ) is a Hausdorff space.
- If for any disjoint subsets $A, B \in X$, there exist disjoint open sets $U, V \in \tau$ such that $A \in U$ and $B \in V$, then we say (X, τ) is a normal space.

Definition 1.34: Let (X, τ) be a topological space. A subset K in X is said to be compact if for any (finite, countable, uncountable) open cover $\{V_\alpha\} \subset \tau$ satisfying $K \subset \bigcup_\alpha V_\alpha$, there exists a finite sub-cover $\{V_{\alpha_i}\}_{i=1}^n \subset \{V_\alpha\}$ such that $K \subset \{V_{\alpha_i}\}_{i=1}^n$.

(X, τ) is said to be locally compact if every point in X has a compact neighborhood.

Theorem 1.35: All metric spaces are Hausdorff.

Proof: Let X be a metric space, then $x, y \in X$ with $x \neq y$ implies $d(x, y) > 0$. Let $\epsilon = \frac{d(x,y)}{2}$, then $N_\epsilon(x) \cap N_\epsilon(y) = \varnothing$. This can be shown by contradiction: If $M = N_\epsilon(x) \cap N_\epsilon(y) \neq \varnothing$, then there exists $z \in M$ such that $d(x, z) < \epsilon$ and $d(x, y) < \epsilon$, thus $d(x, y) \leq d(x, z) + d(y, z) < 2\epsilon = d(x, y)$, which is contradict. Since for we can always find disjoint neighborhoods $N_\epsilon(x), N_\epsilon(y)$ for disjoint points x, y in a metric space, the metric space is Hausdorff. $\square$

Theorem 1.36: The Euclidean space $\mathbb{R}^d$ is a locally compact Hausdorff space.

Proof: For any $x \in \mathbb{R}^d$, for any $\epsilon > 0$, let $M_\epsilon(x) = \{p \in \mathbb{R}^d : \|p - x\|_2 \leq \epsilon\}$, then $M_\epsilon(x)$ is a closed and bounded neighborhood of x. By Heine-Borel Theorem (Theorem 1.14), $M_\epsilon(x)$ is compact. $\square$

Lemma 1.37: If F is a compact subset of a Hausdorff space X and $x \notin F$, then there exist disjoint open sets $U, V \in X$ such that $F \subset U$ and $x \subset V$.

Proof: Since $x \notin F$, for any $y \in F$, $y \neq x$. Since X is Hausdorff, there exist disjoint open sets U_y, V_y such that $y \in U_y$ and $x \in V_y$. Since for each $y \in F$ we can draw an open set V_y, $F \subset \bigcup_{y \in F} V_y$, thus $\{V_y\}_{y \in F}$ is an open cover of F. Since F is compact, there exists a finite subcover $\{V_{y_i}\}_{i=1}^n \subset \{V_y\}_{y \in F}$ such that $F \subset \bigcup_{i=1}^n V_{y_i}$. Since for each V_{y_i}, there exists U_{y_i} such that $x \in U_{y_i}$ and $V_{y_i} \cap U_{y_i} = \varnothing$. Let $V = \bigcup_{i=1}^n V_{y_i}$ and $U = \bigcap_{i=1}^n U_{y_i}$, then $U \cap V = \varnothing$. $\square$

Corollary 1.38: Any compact set in a Hausdorff space is closed.

Proof: By Lemma 1.37, any $x \in F^c$ has an open neighborhood V such that $V \subset F^c$. Thus F^c is open, which implies F is closed. $\square$

Topological Space

Lemma 1.39: Let X be a Hausdorff space and E, F be two disjoint compact sets in X, then there exist open sets U, V in X such that $E \subset U$ and $F \subset V$.

Proof: Let $x \in E$ such that $x \notin F$. By Lemma 1.37, there exists disjoint open sets $U_x, V_x \in X$ such that $x \in U_x$ and $F \subset V_x$. Since E is compact, there exists a finite subcover $\{U_{x_i}\}_{i=1}^n$ of E, and for each U_{x_i} there exists a open set $V_{x_i} \in X$ with $F \subset V_{x_i}$ and $U_{x_i} \cap V_{x_i} = \varnothing$. Take $U = \bigcup_{i=1}^n U_{x_i}$ and $V = \bigcap_{i=1}^n V_{x_i}$. $\square$

Theorem 1.40: If X is a locally compact Hausdorff space, $U \subset X$ is open, and $x \in U$, then there exists a compact neighborhood N of x such that $N \subset U$.

Proof: Assume $\overline{U}$ is compact; otherwise, replace U with $U \cap F^o$ where F is a compact neighborhood of x. Note that $U \cap F^o$ is open and $U \cap F^o \subset F^o \subset F$, this means $\overline{U \cap F^o}$ is a closed subset of F, and by Lemma 1.16 (which holds for topological space as its proof does not involve metric), $\overline{U \cap F^o}$ is compact.

Since $x \in U \subset \overline{U}$ and $\partial U = \overline{U} - U$, we know that $x \notin \partial U$. Since ∂U is a closed subset of $\overline{U}$, by Lemma 1.16, ∂U is compact. By Lemma 1.37, there exist disjoint open sets $V, W \in X$ such that $x \in V$ and $\partial U \subset W$. Take $V_1 = U \cap V$ so that V_1 is an open set in U, containing x, and disjoint with W.

Since $V_1 \subset W^c$, we have $V_1 \subset U \cap W^c \subset \overline{U} \cap W^c$. Since $\overline{U} \cap W^c$ is closed, any limit point of $\overline{U} \cap W^c$ is in $\overline{U} \cap W^c$. Since any limit point of V_1 is also a limit point of $\overline{U} \cap W^c$,

$$\overline{V_1} \subset \overline{U} \cap W^c = (U \cup \partial U) \cap W^c = (U \cap W^c) \cup (\partial U \cap W^c) = U \cap W^c \subset U$$

Also, since $\overline{V_1} \subset \overline{U}$, $\overline{V_1}$ being closed implies that $\overline{V_1}$ is compact. Hence we can take $N = \overline{V_1}$. □

Corollary 1.41: Let X be a locally compact Hausdorff space, $K \subset U \subset X$ where K is compact and U is open, then there exists an open set V such that $K \subset V \subset \overline{V} \subset U$, and $\overline{V}$ is compact.

Proof: By Theorem 1.40, for each $x \in K$, we can choose a compact neighborhood N_x such that $N_x \subset U$. Since $\{N_x^o\}_{x \in X}$ is an open cover of K, there exists a finite subcover $\{N_{x_i}^o\}_{i=1}^n$ of K. Let $V = \bigcup_{i=1}^n N_{x_i}^o$, then $\overline{V} = \bigcup_{i=1}^n N_{x_i}$ (Due to the property $\overline{A \cup B} = \overline{A} \cup \overline{B}$ [13]), and we get $K \subset V \subset \overline{V} \subset U$. $\overline{V}$ is compact because it is a finite union of compact sets. □

[13] https://math.stackexchange.com/questions/1986224

Contents

Weierstrass Theorem

This section shows the proofs of Weierstrass theorem and Stone-Weierstrass theorem. The proof of Weierstrass theorem is referred from [8] and the proof of Stone-Weierstrass theorem is referred from [1]. [14]

Definition 2.1 (Uniform Convergence): Suppose E is a set and $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of real-valued function on it. We say $\{f_n\}_{n\in\mathbb{N}}$ converges to $f : E \to \mathbb{R}$ uniformly if for any $\epsilon > 0$, there exists a natural number N such that for any $n > N$,

$$|f_n(x) - f(x)| < \epsilon \quad \text{for any } x \in E$$

Definition 2.2: Let $f : [0, 1] \to \mathbb{R}$ be a continuous function. For $n \in \mathbb{N}^+$, the nth Bernstein Polynomial of f is defined as

$$B_n(x, f) = \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}x^k(1-x)^{n-k}$$

[14] For Weierstrass theorem, the proof in [8] is much easier to understand. See https://math.stackexchange.com/questions/2254122/an-easy-proof-of-the-stone-weierstrass-theorem

Weierstrass Theorem

Theorem 2.3 (Weierstrass): Let $f : [a, b] \to \mathbb{R}$ be a continuous function, then there is a sequence of polynomials $\{p_n(x)\}$ that converges uniformly to $f(x)$ on $[a, b]$.

Proof: We first consider the continuous function $f : [0, 1] \to \mathbb{R}$. By Theorem 1.14, since $[0, 1]$ is closed and bounded, it is compact. Since f is continuous on $[0, 1]$, by Theorem 1.24, f is uniformly continuous on $[0, 1]$. Thus for any $\epsilon > 0$ there exist $\delta > 0$ such that for any $x, y \in [0, 1]$,

$$|x - y| < \delta \implies |f(x) - f(y)| < \frac{\epsilon}{2} \tag{2}$$

By Theorem 1.22, the range of f is compact. Since the range of f is a compact set on $\mathbb{R}$, by Theorem 1.14, the range of f is bounded. Suppose $\|f\|_\infty < M$, then for any $x, y \in [0, 1]$,

$$|f(x) - f(y)| < 2M \tag{3}$$

Combining Eq (2) and Eq (3) we get that, for any $x, y \in [0, 1]$,

$$|f(x) - f(y)| < 2M \left(\frac{x - y}{\delta}\right)^2 + \frac{\epsilon}{2} \tag{4}$$

Weierstrass Theorem

Now we show that the Bernstein polynomials $B_n(x,f)$ converges to f uniformly as $n \to \infty$. For any $\xi \in [0,1]$ we have

$$
\begin{aligned}
&|B_n(\xi,f) - f(\xi)| = \left|\sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}\xi^k(1-\xi)^{n-k} - f(\xi)\right| \\
&= \left|\sum_{k=0}^{n}\left[f(\frac{k}{n}) - f(\xi)\right]\binom{n}{k}\xi^k(1-\xi)^{n-k}\right| \le \sum_{k=0}^{n}\left|f(\frac{k}{n}) - f(\xi)\right|\binom{n}{k}\xi^k(1-\xi)^{n-k} \\
&\le \sum_{k=0}^{n}\left[2M\left(\frac{\frac{k}{n}-\xi}{\delta}\right)^2 + \frac{\epsilon}{2}\right]\binom{n}{k}\xi^k(1-\xi)^{n-k} \\
&= \frac{2M}{\delta^2}\sum_{k=0}^{n}\left[\frac{k^2}{n^2} - 2\xi\frac{k}{n} + \xi^2\right]\binom{n}{k}\xi^k(1-\xi)^{n-k} + \frac{\epsilon}{2}
\end{aligned}
$$

The second equality above use the fact that

$$\sum_{k=0}^{n}\binom{n}{k}\xi^k(1-\xi)^{n-k} = (\xi + (1-\xi))^n = 1$$

The first inequality above use the triangle inequality and $\binom{n}{k}\xi^k(1-\xi)^{n-k} > 0$. The second inequality plugs in Eq (4).

Weierstrass Theorem

Since

$$\begin{aligned}
\sum_{k=0}^{n} \frac{k^2}{n^2}\binom{n}{k}\xi^k(1-\xi)^{n-k} &= \sum_{k=0}^{n}\left[\frac{n-1}{n}\binom{n-2}{k-2}+\frac{1}{n}\binom{n-1}{k-1}\right]\xi^k(1-\xi)^{n-k} \\
&= \frac{n-1}{n}\xi^2+\frac{1}{n}\xi \\
\sum_{k=0}^{n} \xi\frac{k}{n}\binom{n}{k}\xi^k(1-\xi)^{n-k} &= \xi^2\sum_{k=0}^{n}\binom{n-1}{k-1}\xi^{k-1}(1-\xi)^{n-k} = \xi^2 \\
\sum_{k=0}^{n} \xi^2\binom{n}{k}\xi^k(1-\xi)^{n-k} &= \xi^2
\end{aligned}$$

we have

$$\begin{aligned}
|B_n(\xi,f)-f(\xi)| &\le \frac{2M}{\delta^2}\left[\frac{n-1}{n}\xi^2+\frac{1}{n}\xi-2\xi^2+\xi^2\right]+\frac{\epsilon}{2} \\
&= \frac{2M}{\delta^2 n}(\xi-\xi^2)+\frac{\epsilon}{2} \le \frac{M}{2\delta^2 n}+\frac{\epsilon}{2}
\end{aligned}$$

The last inequality uses the fact that $\xi-\xi^2 \le \frac{1}{4}$ for $\xi \in [0,1]$.

Therefore, let $N = \frac{M}{\delta^2}$, then for any $n > N$,

$$|B_n(\xi, f) - f(\xi)| < \epsilon$$

The above inequality holds for any $\xi \in [0,1]$, thus $B_n(x, f)$ converges to f uniformly on $[0,1]$.
For the continuous function $f : [a,b] \to \mathbb{R}$, we use the Bernstein polynomial $B_n(\frac{x-a}{b-a}, f)$ where $x \in [a,b]$ to approximate f. The convergence can proved in a similar way.

$\square$

In Appendix 2, I will show an alternative proof of Theorem 2.3.

Stone-Weierstrass Theorem

Definition 2.4: A family $\mathcal{A}$ of real functions defined on a set E is said to be an algebra if (1) $f+g \in \mathcal{A}$; (2) $fg \in \mathcal{A}$; (3) $cf \in \mathcal{A}$ for any $f \in \mathcal{A}$, $g \in \mathcal{A}$ and $c \in \mathbb{R}$. That is, the functions in $\mathcal{A}$ is closed under addition, multiplication and scalar multiplication.
$\mathcal{A}$ is said to be uniformly closed if $\mathcal{A}$ has the property that for any f that there exists a sequence $\{f_n\}$ in $\mathcal{A}$ such that $f_n \to f$ uniformly on E, we have $f \in \mathcal{A}$. Let $\mathcal{B}$ be the union of $\mathcal{A}$ and the set of all fs having uniformly convergence sequence in $\mathcal{A}$, then $\mathcal{B}$ is called the uniform closure of $\mathcal{A}$. [15]

Theorem 2.5: Let $\mathcal{B}$ be the uniform closure of an algebra $\mathcal{A}$ of bounded functions, then $\mathcal{B}$ is a uniformly closed algebra.

Proof: Since $\mathcal{B}$ is the uniform closure of $\mathcal{A}$, for any $f, g \in \mathcal{B}$, there exist sequences $\{f_n\}, \{g_n\}$ in $\mathcal{A}$ such that $f_n \to f, g_n \to g$ uniformly. $\mathcal{B}$ is an algebra if for any $f, g \in \mathcal{B}$ and $c \in \mathbb{R}$, we have $f+g, fg, cf \in \mathcal{B}$. We need to show there exist sequences in $\mathcal{A}$ that converge to $f+g, fg, cf$ respectively.
Since $\mathcal{A}$ is an algebra, we know that $\{f_n+g_n\}, \{f_ng_n\}, \{cf_n\} \in \mathcal{A}$. We show that
(1) $f_n + g_n \to f + g$

[15] An algebra is a set with arithmetic operations, and the set is closed under the arithmetic operations. A closed algebra means the set of the algebra is closed.

Stone-Weierstrass Theorem

This is because, $f_n \to f$ implies for any $\epsilon > 0$, there exist N_f such that for any $x \in E$ and any $n > N_f$, $|f_n(x) - f(x)| < \frac{\epsilon}{2}$, and $g_n \to g$ implies for any $\epsilon > 0$, there exist N_g such that for any $x \in E$ and any $n > N_g$, $|g_n(x) - g(x)| < \frac{\epsilon}{2}$. Let $N = \max\{N_f, N_g\}$, then for any $n > N$,

$$|f_n(x) + g_n(x) - (f(x) + g(x))| \leq |f_n(x) - f(x)| + |g_n(x) - g(x)| < \epsilon$$

which implies $f_n + g_n \to f + g$ uniformly.

(2) $f_n g_n \to fg$
Since any function in $\mathcal{A}$ is bounded, we suppose there exists an M such that $|f(x)| < M$ for any $f \in \mathcal{A}$ and $x \in E$. For any $\epsilon > 0$, there exist N_f, N_g such that any $n > N_f$ gives $|f_n(x) - f(x)| < \frac{\epsilon}{2M}$ and any $n > N_g$ gives $|g_n(x) - g(x)| < \frac{\epsilon}{2M}$. Let $N = \max\{N_f, N_g\}$, then for any $n > N$ and $x \in E$,

$$\begin{aligned}
&|f_n(x)g_n(x) - f(x)g(x)| = |f_n(x)g_n(x) - f_n(x)g(x) + f_n(x)g(x) - f(x)g(x)| \\
&\leq |f_n(x)||g_n(x) - g(x)| + |g(x)||f_n(x) - f(x)| \\
&\leq |f_n(x)||g_n(x) - g(x)| + \left(|g_n(x)| + \frac{\epsilon}{2M}\right)|f_n(x) - f(x)| \\
&< M \cdot \frac{\epsilon}{2M} + \left(M + \frac{\epsilon}{2M}\right)\frac{\epsilon}{2M} = \epsilon + \frac{\epsilon^2}{4M}
\end{aligned}$$

Let $\epsilon' = \epsilon + \frac{\epsilon^2}{4M}$. For any $\epsilon' > 0$, we can first solve ϵ then find N that satisfies for any $n > N$ and $x \in E$, $|f_n(x)g_n(x) - f(x)g(x)| < \epsilon'$. This means $f_n g_n \to fg$ uniformly.

(3) $cf_n \to cf$
For any $\epsilon > 0$ and $c \in \mathbb{R}$, we can find N such that for any $n > N$ and $x \in E$,

$$|f_n(x) - f(x)| < \frac{\epsilon}{c} \Longrightarrow |cf_n(x) - cf(x)| < \epsilon$$

Thus $\mathcal{B}$ is an algebra. And by Theorem 1.27, $\mathcal{B}$ is uniformly closed. [16] □

Definition 2.6: Let $\mathcal{A}$ be a family of functions on a set E. $\mathcal{A}$ is said to separate points on E if for every $x_1, x_2 \in E$ and $x_1 \neq x_2$, there exists $f \in \mathcal{A}$ such that $f(x_1) \neq f(x_2)$.
If for any $x \in E$, there exists a function $g \in \mathcal{A}$ such that $g(x) \neq 0$, then we say $\mathcal{A}$ vanishes at no point of E.

[16] Let A be a set of points, the closure is $\overline{A} = A \cup A'$ where A' is the set of all limit points of A. That is, for any point in A', there exists a sequence of points in A that converges to it. Let B be a set of functions, the closure is $\overline{B} = B \cup B'$ where B' is the set of functions with a uniformly convergent sequence of functions in B. A function can be considered as a bunch of points and uniform convergence means all these points converge uniformly.

Stone-Weierstrass Theorem

Theorem 2.7: Suppose $\mathcal{A}$ is an algebra of real valued functions on a set E, $\mathcal{A}$ separates points on E and vanishes at no points of E. Suppose $x_1, x_2 \in E$ and $x_1 \neq x_2$, and $c_1, c_2 \in \mathbb{R}$. Then there exists a function $f \in \mathcal{A}$ such that $f(x_1) = c_1$ and $f(x_2) = c_2$.

Proof: Since $\mathcal{A}$ separates points on E, there exists $g \in \mathcal{A}$ such that $g(x_1) \neq g(x_2)$. Since $\mathcal{A}$ vanishes at no points of E, there exist $h, k \in \mathcal{A}$ such that $h(x_1) \neq 0, k(x_2) \neq 0$.
For $x \in E$, we define

$$u(x) = g(x)k(x) - g(x_1)k(x) \quad v(x) = g(x)h(x) - g(x_2)h(x)$$

Then $u(x_1) = v(x_2) = 0$, $u(x_2) \neq 0$, $v(x_1) \neq 0$. Since $\mathcal{A}$ is an algebra, we have $u, v \in \mathcal{A}$.
Define

$$f(x) = \frac{c_1 v(x)}{v(x_1)} + \frac{c_2 u(x)}{u(x_2)}$$

Then f has the desired properties. $\square$

Theorem 2.8 (Stone-Weierstrass): Let $\mathcal{A}$ be an algebra of real continuous functions on a compact set K. If $\mathcal{A}$ separates points on K and vanishes at no point of K, then the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists all real continuous functions on K.

We divide the proof of Theorem 2.8 into 4 steps.

Step 1: If $f \in \mathcal{B}$, then $|f| \in \mathcal{B}$.

Proof: By the Uniform Limit Theorem [17], any $f \in \mathcal{B}$ is continuous. Since $f : K \to \mathbb{R}$ is continuous and K is compact, by Theorem 1.22 and Theorem 1.14, f is bounded. Let $a = \sup\{|f(x)|\}$ such that $-a \leq f(x) \leq a$ for any x. We define $y = f(x)$. Let $l(y) = |y|$ where $y \in [-a, a]$, then l is a real continuous function. By Theorem 2.3, there exists a sequence of polynomials $\{P_n(y)\}$ that converges to $l(y)$ uniformly on $[-a, a]$. Define $P_n(y) = \sum_{i=0}^{n} c_i y^i$ where $c_0, c_1, ..., c_n \in \mathbb{R}$. Then for any $\epsilon > 0$, there exists N such that for any $n > N$, there exists $c_0, c_1, ..., c_n$ such that for any $y \in [-a, a]$, $\left|\sum_{i=0}^{n} c_i y^i - |y|\right| < \epsilon$. This implies that for any $x \in K$, $\left|\sum_{i=0}^{n} c_i f(x)^i - |f(x)|\right| < \epsilon$.

[17] https://en.wikipedia.org/wiki/Uniform_limit_theorem

Since the functions in $\mathcal{A}$ are bounded, by Theorem 2.5, $\mathcal{B}$ is a uniformly closed algebra. Since $f \in \mathcal{B}$, $P_n(f(x)) = \sum_{i=0}^{n} c_i f(x)^i \in \mathcal{B}$ for any n and $c_0, c_1, ..., c_n$. Thus there exists a sequence of functions $\{P_n(f(x))\}$ in $\mathcal{B}$ which converges to $|f(x)|$ uniformly. By Corollary 1.28, we must have $|f(x)| \in \mathcal{B}$. $\square$

Step 2: If $f \in \mathcal{B}$ and $g \in \mathcal{B}$, then $\max(f, g) \in \mathcal{B}$ and $\min(f, g) \in \mathcal{B}$. Here we define $\max(f, g)(x) = \max\{f(x), g(x)\}$ and $\min(f, g)(x) = \min\{f(x), g(x)\}$ for $x \in K$.

Proof: Since $f \in \mathcal{B}$ and $g \in \mathcal{B}$, we have $f - g \in \mathcal{B}$ and by Step 1, $|f - g| \in \mathcal{B}$. Since

$$\max(f, g) = \frac{f+g}{2} + \frac{|f-g|}{2} \qquad \min(f, g) = \frac{f+g}{2} - \frac{|f-g|}{2}$$

we have $\max(f, g) \in \mathcal{B}$ and $\min(f, g) \in \mathcal{B}$. $\square$

Note that for $h \in \mathcal{B}$, $\max(f, g, h) = \max(\max(f, g), h)$, hence $\max(f, g, h) \in \mathcal{B}$. By induction, if $f_1, f_2, ..., f_n \in \mathcal{B}$, then $\max(f_1, f_2, ..., f_n) \in \mathcal{B}$. Similarly, $\min(f_1, f_2, ..., f_n) \in \mathcal{B}$.

Stone-Weierstrass Theorem

Step 3: Given a real continuous function f on K, a point $x \in K$, and $\epsilon > 0$, there exists a function $g_x \in \mathcal{B}$ such that $g_x(x) = f(x)$ and $g_x(t) > f(t) - \epsilon$ for any $t \in K$.

Proof: Since $\mathcal{A}$ separates points on K and vanishes at no point of K, and $\mathcal{A} \subset \mathcal{B}$, $\mathcal{B}$ separates points on K and vanishes at no point of K. This is because any function in $\mathcal{A}$ satisfying these properties is also in $\mathcal{B}$.
By Theorem 2.7, for any $y \in K$, there exists $h_y \in \mathcal{B}$ such that

$$h_y(x) = f(x) \qquad h_y(y) = f(y)$$

Since h_y is continuous, for any $\epsilon > 0$, there exists $\delta_h > 0$ such that for any $t \in N_{\delta_h}(y)$, $|h_y(t) - h_y(y)| < \frac{\epsilon}{2}$. Since f is continuous, for any $\epsilon > 0$, there exists $\delta_f > 0$ such that for any $t \in N_{\delta_f}(y)$, $|f(t) - f(y)| < \frac{\epsilon}{2}$. Taking $\delta = \min\{\delta_f, \delta_h\}$, then for any $t \in N_\delta(y)$,

$$|h_y(t) - f(t)| \leq |h_y(t) - h_y(y)| + |f(t) - f(y)| < \epsilon$$

Given $\epsilon > 0$, for each $y \in K$, there exists $\delta_y > 0$ satisfying the above inequality, thus $K \subset \bigcup_{y \in K} N_{\delta_y}(y)$. Since K is compact, there exist $y_1, y_2, ..., y_n \in K$ and $\delta_1, \delta_2, ..., \delta_n \in \mathbb{R}$ with finite n such that

$$K \subset N_{\delta_1}(y_1) \cup N_{\delta_2}(y_2) \cup ... \cup N_{\delta_n}(y_n)$$

where $t \in N_{\delta_i}(y_i)$ implies $|h_{y_i}(t) - f(t)| < \epsilon$.

Taking $g_x = \max(h_{y_1}, h_{y_2}, ..., h_{y_n})$, by Step 2, $g_x \in \mathcal{B}$. Since $t \in K$, t must be in at least one open ball $N_{\delta_i}(y_i)$ where $i = 1, 2, ..., n$. For any i, we have

$$g_x(t) - f(t) \geq h_{y_i}(t) - f(t) > -\epsilon \implies g_x(t) > f(t) - \epsilon$$

for $t \in N_{\delta_i}(y_i)$. Thus $g_x(t) > f(t) - \epsilon$ for $t \in K$. $\square$

Step 4: Given a real continuous function f on K, and $\epsilon > 0$, there exists a function $h \in \mathcal{B}$ such that $|h(t) - f(t)| < \epsilon$ for any $t \in K$.

Proof: Since $\max(h_{y_1}, h_{y_2}, ..., h_{y_n})$ is continuous if $h_{y_1}, h_{y_2}, ..., h_{y_n}$ are all continuous, g_x is a continuous function. Thus for any $\epsilon > 0$, there exists $\delta_g > 0$ such that for any $t \in N_{\delta_g}(x)$, $|g_x(t) - g_x(x)| < \frac{\epsilon}{2}$. Since f is a continuous, for any $\epsilon > 0$, there exists $\delta_f > 0$ such that for any $t \in N_{\delta_f}(x)$, $|f(t) - f(x)| < \frac{\epsilon}{2}$. Taking $\delta = \min\{\delta_g, \delta_f\}$, then for any $t \in N_\delta(x)$,

$$|g_x(t) - f(t)| \leq |g_x(t) - g_x(x)| + |f(t) - f(x)| < \epsilon$$

Since K is compact, there exists finite n, $x_1, x_2, ..., x_n \in K$ and $\delta_1, \delta_2, ..., \delta_n \in \mathbb{R}$ such that

$$K \subset N_{\delta_1}(x_1) \cup N_{\delta_2}(x_2) \cup ... \cup N_{\delta_n}(x_n)$$

where $t \in N_{\delta_i}(x_i)$ implies $|g_{x_i}(t) - f(t)| < \epsilon$.

Taking $h = \min(g_{x_1}, g_{x_2}, ..., g_{x_n})$, by Step 2, $h \in \mathcal{B}$. Since $t \in K$, t must be in at least one open ball $N_{\delta_i}(x_i)$ where $i = 1, 2, ..., n$. For any i, we have

$$h(t) - f(t) \leq g_{x_i}(t) - f(t) < \epsilon \implies h(t) < f(t) + \epsilon$$

for $t \in N_{\delta_i}(x_i)$. Thus $h(t) < f(t) + \epsilon$ for $t \in K$.
By Step 3, for any i, $g_{x_i}(t) > f(t) - \epsilon$ for all $t \in K$. Then

$$h(t) = \min\{g_{x_1}(t), g_{x_2}(t), ..., g_{x_n}(t)\} > f(t) - \epsilon$$

for $t \in K$. Therefore, $|h(t) - f(t)| < \epsilon$ for $t \in K$. $\square$

Step 4 implies that for any real continuous function f on K, there exists a sequence of functions in $\mathcal{B}$ that converges to f uniformly, see Lemma C.1. Since $\mathcal{B}$ is an uniform closure, by Corollary 1.28, $f \in \mathcal{B}$. Thus $\mathcal{B}$ contains all real continuous functions on K.

Since $\mathcal{B}$ is the closure of $\mathcal{A}$, by Theorem 1.26, $\mathcal{A}$ is dense in $\mathcal{B}$. Thus, for any real continuous function f on K, we have either $f \in \mathcal{A}$, or for any $\epsilon > 0$, there exists a real continuous function $h \in \mathcal{A}$ such that $|f(t) - h(t)| < \epsilon$ for $t \in K$.

Polynomial Case of Stone-Weierstrass Theorem

If $\mathcal{A}$ is the set of all polynomials with real coefficients on a compact set K, then:

- $\mathcal{A}$ is an algebra. This is because, for any polynomials $f, g \in \mathcal{A}$ and $c \in \mathbb{R}$, $f + g, fg, cf$ are also polynomials, so they are in $\mathcal{A}$.
- $\mathcal{A}$ separates points on K, because there exist polynomial f in $\mathcal{A}$ which is monotonically increasing or decreasing such that for any $x_1, x_2 \in K$ with $x_1 \neq x_2$, $f(x_1) \neq f(x_2)$. For example, let all terms of f be of odd degree and the coefficients be either all positive or all negative.
- $\mathcal{A}$ vanishes at no points of K. If for $x \in K$, there exists $f \in \mathcal{A}$ such that $f(x) = 0$, then there also exists $g(x) = f(x) + c \in \mathcal{A}$ for some $c \neq 0$ such that $g(x) \neq 0$.

Therefore, the set of all polynomials with real coefficients on K is dense in the set of all real continuous functions on K. This means, for any real continuous function, it is either a polynomial, or can be approximated by a polynomial with arbitrary small error on each $x \in K$.

Contents

Lebesgue Integral and Measure Theory

The well known Riemann integral is of the form

$$\int_X f(x)dx$$

where X is an interval in $\mathbb{R}^d$ and $f : X \to \mathbb{R}$ is a function.

However, Riemann integral is not general enough in dealing with limit processes. For example, let $X = \mathbb{R}$ and $f(x) = 1$ if $x \in \mathbb{Q}$ and $f(x) = 0$ if $x \in \mathbb{R} - \mathbb{Q}$. In this case, the Riemann integral is difficult to calculate. Note that every real number a is a closed interval $[a, a]$, so $\mathbb{Q}$ contains infinite number of closed intervals.

At the end of 19th century, it became to many mathematicians that Riemann integral should be replaced by some other type of integral which are more general and flexible to address the limit issues. In this direction, several notable solutions were made by Jordan, Borel, W. H. Young and Lebesgue, and it was turned out that Lebesgue's construction is the most successful.

The the idea of Lebesgue is that, the Riemann integral above can be approximated by a function in the form

$$\sum_{i=1}^{n} f(t_i)m(E_i)$$

as $n \to \infty$, where each E_i is a set such that $\bigcup_{i=1}^{n} E_i = X$, t_i is an element of E_i for any i, and $m(E_i)$ is the length of E_i. However, this function is more general as it can be adapted to any set E_i which may or may not be an interval, and m may not be defined as the geometric length of the interval. For example, in probability theory, if E_i is defined as a set of events, then m can be defined as the total probability of the events in E_i.

Lebesgue discovered a theory showing that the above integral can be defined on most limit situations as long as the following conditions a satisfied: E_is are sets from a measurable space, f is a measurable function and m is a measure. The Lebesgue integral, along with the framework that justifies the validity of its conditions, forms the foundation of measure theory.

Measurable Space

Definition 3.1: Let $\mathcal{R}$ be a family of subsets of X. $\mathcal{R}$ is said to be a σ-algebra (or σ-field) if the following properties hold:

1. $X \in \mathcal{R}$.
2. If $A \in \mathcal{R}$, then $A^c \in \mathcal{R}$, where A^c is the complement of A relative to X.
3. If $A = \bigcup_{n=1}^{\infty} A_n$ and if $A_n \in \mathcal{R}$ for $n = 1, 2, 3, \ldots$, then $A \in \mathcal{R}$.

If $\mathcal{R}$ is a σ-algebra in X, then X is called a measurable space, denoted as $(X, \mathcal{R})$.

Definition 3.2: An alternative definition of σ-algebra is with the properties as follows:

1. $X \in \mathcal{R}$.
2. If $A, B \in \mathcal{R}$, then $A - B \in \mathcal{R}$, where $A - B$ means the set of all elements in A but not in B.
3. If $A = \bigcup_{n=1}^{\infty} A_n$ and if $\{A_n\}$ are mutually disjoint sets where $A_n \in \mathcal{R}$ for $n = 1, 2, 3, \ldots$, then $A \in \mathcal{R}$.

The equivalence of Definition 3.1 and Definition 3.2 of σ-algebra is proved in [18].

In Definition 3.2, if without the property $X \in \mathcal{R}$, then the σ-algebra is reduced to a σ-ring. In a more general case, Lebesgue integral can be defined on σ-ring, see [1], but most of the time we are interested in the σ-algebra case. For example, in probability theory, X represents the set of all possible events and has the measure 1, so X must be included in the σ-ring, which gives an σ-algebra.

The σ in σ-algebra means "closed under countable unions" [19]. If in Definition 3.2 we modify the property 3 to "closed under finite unions", then $\mathcal{R}$ is called an algebra, not a σ-algebra.

Example: Let $\mathcal{F}$ be the family of subsets of $\mathbb{N}$ that is finite or have finite complement, then $\mathcal{F}$ satisfies property 1 and 2. Since for any $i \in \mathbb{N}$ the set $\{2i\}$ is in $\mathcal{F}$, let $A = \bigcup_{i \in \mathbb{N}} \{2i\}$, then both A and A^c are infinite sets, implying that $A \notin \mathcal{F}$. Hence, $\mathcal{F}$ is not closed under countable unions: it is an algebra but not a σ-algebra [20].

[18] https://proofwiki.org/wiki/Equivalence_of_Definitions_of_Sigma-Algebra

[19] https://math.stackexchange.com/questions/2171400.

[20] https://math.stackexchange.com/questions/1010728.

Theorem 3.3: Let $\mathcal{F}$ be any family of subsets of X. There exists a smallest σ-algebra $\mathcal{R}^*$ in X such that $\mathcal{F} \subset \mathcal{R}^*$.

(This $\mathcal{R}^*$ is sometimes called generated by $\mathcal{F}$.)

Proof: Let Ω be the family of all σ-algebras in X which contain $\mathcal{F}$. Ω is non-empty since the set of all subsets of X is a σ-algebra and contained in Ω. Let $\mathcal{R}^*$ be the intersection of all σ-algebras in Ω. $\mathcal{R}^*$ is non-empty for containing $\mathcal{F}$.

Now we show that $\mathcal{R}^*$ is a σ-algebra. Suppose there exists sets $\{A_n\}$ such that $A_n \in \mathcal{R}^*$ for $n = 1, 2, ...$, then for any $\mathcal{R} \in \Omega$, $A_n \in \mathcal{R}$ for all n. Since $\mathcal{R}$ is a σ-algebra, let $A = \bigcup_{n=1}^{\infty} A_n$, then $A \in \mathcal{R}$. Since $A \in \mathcal{R}$ for any $\mathcal{R} \in \Omega$, $A \in \mathcal{R}^*$, which satisfies the property 3 of Definition 3.1. The property 1 and 2 can be proved in a similar way.

Finally, we show that $\mathcal{R}^*$ is the smallest σ-algebra containing $\mathcal{F}$. If there exists a smaller σ-algebra $\mathcal{R}_1$ containing $\mathcal{F}$, then $\mathcal{R}_1 \in \Omega$ and $\mathcal{R}^* \cap \mathcal{R}_1 = \mathcal{R}^*$. Thus, we must have $\mathcal{R}_1 \supset \mathcal{R}^*$. If $\mathcal{R}_1 \neq \mathcal{R}^*$, then $\mathcal{R}_1$ is not the smallest σ-algebra, which is contradict. $\square$

Measurable Function

Definition 3.4 [21]: Let $(X, \mathcal{R})$ be a measurable space and (Y, τ) be a topological space. A function $f : X \to Y$ is measurable if $f^{-1}(E) \in \mathcal{R}$ for every $E \in \tau$.

Definition 3.5: Let (X, τ) be a topological space where τ is the family of all open subsets of X, then by Theorem 3.3, there exists a smallest σ-algebra $\mathcal{B}$ such that $\tau \subset \mathcal{B}$. We say any set in $\mathcal{B}$ is a Borel set, and $\mathcal{B}$ is a Borel σ-algebra.

If an open set is in $\mathcal{B}$, then its complement, which is closed, is also in $\mathcal{B}$. So any open or closed subset of X is a Borel set.

Example: Let (X, τ_1) and (Y, τ_2) be two topological spaces. Let $\mathcal{B}$ be the Borel σ-algebra of X, then $\tau_1 \subset \mathcal{B}$. If the function $f : X \to Y$ is continuous, then for any $E \in \tau_2$, $f^{-1}(E) \in \tau_1 \subset \mathcal{B}$. Therefore, any continuous function f mapping from (X, τ_1) to (Y, τ_2) is measurable, and we call it a Borel function.

Note that $\mathcal{B}$ is generated by a topology τ but it is not a topology. Using $\mathbb{R}$ for example, any subset of $\mathbb{R}$ can be expressed as an uncountable union of closed sets, i.e., the sets with a single element and empty set. If $\mathcal{B}$ is a topology of $\mathbb{R}$, then it is closed under uncountable unions. This implies $\mathcal{B} = \mathcal{P}(\mathbb{R})$, which is impossible.

[21] https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes_ch3.pdf

Measurable Function

Theorem 3.6: Let $\mathcal{R}$ be a σ-algebra in X and Y be a topological space.

(1) If Ω is the family of all $E \subset Y$ such that $f^{-1}(E) \in \mathcal{R}$, then Ω is a σ-algebra.

(2) Let $Y = \mathbb{R}$ and Ω be a σ-algebra of Y, then Ω is a Borel σ-algebra if and only if it is generated by any of the following families of intervals

$$\{(a, \infty) : a \in \mathbb{R}\}, \quad \{[a, \infty) : a \in \mathbb{R}\},$$
$$\{(-\infty, b) : b \in \mathbb{R}\}, \quad \{(-\infty, b] : b \in \mathbb{R}\}$$

(3) Let $f : X \to \mathbb{R}$ be a function. f is measurable if and only if one of the following conditions holds.

$$\{x \in X : f(x) > a\} \in \mathcal{R} \text{ for every } a \in \mathbb{R}$$
$$\{x \in X : f(x) \geq a\} \in \mathcal{R} \text{ for every } a \in \mathbb{R}$$
$$\{x \in X : f(x) < a\} \in \mathcal{R} \text{ for every } a \in \mathbb{R}$$
$$\{x \in X : f(x) \leq a\} \in \mathcal{R} \text{ for every } a \in \mathbb{R}$$

Proof:
(1) First, $f^{-1}(Y) = X \in \mathcal{R}$, showing $Y \in \Omega$. Second, for any $A \in \Omega$, $f^{-1}(Y - A) = X - f^{-1}(A) \in \mathcal{R}$, which means $Y - A \in \Omega$. Third, for any $A_i \in \Omega$, $f^{-1}(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} f^{-1}(A_i) \in \mathcal{R}$, showing that $\bigcup_{i=1}^{\infty} A_i \in \Omega$. Therefore, Ω is a σ-algebra.

Measurable Function

(2) Use $\{(a,\infty) : a \in \mathbb{R}\}$ for example. Let Ω be the σ-algebra generated by all (a,∞) and Ω^* be the Borel σ algebra. Since Ω^* is generated from the family of all open sets and Ω is generated from $\{(a,\infty) : a \in \mathbb{R}\}$, a subset of the family of all open sets, we have $\Omega \subset \Omega^*$.

Conversely, if Ω is a σ-algebra, then for any a, $(a,\infty) \in \Omega$ implies $(-\infty, a] = \mathbb{R} - (a,\infty) \in \Omega$. Since $(-\infty, a) = \bigcup_{i=1}^{\infty}(-\infty, a - \frac{1}{n}]$ where $(-\infty, a - \frac{1}{n}] \in \Omega$ for any $n \in \mathbb{R}$, we have $(-\infty, a) \in \Omega$. For any $b > a$, since $(-\infty, a], [b, \infty) \in \Omega$, $(a, b) = \mathbb{R} - (-\infty, a] \cup [b, \infty) \in \Omega$. Thus we have shown that all open sets $(a,\infty), (-\infty, a), (a, b)$ are contained in Ω. Since Ω^* is the smallest σ-algebra containing all open sets, we have $\Omega \supset \Omega^*$. This shows $\Omega = \Omega^*$. The other three cases can be shown in a similar way.

(3) We prove the first condition. Note that $\{x \in X : f(x) > a\} \in \mathcal{R} \iff f^{-1}((a,\infty)) \in \mathcal{R}$. Since (a,∞) is a subset of $\mathbb{R}$ for any $a \in \mathbb{R}$, by (1), there exists a σ-algebra Ω on $\mathbb{R}$ generated by $\{(a,\infty) : a \in \mathbb{R}\}$. By (2), if a σ-algebra Ω is generated by $\{(a,\infty) : a \in \mathbb{R}\}$, then Ω is a Borel σ-algebra. Again by (1), any Borel set $E \in \Omega$ has $f^{-1}(E) \in \mathcal{R}$. Thus f is measurable.

Also, the other three conditions can be shown in a similar way. $\square$

Theorem 3.6 (3) is a frequently used criterion for the measurability of real valued functions, as we will show in Theorem 3.7 and Theorem 3.9.

Theorem 3.7: Let $f, g : X \to \mathbb{R}$ be real valued measurable functions and $c \in \mathbb{R}$.

(1) $cf, f+g, fg, f/g$ are measurable functions, where we assume $g \neq 0$ in the case f/g.

(2) $|f|, \max(f,g), \min(f,g)$ are measurable functions.

Proof:

(1) cf: By Theorem 3.6 (3), f being measurable implies for every $a \in \mathbb{R}$, $\{x \in X \,:\, f(x) > a\} \in \mathcal{R}$. If $c > 0$, then $\{x \in X \,:\, cf(x) > a\} = \{x \in X \,:\, f(x) > \frac{a}{c}\}$. The cases $c < 0$ and $c = 0$ can be proven similarly.

$f+g$: For any $a \in \mathbb{R}$, we have

$$\{x \in X \,:\, f(x)+g(x) > a\} = \bigcup_{r \in \mathbb{Q}} \{x \in X \,:\, f(x) > r\} \cap \{x \in X \,:\, g(x) > a - r\}$$

Measurable Function

This is because $\{x \in X : f(x) > r\} \cap \{x \in X : g(x) > a - r\}$ $\subset \{x \in X : f(x) + g(x) > a\}$ for every $r \in \mathbb{Q}$. Also, it is impossible to find $x \in X$ such that $f(x) + g(x) > a$ but there is no $r \in \mathbb{Q}$ satisfying $f(x) > r$ and $g(x) > a - r$ [22], because $f(x) + g(x) > a$ implies $f(x) > a - g(x)$, and by Lemma C.4, there must exists $r \in \mathbb{Q}$ giving $f(x) > r > a - g(x) \implies f(x) > r$ and $g(x) > r - a$.

Since f, g are measurable, for any $r \in \mathbb{Q}$, $\{x \in X : f(x) > r\} \in \mathcal{R}$ and $\{x \in X : g(x) > a - r\} \in \mathcal{R}$. Note that if $A_1, A_2, ...A_n$ are in $\mathcal{R}$, then $\bigcap_{i=1}^n A_i = (\bigcup_{i=1}^n A_i^c)^c \in \mathcal{R}$. We have $\{x \in X : f(x) > r\} \cap \{x \in X : g(x) > a - r\} \in \mathbb{R}$. Since $\mathbb{Q}$ is countable (Theorem 1.6), $\bigcup_{r\in\mathbb{Q}} \{x \in X : f(x) > r\} \cap \{x \in X : g(x) > a - r\}$ is a countable union of the members in $\mathcal{R}$, thus $\{x \in X : f(x) + g(x) > a\} \in \mathcal{R}$.

fg: First we show that if f is measurable, then f^2 is measurable. This is because, for any $a \geq 0$,
$\{x \in X : (f(x))^2 > a\} = \{x \in X : f(x) > \sqrt{a}\} \cup \{x \in X : f(x) < -\sqrt{a}\} \in \mathcal{R}$.

[22] https://math.stackexchange.com/questions/2866900

Measurable Function

Also, $\{x \in X : (f(x))^2 > a\} \in \mathcal{R}$ holds for $a < 0$ (this is trivial). Since $fg = \frac{1}{4}[(f+g)^2 - (f-g)^2]$ where both $(f+g)^2$ and $(f-g)^2$ are measurable, fg is measurable.

f/g: Note that $f/g = f \cdot \frac{1}{g}$. We show that $\frac{1}{g}$ is measurable.

$$\{x \in X : 1/g(x) > a\} = \begin{cases} \{x \in X : 0 < g(x) < \frac{1}{a}\} & \text{if } a > 0 \\ \{x \in X : g(x) > 0\} & \text{if } a = 0 \\ \{x \in X : g(x) > 0\} \cup \{x \in X : g(x) < \frac{1}{a}\} & \text{if } a < 0 \end{cases}$$

Note that $\{x \in X : 0 < g(x) < \frac{1}{a}\} = \{x \in X : g(x) > 0\} \cap \{x \in X : g(x) < \frac{1}{a}\}$. Thus, $\{x \in X : 1/g(x) > a\} \in \mathcal{R}$ holds in each of the three cases.

(2) For any $a \in \mathbb{R}$,

$$\{x \in X : \max(f(x), g(x)) > a\} = \{x \in X : f(x) > a\} \cup \{x \in X : g(x) > a\}$$
$$\{x \in X : \min(f(x), g(x)) > a\} = \{x \in X : f(x) > a\} \cap \{x \in X : g(x) > a\}$$

Thus $\max(f,g), \min(f,g)$ are measurable. Take $g = 0$, we have $|f| = \max(f, 0) - \min(f, 0)$. Thus, $|f|$ is measurable. $\square$

Definition 3.8: Let $\{a_n\}$ be a sequence of $(-\infty, \infty)$. Put

$$b_k = \sup\{a_k, a_{k+1}, a_{k+2}, ...\} \quad \text{for } k = 1, 2, 3$$

And

$$\beta = \inf\{b_1, b_2, b_3, ...\}$$

Then we say β is the limit superior of $\{a_n\}$, and write $\beta = \limsup_{n\to\infty} a_n$. It is easy to verify that $b_1 \geq b_2 \geq b_3 ...$, so $b_k \to \beta$ as $k \to \infty$. The limit inferior, denoted as $\liminf_{n\to\infty} a_n$, can be defined analogously by interchanging $\sup$ and $\inf$ above.

It is easy to show that $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$. Also, if $\{a_n\}$ is a bounded sequence, then $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$ if and only if $\{a_n\}$ converges [23]. In this case, $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = \lim_{n\to\infty} a_n$.

Suppose $\{f_n\}$ is a sequence of real valued measurable functions on a set X. Given a fixed $x \in X$, $\{f_n(x)\}$ is a sequence. Thus we define the functions $\sup_n f_n$ and $\limsup_{n\to\infty} f_n$ by

$$\left(\sup_n f_n\right)(x) = \sup_n (f_n(x)), \quad \left(\limsup_{n\to\infty} f_n\right)(x) = \limsup_{n\to\infty} (f_n(x))$$

for every $x \in X$. $\inf_n f_n$ and $\liminf_{n\to\infty} f_n$ can be defined analogously.

[23] https://math.stackexchange.com/questions/1081318/

Measurable Function

Theorem 3.9: If $f_n : X \to \mathbb{R}$ is measurable for $n = 1, 2, 3...$, and

$$g = \sup_{n \geq 1} f_n, \quad h = \lim_{n \to \infty} \sup f_n$$

then g and h are measurable.

Proof: By Theorem 3.6 (3), for any $a \in \mathbb{R}$, $\{x \in X \,:\, f_n(x) > a\} \in \mathcal{R}$ for any f_n, thus

$$\{x \in X \,:\, g(x) > a\} = \bigcup_{n=1}^{\infty} \{x \in X \,:\, f_n(x) > a\} \in \mathcal{R} \tag{5}$$

Thus g is measurable. If replacing $\sup$ with $\inf$ in g, then in Eq (5), $\bigcup$ is replaced by $\bigcap$. In this case, g is still measurable. Since h can be written as

$$h = \inf_{k \geq 1} \left(\sup_{i \geq k} f_i \right)$$

h is measurable. $\square$

Definition 3.10: Let μ be a function of the range $[0, \infty)$ defined on a σ-algebra $\mathcal{R}$. If μ is countably additive, that is, for any countable family of sets $\{A_i\}_{i=1}^{\infty}$ where $A_i \in \mathcal{R}$ for all i and A_i and A_j are disjoint for any $i \neq j$,

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i) \tag{6}$$

then we say μ is a measure on $\mathcal{R}$.

A measure space is a measurable space with a measure defined on it. Let $(X, \mathcal{R})$ be a measurable space and μ be the measure defined on it, we denote the measure space as $(X, \mathcal{R}, \mu)$.

Measure

Theorem 3.11: Let μ be a measure on a σ-algebra $\mathcal{R}$, and $A_1, A_2, ...$ be the sets in $\mathcal{R}$, then

(1) $\mu(\varnothing) = 0$.

(2) If $A_1, A_2, ...$ are mutually disjoint, then for any finite integer $N > 0$, $\mu(\bigcup_{i=1}^N A_i) = \sum_{i=1}^N \mu(A_i)$.

(3) $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$.

(4) If $A_1 \subset A_2$, then $\mu(A_1) \leq \mu(A_2)$.

(5) Suppose $A_1 \subset A_2 \subset A_3 ...$, $A \in \mathcal{R}$, and $A = \bigcup_{i=1}^\infty A_i$. Then, as $n \to \infty$, $\mu(A_n) \to \mu(A)$.

(6) Suppose $A_1 \supset A_2 \supset A_3 ...$, $A \in \mathcal{R}$, $A = \bigcap_{i=1}^\infty A_i$, and $\mu(A_1)$ is finite. Then, as $n \to \infty$, $\mu(A_n) \to \mu(A)$.

Proof:

(1) In Eq (6), let $A_i = \varnothing$ for all i, then $\mu(\varnothing) = \sum_{i=1}^\infty \mu(\varnothing)$, which means $\mu(\varnothing) = 0$.

(2) Since $\varnothing \in \mathcal{R}$ and $A_i \cap \varnothing = \varnothing$ for any $A_i \in \mathcal{R}$, let $A_i = \varnothing$ for all $i > N$ in Eq (6), we get the result.

(3) Since $A_1 \cap A_2, A_1 - A_1 \cap A_2, A_2 - A_1 \cap A_2$ are mutually disjoint, we have

$$\mu(A_1 \cup A_2) = \mu(A_1 \cap A_2) + \mu(A_1 - A_1 \cap A_2) + \mu(A_2 - A_1 \cap A_2)$$

And

$$\begin{aligned} &\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) \\ =& [\mu(A_1 \cap A_2) + \mu(A_1 - A_1 \cap A_2)] + [\mu(A_1 \cap A_2) + \mu(A_2 - A_1 \cap A_2)] \\ =& \mu(A_1) + \mu(A_2) \end{aligned}$$

(4) Since $A_2 = A_1 \cup (A_2 - A_1)$ and $\mu(A_2 - A_1) > 0$ due to the non-negativity of μ,

$$\mu(A_2) = \mu(A_1) + \mu(A_2 - A_1) \geq \mu(A_1)$$

(5) Define $B_1 = A_1$ and $B_i = A_i - A_{i-1}$ for $i = 2, 3, ...$ Clearly $B_i \cup B_j = \varnothing$ for any $i \neq j$. Then $A_n = \bigcup_{i=1}^n B_i$, $A = \bigcup_{i=1}^\infty B_i$ and

$$\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \mu\left(\bigcup_{i=1}^n B_i\right) = \lim_{n\to\infty} \sum_{i=1}^n \mu(B_i) = \sum_{i=1}^\infty \mu(B_i) = \mu\left(\bigcup_{i=1}^\infty B_i\right) = \mu(A)$$

(6) Put $C_n = A_1 - A_n$, then $C_1 \subset C_2 \subset C_3 \ldots$, and $\mu(C_n) = \mu(A_1) - \mu(A_n)$. Since $A_1 - A = \bigcup_{n=1}^{\infty} C_n$, by (5) we have

$$\mu(A_1) - \mu(A) = \mu(A_1 - A) = \mu\left(\bigcup_{n=1}^{\infty} C_n\right) = \lim_{n\to\infty} \mu(C_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$$

Since $\mu(A_1)$ is finite, we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

□

Theorem 3.11(2) shows that countable additivity implies finite additivity. However, the conversion is not true. For example, let $\mathcal{P}(\mathbb{N})$ be the power set of $\mathbb{N}$ so that it is a σ-algebra, and μ be a function defined on $\mathcal{P}(\mathbb{N})$ that for any $A \in \mathcal{P}(\mathbb{N})$, $\mu(A) = 0$ if A is a finite set and $\mu(A) = 1$ if A is an infinite set, then μ is finitely additive. Let $A_i = \{i\}$, then

$$1 = \mu(\mathbb{N}) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \neq \sum_{i=1}^{\infty} \mu(A_i) = 0$$

which means μ is not countably additive. Thus, μ is not a measure on $\mathcal{P}(\mathbb{N})$.

Lebesgue Integral

Definition 3.12: Let s be a real valued function defined on X. If the range of s is finite, we say s is a simple function.
Let $A \subset X$ and denote

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

χ_A is called the characteristic function of E. Suppose the range of s is a finite set of real numbers $\{c_i\}_{i=1}^n$ and define $A_i = \{x : s(x) = c_i\}$, then $s = \sum_{i=1}^n c_i \chi_{A_i}$. That is, every simple function is a finite linear combination of characteristic functions.

Theorem 3.13: For any non-negative measurable function f with finite range, there exists simple functions $s_1, s_2, ..., s_n, ...$ such that $0 \leq s_1 \leq s_2 ... \leq s_n ... \leq f$, and $s_n \to f$ uniformly as $n \to \infty$.

Proof: We define s_n as

$$s_n(x) = \begin{cases} \frac{i-1}{2^n} & \text{if } x \in \left\{ x \,\middle|\, \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n} \right\} \text{ for } i = 1, 2, ..., n2^n \\ n & \text{if } x \in \{x \,|\, f(x) \geq n\} \end{cases}$$

Thus for any x, $s_n(x) \leq f(x)$. We also have

$$s_{n+1}(x) - s_n(x) =$$

$$\begin{cases} 1 & \text{if } x \in \{x \,|\, f(x) \geq n+1\} \\ 0 & \text{if } x \in \left\{x \,\middle|\, \frac{2i-2}{2^{n+1}} \leq f(x) < \frac{2i-1}{2^{n+1}}\right\} \text{ for } i = 1, 2, ..., n2^n \\ \frac{2i-1}{2^{n+1}} & \text{if } x \in \left\{x \,\middle|\, \frac{2i-1}{2^{n+1}} \leq f(x) < \frac{2i}{2^{n+1}}\right\} \text{ for } i = 1, 2, ..., n2^n \\ \frac{i-1}{2^{n+1}} - n & \text{if } x \in \left\{x \,\middle|\, \frac{i-1}{2^{n+1}} \leq f(x) < \frac{i}{2^{n+1}}\right\} \text{ for } i = n2^{n+1}+1, ..., (n+1)2^{n+1} \end{cases}$$

Thus for any x, $s_n(x) \leq s_{n+1}(x)$.
For any $n \in \mathbb{N}^+$, $\max_x\{f(x)\} \leq n \implies f(x) - s_n(x) \leq \frac{1}{2^n}$ for any x. If $f(x)$ is finite, then there exists an $N \in \mathbb{N}^+$ such that for all $n > N$, $\sup\{|f(x) - s_n(x)|\} = \sup\{f(x) - s_n(x)\} = \frac{1}{2^n}$. Therefore,

$$\forall x, \lim_{n\to\infty} \sup\{|f(x) - s_n(x)|\} = \lim_{n\to\infty} \frac{1}{2^n} = 0$$

which implies $s_n \to f$ uniformly as $n \to \infty$.

$\square$

Lebesgue Integral

Definition 3.14 (Lebesgue Integral): Let $(X, \mathcal{R}, \mu)$ be a measure space, $E \in \mathcal{R}$ be a set, and $A_i \in \mathcal{R}$ for any i. Suppose $s(x) = \sum_{i=1}^{n} c_i \chi_{A_i}$ is measurable simple function, and define $I_E(s) = \sum_{i=1}^{n} c_i \mu(E \cap A_i)$ ($I_E(s)$ can be considered as the inner product of $s(x)$ and the measure vector μ). If $f : X \to [0, \infty)$ is measurable, we define

$$\int_E f \, d\mu = \sup I_E(s) \tag{7}$$

where the supremum is taken over all measurable simple functions s such that $0 \leq s \leq f$. We call the left hand side of Eq (7) as the Lebesgue Integral.

It is easy to verify that for every non-negative simple function s,

$$\int_E s \, d\mu = I_E(s)$$

If f is ranged in $\mathbb{R}$, we define $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ such that both f^+ and f^- are non-negative. f^+ is called the positive part of f and f^- is called the negative part of f. We have $|f| = f^+ + f^-$ and $f = f^+ - f^-$. By Theorem 3.7 (2), if f is measurable, then both f^+ and f^- are measurable.

Definition 3.15: Let $(X, \mathcal{R}, \mu)$ be a measure space, $f : X \to \mathbb{R}$ be a measurable function and $E \in \mathcal{R}$ be a set. Consider the two Lebesgue integrals:

$$\int_E f^+ \, d\mu, \quad \int_E f^- \, d\mu$$

If at least one of them is finite (in order to avoid the $\infty - \infty$ case, which is undefined), we define the Lebesgue integral of f as

$$\int_E f \, d\mu = \int_E f^+ \, d\mu - \int_E f^- \, d\mu$$

If both of them are finite, then we say f is Lebesgue integrable on E with respect to the measure μ.
If f is Lebesgue integrable, then

$$\int_E |f| \, d\mu = \int_E f^+ \, d\mu + \int_E f^- \, d\mu < \infty \tag{8}$$

we define the set of all f satisfying Eq (8) as $L^1(\mu)$. That is, $L^1(\mu)$ is the family of all Lebesgue integrable functions on E. (By default, we assume $L^1(\mu)$ as a family of all Lebesgue integrable functions on X, unless otherwise specified.)

Lebesgue Integral

Assume f, g are measurable functions, the following properties hold for Lebesgue integral:

- If $0 \le f \le g$, then $\int_E f\, d\mu \le \int_E g\, d\mu$. (Monotonicity)
- If $A \subset B$ and $f \ge 0$, then $\int_A f\, d\mu \le \int_B f\, d\mu$.
- If $f \ge 0$, c is a positive constant, then $\int_E cf\, d\mu = c \int_E f\, d\mu$.

Theorem 3.16: Let $f : X \to [0, \infty)$ be a measurable function, for $E \in \mathcal{R}$, define

$$\phi(E) = \int_E f\, d\mu$$

Then ϕ is a measure on $\mathcal{R}$.

Proof: We need to show that $\phi(E)$ is non-negative and countably additive. Since $0 \le f$ implies $\phi(E) = \int_E f\, d\mu \ge \int_E 0\, d\mu = 0$, ϕ is non-negative.

To prove the countable additivity of ϕ, we need to show that

$$\phi(E) = \sum_{i=1}^{n} \phi(E_i)$$

where $E_i \in \mathcal{R}$ for $i = 1, 2, ...$, $E_i \cap E_j = \varnothing$ for $i \ne j$ and $E = \bigcup_{i=1}^{\infty} E_i$.

Lebesgue Integral

If f is simple, the conclusion holds, since for any $A_1, A_2, ..., A_n \in \mathcal{R}$ such that $f(x) = \sum_{i=1}^{n} c_i \chi_{A_i}$,

$$\phi(E) = \sum_{i=1}^{n} c_i \mu(E \cap A_i) = \sum_{i=1}^{n} c_i \sum_{j=1}^{\infty} \mu(E_j \cap A_i) = \sum_{j=1}^{\infty} \sum_{i=1}^{n} c_i \mu(E_j \cap A_i) = \sum_{j=1}^{\infty} \phi(E_j)$$

In general case, for any measurable simple function s such that $0 \le s \le f$, we have

$$\int_E s\, d\mu = \sum_{i=1}^{\infty} \int_{E_i} s\, d\mu \le \sum_{i=1}^{\infty} \int_{E_i} f\, d\mu = \sum_{i=1}^{\infty} \phi(E_i) \tag{9}$$

Note that the first equality in Eq (9) uses the countable additivity of the measure $\int_E s\, d\mu$ as s is simple. Thus

$$\phi(E) = \int_E f\, d\mu = \sup_s \left(\int_E s\, d\mu \right) \le \sum_{i=1}^{\infty} \phi(E_i) \tag{10}$$

Note that if $\phi(E_i) = \infty$ for some i, the equality of Eq (10) holds since $\phi(E) \ge \phi(E_i)$. We consider the case that all E_i are finite.

Lebesgue Integral

For any $\epsilon > 0$, we can find s such that

$$\int_{E_1} s\ d\mu \geq \int_{E_1} f\ d\mu - \epsilon \quad \text{and} \int_{E_2} s\ d\mu \geq \int_{E_2} f\ d\mu - \epsilon$$

Such s always exists. This is because, by Theorem 3.13, there exists a sequence $\{s_n\}$ such that $0 \leq s_1 \leq s_2 ... \leq f$ and $s_n \to f$ uniformly as $n \to \infty$. By the monotonicity of Lebesgue integral we have $0 \leq \int s_1\ d\mu \leq \int s_2\ d\mu \leq ... \leq \int f\ d\mu$. Suppose there exist s_i on E_1 satisfying $\int_{E_1} s_i\ d\mu \geq \int_{E_1} f\ d\mu - \epsilon$ and s_j on E_2 satisfying $\int_{E_2} s_j\ d\mu \geq \int_{E_2} f\ d\mu - \epsilon$, we take $s = \max(s_i, s_j)$.

Therefore,

$$\phi(E_1 \cup E_2) \geq \int_{E_1 \cup E_2} s\ d\mu = \int_{E_1} s\ d\mu + \int_{E_2} s\ d\mu = \phi(E_1) + \phi(E_2) - 2\epsilon$$

Since the above inequality holds for any $\epsilon > 0$, by Lemma 3.23 we have

$$\phi(E_1 \cup E_2) \geq \phi(E_1) + \phi(E_2)$$

Thus

$$\phi(E) = \phi(\bigcup_{i=1}^{\infty} E_i) \geq \sum_{i=1}^{\infty} \phi(E_i) \tag{11}$$

Combining Eq (10) and Eq (11), we get

$$\phi(E) = \sum_{i=1}^{\infty} \phi(E_i)$$

(In general, for any measurable $f : X \to \mathbb{R}$, the countable additivity of $\phi(E)$ holds, but we may not have $\phi(E) \geq 0$.) $\square$

Theorem 3.17 (Lebesgue's Monotone Convergence): Let $\{f_n\}$ be a sequence of real valued measurable functions on X. Suppose that

(1) $0 \leq f_1(x) \leq f_2(x) \leq ...$ for every $x \in X$.

(2) $f_n(x) \to f(x)$ for every $x \in X$.

Then f is measurable, and

$$\int_X f_n \, d\mu \to \int_X f \, d\mu \quad \text{as } n \to \infty \tag{12}$$

Proof: Since $f = \sup_n \{f_n\}$, by Theorem 3.9, f is measurable. Since $0 \leq f_n \leq f$, by the monotonicity of Lebesgue integral, $\int_X f_n \, d\mu \leq \int_X f \, d\mu$ for all n, thus

$$\sup_n \left(\int_X f_n \, d\mu \right) \leq \int_X f \, d\mu \tag{13}$$

Lebesgue Integral

Let s be any simple function such that $0 \le s \le f$, and c be a constant with $0 < c < 1$ [24]. Define

$$E_n = \{x \in X : f_n(x) > cs(x)\} \quad \text{for } n = 1, 2, 3, ...$$

Then each $E_n \in \mathcal{R}$, because

$$E_n = \{x \in X : f_n(x) > 0\} \cup \{x \in X : f_n(x)/c(x) > c, c(x) \neq 0\}$$

And by Theorem 3.6 (3) and Theorem 3.7 (1), both $\{x \in X : f_n(x) > 0\}$ and $\{x \in X : f_n(x)/s(x) > c, s(x) \neq 0\}$ are in $\mathcal{R}$.

Also we have $E_1 \subset E_2 \subset E_3 \subset ...$ and $\bigcup_{n=1}^{\infty} E_n = \{x \in X : f(x) > cs(x)\} = \{x \in X : f(x) > s(x)\} = X$. Thus

$$\int_X f_n \, d\mu \ge \int_{E_n} f_n \, d\mu \ge c \int_{E_n} s \, d\mu \tag{14}$$

Since by Theorem 3.16, $\phi(E_n) = \int_{E_n} s \, d\mu$ is a measure, and by Theorem 3.11, $\phi(E_n) \to \phi(X)$, we have

$$\int_X s \, d\mu = \phi(X) = \sup_n \phi(E_n) = \sup_n \left(\int_{E_n} s \, d\mu \right)$$

[24] The reason not taking $c = 1$. https://math.stackexchange.com/questions/2916756

Thus, taking the supremum with respect to n on the left and right side of Eq (14), we get

$$\sup_n \left(\int_X f_n \, d\mu \right) \geq c \int_X s \, d\mu$$

Since the above inequality holds for any s and any $0 < c < 1$, taking the supremum with respect to s and c on the right hand side of the above inequality, we get

$$\sup_n \left(\int_X f_n \, d\mu \right) \geq \sup_{s,c} \left(c \int_X s \, d\mu \right) = \int_X f \, d\mu \tag{15}$$

Combining Eq (13) and Eq (15),

$$\int_X f \, d\mu = \sup_n \left(\int_X f_n \, d\mu \right)$$

Since f_n increases monotonically with n on each $x \in X$, the above equality is equivalent to Eq (12).

$\square$

Lebesgue Integral

Theorem 3.18: Suppose $f, g \in L^1(\mu)$ and α, β are two real numbers, then $\alpha f + \beta g \in L^1(\mu)$, and

$$\int_X (\alpha f + \beta g)\, d\mu = \alpha \int_X f\, d\mu + \beta \int_X g\, d\mu \tag{16}$$

Proof: The measurability of $\alpha f + \beta g$ follows from Theorem 3.7 (1). And

$$\int_X |\alpha f + \beta g|\, d\mu \leq \int_X (|\alpha||f| + |\beta||g|)\, d\mu = |\alpha| \int_X |f|\, d\mu + |\beta| \int_X |g|\, d\mu < \infty$$

Thus $\alpha f + \beta g \in L^1(\mu)$.

Eq (16) can be proved by showing

$$\int_X \alpha f\, d\mu = \alpha \int_X f\, d\mu \tag{17}$$

$$\int_X (f + g)\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu \tag{18}$$

Eq (17) holds by the property of Lebesgue integral. We will show that Eq (18) holds.

Lebesgue Integral

If f and g are non-negative simple measurable functions, let $f(x) = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ where $A_i = \{x \in X : f(x) = \alpha_i\}$, and let $g(x) = \sum_{j=1}^{m} \beta_j \chi_{B_j}$ where $B_j = \{x \in X : g(x) = \beta_j\}$. Since $X = \bigcup_{i=1}^{n} A_i = \bigcup_{j=1}^{m} B_j$, each $x \in X$ must be at least in one A_i and one B_j. Thus $f(x) + g(x) = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) \chi_{A_i \cap B_j}$.

$$\begin{aligned}
\int_E (f+g)\, d\mu = f(x) + g(x) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha_i + \beta_j) \mu(A_i \cap B_j \cap E) \\
&= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} \mu(A_i \cap B_j \cap E) + \sum_{j=1}^{m} \beta_j \sum_{i=1}^{n} \mu(A_i \cap B_j \cap E) \\
&= \sum_{i=1}^{n} \alpha_i \mu(A_i \cap E) + \sum_{j=1}^{m} \beta_j \mu(B_j \cap E) = \int_E f\, d\mu + \int_E g\, d\mu
\end{aligned}$$

If f and g are non-negative measurable functions, then by Theorem 3.13 there exist a sequence of simple functions $\{s'_n\}$ such that $s_n \to f$ as $n \to \infty$, and another sequence $\{s''_n\}$ such that $s''_n \to g$. Let $s_n = s'_n + s''_n$, then $s_n \to f + g$ as $n \to \infty$.

Thus by Theorem 3.17,

$$\int_E (f+g)\, d\mu = \sup_n \left(\int_E s_n \, d\mu \right) = \sup_n \left(\int_E s_n' \, d\mu \right) + \sup_n \left(\int_E s_n'' \, d\mu \right)$$
$$= \int_E f \, d\mu + \int_E g \, d\mu \tag{19}$$

If f and g are any measurable functions, define $h = f + g$, then $h^+ - h^- = (f^+ - f^-) + (g^+ - g^-)$, which is equivalent to $h^+ + f^- + g^- = f^+ + g^+ + h^-$. By Eq (19) we have $\int h^+ + \int f^- + \int g^- = \int f^+ + \int g^+ + \int h^-$. Reorganizing the equation we get $\int h^+ - \int h^- = (\int f^+ - \int f^-) + (\int g^+ - \int g^-) \iff \int h = \int f + \int g$.

$\square$

Lemma 3.19 (Fatou's): Let $\{f_n\}$ be a sequence of measurable functions where each $f_n : X \to [0, \infty)$, then

$$\int_X \left(\liminf_{n\to\infty} f_n \right) \, d\mu \le \liminf_{n\to\infty} \int_X f_n \, d\mu$$

Proof: Since $\liminf_{n\to\infty} f_n = \sup_{k\geq 1}(\inf_{n\geq k} f_n)$, let $g_k = \inf_{n\geq k} f_n$, then each g_k is measurable by Theorem 3.9. Also, $0 \leq g_1 \leq g_2 \leq g_3 \leq ...$ and $g_k \to \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \leq f_n$ for any $n \geq k$, we have $\int_X g_k \, d\mu \leq \int_X f_n \, d\mu$ for any $n \geq k$, i.e., $\int_X g_k \, d\mu \leq \inf_{n\geq k}\left(\int_X f_n \, d\mu\right)$. Then by Lebesgue's Monotone Convergence Theorem (Theorem 1.37),

$$\int_X \left(\liminf_{n\to\infty} f_n\right) \, d\mu = \int_X \left(\lim_{k\to\infty} g_k\right) \, d\mu = \lim_{k\to\infty}\int_X g_k \, d\mu$$
$$\leq \lim_{k\to\infty}\inf_{n\geq k}\left(\int_X f_n \, d\mu\right) = \liminf_{n\to\infty}\int_X f_n \, d\mu$$

□

Theorem 3.20 (Lebesgue's Dominated Convergence): Suppose $\{f_n\}$ is a sequence of real measurable functions on X, and let f be a function such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in X$. If there exists a function $g \in L^1(\mu)$ such that $|f_n(x)| \leq g(x)$ for any n and $x \in X$, then $f \in L^1(\mu)$ and

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

Lebesgue Integral

Proof: Since $|f_n| \le g$ holds when $n \to \infty$, we have $|f| \le g$. Since $|f| = f^+ + f^-$ where $f^+ \le g$ and $f^- \le g$, both f^+ and f^- are in $L^1(\mu)$, and by Theorem 3.18, $f = f^+ - f^- \in L^1(\mu)$.

$|f_n| \le g$ implies that $2g \ge g + f_n \ge 0$ and $2g \ge g - f_n \ge 0$ for any n. By Fatou's Lemma (Lemma 3.19),

$$\int_X g\ d\mu + \int_X f\ d\mu = \int_X (g+f)\ d\mu = \int_X \lim_{n\to\infty} (g+f_n)\ d\mu$$

$$= \int_X \liminf_{n\to\infty} (g+f_n)\ d\mu \le \liminf_{n\to\infty} \int_X (g+f_n)\ d\mu = \int_X g\ d\mu + \liminf_{n\to\infty} \int_X f_n\ d\mu$$

Similarly,

$$\int_X g\ d\mu - \int_X f\ d\mu = \int_X \liminf_{n\to\infty} (g-f_n)\ d\mu \le \liminf_{n\to\infty} \int_X (g-f_n)\ d\mu$$

$$= \int_X g\ d\mu + \liminf_{n\to\infty} \int_X (-f_n)\ d\mu = \int_X g\ d\mu - \limsup_{n\to\infty} \int_X f_n\ d\mu$$

Thus, $\liminf\limits_{n\to\infty} \int_X f_n\ d\mu \ge \int_X f\ d\mu \ge \limsup\limits_{n\to\infty} \int_X f_n\ d\mu$, implying that $\lim_{n\to\infty} \int_X f_n\ d\mu = \int_X f\ d\mu$. $\square$

Complete Measure Space

Let $\mathcal{R}$ be an σ-algebra of X. It is possible that there exists some subsets of X that not in $\mathcal{R}$. If $A \subset E \subset B$ where $A, B \in \mathcal{R}$ and $\mu(A) = \mu(B)$, it is reasonable to define $\mu(E) = \mu(A) = \mu(B)$ (As a special case, let $A = \varnothing$ and B be a set with $\mu(B) = 0$, this implies every subset of a zero measure set has zero measure). However, it is possible that $E \notin \mathcal{R}$, so μ is undefined on E. We would like to extend $\mathcal{R}$ by including all such Es to allow the measure defined on them.

Theorem 3.21: Let $(X, \mathcal{R}, \mu)$ be a measure space. Let $\mathcal{R}^*$ be the family of all $E \subset X$ such that there exists $A, B \in \mathcal{R}$ with $\mu(A) = \mu(B)$ such that $A \subset E \subset B$, and define $\mu(E) = \mu(A)$. Then $\mathcal{R}^*$ is a σ-algebra, and μ is a measure on $\mathcal{R}^*$.

Proof: It is easy to see that $\mathcal{R}$ is a subset of $\mathcal{R}^*$, and the above conditions hold for any $E \in \mathcal{R}$, which can be shown by taking $A = B = E$. So we only discuss the case when $E \notin \mathcal{R}$.

We first check that μ is well defined for every $E \in \mathcal{R}^*$. Suppose $A \subset E \subset B$ for some $A, B \in \mathcal{R}$ with $\mu(A) = \mu(B)$, and $A_1 \subset E \subset B_1$ for some $A_1, B_1 \in \mathcal{R}$ with $\mu(A_1) = \mu(B_1)$. We use the property that if the two sets C, D have $C \subset D$, then $\mu(C) \leq \mu(D)$ ($\mu(D) = \mu(C) + \mu(D - C)$ and $\mu(D - C)$ is non-negative).

$A \subset E \subset B$ and $A_1 \subset E \subset B_1$ imply $A \subset E \subset B_1$ and $A_1 \subset E \subset B$, which give $\mu(A) \leq \mu(B_1)$ and $\mu(A_1) \leq \mu(B)$. Since $\mu(A_1) = \mu(B_1)$, we have $\mu(A) \leq \mu(A_1) \leq \mu(B)$, which implies $\mu(A) = \mu(A_1)$. Thus, the value of $\mu(E)$ in the cases $A \subset E \subset B$ and $A_1 \subset E \subset B_1$ are equivalent, which means the definition of μ is unambiguous.

Next we show that $\mathcal{R}^*$ is a σ-algebra by verifying the three properties of Definition 3.1:

(1) $X \in \mathcal{R}^*$ because $\mathcal{R} \subset \mathcal{R}^*$.

(2) If $A \subset E \subset B$ then $B^c \subset E^c \subset A^c$. $\mu(A^c) = \mu(X) - \mu(A) = \mu(X) - \mu(B)$ $= \mu(B^c)$ when $\mu(X)$ is either finite or infinite. Thus $E^c \in \mathcal{R}$.

(3) Suppose we have a sequence $\{E_i\}$ where for each i, there exist $A_i, B_i \in \mathcal{R}$ such that $A_i \subset E_i \subset B_i$ and $\mu(A_i) = \mu(B_i)$. Let $A = \bigcup_{i=1}^{\infty} A_i, B = \bigcup_{i=1}^{\infty} B_i$ and $E = \bigcup_{i=1}^{\infty} E_i$, then $A, B \in \mathcal{R}$. We show $\mu(A) = \mu(B)$ as follows: Since $A_1 \subset B_1$ and $A_2 \subset B_2$, $A_1 \cap A_2 \subset B_1 \cap B_2$. Thus $\mu(A_1 \cap A_2) \leq \mu(B_1 \cap B_2)$. Given that $\mu(A_1) = \mu(B_1)$ and $\mu(A_2) = \mu(B_2)$,

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2) \geq \mu(B_1) + \mu(B_2) - \mu(B_1 \cap B_2) = \mu(B_1 \cup B_2)$$

Complete Measure Space

Since $A_1 \cup A_2 \subset B_1 \cup B_2$, $\mu(A_1 \cup A_2) \le \mu(B_1 \cup B_2)$, thus $\mu(A_1 \cup A_2) = \mu(B_1 \cup B_2)$. Extending this property to countable number of is, we get

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

Thus $E \in \mathcal{R}^*$.

Finally, if E_is are disjoint in step (3) above, then A_is are disjoint, and

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

which shows μ is countably additive.

$\square$

Definition 3.22: A measure μ on a σ-algebra $\mathcal{R}$ is called complete if for any $N \in \mathcal{R}$ and $S \subset N$, $\mu(N) = 0$ implies $S \in \mathcal{R}$ and $\mu(S) = 0$.

Theorem 3.21 says every measure μ can be completed by extending $\mathcal{R}$ to $\mathcal{R}^*$. Such $\mathcal{R}^*$ is called the μ-completion of $\mathcal{R}$.

Motivations on σ-algebra

The measure μ is a nonnegative set function defined on the subsets of a set X. However, there may exist some subsets in X that μ cannot be defined on. For example, if X is the Euclidean space $\mathbb{R}$ and μ is the Lebesgue measure, let $V \subset \mathbb{R}$ be a Vitali set [25], then μ cannot be defined on V [26].

Let $\mathcal{P}(X)$ be the power set of X, we would like to find a subset $\mathcal{R} \subset \mathcal{P}(X)$ such that μ can be defined on every member of $\mathcal{R}$. We wish μ satisfies the property that for any disjoint A, B, $\mu(A \cup B) = \mu(A) + \mu(B)$. Thus $A \cup B \in \mathcal{R}$. Also, we wish that for any $A, B \in \mathcal{R}$, $A \cup B$ can be written as the union of two disjoint sets $A - B$ and B such that $\mu(A \cup B) = \mu(A - B) + \mu(B)$. So $A - B \in \mathcal{R}$. Therefore, $\mathcal{R}$ must be closed under disjoint set union and set difference. We call such $\mathcal{R}$ an algebra.

The algebra can be defined in alternative ways. For example, since $A - B = (A^c \cup B)^c$, $\mathcal{R}$ is also an algebra if it is closed under set union (no need to be disjoint) and set complement.

---

[25] https://en.wikipedia.org/wiki/Vitali_set

[26] https://e.math.cornell.edu/people/belk/measuretheory/NonMeasurableSets.pdf

Motivations on σ-algebra

However, as we mentioned, the measure μ also need to satisfy the countably additivity. Consider such a case that we calculate the volume of a unit ball in $\mathbb{R}^d$ by splitting it into disjoint intervals and sum up the volumes of these intervals. The number of such intervals cannot be finite. If we define the volume of each individual interval E as $\mu(E)$, then to calculate volume of the ball, μ should be defined on a countable union of intervals.

As we have mentioned, let $\{A_i\}_{i=1}^{\infty}$ be an arbitrary subset of members of $\mathcal{R}$, the statement $\bigcup_{i=1}^{n} A_i \in \mathcal{R}$ for any $n \in \mathbb{N}^+ \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$ is false, while the converse is true. Since we need $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$ to guarantee the countably additivity of μ, $\mathcal{R}$ should be a σ-algebra, rather than simply an algebra.

More generally, consider the principles of mathematical induction: Let P(n) be a statement related to an integer $n \in \mathbb{N}$. If

- P(0) is true.
- If P(n) is true, then P$(n+1)$ is true.

then P(n) is true for all $n \in \mathbb{N}$. However, this does not imply P(n) is true when $n \to \infty$. Here are some counterexamples.

Motivations on σ-algebra

- $\frac{1}{n} > 0$ for all $n \in \mathbb{N}^+$, but $\lim_{n\to\infty} \frac{1}{n} = 0$.
- $\sum_{i=1}^{n} \frac{1}{i}$ is finite for all $n \in \mathbb{N}^+$, but $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i}$ is infinite.
- $\sum_{i=1}^{n} \frac{1}{i^2} \in \mathbb{Q}$ for all $n \in \mathbb{N}^+$, but $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i^2} = \frac{\pi^2}{6} \notin \mathbb{Q}$.

However, for the last example above, $\sum_{i=1}^{n} \frac{1}{i^2}$ is finite does imply $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i^2}$ is finite. Thus, $\mathsf{P}(n)$ is true for all $n \in \mathbb{N} \implies \mathsf{P}(n)$ is true when $n \to \infty$ holds only for some statements P, which needs proof.

Another important thing is that infinity can be either countable or uncountable, and they make huge differences. For example, compare $\lim_{n\to\infty}(n \cdot 0)$ and $(\lim_{n\to\infty} n) \cdot 0$ [27]. Since $\lim_{n\to\infty}(n \cdot 0) = \lim_{n\to\infty}(\sum_{i=1}^{n} 0) = \sum_{i=1}^{\infty} 0$, it is a sum of countably many zeros. We can prove that the sum of countably many zeros is zero: For any $\epsilon > 0$, let the ith 0 be bounded by $-\frac{\epsilon}{2^i} < 0 < \frac{\epsilon}{2^i}$, then $-\epsilon < \sum_{i=1}^{\infty} 0 < \epsilon$, which implies $\sum_{i=1}^{\infty} 0 = 0$. However, $(\lim_{n\to\infty} n) \cdot 0 = \infty \cdot 0$, where $\infty \cdot 0$ is undefined as we do not know whether ∞ is countable or uncountable.

[27] https://math.stackexchange.com/questions/28940/

Motivations on σ-algebra

The sum of uncountably many zeros can be any real number. For example, let (a, b) be an interval on $\mathbb{R}$ where $b > a$ and μ be the Lebesgue measure, then $\mu((a, b)) = b - a$. We have $b - a = \mu(\bigcup_{x \in (a,b)} x) = \sum_{x \in (a,b)} \mu(\{x\})$. For any a and b, there are uncountably many elements x in (a, b) [28], and $\mu(\{x\}) = 0$ on each x (Lemma C.9).

Finally, the power set $\mathcal{P}(X)$ is a σ-algebra for sure, and we have shown that there exist some members in it that we cannot define the Lebesgue measure on. The question is, given a specific measure μ, can we find a subset $\mathcal{R}$ of $\mathcal{P}(X)$ such that $\mathcal{R}$ is a σ-algebra and μ can be defined on all members of $\mathcal{R}$? We will show that the existence of such $\mathcal{R}$ can be proven by construction, based on Caratheodory's Extension Theorem.

[28] $\mathbb{R}$ is isomorphic to $(-\frac{\pi}{2}, \frac{\pi}{2})$ since $f(x) = \tan x$ is a bijection from $(-\frac{\pi}{2}, \frac{\pi}{2})$ to $\mathbb{R}$. Also, $(-\frac{\pi}{2}, \frac{\pi}{2})$ is isomorphic to any (a, b) since $g(x) = a + (x + \pi/2)(b - a)/\pi$ is a bijection from $(-\frac{\pi}{2}, \frac{\pi}{2})$ to (a, b).

Next we will introduce how to construct a measure. We first show some useful inequalities. Let a and b be two constants.

Lemma 3.23: $a \leq b + \epsilon$ for any $\epsilon > 0$ if and only if $a \leq b$.

Proof: $\Longleftarrow$ is obviously true. We prove $\Longrightarrow$ by contradiction. Suppose $a > b \Longleftrightarrow \frac{a-b}{2} > 0$. Let $\epsilon = \frac{a-b}{2} > 0$, then

$$a \leq b + \frac{a-b}{2} \Longleftrightarrow b \geq a$$

which is contradict.

Prove by contradiction is equivalent to showing the negation is wrong [29]. We can also prove it by showing the contraposition is true. The contraposition is: if $a > b$, then there exists $\epsilon > 0$ such that $a > b + \epsilon$. Such ϵ exists, e.g., $\epsilon = \frac{a-b}{2}$. $\square$

Corollary 3.24: $a < b + \epsilon$ for any $\epsilon > 0$ if and only if $a \leq b$.

[29] The negation of $P \to Q$ is $\neg(P \to Q) \Longleftrightarrow P \wedge \neg Q$. Prove by contradiction is to show that $\neg Q$ and P cannot be both true. Especially, $P \wedge \neg Q \to Q \Longleftrightarrow \neg(P \wedge \neg Q) \vee Q \Longleftrightarrow \neg(P \wedge \neg Q)$, which implies that if $P \wedge \neg Q \to Q$ is true, then $P \wedge \neg Q$ must be false.

Preliminaries

Here is a trick in dealing with supremum and infinimum related inequalities, which will be frequently used in the following proofs. The main idea is that leave out some errors, and show the error can be eliminated.

Let $f : X \to \mathbb{R}$ be a continuous function. Suppose $A \leq f(x)$ for any $x \in X$ where $A \in \mathbb{R}$ is independent of x. We show that $A \leq \inf_x\{f(x)\}$.

- By the definition of infinimum, for any $\epsilon > 0$, there exists $x \in X$ such that $f(x) < \inf_x\{f(x)\} + \epsilon$.
- For all $x \in X$, $A \leq f(x)$.
- Thus $A < \inf_x\{f(x)\} + \epsilon$ holds for any $\epsilon > 0$. Since both A and $\inf_x\{f(x)\}$ are independent of x, they can be treated as constants. By Corollary 3.24, $A \leq \inf_x\{f(x)\}$.

Construction of Measures

We have shown that the measure μ is an nonnegative countably additive function defined on a σ-algebra $\mathcal{R}$, but we have not shown whether such μ and $\mathcal{R}$ exist. The Caratheodory's Extension Theorem shows that we can extend a nonnegative additive function m (called premeasure) on an algebra $\mathcal{A}$ to a measure μ on a σ-algebra $\mathcal{R}$. In other words, $\mathcal{A} \subset \mathcal{R}$, and $\mu = m$ on $\mathcal{A}$.

Definition 3.25: Let $\mathcal{A}$ be an algebra of the subsets of X, a function $m : \mathcal{A} \to [0, \infty)$ is a premeasure of $\mathcal{A}$ if

(1) $m(\varnothing) = 0$.

(2) For any finite numbers of disjoint sets $A_1, A_2, ..., A_n \in \mathcal{A}$,

$$m\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} m(A_i)$$

Definition 3.26: An outer measure on a set X is a function $\mu^* : \mathcal{P}(X) \to [0, \infty)$ satisfying

(1) $\mu^*(\varnothing) = 0$.

(2) (Monotonicity) If $B_1 \subset B_2$, then $\mu^*(B_1) \leq \mu^*(B_2)$.

(3) (Subadditivity) If $\{B_j\}$ is any sequence of subsets of X, then

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

Theorem 3.27 (Caratheodory's Extension): Let μ^* be an outer measure on X and $\mathcal{R}$ be a family of subsets of X defined as follows

$$\mathcal{R} = \{A \subset X \ : \ \text{for all } E \subset X,\ \mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)\}$$

Then $\mathcal{R}$ is a σ-algebra, and μ^* is a measure on $\mathcal{R}$.

Proof: We carry out the proof in several steps.

Step 1: $\mathcal{R}$ is an algebra.

We show that $\mathcal{R}$ is closed under set complement and set union.

Construction of Measures

Let $A = X$, obviously $\mu^*(E) = \mu^*(X \cap E) + \mu^*(X^c \cap E)$ holds, so $X \in \mathcal{R}$.

For any $E \subset X$, if $A \in \mathcal{R}$, then $\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)$, of which A^c also satisfies. Thus, $A \in \mathcal{R}$ implies $A^c \in \mathcal{R}$.

If $A, B \in \mathcal{R}$, we show $A \cup B \in \mathcal{R}$. For any $E \subset X$, denote $F = A \cap E$ and $G = A^c \cap E$, then $F, G \subset X$, which implies

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

Therefore, by the subadditivity of μ^*,

$$\begin{aligned}
&\mu^*(E) = \mu^*(F) + \mu^*(G) \\
&= \mu^*(A \cap B \cap E) + \mu^*(A^c \cap B \cap E) + \mu^*(A \cap B^c \cap E) + \mu^*(A^c \cap B^c \cap E) \\
&= [\mu^*(A \cap B \cap E) + \mu^*(A^c \cap B \cap E) + \mu^*(A \cap B^c \cap E)] + \mu^*((A \cup B)^c \cap E) \\
&\geq \mu^*([(A \cap B) \cup (A^c \cap B) \cup (A \cap B^c)] \cap E) + \mu^*((A \cup B)^c \cap E) \\
&= \mu^*((A \cup B) \cap E) + \mu^*((A \cup B)^c \cap E)
\end{aligned}$$

Also by the subadditivity of μ^*,

$$\mu^*(E) = \mu^*([(A \cup B) \cup (A \cup B)^c] \cap E) \leq \mu^*((A \cup B) \cap E) + \mu^*((A \cup B)^c \cap E)$$

Therefore,

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

which implies $A \cup B \in \mathcal{R}$.

Note that if $\mathcal{R}$ is closed under set complement and set union, then it is closed under set difference: If $A, B \in \mathcal{R}$ implies $A^c, B^c \in \mathcal{R}$, then $A - B = A \cap B^c = (A^c \cup B)^c \in \mathcal{R}$; similarly we have $B - A \in \mathcal{R}$. By Definition 3.2, to prove $\mathcal{R}$ is a σ-algebra, we need to show that $\mathcal{R}$ is closed under countable unions of disjoint sets.

Step 2: μ^* is finitely additive on $\mathcal{R}$.

For any $A, B \in \mathcal{R}$, we have $E \cap (A \cup B) \subset X$. Thus, replacing E with $E \cap (A \cup B)$ in $\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)$, we get

$$\begin{aligned}\mu^*\left(E \cap (A \cup B)\right) &= \mu^*(E \cap (A \cup B) \cap A) + \mu^*(E \cap (A \cup B) \cap A^c) \\ &= \mu^*(E \cap A) + \mu^*(E \cap A^c \cap B)\end{aligned}$$

If $A \cap B = \varnothing$, then $B \subset A^c$. In this case,

$$\mu^*\left(E \cap (A \cup B)\right) = \mu^*(E \cap A) + \mu^*(E \cap B) \tag{20}$$

In particular, taking $E = X$, then $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$, which means E is finitely additive on the disjoint sets in $\mathcal{R}$.

Step 3: $\mathcal{R}$ is a σ-algebra.

Let $\{A_i\}$ be a sequence of disjoint sets in $\mathcal{R}$, and let $A = \bigcup_{i=1}^{\infty} A_i$ and $B_n = \bigcup_{i=1}^{n} A_i$. Since $\mathcal{R}$ is closed under set union, we have $B_n \subset X$, and for any $E \subset X$,

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c)$$

Since $A^c \subset B_n^c$, by the monotonicity of μ^*,

$$\mu^*(E) \geq \mu^*(E \cap B_n) + \mu^*(E \cap A^c) \tag{21}$$

By monotonicity, $\mu^*(E \cap A) = \mu^*(E \cap (\bigcup_{i=1}^{\infty} A_i)) \geq \mu^*(E \cap (\bigcup_{i=1}^{n} A_i)) = \mu^*(E \cap B_n)$ holds for any n. Thus $\mu^*(E \cap A) \geq \lim_{n\to\infty} \mu^*(E \cap B_n)$. Note that $\mu^*(E \cap B_n) = \sum_{i=1}^{n} \mu^*(E \cap A_i)$. By countable subadditivity, $\mu^*(E \cap A) \leq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) = \lim_{n\to\infty} \mu^*(E \cap B_n)$. Therefore, $\mu^*(E \cap A) = \lim_{n\to\infty} \mu^*(E \cap B_n)$ [30].

[30] This proof comes from [19], which shows the induction $\mu^*(E \cap B_n) = \sum_{i=1}^{n} \mu^*(E \cap A_i)$ holds for any $n \in \mathbb{N}^+$ also holds when $n \to \infty$. Note that in general the induction from n to ∞ does not always hold, see `https://math.stackexchange.com/questions/98093/`. (Formally, we may not have $f(\lim_{n\to\infty} a_n) = \lim_{n\to\infty} f(a_n)$ in general.)

Thus, for any $\epsilon > 0$, there exists n such that $\mu^*(E \cap B_n) \geq \mu^*(E \cap A) - \epsilon$. Since Eq (21) holds for any n,

$$\mu^*(E) \geq \mu^*(E \cap B_n) + \mu^*(E \cap A^c) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) - \epsilon$$

implying that $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$. Since $\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c)$ by subadditivity, we have $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$. Thus $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$, and $\mathcal{R}$ is an σ-algebra.

Step 4: μ^* is countably additive on $\mathcal{R}$.

In Step 3, we get $\mu^*(E) = \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap A^c)$. Since $A \in \mathcal{R}$, let $E = A$, then

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

This shows μ^* is countably additive, thus it is a measure on $\mathcal{R}$.

□

We call the $\mathcal{R}$ obtained in Theorem 3.27 as Caratheodory σ-algebra.

Construction of Measures

The following theorem shows a general way to construct the outer measure μ^*, using a family of subsets $\mathcal{E}$ of X and a nonnegative set function ρ defined on it. Note that $\mathcal{E}$ is not necessary an algebra and ρ is not necessary a premeasure.

Theorem 3.28: Let $\mathcal{E}$ be a family of subsets of X with $\varnothing \in \mathcal{E}$ and $X \in \mathcal{E}$, $\rho : \mathcal{E} \to [0, \infty)$ be a set function with $\rho(\varnothing) = 0$. Define μ^* as

$$\mu^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \rho(A_i) \; : \; E \subset \bigcup_{i=1}^{\infty} A_i, A_i \in \mathcal{E} \text{ for all } i \right\} \tag{22}$$

for any $E \subset X$, then μ^* is an outer measure on X, and $\mu^*(A) \leq \rho(A)$ for any $A \in \mathcal{E}$.

Proof: We show μ^* defined in Eq (22) satisfies the three properties of Definition 3.26. Let $A_i = \varnothing$ for all i in Eq (22), since $\rho(\varnothing) = 0$, we have $\mu(\varnothing) = 0$.

If $E_1 \subset E_2$ and $E_2 \subset \bigcup_{i=1}^{\infty} A_i$ for some sequence $\{A_i\}$, then $E_1 \subset \bigcup_{i=1}^{\infty} A_i$, which implies $\mu^*(E_1) \leq \mu^*(E_2)$.

The subadditivity can be shown as follows: Let $\{E_i\}$ be any sequence of subsets of X, then $\mu^*(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu^*(E_i)$. Let $\{A_{i,j}\}$ be the sequence of sets in $\mathcal{A}$ such that $E_i \subset \bigcup_{j=1}^{\infty} A_{i,j}$ (if E_i is covered by the finite union of sets $\bigcup_{j=1}^{n} A_{i,j}$, let $A_{i,j} = \varnothing$ for $j > n$ in the sequence $\{A_{i,j}\}$).

By the definition of $\inf$, on each i, for any $\epsilon > 0$, there exists a sequence $\{A_{i,j}\}$ such that

$$\mu^*(E_i) + \frac{\epsilon}{2^i} > \sum_{j=1}^{\infty} \rho(A_{i,j})$$

Since $\bigcup_{i=1}^{\infty} E_i \subset \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} A_{i,j}$,

$$\mu^* \left(\bigcup_{i=1}^{\infty} E_i \right) \le \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \rho(A_{i,j}) < \sum_{i=1}^{\infty} \left(\mu^*(E_i) + \frac{\epsilon}{2^i} \right) = \sum_{i=1}^{\infty} \mu^*(E_i) + \epsilon$$

The above ineqaulity holds for any $\epsilon > 0$. By Corollary 3.24, $\mu^* \left(\bigcup_{i=1}^{\infty} E_i \right) \le \sum_{i=1}^{\infty} \mu^*(E_i)$.
The relationship $\mu^*(A) \le \rho(A)$ for any $A \in \mathcal{E}$ can be shown as follows. Since $A \subset X$, let $\{A_i\}$ be any sequence in $\mathcal{E}$ such that $A \subset \bigcup_{i=1}^{\infty} A_i$, then $\mu^*(A) \le \sum_{i=1}^{\infty} \rho(A_i)$. Let $A_1 = A$ and $A_i = \varnothing$ for all $i \ge 2$, we get the result.

□

Now we consider the special case that $\mathcal{E}$ is an algebra and ρ is a premeasure. We let $\mathcal{E}$ be $\mathcal{A}$ and ρ be m.

Theorem 3.29: Let $\mathcal{R}$ be the Caratheodory σ-algebra generated by μ^* according to Theorem 3.27, and take $\mathcal{E} = \mathcal{A}$, $\rho = m$ in Theorem 3.28, then $\mathcal{A} \subset \mathcal{R}$.

Proof: We need to show that for any $A \in \mathcal{A}$, $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$ for all $E \subset X$. By Eq (22), for any $\epsilon > 0$, there exists a sequence $\{A_i\}_{i=1}^{\infty}$ in $\mathcal{A}$ such that $E \subset \bigcup_{i=1}^{n} A_i$ and

$$\mu^*(E) + \epsilon > \sum_{i=1}^{\infty} m(A_i)$$

Since m is additive on $\mathcal{A}$, for any $A \in \mathcal{A}$, $m(A_i) = m(A_i \cap A) + m(A_i \cap A^c)$. Since $\mathcal{A}$ is closed under union and complement, we have $A_i \cap A = (A_i^c \cup A^c)^c \in \mathcal{A}$ and $A_i \cap A^c \in \mathcal{A}$. Thus

$$\mu^*(E) + \epsilon > \sum_{i=1}^{\infty} m(A_i \cap A) + m(A_i \cap A^c) \geq \sum_{i=1}^{\infty} \mu^*(A_i \cap A) + \mu^*(A_i \cap A^c)$$

$$\geq \mu^*(\left(\bigcup_{i=1}^{n} A_i\right) \cap A) + \mu^*(\left(\bigcup_{i=1}^{n} A_i\right) \cap A^c) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

Thus $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$. By the subadditivity of μ^*, $\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c)$. Thus $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$, which means $A \in \mathcal{R}$. $\square$

Construction of Measures

So far we have shown that μ^* is a measure on $\mathcal{R}$, $\mathcal{A} \subset \mathcal{R}$ and $\mu^* \leq m$ on $\mathcal{A}$. It remains to be shown that $\mu^* = m$ on $\mathcal{A}$. However, there exists some ill behaved m giving that $\mu^* < m$ for some $A \in \mathcal{A}$.

For example, let $\mathcal{A}$ be the algebra of all half open intervals on $\mathbb{R}$ [31], and $F(x)$ be the function that $F(x) = 1$ for all $x > 0$ and $F(x) = 0$ for all $x \leq 0$. Denote $m((a, b]) = F(b) - F(a)$ [32], then by Definition 3.25, m is a premeasure (For property 1, $m(\varnothing) = m((a, a]) = 0$). Thus $m((-1, 1]) = 1$.

Since $(-1, 1] = (-1, 0] \cup (0, 1] = (-1, 0] \cup (\bigcup_{i=2}^{\infty} (\frac{1}{i}, \frac{1}{i-1}])$. Since $(-1, 0] \in \mathcal{A}$ and $(\frac{1}{i}, \frac{1}{i-1}] \in \mathcal{A}$ for any $i \geq 2$, we have $\mu^*((-1, 1]) \leq m((-1, 0]) + \sum_{i=2}^{\infty} m((\frac{1}{i}, \frac{1}{i-1}]) = 0$ [33]. Thus $\mu^*((-1, 1]) < m((-1, 1])$.

We need to add extra constraints on m to enable $\mu^* = m$ on $\mathcal{A}$.

[31] https://math.stackexchange.com/questions/735621/

[32] Note that the complement of $(a, b]$ is $(-\infty, a] \cup (b, \infty)$, which is in $\mathcal{A}$, but the premeasure m is undefined on it. See https://math.stackexchange.com/questions/4086027/.

[33] $(0, 1] = \bigcup_{i=2}^{\infty} (\frac{1}{i}, \frac{1}{i-1}] \in \mathcal{A}$, but m is not countably additive, so $1 = m((0, 1]) \neq \sum_{i=2}^{\infty} m((\frac{1}{i}, \frac{1}{i-1}]) = 0$. Besides, since $\mathcal{A}$ is an algebra but not a sigma-algebra, $A = \bigcup_{i=1}^{\infty} A_i$ where each $A_i \in \mathcal{A}$ does not necessary give $A \in \mathcal{A}$. For example, $(-1, 1) = \bigcup_{i=1}^{\infty} (-1, 1 - \frac{1}{i}]$, but $(-1, 1) \notin \mathcal{A}$.

Definition 3.30: Let $\mathcal{A}$ be an algebra of subsets of X and m be a premeasure on $\mathcal{A}$. We say m is continuous at the empty set if and only if for any sequence of sets $\{A_i\}$ in $\mathcal{A}$ satisfying $A_{i-1} \supset A_i$ for all $i \geq 1$, $m(A_1) < \infty$ and $\bigcap_{i=1}^{\infty} A_i = \varnothing$ implies $\lim_{i\to\infty} m(A_i) = 0$. We say m is semifinite if and only if whenever $A \in \mathcal{A}$ has $m(A) = \infty$, for any $r > 0$, there exists $B \in \mathcal{A}$ such that $B \subset A$ and $r < m(B) < \infty$.

Theorem 3.31: Let $\mathcal{A}$ be an algebra of subsets of X, and m be a semifinite measure on $\mathcal{A}$ that is continuous at empty set. Let μ^* be the outer measure on X defined by Eq (22). Then for all $A \in \mathcal{A}$, $\mu^*(A) = m(A)$, so that μ^* extends m.

Proof: We only need to show that $\mu^*(A) \geq m(A)$ for all $A \in \mathcal{A}$ with $m(A) < \infty$.

If $m(A) = \infty$, by the semifinite property of m, for any $r > 0$, there exists $B \in \mathcal{A}$, $B \subset A$ such that $r < m(B) < \infty$. If $\mu^*(B) \geq m(B)$ is true, then by the monotonicity of μ^*, $\mu^*(A) \geq \mu^*(B) \geq m(B) > r$, which holds for any $r > 0$. This implies $\mu^*(A) = \infty$, thus $\mu^*(A) = m(A)$.

Consider the $m(A) < \infty$. By Eq (22), for any $\epsilon > 0$, there exists a sequence $\{A_i\}$ in $\mathcal{A}$ such that $A \subset \bigcup_{i=1}^{\infty} A_i$ with $m(A_i) < \infty$ for each i (In the worst case we can take $A_1 = A$ and $A_i = \varnothing$ for all $i \neq 1$) and

$$\mu^*(A) + \epsilon > \sum_{i=1}^{\infty} m(A_i)$$

Define the sequences $\{B_i\}$ and $\{C_i\}$ as follows: $B_1 = A \cap A_1$ and $B_n = A \cap A_n - \bigcup_{i=1}^{n-1} B_i$. It can be shown that $B_n \in \mathcal{A}$ for any n and $B_1, B_2, ..., B_n$ are disjoint: Since $\mathcal{A}$ is closed under set union and set complement, it is closed under set intersection, so $B_1 \in \mathcal{A}$. For any $n \geq 2$, if $B_1, B_2, ..., B_{n-1} \in \mathcal{A}$, then $B_n \in \mathcal{A}$ and $B_n \cap (\bigcup_{i=1}^{n-1} B_i) = \varnothing$.

Moreover, since $\bigcup_{i=1}^{n} B_i = A \cap (\bigcup_{i=1}^{n} A_i)$, $\bigcup_{i=1}^{\infty} B_i = A \cap (\bigcup_{i=1}^{\infty} A_i) = A$.

Define $C_n = A \cap (\bigcup_{i=1}^{n} B_i)^c$ for $i \geq 1$, then $C_n \in \mathcal{A}$, $C_n \supset C_{n+1}$ for all n and $\lim_{n\to\infty} C_n = A \cap (\bigcup_{i=1}^{\infty} B_i)^c = A \cap A^c = \varnothing$. Hence

$$A = A \cap X = A \cap \left((\bigcup_{i=1}^{n} B_i) \cup (\bigcup_{i=1}^{n} B_i)^c \right) = (\bigcup_{i=1}^{n} B_i) \cup \left(A \cap (\bigcup_{i=1}^{n} B_i)^c \right)$$

$$= (\bigcup_{i=1}^{n} B_i) \cup C_n$$

Since $(\bigcup_{i=1}^n B_i) \cap C_n = \varnothing$, we have

$$m(A) = \sum_{i=1}^n m(B_i) + m(C_n)$$

which holds for any n. Note that we still have $B_n, C_n \in \mathcal{A}$ as $n \to \infty$ because of induction. Since we assume m is continuous at empty set, $\lim_{n\to\infty} m(C_n) = 0$. Therefore,

$$m(A) = \lim_{n\to\infty} \sum_{i=1}^n m(B_i) + m(C_n) = \sum_{i=1}^\infty m(B_i) \le \sum_{i=1}^\infty m(A_i) \le \mu^*(A) + \epsilon$$

which holds for any $\epsilon > 0$, implying that $m(A) \le \mu^*(A)$.

□

So far we have shown that the extension from m on $\mathcal{A}$ to μ^* on $\mathcal{R}$ holds as long as m is semifinite and continuous at empty set. We say m is finite if $m(X) < \infty$. The following theorem shows that such extension is unique if m is finite.

Theorem 3.32 [34]: When m is a finite measure, its extension to μ^* is unique.

Proof: Suppose there exists another measure ν^* from m such that $m(A) = \nu^*(A)$ for all $A \in \mathcal{A}$. For any $E \in \mathcal{R}$, let $\{A_i\}$ be a sequence of sets in $\mathcal{A}$ such that $E \subset \bigcup_{i=1}^{\infty} A_i$. Since ν^* is an outer measure, we have

$$\nu^*(E) \leq \nu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \nu^*(A_i) = \sum_{i=1}^{\infty} m(A_i)$$

Since the above inequality holds for any $\{A_i\}$, and by Eq (22), $\mu^*(E) = \inf\{\sum_{i=1}^{\infty} m(A_i)\}$. Thus $\nu^*(E) \leq \mu^*(E)$. Note that when taking $A_1 = X$ and $A_i = \varnothing$ for $i \geq 2$, we have $E \subset X$ and $\mu^*(E) \leq m(X) < \infty$, thus $\mu^*(E)$ is finite.

Replacing E with E^c, we get $\nu^*(E^c) \leq \mu^*(E^c)$. Since

$$\nu^*(E) + \nu^*(E^c) = \nu^*(X) = m(X) = \mu^*(X) = \mu^*(E) + \mu^*(E^c)$$

we must have $\nu^*(E) = \mu^*(E)$.

□

[34] https://sites.stat.washington.edu/jaw/COURSES/520s/521/HO.521/Caratheodory-unique.pdf

Lebesgue Measure

We have shown that a premeasure m on an algebra $\mathcal{A}$ can be extended to a measure μ on a σ-algebra $\mathcal{R}$ by Caratheodory's extension theorem. Now we introduce Lebesgue measure, an important special case constructed under this method.

Definition 3.33: Let $\mathbb{R}^d$ be the d dimensional Euclidean space. Given a set of pairs $\{(a_i, b_i)\}_{i=1}^d$ with $a_i, b_i \in \mathbb{R}, a_i \leq b_i$ for all i, we define the interval as

$$I = \{x \in \mathbb{R}^d : a_i \leq x_i \leq b_i \text{ for } i = 1, 2, ..., d\}$$

and the volume of the interval is defined by

$$m(I) = \prod_{i=1}^{d} (b_i - a_i)$$

Let $\mathcal{A}$ be the family of sets which can be expressed as the union finite number of intervals, then $\mathcal{A}$ is an algebra but not a σ-algebra [35]. Let $I_1, I_2, ..., I_n$ be disjoint intervals on $\mathcal{A}$, we set

[35] $\mathbb{R}^d$ itself is an interval with $m(\mathbb{R}^d) = \infty$, so $\mathbb{R}^d \in \mathcal{A}$. Also, since $\mathcal{A}$ is not closed under countable unions, it is not a σ-algebra.

Lebesgue Measure

$$m\left(\bigcup_{i=1}^{n} I_i\right) = \sum_{i=1}^{n} m(I_i)$$

such that m is a premeasure on $\mathcal{A}$. It is easy to verify that m is semifinite and continuous at the empty set.

By Caratheodory's extension theorem, m on $\mathcal{A}$ can be extended to a measure μ^* on a σ-algebra $\mathcal{R}$. Such μ^* is called Lebesgue measure, $\mathcal{R}$ is called Lebesgue σ-algebra and any set in $\mathcal{R}$ is said to be Lebesgue measurable. Since $\mathcal{A} \subset \mathcal{R}$ and $\mathcal{R}$ is closed under countable unions, any set that can be expressed as countable union of intervals is in $\mathcal{R}$ [36].

In detail, let E be a set that can be expressed by countable union of intervals, i.e., there exists $\{I_i\}_{i=1}^{\infty}$ in $\mathcal{A}$ such that $E = \bigcup_{i=1}^{\infty} I_i$, then $E \in \mathcal{R}$. Denote $J_n = I_n - \bigcup_{j=1}^{n-1} I_j$, then $\{J_i\}_{i=1}^{\infty}$ is a sequence of disjoint sets in $\mathcal{R}$ (since each J_i includes countable union of intervals) and $\bigcup_{i=1}^{\infty} I_i = \bigcup_{i=1}^{\infty} J_i$. Thus $\mu(E) = \mu\left(\bigcup_{i=1}^{\infty} I_i\right) = \mu\left(\bigcup_{i=1}^{\infty} J_i\right) = \sum_{i=1}^{\infty} \mu(J_i)$. Since μ is well defined on each J_i, $\mu(E)$ is well defined.

[36] Note that those sets that do not equivalent to but can be bounded by countable unions of intervals may not be in $\mathcal{R}$, and μ^* is an outer measure but may not be a measure on them. Also, $\mathcal{R}$ may contain the sets that cannot be expressed as countable union of intervals.

Lebesgue Measure

By Corollary C.7, every open set on $\mathbb{R}^d$ can be expressed as countable unions of intervals, thus every open set is Lebesgue measurable. Since every Borel set is obtained by countable unions and complements of open sets, it must be in $\mathcal{R}$. Thus every Borel set on $\mathbb{R}^d$ is Lebesgue measurable. Moreover, let $\mathcal{B}$ be the Borel σ-algebra on $\mathbb{R}^d$, then $\mathcal{B} \subset \mathcal{R}$.

Contents

Definition 4.1: A real vector space (or a vector space over the real field) is a set V. The elements in V are called vectors. There are two operations defined on V: addition and scalar multiplication, which satisfy the following axioms for any vectors $x, y, z \in V$ and scalars α, β:

- **Addition**: (1) The addition satisfies Commutativity $x + y = y + x$ and Associativity $x + (y + z) = (x + y) + z$. (2) There exists an identity element of addition in V, called the zero vector, denoted as 0, such that $x + 0 = x$ for any $x \in V$. (3) For every $x \in V$, there exists an vector $-x \in V$, called the additional inverse, such that $x + (-x) = 0$.
- **Scalar Multiplication**: (1) The scalar multiplication satisfies Commutativity $\alpha(\beta x) = (\alpha\beta)x$, Distributivity with respect to scalars $(\alpha + \beta)x = \alpha x + \beta x$ and Distributivity with respect to vectors $\alpha(x + y) = \alpha x + \alpha y$. (2) There exists an identity element in the field of scalar, called multiplicative identity, denoted as 1, such that $1x = x$ for any $x \in V$.

Theorem 3.18 shows that $L^1(\mu)$ is a vector space. We consider each function f as an element in the space $L^1(\mu)$ such that the operations between the elements satisfy all the axioms in Definition 4.1.

Definition 4.2: A linear transformation from a vector space V into a vector space W is a mapping $\Lambda : V \to W$ such that

$$\Lambda(\alpha x + \beta y) = \alpha \Lambda x + \beta \Lambda y$$

for any $x, y \in V$ and any scalars α, β [37]. If $W = \mathbb{R}$, then Λ is called a linear functional. We say Λ is a positive linear functional if $\Lambda f \geq 0$ whenever $f \geq 0$.

Integration as a Linear Functional: By Theorem 3.18, the mapping

$$f \to \int_X f \, d\mu$$

is a linear functional on $L^1(\mu)$ for any μ.

[37] We write $\Lambda(x)$ as Λx for simplicity.

Definition 4.3: The support of a function f on a topological space X, denoted as $\text{supp}(f)$, is the closure of the set $\{x \in X : f(x) \neq 0\}$, i.e., the set of points in X where f is non-zero.

The family of all continuous functions $f : X \to \mathbb{R}$ whose support is compact is denoted as $C_c(X)$.

Observe that $C_c(X)$ is a vector space, This is due to the following facts: If f and g are continuous and of compact support, then

- $f + g$ is continuous [38] and of compact support. Define $\text{cl}(A)$ as the closure of the set A. The support of $f+g$ is $\text{cl}\{x \in X : f(x) + g(x) \neq 0\} = \text{cl}\{x \in X : f(x) \neq 0\} \cup \text{cl}\{x \in X : g(x) \neq 0\}$, because the closure of the union equals to the union of closures [39]. Since both $\text{cl}\{x \in X : f(x) \neq 0\}$ and $\text{cl}\{x \in X : g(x) \neq 0\}$ are compact, $\text{cl}\{x \in X : f(x) + g(x) \neq 0\}$ is compact, because the finite unions of compact sets are compact [40].
- αf is continuous and compact for any $\alpha \in \mathbb{R}$. (Note that when $\alpha = 0$, the support of αf is $\varnothing$. $\varnothing$ is compact because it is the subset of any open ball.)

[38] https://math.stackexchange.com/questions/2750160/
[39] https://math.stackexchange.com/questions/1986224/
[40] https://math.stackexchange.com/questions/1973340/

Normed Space and Banach Space

Definition 4.4: A real vector space X is said to be a normed space [41] if to each x there is associated a nonnegative real number $\|x\|$, called the norm of x, such that

- $\|x+y\| \leq \|x\| + \|y\|$ for $x, y \in X$.
- $\|\alpha x\| = |\alpha| \, \|x\|$ for $x \in X$ and $\alpha \in \mathbb{R}$.
- $\|x\| = 0$ implies $x = 0$.

Since $\|x+y\| \leq \|x\| + \|y\|$ and $\|x\| = \|-x\|$, we have

$$\|x-y\| = \|x-z+z-y\| \leq \|x-z\| + \|z-y\|$$

This shows every normed space can be regarded as a metric space by taking $d(x,y) = \|x-y\|$.

A Banach Space is a normed space which is complete in the metric defined by its norm.

[41] Also called "normed vector space" or "normed linear space".

Example: A real vector space H is said to be an inner product space if to any $x, y \in H$, there is associated a real number $\langle x, y\rangle$, called the inner product of x, y, such that for all $x, y, z \in H$ and $\alpha \in \mathbb{R}$: (1) $\langle x, y\rangle = \langle y, x\rangle$; (2) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$; (3) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$; (4) $\langle x, x\rangle \geq 0$; (5) $\langle x, x\rangle = 0$ if and only if $x = 0$.

With inner product, we can define norm by $\|x\| = \sqrt{\langle x, x\rangle}$. The square root is used here in order to satisfy the scalar axiom of norm: $|\alpha|\,\|x\| = \|\alpha x\| = \sqrt{\langle \alpha x, \alpha x\rangle} = |\alpha|\sqrt{\langle x, x\rangle}$.

A Hilbert Space is an inner product space which is complete in the metric defined by its inner product (inner product defines norm, and norm is a metric). Hence, inner product space is a special case of normed space, and Hilbert space is a special case of Banach space.

Normed Space and Banach Space

Remark: Here we review the relationships between vector space, metric space, topological space and normed space.

In general, an abstract space is a set of (unspecified) elements satisfying certain axioms.

A vector space is an abstract space satisfying the axioms of algebraic operations (addition, multiplication and identity), as given in Definition 4.1.

A metric space is an abstract space satisfying the axioms of metric (Definition 1.8), on which the geometric properties of the sets like openness and closeness can be defined.

A topological space defines the openness of the sets (Definition 1.29) with more general axioms, showing that the geometric properties of sets can be analyzed even without a metric. A metric space is a special case of a topological space.

A metric space may not be a vector space and vice versa. A normed space is both a vector space and a metric space.

Normed Space and Banach Space

Definition 4.5: Let X and Y be normed spaces and $\Lambda : X \to Y$ be a linear transformation. The norm of Λ is defined by

$$\|\Lambda\| = \sup\{\|\Lambda x\| \, : \, x \in X, \, \|x\| = 1\}$$

Λ is said to be bounded if $\|\Lambda\| < \infty$.

It is easy to show that $\|\Lambda x\| \leq \|\Lambda\|\|x\|$ for any $x \in X$.

Theorem 4.6: Let X and Y be normed spaces and $\Lambda : X \to Y$ be a linear transformation, then the following three conditions are equivalent:
(1) Λ is bounded.
(2) Λ is continuous.
(3) Λ is continuous at any point of X.

Proof: (1) $\Longrightarrow$ (2): Since $\|\Lambda(x_1 - x_2)\| \leq \|\Lambda\|\|x_1 - x_2\|$, for any $\epsilon > 0$, let $\delta = \frac{\epsilon}{\|\Lambda\|}$, then $\|x_1 - x_2\| < \delta \Longrightarrow \|\Lambda(x_1 - x_2)\| < \epsilon$. (2) $\Longrightarrow$ (3) is trivial. (3) $\Longrightarrow$ (1): Suppose Λ is continuous at x_0, then for each $\epsilon > 0$, there exists $\delta > 0$ such that $\|x - x_0\| < \delta \Longrightarrow \|\Lambda(x - x_0)\| < \epsilon$. Let $y = x - x_0$, then $y \in X$ and $\|y\| < \delta \Longrightarrow \|\Lambda y\| < \epsilon$, thus $\|\Lambda \frac{y}{\|y\|}\| < \frac{\epsilon}{\|y\|} < \frac{\epsilon}{\delta}$ holds for any Λ. Therefore, $\|\Lambda\| \leq \frac{\epsilon}{\delta}$. $\square$

Hahn-Banach Theorem

The Hahn-Banach Theorem states that a bounded linear functional of a subspace can be extended to the whole space with its norm unchanged.

Theorem 4.7 (Hahn-Banach): If M is a subspace of a normed space X and if f is a bounded linear functional on M, then there exists a function F on X such that $F(x) = f(x)$ for all $x \in M$, and $\|F\| = \|f\|$, where we define

$$\|f\| = \sup\{|f(x)| : x \in M, \|x\| = 1\} \qquad \|F\| = \sup\{|F(x)| : x \in X, \|x\| = 1\}$$

Proof: If $\|f\| = 0$, then the desired F satisfies $\|F\| = 0$, which implies $F(x) = 0$ for any $x \in X$. Such F is a linear functional. We consider the case that $\|f\| > 0$.

If $M = X$, then $f = F$, which is trivial. If $M \subset X$, then there exists $x_0 \in X, x_0 \notin M$. We define the space M' as

$$M' = \{x + \lambda x_0 : x \in M, \lambda \in \mathbb{R}\}$$

Here M' can be considered we extend M by adding a new base x_0 to it. Obviously, $M \subset M'$. Then we have the following lemma:

Lemma 4.8: There exists a linear functional f_1 on M' such that $f_1(x) = f(x)$ for all $x \in M$ and $\|f_1\| = \|f\|$.

Hahn-Banach Theorem

Proof: Let $f_1(x + \lambda x_0) = f(x) + \lambda\alpha$ where $\alpha \in \mathbb{R}$, then f_1 is a linear functional, because for any $x_1, x_2 \in M$ and $\lambda_1, \lambda_2, \beta_1, \beta_2 \in \mathbb{R}$, we have $x_1 + \lambda_1 x_0, x_2 + \lambda_2 x_0 \in M'$ and

$$\begin{aligned} &f_1(\beta_1(x_1 + \lambda_1 x_0) + \beta_2(x_2 + \lambda_2 x_0)) = f(\beta_1 x_1 + \beta_2 x_2) + \beta_1\lambda_1\alpha + \beta_2\lambda_2\alpha \\ &= \beta_1 f(x_1) + \beta_2 f(x_2) + \beta_1\lambda_1\alpha + \beta_2\lambda_2\alpha = \beta_1(f(x_1) + \lambda_1\alpha) + \beta_2(f(x_2) + \lambda_2\alpha) \\ &= \beta_1 f_1(x_1 + \lambda_1 x_0) + \beta_2 f_1(x_2 + \lambda_2 x_0) \end{aligned}$$

Also, $f_1(x) = f(x)$ for $x \in M$, as shown by taking $\lambda = 0$.
Now we show that there exists α such that $\|f_1\| = \|f\|$. We prove this by showing

(1) For any α, $\|f\| \leq \|f_1\|$.

(2) There exists α such that $\|f_1\| \leq \|f\|$.

For (1), since $M \subset M'$, we have [42]

$$\begin{aligned} \|f\| &= \sup\{|f(x)| : x \in M, \|x\| = 1\} = \sup\{|f_1(x)| : x \in M, \|x\| = 1\} \\ &\leq \sup\{|f_1(x)| : x \in M', \|x\| = 1\} = \|f_1\| \end{aligned}$$

This shows that under an extension the norm cannot decrease, which holds for any α.

[42] Here we use the property of $\sup$ that if A and B are two sets of real numbers and $A \subset B$, then $\sup A \leq \sup B$.

Hahn-Banach Theorem

For (2), the $\lambda = 0$ case is trivial. We consider $\lambda \neq 0$. Then $x \in M \iff -\lambda x \in M$, and there exists α such that

$$
\begin{aligned}
\|f_1\| \leq \|f\| &\iff \sup_{\lambda \in \mathbb{R}/\{0\}, x \in M} \left\{ \left| f_1\left(\frac{x + \lambda x_0}{\|x + \lambda x_0\|} \right) \right| \right\} \leq \|f\| \\
&\iff \sup_{\lambda \in \mathbb{R}/\{0\}, x \in M} \left\{ \frac{|f(x) + \lambda \alpha|}{\|x + \lambda x_0\|} \right\} \leq \|f\| \\
&\iff \sup_{\lambda \in \mathbb{R}/\{0\}, x \in M} \left\{ \frac{|f(-\lambda x) + \lambda \alpha|}{\|-\lambda x + \lambda x_0\|} \right\} \leq \|f\| \\
&\iff \forall x \in M,\ |f(x) - \alpha| \leq \|f\| \|x - x_0\| \\
&\iff \forall x \in M,\ f(x) - \|f\| \|x - x_0\| \leq \alpha \leq f(x) + \|f\| \|x - x_0\| \\
&\Longleftarrow \forall x, y \in M,\ f(x) - \|f\| \|x - x_0\| \leq \alpha \leq f(y) + \|f\| \|y - x_0\| \qquad (23)
\end{aligned}
$$

Now we show that there exists α such that Eq (23) holds. Since for any $x, y \in M$,

$$
\begin{aligned}
& f(x) - f(y) = f(x - y) \leq |f(x - y)| \leq \|f\| \|x - y\| \leq \|f\| (\|x - x_0\| + \|y - x_0\|) \\
\iff & f(x) - \|f\| \|x - x_0\| \leq f(y) + \|f\| \|y - x_0\|
\end{aligned}
$$

Eq (23) must hold, thus $\|f_1\| \leq \|f\|$. $\square$

Hahn-Banach Theorem

Hence we show that the f on M can be extended to f_1 on M'. We would like to show that by induction, f can be extended to F on X.

For any M' with $M \subset M' \subset X$ and any f' on M', we construct a pair (M', f'). Let $\mathcal{P}$ be the collection of all such pairs. We define the partial order $<$ for the elements in $\mathcal{P}$: $(M', f') < (M'', f'')$ if $M' \subset M''$, $\|f'\| = \|f''\|$ and $f''(x) = f'(x)$ for all $x \in M'$ [43]. $\mathcal{P}$ is non-empty since it at least contains (M, f), and by the Hausdorff Maximality Theorem (Theorem B.3), there exists a maximal element $(\tilde{M}, F) \in \mathcal{P}$, i.e., there is no $(\tilde{M}_1, F_1) \in \mathcal{P}$ satisfying $\tilde{M} \subset \tilde{M}_1, \|F_1\| = \|F\|$ and $F_1(x) = F(x)$ for any $x \in \tilde{M}$.

Let Ω be a subchain of $\mathcal{P}$ that contains $(\tilde{M}, F)$, then we can write

$$\tilde{M} = \bigcup_{(M', f') \in \Omega} M' \qquad F = \bigcup_{(M', f') \in \Omega} f'$$

It is clear that $\tilde{M}$ is a subspace of X [44].

[43] Here we do not specify how (M'', f'') is obtained from (M', f'). This is a non-constructive process.

[44] If A and B are two subspaces, $A \cup B$ may not be a subspace unless $A \subset B$ or $B \subset A$.

F is well-defined on $\tilde{M}$, because for any $x \in \tilde{M}$, there must exist a pair $(M', f') \in \Omega$ such that $x \in M'$, and $F(x) = f'(x)$ in this case.

For any $x, y \in \tilde{M}$, suppose there are two pairs $(M_x, f_x), (M_y, f_y) \in \Omega$ and $x \in M_x, y \in M_y$, then we have either $M_x \subset M_y$ or $M_y \subset M_x$. Suppose $M_x \subset M_y$, then for any $\beta_1, \beta_2 \in \mathbb{R}$, $\beta_1 x + \beta_2 y \in M_y$ and

$$F(\beta_1 x + \beta_2 y) = f_y(\beta_1 x + \beta_2 y) = \beta_1 f_y(x) + \beta_2 f_y(y) = \beta_1 F(x) + \beta_2 F(y)$$

This shows F is a linear functional on $\tilde{M}$ with $\|F\| = \|f\|$ and $F(x) = f(x)$ for all $x \in M$.

Clearly $\tilde{M} \subset X$. If $\tilde{M} \subset X$, then there exists $x_0 \in X, x_0 \notin \tilde{M}$. Let $\tilde{M}_1 = \{x + \lambda x_0 \,:\, x \in \tilde{M}, \lambda \in \mathbb{R}\}$, then $\tilde{M} \subset \tilde{M}_1$, and by Lemma 4.8, there exists a linear functional F_1 on $\tilde{M}_1$ satisfying $\|F_1\| = \|F\|$ and $F_1(x) = F(x)$ for any $x \in \tilde{M}$. Consequently, there will be a pair $(\tilde{M}_1, F_1) \in \mathcal{P}$ satisfying $(\tilde{M}, F) < (\tilde{M}_1, F_1)$, which is contradict. Thus $\tilde{M} = X$.

$\square$

Hahn-Banach Theorem

The main idea of the proof is as follows: We define an induction by partial order from a start point, and show that there exists an end point for the induction to stop (stated by Hausdorff Maximal Theorem). If the end point is not X, then starting from the end point, we can show by construction that the induction will continue, which causes contradiction [45].

Theorem 4.9: Let M be a subspace of a normed space X, $\overline{M}$ be the closure of M, and $x_0 \in X$. Then $x_0 \in \overline{M}$ if and only if there is no bounded linear functional Λ such that $\Lambda x = 0$ for all $x \in M$ but $\Lambda x_0 \neq 0$.

Proof: $\Longrightarrow$: Suppose $x_0 \in \overline{M}$ and Λ is a bounded linear functional satisfying $\Lambda x = 0$ for all $x \in M$. A bounded Λ is continuous (Theorem 4.6), thus for any $\epsilon > 0$, there exists $\delta > 0$ such that for any x satisfying $\|x_0 - x\| < \delta$, we have $\|\Lambda x_0 - \Lambda x\| < \epsilon$. Also, for any $\delta > 0$ there exists $x \in M$ such that $\|x_0 - x\| < \delta$. Thus for any $\epsilon > 0$, we can always find an $x \in M$ to make $\|\Lambda x_0 - \Lambda x\| = \|\Lambda x_0\| < \epsilon$. Thus $\|\Lambda x_0\| = 0$.

---

[45] See also Transfinite Induction
https://mathworld.wolfram.com/TransfiniteInduction.html

Hahn-Banach Theorem

$\Longleftarrow$: We show that if $x_0 \notin \overline{M}$, then there exists a Λ such that $\Lambda x = 0$ for all $x \in M$ but $\Lambda x_0 \neq 0$. Let M' be the linear subspace generated by M and x_0, i.e., $M' = \{x + \lambda x_0 : x \in M, \lambda \in \mathbb{R}\}$. Define $\Lambda(x + \lambda x_0) = \lambda$ if $x \in M$ and $\lambda \in \mathbb{R}$, then it is easy to verify that Λ is a linear functional satisfying $\Lambda x = 0$ for all $x \in M$ and $\Lambda x_0 = 1$.

Since x_0 is not a limit point of M, there exists $\delta > 0$ such that $\|x_0 - x\| > \delta$ for all $x \in M$. $x \in M$ implies $-\lambda x \in M$. Thus

$$\|x + \lambda x_0\| = |\lambda| \|\frac{1}{\lambda} x + x_0\| = |\lambda| \|\frac{1}{\lambda}(-\lambda x) + x_0\| > \delta |\lambda|$$

Therefore,

$$\|\Lambda\| = \sup\{\frac{|\lambda|}{\|x + \lambda x_0\|}\} < \frac{1}{\delta}$$

which implies that Λ is bounded. $\square$

The following Theorem shows that a bounded linear functional always exists in a normed space.

Theorem 4.10: If X is a normed space and if $x_0 \in X$, $x_0 \neq 0$, then there is a bounded linear functional Λ on X with $\|\Lambda\| = 1$ and $\Lambda x_0 = \|x_0\|$.

Proof: Take $M = \{\lambda x_0\}_{\lambda \in \mathbb{R}}$, then M is a subspace of X. Define Λ on M by $\Lambda(\lambda x_0) = \lambda \|x_0\|$, then $\|\Lambda\| = 1$. By Hahn-Banach Theorem, Λ can be extended to X with its norm unchanged. $\square$

Theorem 4.10 implies that, if $x \in X$ and $\|x\|$ is finite, then $|\Lambda x|$ is finite, because $|\Lambda x| \leq \|\Lambda\| \|x\|$.

Riesz Representation Theorem

The Riesz Representation Theorem, also called Riesz–Markov–Kakutani representation theorem [46], states that any positive linear functional for the functions in $C_c(X)$ can be expressed as a Lebesgue integral. The proof of Riesz Representation Theorem is mainly referred from [17] and [18], which relates to Uryshon's Lemma, Caratheodory's Extension Theorem and Radon measure.

Lemma 4.11: Let X be a normal space and A, B be two disjoint closed subsets in X. Define Δ as the set of all rational numbers in $(0,1)$ that can be expressed in the form $k2^{-n}$, i.e., $\Delta = \{k2^{-n} : 0 < k < 2^n, k, n \in \mathbb{N}^+\}$. Then there exist a family of open sets $\{U_r : r \in \Delta\}$ such that $A \subset U_r \subset B^c$ for all $r \in \Delta$ and $\overline{U_r} \subset U_s$ for $r < s$.

Proof: By nomality, there exist disjoint open sets $V, W \subset X$ such that $A \subset V$ and $B \subset W$. $V \cap W = \varnothing$ implies $V \subset W^c$. Since W^c is closed, and any limit point of V is also a limit point of W^c, we have $V \subset \overline{V} \subset W^c$. Take $\overline{U}_0 = A$, $U_1 = B^c$ and $U^{1/2} = V$, we have $A = \overline{U}_0 \subset U_{1/2} \subset \overline{U}_{1/2} \subset U_1 = B^c$. This process can be conducted recursively: For any $\overline{U}_{k2^{-n}}$ and $U_{(k+1)2^{-n}}$ where $0 < k < 2^n$, there exists an open set $U_{(2k+1)2^{-(n+1)}}$ such that $\overline{U}_{k2^{-n}} \subset U_{(2k+1)2^{-(n+1)}} \subset \overline{U}_{(2k+1)2^{-(n+1)}} \subset U_{(k+1)2^{-n}}$. □

[46] https://en.wikipedia.org/wiki/Riesz-Markov-Kakutani_representation_theorem

We denote $C(X,[0,1])$ as the set of all continuous functions mapped from the topological space X to $[0,1]$, and denote $C_c(X,[0,1])$ as all such functions whose supports are compact in X. Clearly $C_c(X,[0,1])$ is a subset of $C_c(X)$.

Lemma 4.12 (Uryshon's): Let X be a normal space. If A and B are disjoint closed subsets on X, then there exists $f \in C(X,[0,1])$ such that $f=0$ on A and $f=1$ on B.

Proof: Let $\{U_r\}$ be the family of open sets defined in Lemma 4.11 such that $A \subset U_r \subset B^c$ for any $r \in \Delta$ (Note that $0 < r < 1$). Redefine $U_1 = X$ (Remember that X is an open set, and U_1 is not in $\{U_r\}$), then $A \subset U_r \subset B^c \subset U_1 = X$. Define $f(x) = \inf\{r : x \in U_r\}$. We will show such f is a desired function.

Since $A \subset U_{2^{-n}}$ holds for any $n \in \mathbb{N}^+$, for any $x \in A$, $f(x) = \inf\{2^{-n}\} = 0$. Also, $x \in B$ implies $x \in U_1$ and $x \notin B^c$, thus for any $r < 1$, $x \notin U_r$. We must have $f(x) = 1$ on B.

Riesz Representation Theorem

It remains to show that f is a continuous. Observe that $f(x) < \alpha$ if and only if $x \in \bigcup_{r<\alpha} U_r$. $\Longleftarrow$: $x \in \bigcup_{r<\alpha} U_r$ implies there exists some r with $r < \alpha$ such that $x \in U_r$, hence $f(x) = \inf\{r\} \le r < \alpha$. $\Longrightarrow$: There exists $r^* \in \Delta$ such that $f(x) < r^* < \alpha$ (Because the gap $\alpha - f(x)$ is finite and the gap of neighborhood elements in Δ can be arbitrary small), thus $x \in U_{r^*} \subset \bigcup_{r<\alpha} U_r$. Therefore, $f^{-1}((-\infty, \alpha)) = \bigcup_{r<\alpha} U_r$.

Also, $f(x) > \alpha$ if and only if $x \in \bigcup_{s>\alpha} \overline{U}_s^c$. $\Longleftarrow$: $x \in \bigcup_{s>\alpha} \overline{U}_s^c$ implies there exists some s with $s > \alpha$ such that $x \in \overline{U}_s^c$, that is, $x \notin U_s$. So it is impossible for $f(x) < s$, we have $f(x) \ge s > \alpha$. $\Longrightarrow$: There exist $r^*, s^* \in \Delta$ with $\alpha < s^* < r^* < f(x)$, thus $\overline{U}_{s^*} \subset U_{r^*}$. $x \notin U_{r^*}$ implies $x \in \overline{U}_{s^*}^c \subset \bigcup_{s>\alpha} \overline{U}_s^c$. Therefore, $f^{-1}((\alpha, \infty)) = \bigcup_{r<\alpha} \overline{U}_s^c$.

Since both $f^{-1}((-\infty, \alpha))$ and $f^{-1}((\alpha, \infty))$ are open, let $\alpha < \beta$, then $f^{-1}(\alpha, \beta) = f^{-1}((\alpha, \infty)) \cap f^{-1}((-\infty, \beta))$ is open. Thus the inverse image of any interval on $\mathbb{R}$ is open. Let V be an arbitrary open set on $[0, 1]$, then V can be written as countable union of open intervals (Lemma C.6), which means $f^{-1}(V)$ is open. By definition, f is continuous. $\square$

Lemma 4.13 (Uryshon's, Locally Compact Hausdorff version): Let X be a locally compact Hausdorff space, $K \subset U \subset X$ where K is compact and U is open, then there exists a function $f \in C(X, [0,1])$ such that $f = 1$ on K and $f = 0$ outside a compact subset of U.

Proof: By Corollary 1.41, there exists open set V such that $K \subset V \subset \overline{V} \subset U$ where $\overline{V}$ is compact. Note that $\partial V \subset \overline{V}$ and ∂V is closed, so ∂V is compact. Since V is open, $V \cap \partial V = \varnothing$, thus $K \subset V$ implies $K \cap \partial V = \varnothing$. By Lemma 1.39, there exist disjoint open sets $W_1, W_2 \in X$ such that $K \subset W_1$ and $\partial V \subset W_2$, so the nomality holds for K and ∂V. By Lemma 4.12, there exists $f \in C(X, [0,1])$ such that $f = 1$ on K and $f = 0$ on ∂V.

Let $\overline{V}$ be the desired compact subset in U. We choose an arbitrary $f \in C(X, [0,1])$ satisfying $f = 1$ on K and $f = 0$ on ∂V, and set $f(x) = 0$ for all $x \in \overline{V}^c$, and show that this f is still in $C(X, [0,1])$. It is obvious that the range of f is in $[0,1]$, so we only need to show that f is continuous.

Note that f is still a continuous function on $\overline{V}$. Let S be an open set on $\mathbb{R}$. If S does not contain 0, then $f^{-1}(S) \subset \overline{V}$, thus $f^{-1}(S)$ is an open subset due to the continuity of f. If S contains 0, then S^c is a closed set that does not contain 0, we have $f^{-1}(S^c) \subset \overline{V}$, and by the continuity of f, $f^{-1}(S^c)$ is closed (Corollary 1.21). This implies $f^{-1}(S^c)^c = f^{-1}(S)$ is open. Thus we have shown that the continuity of f on $\overline{V}$ can be extended to X. □

It is obvious that the function f in Lemma 4.13 satisfies $K \subset$ supp$(f) \subset \overline{V}$. Note that Lemma 4.13 holds for the case that there exist $E \subset V - K$ such that $f(E) = \{0\}$. This means supp(f) is not necessary equivalent to $\overline{V}$.

Notation: Let X be a locally compact Hausdorff space, U be a open set in X, K be a compact set in X, and f be a function in $C_c(X, [0,1])$. If f satisfies supp$(f) \subset U$, we use the notation $f \prec U$. If f satisfies $f = 1$ on K, we use the notation $K \prec f$. By Lemma 4.13, for any $K \subset U \subset X$, we can always find an f satisfying $K \prec f \prec U$. This can be carried out by choosing a function in $C(X, [0,1])$ with $f = 1$ on K and supp$(f) = \overline{V}$.

Riesz Representation Theorem

Lemma 4.14: Let X be an locally compact Hausdorff space, $K \subset X$ be compact, $\{V_i\}_{i=1}^n$ be an open cover of K. Then there exists functions $h_1, h_2, ..., h_n$ on X such that $h_i \prec V_i$ for $i = 1, 2, ..., n$, and $\sum_{i=1}^n h_i(x) = 1$ for any $x \in K$.

(We say the set $\{h_i\}_{i=1}^n$ is a partition of unity on K, subordinate to the cover $\{V_i\}_{i=1}^n$.)

Proof: Since each $x \in K$ is contained in at least one open set V_i for some i, by Theorem 1.40, there exists a compact neighborhood N_x such that $x \in N_x \subset V_i$. Since $\{N_x^o\}_{x \in K}$ is an open cover of K, there exists a subcover $\{N_{x_j}^o\}_{j=1}^m$ of K. Hence $K \subset \bigcup_{j=1}^m N_{x_j}$. Since each N_{x_j} must be in at least one open cover V_i for some i, let F_i be the union of the N_{x_j}'s that are in V_i, then F_i is a compact subset of V_i. By Uryshon's Lemma (Lemma 4.13), there exists $g_1, g_2, ..., g_n$ on X such that $F_i \prec g_i \prec V_i$ for all i. Thus $\sum_{i=1}^n g_i \geq 1$ on $\bigcup_{i=1}^n F_i$. Since $K \subset \bigcup_{i=1}^n F_i$, we have $\sum_{i=1}^n g_i(x) \geq 1$ for any $x \in K$. Since $K \subset \bigcup_{j=1}^m N_{x_j}^o \subset \bigcup_{i=1}^n F_i$, again by Uryshon's Lemma, there exists f such that $K \prec f \prec \bigcup_{j=1}^m N_{x_j}^o$. Let $g_{n+1} = 1 - f$, then $g_{n+1} = 0$ on K and $g_{n+1} = 1$ on $(\bigcup_{j=1}^m N_{x_j}^o)^c$. Since $(\bigcup_{j=1}^m N_{x_j}^o)^c \supset (\bigcup_{i=1}^n F_i)^c$, $\sum_{i=1}^{n+1} g_i \geq 1$ for all $x \in X$, and $\sum_{i=1}^{n+1} g_i = \sum_{i=1}^n g_i$ for all $x \in K$.

Let $h_i = \frac{g_i}{\sum_{j=1}^{n+1} g_j}$ for $j = 1, 2, ..., n$, then each h_i is a continuous function with range in $[0, 1]$, $\sum_{i=1}^{n} h_i(x) = 1$ for all $x \in K$, and $\text{supp}(h_i) = \text{supp}(g_i) \subset V_i$. $\square$

Definition 4.15: Let X be a locally compact Hausdorff space, μ be a Borel measure on X, and E be a Borel subset of X.

μ is called outer regular on E if

$$\mu(E) = \inf\{\mu(U) : E \subset U, U \text{ open in } X\}$$

and inner regular on E if

$$\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compact in } X\}$$

μ is called a Radon measure if it is finite on all compact sets [47], outer regular on all Borel sets, and inner regular on all open sets.

[47] Remember that by Corollary 1.38, a compact set in a Hausdorff space is closed, so it is a Borel set.

Motivations of Radon measure: Radon measure is a Borel measure defined on all Borel sets, but it satisfies the regularities: The outer regularity states that the measure of any Borel set can be approximated from above by the measure of open sets (If the Borel set is open, no approximation is needed; if the Borel set is compact, it can be covered by finite open sets, so it can be approximated by open sets from above), and the inner regularity states that the measure of any open set can be approximated from below by the measure of compact sets (Remember that in a locally compact Hausdorff space, every non-empty open set has a compact subset). Therefore, the regularity guarantees that the measure is continuous and interacts nicely with the underlying topology [48].

For example, Lebesgue measure on Euclidean space $\mathbb{R}^d$ is a Radon measure. The proof can be found in Section 2.7 of [49].

[48] https://sites.math.washington.edu/~farbod/teaching/cornell/math6210pdf/math6210Radon.pdf

[49] https://www.math.ucdavis.edu/~hunter/m206/ch1_measure.pdf

Theorem 4.16 (Riesz Representation): Let X be a locally compact Hausdorff space, and Λ be a bounded positive linear functional on $C_c(X)$, then there is a unique Radon measure μ on X such that $\Lambda f = \int f \, d\mu$ for all $f \in C_c(X)$. Moreover, μ satisfies

$$\mu(U) = \sup\{\Lambda f \, : \, f \prec U\} \quad \text{for all open } U \subset X \tag{24}$$

and

$$\mu(K) = \inf\{\Lambda f \, : \, K \prec f\} \quad \text{for all compact } K \subset X \tag{25}$$

Proof: $f \in C_c(X)$ implies supp(f) is compact, thus $\|f\|_\infty$ is finite. Also, any $f \in C_c(X)$ can be written as $f = f^+ - f^- = \|f\|_\infty \left(\frac{f^+}{\|f\|_\infty}\right) - \|f\|_\infty \left(\frac{f^-}{\|f\|_\infty}\right)$ where both $\frac{f^+}{\|f\|_\infty}$ and $\frac{f^-}{\|f\|_\infty}$ are in $C_c(X,[0,1])$ (Note that both supp(f^+) and supp(f^-) are closed subsets of supp(f), so they are compact). Hence, if $\Lambda f = \int f \, d\mu$ holds for all $f \in C_c(X,[0,1])$, then it holds for all $f \in C_c(X)$.

We first prove the uniqueness. Suppose μ is a Radon measure such that $\Lambda f = \int f \, d\mu$ for all $f \in C_c(X,[0,1])$. Since supp(f) is compact, it must be covered by an open set. Let $U \subset X$ be an open set. For any $f \prec U$, we have $f \leq \chi_U$, thus

Riesz Representation Theorem

$$\Lambda f = \int f \, d\mu \leq \int \chi_U \, d\mu = \mu(U) = \sup\{\mu(K) \,:\, K \subset U, \, K \text{ compact in } X\}$$

The last equality is by the inner regularity of μ on open sets. Thus,

$$\sup\{\Lambda f \,:\, f \prec U\} \leq \sup\{\mu(K) \,:\, K \subset U, \, K \text{ compact in } X\} \tag{26}$$

Also, for any compact $K \subset U$, by Uryshon's Lemma (Lemma 4.13), there exists f such that $K \prec f \prec U$. This f satisfies $f \geq \chi_K$ on U, thus

$$\mu(K) = \int \chi_K \, d\mu \leq \int f \, d\mu = \Lambda f \leq \sup\{\Lambda f \,:\, f \prec U\}$$

which implies

$$\sup\{\mu(K) \,:\, K \subset U, \, K \text{ compact in } X\} \leq \sup\{\Lambda f \,:\, f \prec U\} \tag{27}$$

By Eq (26) and Eq (27), we have

$$\sup\{\Lambda f \,:\, f \prec U\} = \sup\{\mu(K) \,:\, K \subset U, \, K \text{ compact in } X\} = \mu(U)$$

This implies Eq (24) holds, and the measure μ on any open set U is uniquely determined by Λ. The outer regularity of μ states that, for any Borel set E, $\mu(E) = \inf\{\mu(U) \,:\, E \subset U, \, U \text{ open in } X\}$. This implies μ on any Borel set is uniquely determined by μ on open sets. Thus Λ uniquely determines μ on all Borel sets.

We will divide the proof into steps and outline them as follows.

Let us begin by defining a set function

$$\mu(U) = \sup\{\Lambda f \,:\, f \prec U\} \tag{28}$$

for any open set U. μ is not necessary a measure since its countably additivity is unknown, also μ is undefined on closed sets. We are going to show that μ can be extended to a Radon measure.

Remember that the Borel σ-algebra $\mathcal{B}$ includes all open sets. We show that μ can be extended to a Borel measure using Caratheodory's Extension Theorem: First, show that μ can be extended to a measure μ^* on a σ-algebra $\mathcal{R}$. Second, show that $\mathcal{B} \subset \mathcal{R}$. Hence, μ^* is a Borel measure on $\mathcal{B}$, and $\mu = \mu^*$ on all open sets.

For any $E \subset X$, define

$$\mu^*(E) = \inf\{\mu(U) \,:\, E \subset U, U \text{ open in } X\} \tag{29}$$

Clearly $\mu^*(U) = \mu(U)$ for any open set U.

Step 1: μ^* is an outer measure.

Then by Caratheodory's Extension Theorem (Theorem 3.27), there exists a σ-algebra $\mathcal{R}$ such that μ^* is a measure on $\mathcal{R}$.

Step 2: Every open set is in $\mathcal{R}$.

Since $\mathcal{R}$ is closed under complement and countable union, the Borel set $\mathcal{B}$ is generated by open sets under complement and countable union, we have $\mathcal{B} \subset \mathcal{R}$. Thus μ^* is a Borel measure. We replace the symbol μ^* with μ for simplicity.

Thus, we can rewritten Eq (29) as $\mu(E) = \inf\{\mu(U) : E \subset U, U \text{ open in } X\}$ for any Borel set E, thus the outer regularity of μ holds.

Every compact set in a Hausdorff space is a Borel set because it is closed (Corollary 1.38), so μ can be defined on compact sets.

Step 3: μ satisfies Eq (25).

We attach the infinite norm $\|\cdot\|_\infty$ to $C_c(X)$ to make it a normed space, and define $\|\Lambda\|_\infty = \sup\{\Lambda f : f \in C_c(X), \|f\|_\infty = 1\}$. Eq (25) implies that there exists some $f \in C_c(X, [0,1])$ such that $\mu(K) < \Lambda f \leq \|\Lambda\|_\infty \|f\|_\infty \leq \|\Lambda\|_\infty$. Since we assume Λ is bounded, $\mu(K)$ is finite. This shows μ is finite on compact sets.

Riesz Representation Theorem

Let U be an arbitrary open set. By Eq (28), for any $\epsilon > 0$, there exists $f \prec U$ such that $\Lambda f > \mu(U) - \epsilon$. Let $K = \mathrm{supp}(f)$, then K is compact and $K \subset U$. By Uryshon's Lemma, there exists g such that $K \prec g \prec U$. Since $f \leq \chi_K \leq g$, $\Lambda g - \Lambda f = \Lambda(g - f) \geq 0$, thus $\Lambda f \leq \Lambda g$ for any $K \prec g \prec U$. By Eq (25), $\mu(K) = \inf\{\Lambda g : K \prec g \prec U\}$, thus $\Lambda f \leq \mu(K)$. Since $K \subset U \implies \mu(K) \leq \mu(U)$, we have $\mu(U) - \epsilon \leq \Lambda f \leq \mu(K) \leq \mu(U)$. This implies $\mu(U) = \sup\{\mu(K) : K \subset U, K \text{ compact in } X\}$. So the inner regularity of μ holds. So far we have shown that μ is a Radon measure.

Step 4: $\Lambda f = \int f \, d\mu$ for all $f \in C_c(X)$.

This completes the proof. The uniqueness of μ follows from here, as we mentioned at the beginning of the proof.

Proof of Step 1: Λ satisfies $\Lambda 0 = 0$, because for any f, $\Lambda 0 = \Lambda(0f) = 0\Lambda f$ where Λf is finite. Thus in Eq (28), $\mu(\varnothing) = \sup\{\Lambda f : f \prec \varnothing\} = \Lambda 0 = 0$.

μ is nonnegative because Λf is nonnegative. Let $\mathcal{E}$ be the family of all open sets of X and μ be the nonnegative set function on $\mathcal{E}$. Clearly we have $\varnothing, X \in \mathcal{E}$. Let $\{U_i\}_{i=1}^{\infty}$ be a subset of $\mathcal{E}$ such that $U = \bigcup_{i=1}^{\infty} U_i$. Such subset always exists as we can take $U_1 = U$ and $U_i = \varnothing$ for $i \geq 2$.

Riesz Representation Theorem

If we can show $\mu(U) \leq \sum_{i=1}^{\infty} \mu(U_i)$, then for any $E \subset X$, Eq (29) has

$$\mu^*(E) = \inf\{\mu(U) : E \subset U, U \text{ open in } X\}$$
$$\leq \inf\{\sum_{i=1}^{\infty} \mu(U_i) : E \subset \bigcup_{i=1}^{\infty} U_i, U_i \subset \mathcal{E} \text{ for all } i\}$$

The $\leq$ above can be replaced with $=$ because we can take $U_1 = U$ and $U_i = \varnothing$ for $i \geq 2$. Thus by Theorem 3.28, μ^* is an outer measure.

Now we show $\mu(U) \leq \sum_{i=1}^{\infty} \mu(U_i)$. If $f \prec U$, let $K = \text{supp}(f)$, then K is compact. Since $K \subset U = \bigcup_{i=1}^{\infty} U_i$, $\{U_i\}_{i=1}^{\infty}$ is an open cover of K, thus there exists finite n such that $K \subset \bigcup_{i=1}^{n} U_i$. By Lemma 4.14, there exist $h_1, h_2, ..., h_n$ on X such that $h_i \prec U_i$ and $\sum_{i=1}^{n} h_i = 1$ on K. Then $f = f(\sum_{i=1}^{n} h_i) = \sum_{i=1}^{n} fh_i$ (Note that $f = 0$ on K^c). Also, $0 \leq fh_i \leq 1$ and $\text{supp}(fh_i) = \text{supp}(f) \cap \text{supp}(h_i) \subset \text{supp}(h_i) \subset U_i$ imply that $fh_i \prec U_i$. Hence

$$\Lambda f = \sum_{i=1}^{n} \Lambda(fh_i) \leq \sum_{i=1}^{n} \sup\{\Lambda(fh_i) : fh_i \prec U_i\} = \sum_{i=1}^{n} \mu(U_i)$$

which holds for any $f \prec U$. Thus $\mu(U) = \sup\{\Lambda f : f \prec U\} \leq \sum_{i=1}^{n} \mu(U_i)$.

Proof of Step 2: By Caratheodory's Extension Theorem (Theorem 3.27), we need to show that for any open set U, $\mu^*(E) = \mu^*(E \cap U) + \mu^*(E \cap U^c)$. Since the subadditivity implies $\mu^*(E) \le \mu^*(E \cap U) + \mu^*(E \cap U^c)$ for any U, we only need to show for all $E \subset X$ with $\mu^*(E) < \infty$, $\mu^*(E) \ge \mu^*(E \cap U) + \mu^*(E \cap U^c)$.

First we assume E is open, then $E \cap U$ is open. Remember that $\mu^* = \mu$ on all open sets. Thus for any $\epsilon > 0$, there exists $f \prec E \cap U$ such that $\Lambda f > \mu^*(E \cap U) - \epsilon$. Since $\mathrm{supp}(f)$ is a compact subset in E, $E - \mathrm{supp}(f)$ is open, thus there is $g \prec E - \mathrm{supp}(f)$ such that $\Lambda g > \mu^*(E \cap (E - \mathrm{supp}(f))) - \epsilon$. Since $\mathrm{supp}(f) \cap \mathrm{supp}(g) = \varnothing$, we have $0 \le f + g \le 1$. Also, $\mathrm{supp}(f+g) = \mathrm{supp}(f) \cup \mathrm{supp}(g) \subset E$, thus $f + g \prec E$. Therefore,

$$\begin{aligned} \mu^*(E) &\ge \Lambda(f+g) = \Lambda f + \Lambda g > \mu^*(E \cap U) + \mu^*(E \cap (E - \mathrm{supp}(f))) - 2\epsilon \\ &\ge \mu^*(E \cap U) + \mu^*(E \cap U^c) - 2\epsilon \end{aligned}$$

where the last inequality is because $E - \mathrm{supp}(f) = E \cap \mathrm{supp}(f)^c \supset E \cap U^c$. Thus by Corollary 3.24, $\mu^*(E) \ge \mu^*(E \cap U) + \mu^*(E \cap U^c)$.

Now we consider the general case where E is an arbitrary set of X. By Eq (29), for any $\epsilon > 0$, there exists open set V with $E \subset V$ such that $\mu(V) \le \mu^*(E) + \epsilon$. Hence

$$\mu^*(E) + \epsilon \geq \mu(V) = \mu^*(V) \geq \mu^*(V \cap U) + \mu^*(V \cap U^c)$$
$$\geq \mu^*(E \cap U) + \mu^*(E \cap U^c)$$

Thus $\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \cap U^c)$ holds.

Proof of Step 3: Let K be a compact set. For any f such that $K \prec f$, given an arbitrary $\epsilon \in (0, 1)$, we have

$$K \subset \{x : f(x) = 1\} \subset \{x : f(x) > 1 - \epsilon\} \subset \{x : f(x) > 0\} \subset \text{supp}(f)$$

Let $U_\epsilon = \{x : f(x) > 1 - \epsilon\}$, then U_ϵ is open, since f is continuous and $U_\epsilon = f^{-1}((1 - \epsilon, \infty))$. By Uryshon's Lemma, there exists g such that $K \prec g \prec U_\epsilon$. Since $(1 - \epsilon)^{-1} f > 1$ on U_ϵ and $g \leq 1$ on U_ϵ, we have $(1 - \epsilon)^{-1} f \geq g \implies (1 - \epsilon)^{-1} \Lambda f \geq \Lambda g$ on U_ϵ. Since $\mu(U_\epsilon) = \sup\{\Lambda g : g \prec U_\epsilon\}$, we have $\mu(K) \leq \mu(U_\epsilon) \leq (1 - \epsilon)^{-1} \Lambda f$. This holds for any $\epsilon > 0$, thus $\mu(K) \leq \Lambda f$ [50]. This implies $\mu(K) \leq \inf\{\Lambda f : K \prec f\}$.

[50] For any positive constants a and b, $a \leq b$ if and only if $a \leq \lambda b$ for any $\lambda > 1$. This can be proven in a similar way as Theorem 3.21: Assume both $a > b$ and $a \leq \lambda b$ for any $\lambda > 1$ are true, take $\lambda = \sqrt{a/b}$, then we get $a \leq b$, which is contradict.

On the other hand, the outer regularity of μ states that $\mu(K) = \inf\{\mu(U) : K \subset U, U \text{ open in } X\}$. This implies for any $\epsilon > 0$, there exists open U such that $K \subset U$ and $\mu(U) < \mu(K) + \epsilon$. For any open U such that $K \subset U$, by Uryshon's Lemma, there exists f such that $K \prec f \prec U$. So $\inf\{\Lambda f : K \prec f\} \leq \inf\{\Lambda f : K \prec f \prec U\} \leq \Lambda f \leq \mu(U)$. Since $\inf\{\Lambda f : K \prec f\} < \mu(K) + \epsilon$ holds for any $\epsilon > 0$, we have $\inf\{\Lambda f : K \prec f\} \leq \mu(K)$.

Therefore, $\mu(K) = \inf\{\Lambda f : K \prec f\}$.

Proof of Step 4: As we mentioned, it suffices to show that $\Lambda f = \int f \, d\mu$ for all $f \in C_c(X, [0,1])$.

For any $f \in C_c(X, [0,1])$, choose $N \in \mathbb{N}^+$, and set $K_i = \{x : f(x) \geq \frac{i}{N}\}$ for $i \in \{1, 2, ..., N\}$ and $K_0 = \text{supp}(f)$, then $K_N \subset K_{N-1} \subset ... \subset K_1 \subset K_0$. K_0 is compact, and each K_i is closed since $K_i^c = f^{-1}((-\infty, \frac{i}{N}))$ is open, so each K_i is compact.

Riesz Representation Theorem

Consider a sequence of functions $f_1, f_2, ..., f_N$. We define f_i as follows: $f_i(x) = 0$ if $x \in K_{i-1}^c$, $f_i(x) = f(x) - \frac{i-1}{N}$ if $x \in K_{i-1} - K_i$, and $f_i(x) = \frac{1}{N}$ if $x \in K_i$. In other words,

$$f_i = \min\left\{\max\left\{f - \frac{i-1}{N}, 0\right\}, \frac{1}{N}\right\}$$

So each f_i is a continuous function. Also, since supp$(f_i) = $ cl$(\{x : f_i(x) = 0\}) \subset K_{i-1}$, supp$(f_i)$ is compact. Thus $f_i \in C_c(X, [0,1])$ for all i.

Since f_i satisfies $\chi_{K_i} \le N f_i \le \chi_{K_{i-1}}$, we have

$$\frac{1}{N}\mu(K_i) = \frac{1}{N}\int \chi_{K_i}\, d\mu \le \int f_i\, d\mu \le \frac{1}{N}\int \chi_{K_{i-1}}\, d\mu = \frac{1}{N}\mu(K_{i-1}) \tag{30}$$

Since $K_i \prec N f_i$, by Eq (25), $\mu(K_i) \le \Lambda(N f_i)$. Also, for any open U such that $K_{i-1} \subset U$, we have $N f_i \prec U$, thus $\Lambda(N f_i) \le \mu(U)$. By the outer regularity of μ, for any $\epsilon > 0$, there exists open U such that $K_{i-1} \subset U$ and $\mu(U) < \mu(K_{i-1}) + \epsilon$. Thus $\Lambda(N f_i) \le \mu(K_{i-1}) + \epsilon$ holds for any $\epsilon > 0$, implying that $\Lambda(N f_i) \le \mu(K_{i-1})$. Therefore,

$$\frac{1}{N}\mu(K_i) \le \Lambda f_i \le \frac{1}{N}\mu(K_{i-1}) \tag{31}$$

It is easy to verify that $f = \sum_{i=1}^{N} f_i$, thus Eq (30) and Eq (31) can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\mu(K_i) \leq \int f\ d\mu \leq \frac{1}{N}\sum_{i=1}^{N}\mu(K_{i-1})$$

$$\frac{1}{N}\sum_{i=1}^{N}\mu(K_i) \leq \Lambda f \leq \frac{1}{N}\sum_{i=1}^{N}\mu(K_{i-1})$$

Therefore,

$$\left|\int f\ d\mu - \Lambda f\right| \leq \frac{1}{N}\sum_{i=1}^{N}\mu(K_{i-1}) - \frac{1}{N}\sum_{i=1}^{N}\mu(K_i) = \frac{\mu(K_0) - \mu(K_N)}{N} \leq \frac{\mu(K_0)}{N}$$

which holds for any N. Since $K_0 = \text{supp}(f)$ is compact, $\mu(K_0)$ is finite. Thus by taking $N \to \infty$, we get $\left|\int f\ d\mu - \Lambda f\right| = 0 \Longrightarrow \Lambda f = \int f\ d\mu$.

$\square$

Contents

The Framework of Statistical Learning

The probability theory is built upon the measure theory. A measure space $(X, \mathcal{R}, \mu)$ is called a probability space if μ satisfies $\mu(X) = 1$. Such μ is called a probability measure.

Probability theory uses different terminologies to describe the same mathematical objects in measure theory, as shown in the following table [17].

Measure Theory Terms	Probability Theory Terms
Measure space $(X, \mathcal{R}, \mu)$ with $\mu(X) = 1$	Probability space $(\Omega, \mathcal{B}, P)$
Measurable function f	Random variable X
Integral of f, $\int f \, d\mu$	Expectation or mean of X, $\mathbb{E}[X]$
Borel probability measure on $\mathbb{R}^d$	Distribution

Let (x, y) be the data pair where $x \in \mathbb{R}^d$ is the feature vector and $y \in \mathbb{R}$ is the label. Suppose (x, y) is drawn from a probability space $(\mathbb{R}^{d+1}, \mathcal{B}, P)$ where the distribution P is unknown. Let a real valued function f be a machine learning model. It takes a feature vector x as input and output a prediction $f(x)$, which is supposed to be as close to the true label y as possible.

The Framework of Statistical Learning

The error between $f(x)$ and y is measured by a non-negative function $L(y, f(x))$, called the loss function. Some commonly used loss functions are:

- $0-1$ loss: $L(y, f(x)) = 0$ if $y = f(x)$ and $L(y, f(x)) = 1$ if $y \neq f(x)$.
- Squared loss: $L(y, f(x)) = (y - f(x))^2$.
- Cross entropy loss: If $y \in \{0, 1\}$, then $L(y, f(x)) = -y \log f(x) - (1-y) \log(1 - f(x))$.

The squared loss are typically used in regression problems while the cross entropy loss are typically used in classification problems.

The true risk $R^{\text{true}}(f)$ is defined as the expectation of the loss over the data distribution:

$$R^{\text{true}}(f) = \int L(y, f(x)) \, dP(x, y) \tag{32}$$

By Riesz Representation Theorem, we know that $R^{\text{true}}(f)$ is a linear functional.

Let $f^* = \text{argmin}_f R^{\text{true}}(f)$, then f^* is called the optimal machine learning model. If L is squared loss, then f^* can be expressed as $f^*(x) = \int y \, dP(y|x) = \mathbb{E}_{y|x}[y]$ [21].

In this case,

$$R^{\text{true}}(f^*) = \int (y - \mathbb{E}_{y|x}[y])^2 \, dP(x,y)$$

Given any fixed x, $\mathbb{E}_{y|x}[y]$ is fixed, while y varies with different probabilities. Thus $R^{\text{true}}(f^*) > 0$ typically. This implies that we cannot find a machine learning model that makes zero mistakes in prediction. We call $R^{\text{true}}(f^*)$ the Bayes error.

Moreover, solving f^* by minimizing $R^{\text{true}}(f)$ is an almost impossible task. There are two main obstacles.

First, let X be the set of all possible values of x and Y be the set of all possible values of Y, then the set of all functions mapping from X to Y is denoted as Y^X [51]. It is obvious that $f^* \in Y^X$, however, Y^X is too large, and searching f^* is very difficult. Usually we choose a subspace $\mathcal{H} \subset Y^X$ called hypothesis space. $\mathcal{H}$ can be the space of all polynomials, or the space of all multilayer perceptrons. We solve

$$f^*_{\mathcal{H}} = \operatorname*{argmin}_{f \in \mathcal{H}} R^{\text{true}}(f)$$

such that $f^*_{\mathcal{H}}$ is a sub-optimal model.

[51] https://math.stackexchange.com/questions/4598731/

The Framework of Statistical Learning

It is obvious that $R^{\text{true}}(f^*_{\mathcal{H}}) \geq R^{\text{true}}(f^*)$. We call $R^{\text{true}}(f^*_{\mathcal{H}}) - R^{\text{true}}(f^*)$ the approximation error. If $\mathcal{H}_1 \subset \mathcal{H}_2 \subset Y^X$, then $\inf_{f \in \mathcal{H}_1} R^{\text{true}}(f) \geq \inf_{f \in \mathcal{H}_2} R^{\text{true}}(f) \geq \inf_{f \in Y^X} R^{\text{true}}(f) = R^{\text{true}}(f^*)$. We say the $\mathcal{H}$ that gives smaller approximation error has stronger expressivity.

Second, even $\mathcal{H}$ is given, it is still hard to obtain $f^*_{\mathcal{H}}$. Remember that the expression of $R^{\text{true}}(f)$ contains a unknown distribution P (see Eq (32)). We cannot compute the integral of $R^{\text{true}}(f)$ unless P is given, thus it is not possible to optimize $R^{\text{true}}(f)$ directly.

Suppose we have a dataset $S = \{(x_i, y_i)\}_{i=1}^n$ where each (x_i, y_i) is a random variable i.i.d. sampled from $(\mathbb{R}^{d+1}, \mathcal{B}, P)$. We define the empirical risk $R^{\text{emp}}(f)$ as

$$R^{\text{emp}}(f) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i))$$

$R^{\text{emp}}(f)$ is a random variable depending on S, n and f. Let s be an observation of S, and denote $R^{\text{emp}}_s(f)$ as $R^{\text{emp}}(f)$ conditioned on $S = s$, then $R^{\text{emp}}_s(f)$ depends on f. Let $f^n_{\mathcal{H},s} = \operatorname{argmin}_{f \in \mathcal{H}} R^{\text{emp}}_s(f)$. Define $f^n_{\mathcal{H},S} = \{f^n_{\mathcal{H},s} : s \in \operatorname{supp}(S)\}$ and $R^{\text{emp}}(f^n_{\mathcal{H},S}) = \{R^{\text{emp}}_s(f^n_{\mathcal{H},s}) : s \in \operatorname{supp}(S)\}$, then both $f^n_{\mathcal{H},S}$ and $R^{\text{emp}}(f^n_{\mathcal{H},S})$ are random variables depending on S and n.

Definition 5.1: A sequence of random variables $\{X_n\}$ is said to converge to a random variable X in probability if for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(|X_n - X| > \epsilon\right) = 0$$

If $\{X_n\}$ converges to X in probability, we write $X_n \xrightarrow{p} X$.

Consistency of Empirical Risk Minimization (ERM): We say the ERM is consistent if the sequence of random variables $\{f^n_{\mathcal{H},S}\}_{n=1}^{\infty}$ satisfies

$$R^{\text{true}}(f^n_{\mathcal{H},S}) \xrightarrow{p} R^{\text{true}}(f^*_{\mathcal{H}}) \quad \text{and} \quad R^{\text{emp}}(f^n_{\mathcal{H},S}) \xrightarrow{p} R^{\text{true}}(f^*_{\mathcal{H}})$$

as $n \to \infty$.

The consistency of empirical risk minimization states that, if n is larger, the probability of finding a $f^n_{\mathcal{H},s}$ in $f^n_{\mathcal{H},S}$ that is close to $f^*_{\mathcal{H}}$ will be higher.

Note that $R^{\text{true}}(f^n_{\mathcal{H},S})$ is a random variable depending on S and n. The following theorem shows a sufficient condition that guarantees the consistency of ERM.

Theorem 5.2: If $\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\} \xrightarrow{p} 0$ as $n \to \infty$, then $R^{\text{true}}(f^n_{\mathcal{H},S}) \xrightarrow{p} R^{\text{true}}(f^*_{\mathcal{H}})$ and $R^{\text{emp}}(f^n_{\mathcal{H},S}) \xrightarrow{p} R^{\text{true}}(f^*_{\mathcal{H}})$ as $n \to \infty$.

Proof: Given any n and any observation $S = s$,

$$
\begin{aligned}
&|R^{\text{true}}(f^n_{\mathcal{H},s}) - R^{\text{true}}(f^*_{\mathcal{H}})| = R^{\text{true}}(f^n_{\mathcal{H},s}) - R^{\text{true}}(f^*_{\mathcal{H}}) \\
&= R^{\text{true}}(f^n_{\mathcal{H},s}) - R^{\text{emp}}_s(f^n_{\mathcal{H},s}) + R^{\text{emp}}_s(f^n_{\mathcal{H},s}) - R^{\text{emp}}_s(f^*_{\mathcal{H}}) + R^{\text{emp}}_s(f^*_{\mathcal{H}}) - R^{\text{true}}(f^*_{\mathcal{H}}) \\
&\leq |R^{\text{true}}(f^n_{\mathcal{H},s}) - R^{\text{emp}}_s(f^n_{\mathcal{H},s})| + |R^{\text{emp}}_s(f^*_{\mathcal{H}}) - R^{\text{true}}(f^*_{\mathcal{H}})| \\
&\leq 2 \sup_{f\in\mathcal{H}}\{|R^{\text{emp}}_s(f) - R^{\text{true}}(f)|\}
\end{aligned}
$$

Here we use the fact that $R^{\text{emp}}_s(f^n_{\mathcal{H},s}) - R^{\text{emp}}_s(f^*_{\mathcal{H}}) \leq 0$. This is because $f^n_{\mathcal{H},s}$ is the minimizer of $R^{\text{emp}}_s(f)$ while $f^*_{\mathcal{H}}$ is not. Hence,

$$|R^{\text{true}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}})| = R^{\text{true}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}}) \leq 2 \sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\}$$

Therefore, for any $\epsilon > 0$,

$$P\left(R^{\text{true}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}}) > \epsilon\right) \leq P\left(\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\} > \frac{\epsilon}{2}\right)$$

The Framework of Statistical Learning

Since $\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\} \xrightarrow{p} 0$, taking limits on both sides, we get

$$\lim_{n\to\infty} P\left(R^{\text{true}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}}) > \epsilon\right) \leq \lim_{n\to\infty} P\left(\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\} > \frac{\epsilon}{2}\right) = 0$$

Similarly, for any n and $S = s$,

$$\begin{aligned}|R^{\text{emp}}_s(f^n_{\mathcal{H},s}) - R^{\text{true}}(f^*_{\mathcal{H}})| &\leq |R^{\text{emp}}_s(f^n_{\mathcal{H},s}) - R^{\text{true}}(f^n_{\mathcal{H},s})| + |R^{\text{true}}(f^n_{\mathcal{H},s}) - R^{\text{true}}(f^*_{\mathcal{H}})| \\ &\leq 3\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}_s(f) - R^{\text{true}}(f)|\}\end{aligned}$$

which implies

$$|R^{\text{emp}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}})| \leq 3\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\}$$

$\square$

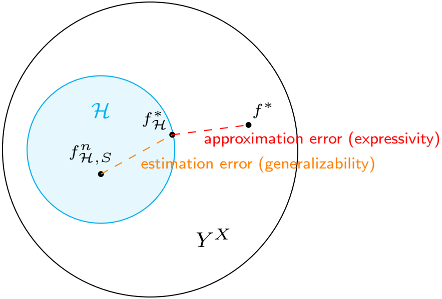
In statistical learning theory, the convergence $\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\} \xrightarrow{p} 0$ is shown by the PAC bounds [21]. We call $\sup_{f\in\mathcal{H}}\{|R^{\text{emp}}(f) - R^{\text{true}}(f)|\}$ the generalization error, and $R^{\text{true}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}})$ the estimation error. Clearly the estimation error is upper bounded by the generalization error, scaled by a constant factor.

Besides, given any n and $S = s$, $R_s^{\text{emp}}(f)$ is usually a non-convex or discrete function, of which finding the global minimizer $f_{\mathcal{H},s}^n$ is typically NP-hard. We can only get a sub-optimal solution $\hat{f}_{\mathcal{H},s}^n$ through optimizing $R_s^{\text{emp}}(f)$. Since $R_s^{\text{emp}}(\hat{f}_{\mathcal{H},s}^n) \geq R_s^{\text{emp}}(f_{\mathcal{H},s}^n)$ for any s, $R^{\text{emp}}(\hat{f}_{\mathcal{H},S}^n) - R^{\text{emp}}(f_{\mathcal{H},S}^n)$ is a non-negative random variable, and we call it the optimization error.

It should be noticed that we may not have $R^{\text{true}}(\hat{f}_{\mathcal{H},s}^n) \geq R^{\text{true}}(f_{\mathcal{H},s}^n)$ on every s. This implies that the global minimizer of empirical risk does not necessary give smaller true risk. Also, the sequence $\{\hat{f}_{\mathcal{H},S}^n\}_{n=1}^{\infty}$ may not give a consistent ERM.

The Framework of Statistical Learning

The relationships between hypothesis space, approximation error and estimation error are shown in the following figure.



Hence, $R^{\text{true}}(f^n_{\mathcal{H},S})$ can be split into 3 parts:

$$R^{\text{true}}(f^n_{\mathcal{H},S}) = \underbrace{R^{\text{true}}(f^n_{\mathcal{H},S}) - R^{\text{true}}(f^*_{\mathcal{H}})}_{\text{estimation error}} + \underbrace{R^{\text{true}}(f^*_{\mathcal{H}}) - R^{\text{true}}(f^*)}_{\text{approximation error}} + \underbrace{R^{\text{true}}(f^*)}_{\text{Bayes error}}$$

Statistical learning theory primarily focuses on analyzing generalization error, which is the upper bound of estimation error. Universal approximation theory mainly examines approximation error.

References

[1] Walter Rudin. Principles of mathematical analysis. New York: McGraw-hill, 1964. (Important!)

[2] Differences between ring and σ-ring.
`https://math.stackexchange.com/questions/1178317/`

[3] Additive but not σ-additive function.
`https://math.stackexchange.com/questions/929528/`

[4] Another method of Integration: Lebesgue Integral.
`https://www.whitman.edu/Documents/Academics/Mathematics/2017/Wang.pdf`

[5] Approximation by Simple Functions.
`https://heil.math.gatech.edu/6337/spring11/section4.1.pdf`

[6] George Cybenko. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals and Systems, 1989.
`https://web.njit.edu/~usman/courses/cs677/10.1.1.441.7873.pdf`

References

[7] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. Neural Networks, 1989.
https://cognitivemedium.com/magic_paper/assets/Hornik.pdf

[8] Matt Young. The Stone-Weierstrass Theorem.
https://wiki.math.ntnu.no/_media/ma8106/2020v/stoneweierstrass.pdf

[9] Bernstein Polynomial. https://en.wikipedia.org/wiki/Bernstein_polynomial

[10] Tiangang Cui and Friedrich Pillichshammer. Bernstein approximation and beyond: proofs by means of elementary probability theory.
https://arxiv.org/pdf/2307.11533

[11] Walter Rudin. Real and Complex Analysis. McGraw-hill, 1987. (Important!)

[12] Erwin Kreyszig. Introductory Functional Analysis with Applications. Wiley, 1978.

[13] Proof of Hahn-Banach Theorem.
https://www.ucl.ac.uk/~ucahad0/3103_handout_6.pdf

References

[14] Proof of Monotone Convergence Theorem.
https://planetmath.org/proofofmonotoneconvergencetheorem

[15] Caratheodory's Extension Theorem.
https://sites.math.rutgers.edu/~carlen/501F14/Measconst.pdf

[16] Proof of Inner and Outer Measures Equality. https://math.stackexchange.com/questions/2044580/inner-and-outer-measures-equality

[17] Gerald B. Folland. Real Analysis: Modern Techniques and Their Applications. Wiley, 2007. (Important!)

[18] Danny Espejo. The Riesz-Markov-Kakutani Representation Theorem.
https://math.uchicago.edu/~may/REU2023/REUPapers/Espejo.pdf

[19] Patrick Billingsley. Probability and Measure. Wiley, 1986. https://www.colorado.edu/amath/sites/default/files/attached-files/billingsley.pdf

References

[20] Felipe Cucker and Ding Xuan Zhou. Learning Theory: an Approximation Theory Viewpoint. Cambridge University Press, 2007.

[21] Vladimir Vapnik. The Nature of Statistical Learning Theory. Springer, 1995.

Appendix 1: Alternative Proof of Weierstrass Theorem

Here we show the original proof of Bernstein on Weierstrass Theorem, which is an probabilistic proof [9]. This proof shows how the Bernstein polynomial is obtained.

Proof of Theorem 2.3:
Let $[a, b] = [0, 1]$, for any $x \in [0, 1]$, we consider it as the parameter of Bernoulli distribution Bern(x). Suppose we make n independent random trials $X_1, X_2, ..., X_n$ from Bern(x) and let $k = \sum_{i=1}^{n} X_i$ be the number of successes, then k is an random variable satisfying binomial distribution Binom(n, x), i.e.,

$$p(k) = \binom{n}{k} x^k (1-x)^{n-k} \quad \text{for } k = 0, 1, ..., n$$

Also, by Weak Law of Large Numbers, the sample mean $\frac{k}{n}$ approximates the expectation $\mathbb{E}\left[\frac{k}{n}\right] = x$ in probability as $n \to \infty$. That is, for any $\delta > 0$,

$$\lim_{n \to \infty} P\left(\left|\frac{k}{n} - x\right| > \delta\right) = 0 \tag{33}$$

Appendix 1: Alternative Proof of Weierstrass Theorem

Since f is uniformly continuous on $[0,1]$, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\left|\frac{k}{n} - x\right| < \delta \implies \left|f\left(\frac{k}{n}\right) - f(x)\right| < \epsilon$$

which implies

$$\left|f\left(\frac{k}{n}\right) - f(x)\right| \geq \epsilon \implies \left|\frac{k}{n} - x\right| \geq \delta$$

Thus

$$P\left(\left|f\left(\frac{k}{n}\right) - f(x)\right| \geq \epsilon\right) \leq P\left(\left|\frac{k}{n} - x\right| \geq \delta\right)$$

By Eq (33) and sandwich theorem, we have

$$\lim_{n\to\infty} P\left(\left|f\left(\frac{k}{n}\right) - f(x)\right| \geq \epsilon\right) = 0 \tag{34}$$

for any $\epsilon > 0$.

Appendix 1: Alternative Proof of Weierstrass Theorem

Since f is bounded, we assume there exists M such that $|f(x)| < M$ for any x, then $\left|f\left(\frac{k}{n}\right) - f(x)\right| < 2M$ and

$$\begin{aligned}
\mathbb{E}\left[\left|f\left(\frac{k}{n}\right) - f(x)\right|\right] &= \sum_{k=0}^{n} p\left(\left|f\left(\frac{k}{n}\right) - f(x)\right|\right)\left|f\left(\frac{k}{n}\right) - f(x)\right| \\
&\leq \sum_{k=0}^{n} p\left(\left|f\left(\frac{k}{n}\right) - f(x)\right|\right)\left(2M\mathbf{1}_{\left[\left|f\left(\frac{k}{n}\right)-f(x)\right|\geq\epsilon\right]} + \epsilon\mathbf{1}_{\left[\left|f\left(\frac{k}{n}\right)-f(x)\right|<\epsilon\right]}\right) \\
&\leq 2MP\left(\left|f\left(\frac{k}{n}\right) - f(x)\right| \geq \epsilon\right) + \epsilon
\end{aligned}$$

Thus for any $\epsilon > 0$, there exists $N \in \mathbb{N}^{+}$ such that for any $n > N$,

$$P\left(\left|f\left(\frac{k}{n}\right) - f(x)\right| \geq \epsilon\right) < \frac{\epsilon}{2M}$$

Consequently,

$$\mathbb{E}\left[\left|f\left(\frac{k}{n}\right) - f(x)\right|\right] < 2\epsilon$$

Appendix 1: Alternative Proof of Weierstrass Theorem

Since $g(x) = |x|$ is a convex function, by Jensen's inequality, $g\mathbb{E}[x] \leq \mathbb{E}[g(x)]$. Thus

$$\left| \mathbb{E}\left[f\left(\frac{k}{n}\right) \right] - f(x) \right| \leq \mathbb{E}\left[\left| f\left(\frac{k}{n}\right) - f(x) \right| \right] < 2\epsilon$$

This implies $\mathbb{E}\left[f\left(\frac{k}{n}\right) \right]$ converges to $f(x)$ uniformly as $n \to \infty$, and

$$\mathbb{E}\left[f\left(\frac{k}{n}\right) \right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

is the Bernstein polynomial.

$\square$

Appendix 1: Alternative Proof of Weierstrass Theorem

The main idea of the proof is as follows:

(1) For any x, show there exists a sequence of random variables $\{X_n\}$ parameterized by x that $X_n \to x$ in probability.

(2) Show that for any uniformly continuous f, $f(X_n) \to f(x)$ in probability.

(3) Show that $\mathbb{E}[f(X_n)] \to f(x)$ [52].

$\mathbb{E}[f(X_n)]$ can be expressed as a polynomial of x depending on n if X_n is of binomial distribution. In general, X_n does not have to be binomial, the cases of other distributions can be found at [10].

[52] See the portmanteau lemma.
https://en.wikipedia.org/wiki/Convergence_of_random_variables#Properties

Appendix 2: Hausdorff's Maximality Theorem

Definition B.1: A partially ordered set is a set X with a binary operation called partial ordering, written as $\leq$, which satisfies the following conditions:

- $a \leq a$ for every $a \in X$.
- If $a \leq b$ and $b \leq a$, then $a = b$.
- If $a \leq b$ and $b \leq c$, then $a \leq c$.

"Partially" means there may exist $a, b \in X$ such that neither $a \leq b$ nor $b \leq a$ holds, that is, a and b are incomparable. In contrast, if $a, b \in X$ have either $a \leq b$ or $b \leq a$, then we say a, b are comparable.

A totally ordered set or chain is a partially ordered set, of which any two elements are comparable. That is, there is no incomparable elements in a totally ordered set.

Let A be a subset of X. An upper bound of A is an element $u \in X$ such that $a \leq u$ for any $a \in A$. A maximal element of A is an element $m \in A$ such that $a \leq m$ for any $a \in A$.

Depending on A and X, A may or may not have upper bound or maximal elements.

Examples:

- **Real Numbers**: Let $\mathbb{R}$ be the set of all real numbers and let $\leq$ has its usual meaning, then $\mathbb{R}$ is totally ordered and has no maximal element.
- **Power Set**: Let $\mathcal{P}(X)$ be the power set of X, and let $A \leq B$ mean $A \subset B$ for any $A, B \in \mathcal{P}(X)$, then $\mathcal{P}(X)$ is partially ordered. For example, let $X = \{a, b, c\}$, then both $\{a, b\}$ and $\{b, c\}$ are elements of $\mathcal{P}(X)$ but they are not comparable. The only maximal element of $\mathcal{P}(X)$ is X.
- **Positive Integers**: Let $\mathbb{N}$ be the set of all positive integers, and let $m \leq n$ mean m divides n for any $m, n \in \mathbb{N}$, then $\mathbb{N}$ is partially ordered.

Definition B.2: Let $\mathcal{F}$ be a family of sets and Φ be a subset of $\mathcal{F}$. We say Φ is a subchain of $\mathcal{F}$ if Φ is totally ordered by set inclusion, i.e., for any sets $A, B \in \Phi$, either $A \subset B$ or $B \subset A$.

Appendix 2: Hausdorff's Maximality Theorem

Let P be a partially ordered set and $\mathcal{F}$ be the family of subsets of P. The maximal totally ordered subset of P is the maximal element in $\mathcal{F}$ with partial order defined by set inclusion, and the "totally ordered" means the elements in the subset is totally ordered, but the order can be defined in any way.

Theorem B.3 (Haussdorff's Maximality): Every non-empty partially ordered set P contains a maximal totally ordered subset.

Proof: The proof of Theorem B.3 utilizes Lemma B.4 and The Axiom of Choice.

Lemma B.4: Suppose $\mathcal{F}$ is a non-empty family of subsets of a set X such that for every subchain of $\mathcal{F}$, the union of all its members belongs to $\mathcal{F}$. Let g be a function which associates each $A \in \mathcal{F}$ a set $g(A) \in \mathcal{F}$ such that $A \subset g(A)$ and $g(A) - A$ consists of at most one element. Then there exists an $A \in \mathcal{F}$ for which $g(A) = A$.

Proof: Fix $A_0 \in \mathcal{F}$, we call a subset $\mathcal{F}'$ of $\mathcal{F}$ a tower if $\mathcal{F}'$ has the following three properties:
(1) $A_0 \in \mathcal{F}'$.
(2) For every subchain of $\mathcal{F}'$, the union of all its members belongs to $\mathcal{F}'$.
(3) If $A \in \mathcal{F}'$, then also $g(A) \in \mathcal{F}'$.

Appendix 2: Hausdorff's Maximality Theorem

The family of all towers of $\mathcal{F}$ is non-empty. This is because, let $\mathcal{F}_1'$ be the family of all sets $A \in \mathcal{F}$ with $A_0 \subset A$, then $\mathcal{F}_1'$ is a tower and contains at least A_0. Thus, the intersection of all towers of $\mathcal{F}$, denoted as $\mathcal{F}_0$, is non-empty. It can be verified that $\mathcal{F}_0$ satisfies the three properties of a tower:
(1) Let $\mathcal{F}_i$ $(i = 1, 2, ...)$ be towers of $\mathcal{F}$. $A_0 \in \mathcal{F}_i$ for any i implies $A_0 \in \bigcap_i \mathcal{F}_i = \mathcal{F}_0$.
(2) Any subchain in $\mathcal{F}_0$ is a subchain in $\mathcal{F}_i$ for $i = 1, 2, ...$, thus the union of all members of the subchain is a common element in all $\mathcal{F}_i$, which shows that it is also in $\mathcal{F}_0$.
(3) $A \in \mathcal{F}_0$ implies $A \in \mathcal{F}_i$ for $i = 1, 2, ...$, thus $g(A) \in \mathcal{F}_i$ for $i = 1, 2, ...$, which means $g(A) \in \mathcal{F}_0$.
Thus $\mathcal{F}_0$ is a tower. Also, no proper subset of $\mathcal{F}_0$ is a tower. Since $\mathcal{F}_0$ is the intersection of all towers of $\mathcal{F}$, if $\mathcal{F}_0' \in \mathcal{F}$ is a tower and $\mathcal{F}_0' \subset \mathcal{F}_0$, then $\mathcal{F}_0' \cap \mathcal{F}_0 = \mathcal{F}_0' \neq \mathcal{F}_0$, which is contradict.

Appendix 2: Hausdorff's Maximality Theorem

Now we would like to show that $\mathcal{F}_0$ is totally ordered. Let Γ be the family of all $C \in \mathcal{F}_0$ such that every $A \in \mathcal{F}_0$ satisfies either $C \subset A$ or $A \subset C$. For each $C \in \Gamma$, let $\Phi(C)$ be the family of all $A \in \mathcal{F}_0$ such that either $A \subset C$ or $g(C) \subset A$. It is easy to verify that both Γ and $\Phi(C)$ satisfy the properties (1) and (2) of a tower.

For any $C \in \Gamma$, we would like to show $\Phi(C)$ satisfies property (3): If $A \in \Phi(C)$, then $g(A) \in \Phi(C)$. Suppose $A \in \Phi(C)$, then there are three cases: (a) $A \subset C$; (b) $A = C$; (c) $C \subset A$. Since $C \in \Gamma$ and $g(A) \in \mathcal{F}_0$, by the definition of $\Phi(C)$, we have either $g(A) \subset C$ or $C \subset g(A)$. For (a), $A \subset C$ and $g(A) \subset C$ imply $g(A) \in \Phi(C)$ (The combination of $A \subset C$ and $C \subset g(A)$ is impossible since $g(A) - A$ would contain at least two elements). For (b), $A = C$ means $g(A) = g(C)$, thus $g(C) \subset g(A)$ holds and $g(A) \in \Phi(C)$. For (c), $g(C) \subset A \subset g(A)$ gives $g(A) \in \Phi(C)$. Therefore, $\Phi(C)$ is a tower for any $C \in \Gamma$.

Since $\Phi(C)$ is a subset of $\mathcal{F}_0$ and no proper subset of $\mathcal{F}_0$ is a tower, we have $\Phi(C) = \mathcal{F}_0$ for any $C \in \Gamma$. This implies that any $A \in \mathcal{F}_0$ satisfies either $A \subset C$ or $g(C) \subset A$ for any $C \in \Gamma$. Since $A \subset C$ implies $A \subset g(C)$, which shows $g(C) \in \mathcal{F}_0$. Thus Γ satisfies the property (3) and is a tower. Also by the minimality of $\mathcal{F}_0$, $\Gamma = \mathcal{F}_0$.

By the definition of Γ, for any $A \in \mathcal{F}_0$ and $C \in \mathcal{F}_0$, either $A \subset C$ or $C \subset A$. Thus $\mathcal{F}_0$ is totally ordered.

Since all elements of $\mathcal{F}_0$ make a subchain of $\mathcal{F}_0$, let A be the union of all elements in $\mathcal{F}_0$, then by property (2), $A \in \mathcal{F}_0$. By property (3), $g(A) \in \mathcal{F}_0$. Since A is the largest member in $\mathcal{F}_0$ and $A \subset g(A)$, it follows that $A = g(A)$.

□

Definition B.5: A choice function is a function f which associates every subset E of X an element of E, i.e., $f(E) \in E$.

The Axiom of Choice: For every set there is a choice function.

Now we return to the proof of Theorem B.3. Let $\mathcal{F}$ be the family of all totally ordered subsets of P. $\mathcal{F}$ is non-empty, because every subset of P consisting a single element is totally ordered and thus in $\mathcal{F}$. Also, the union of all members of every subchain of $\mathcal{F}$ is in $\mathcal{F}$, because the union of any chain of totally ordered sets is still totally ordered.

Appendix 2: Hausdorff's Maximality Theorem

Now we define g. Let f be the choice function of P. Such f exists due to the Axiom of Choice. For any $A \in \mathcal{F}$, let A^* be the set of all x in $P - A$ such that $A \cup \{x\} \in \mathcal{F}$. If $A^* \neq \varnothing$, put $g(A) = A \cup f(A^*)$; if $A^* = \varnothing$, put $g(A) = A$.

By Lemma B.4, there is at least one $A \in \mathcal{F}$ such that $A = g(A)$. Since A is not included in any other totally ordered set, it is a maximal element of $\mathcal{F}$.

$\square$

Appendix 3: Supplemental Proofs

Lemma C.1: Let $\mathcal{A}$ be a family of real continuous functions on K, f be a real continuous function on K. The following two statements are equivalent:
(1) For any $\epsilon > 0$, there exists $h \in \mathcal{A}$ such that $|h(t) - f(t)| < \epsilon$ for all $t \in K$.
(2) There exists a sequence of functions $\{h_n\}$ in $\mathcal{A}$ such that for any $\epsilon' > 0$, there exists N such that for any $n > N$, $|h_n(t) - f(t)| < \epsilon'$ for all $t \in K$.

Proof:
$(1) \implies (2)$: Let $\epsilon = \frac{1}{n}$, and h_n be the function satisfying $|h_n(t) - f(t)| < \frac{1}{n}$ for all $t \in K$. Then the sequence $\{h_n\}$ and $N = \lceil 1/\epsilon' \rceil$ satisfy the condition of (2).

$(2) \implies (1)$: Let $h = h_{N+1}$, so the h exists.

□

Definition C.2: A sequence of functions $\{f_n\}$ converges to f on X pointwisely if for any $x \in X$, $\lim_{n\to\infty} f_n(x) = f(x)$. That is, for any $x \in X$ and any $\epsilon > 0$, there exists N such that $|f_n(x) - f(x)| < \epsilon$ for all $n > N$.

Lemma C.2: Pointwise convergence does not imply uniform convergence.

Proof: Let $f_n(x) = x^n$ be a sequence of functions on $(0, 1)$ and $f(x) = 0$, then f_n converges to f pointwisely, because $\lim_{n\to\infty} x^n = 0$ for any $x \in (0, 1)$. However, given any $0 < \epsilon < 1$,

$$|x^n - 0| < \epsilon \Longleftrightarrow n > \frac{\ln \epsilon}{\ln x}$$

when $x \to 1$, $\ln \epsilon / \ln x \to +\infty$. Thus there does not exist a finite N such that $|x^n - 0| < \epsilon$ for all $n > N$. $\square$

Lemma C.3: Continuity does not imply uniform continuity.

Proof: Let $f(x) = \frac{1}{x}$ for $x \in (0, 1)$. Given any x, for any $0 < \epsilon < 1/x - 1$,

$$\left|\frac{1}{x} - \frac{1}{y}\right| < \epsilon \Longrightarrow \frac{1}{\frac{1}{x} + \epsilon} < y < \frac{1}{\frac{1}{x} - \epsilon}$$

Since $x - \frac{1}{\frac{1}{x}+\epsilon} < \frac{1}{\frac{1}{x}-\epsilon} - x$, let $\delta = x - \frac{1}{\frac{1}{x}+\epsilon}$, then $|x - y| < \delta$ implies $\left|\frac{1}{x} - \frac{1}{y}\right| < \epsilon$, which means f is continuous on $(0, 1)$. Note that δ depends on x.

Appendix 3: Supplemental Proofs

If f is uniformly continuous on $(0,1)$, then for any $\epsilon > 0$, there exists $\delta > 0$ such that for any $x, y \in (0,1)$, $|x-y| < \delta \implies |\frac{1}{x} - \frac{1}{y}| < \epsilon$.
Suppose ϵ is given. Since $|\frac{1}{x} - \frac{1}{y}| < \epsilon \iff |x-y| < \epsilon xy$, if $|x-y| < \delta \implies |\frac{1}{x} - \frac{1}{y}| < \epsilon$, then $\delta < \epsilon xy$. We take $y = x + \frac{\delta}{2}$ such that $|x-y| < \delta$ is satisfied, then $\delta < \epsilon x(x + \frac{\delta}{2}) \iff \delta < \frac{\epsilon x^2}{1-\frac{x}{2}}$. It is impossible to find $\delta > 0$ such that $\delta < \frac{\epsilon x^2}{1-\frac{x}{2}}$ holds for any $x \in (0,1)$, because when $x \to 0^+$, $\frac{\epsilon x^2}{1-\frac{x}{2}} \to 0$. Thus f is not uniformly continuous on $(0,1)$. $\square$

Lemma C.4: For any $x, y \in \mathbb{R}$ and $x < y$, there exists $q \in \mathbb{Q}$ such that $x < q < y$.

Proof: Since $y - x > 0$, $1/(y-x)$ is finite, thus there exists $n \in \mathbb{Z}$ such that $n(y-x) > 1$. Since $nx \in \mathbb{R}$ is finite, there must exists $m \in \mathbb{Z}$ such that $m - 1 \leq nx < m$. Thus $nx < m \leq nx + 1 < ny \implies x < \frac{m}{n} < y$, and $\frac{m}{n} \in \mathbb{Q}$. $\square$

Lemma C.5: $\mathbb{Q}$ is dense in $\mathbb{R}$.

Proof: For any $p \in \mathbb{R}$ and real $\epsilon > 0$, there exists $q \in \mathbb{R}$ such that $|p - q| < \epsilon$. By Lemma C.4, there exists $r \in \mathbb{Q}$ such that $|p - q| = |p - r| + |r - q|$, which implies $|p - r| < \epsilon$. $\square$

Lemma C.6 [53]: Every open set in the Euclidean space $\mathbb{R}^d$ can be written as a countable union of open balls.

Proof: Let $X \subset \mathbb{R}^d$ be an open set. Then for any $x \in X$, there exists a rational number $\epsilon_x > 0$ such that $N_{2\epsilon_x}(x) \subset X$ (If ϵ_x is real, let q be a rational number satisfying $0 < q < \epsilon_x$, we assign q to ϵ_x). By Lemma C.5, $\mathbb{Q}^d$ is dense in $\mathbb{R}^d$, thus for any $\epsilon_x > 0$, there exists $p_x \in \mathbb{Q}^d$ such that $p_x \in N_{\epsilon_x}(x)$, thus $p_x \in X$. Since $p_x \in N_{\epsilon_x}(x)$ if and only if $x \in N_{\epsilon_x}(p_x)$, we have $X \subset \bigcup_{x \in X} N_{\epsilon_x}(p_x)$. Also, since $N_{\epsilon_x}(p_x) \subset N_{2\epsilon_x}(x) \subset X$ for any x, we have $\bigcup_{x \in X} N_{\epsilon_x}(p_x) \subset X$. Hence $X = \bigcup_{x \in X} N_{\epsilon_x}(p_x)$. Since both ϵ_x and p_x are rational numbers, $\bigcup_{x \in X} N_{\epsilon_x}(p_x)$ is a countable union of open balls. $\square$

[53] https://metaphor.ethz.ch/x/2023/hs/401-3601-00L/files/4-recap-topology.pdf

Corollary C.7 [54]: Every open set in the Euclidean space $\mathbb{R}^d$ can be written as a countable union of intervals.

Proof: For any real $x \in X$, there exists a real $\epsilon_x > 0$ such that $N_{2\epsilon_x}(x) \subset X$. We show that there exists an interval $I \subset N_{2\epsilon_x}(x)$ such that all edges of I_x are rational and $x \in I_x$. Let $x = [x_1, x_2, ..., x_d]^T$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there exist $a_i, b_i \in \mathbb{Q}$ such that $x_i - \frac{\epsilon_x}{2\sqrt{d}} < a_i < x_i < b_i < x_i + \frac{\epsilon_x}{2\sqrt{d}}$. Let $I_x = \{p \in \mathbb{R}^d : a_i \leq p \leq b_i \text{ for } i = 1, 2, ..., d\}$, then I_x is an interval and $x \in I_x$. Let $a = [a_1, a_2, ..., a_d]^T$ and $b = [b_1, b_2, ..., b_d]^T$, since $\|b - a\|_2 < \epsilon_x$, we have $I_x \subset N_{2\epsilon_x}(x)$. Therefore, $X = \bigcup_{x \in X} I_x$ and the set of I_x is countable. $\square$

An example of Corollary C.7 is that the open set $(0, 1)$ on $\mathbb{R}$ can be expressed as a countable union of (closed) intervals by $(0, 1) = \bigcup_{n=2}^{\infty} [\frac{1}{n}, 1 - \frac{1}{n}]$.

[54] https://math.stackexchange.com/questions/3364802/

Lemma C.8 [55]: Any countable union of intervals in $\mathbb{R}^d$ can be expressed as a countable union of disjoint intervals.

Proof: Let $\{I_n\}_{n=1}^{\infty}$ be a set of intervals on $\mathbb{R}^n$, and let $J_n = I_n - (\bigcup_{i=1}^{n-1} I_i)$, then

$$\bigcup_{n=1}^{\infty} I_n = \bigcup_{n=1}^{\infty} \left(I_n - (\bigcup_{i=1}^{n-1} I_i) \right) = \bigcup_{n=1}^{\infty} J_n$$

$\{J_n\}_{n=1}^{\infty}$ is a family of disjoint sets since each J_n is disjoint with J_i for all $i \leq n-1$. Let $[a_j^{(i)}, b_j^{(i)}]$ be the jth edge of I_i where $j = 1, ..., d$ and $i = 1, ..., n$, we partition the jth edge of I_n with respect to $\{a_j^{(i)}\}_{i=1}^{n-1}$ and $\{b_j^{(i)}\}_{i=1}^{n-1}$, then there can be at most $2n-1$ partitions. With the partitions on each edge, I_n can be split into at most $(2n-1)^d$ disjoint intervals, and whenever I_n intersects with I_i $(i < n)$, the intersection must be the union of some of these intervals. Thus for any n, J_n can be expressed as no more than $(2n-1)^d$ intervals. When $n \to \infty$, the number of intervals in J_n is countable, and by Theorem 1.5, $\bigcup_{n=1}^{\infty} J_n$ is a countable union of disjoint intervals. $\square$

By Corollary C.7 and Lemma C.8, every open set in $\mathbb{R}^d$ can be written as countable union of disjoint intervals.

[55] https://arxiv.org/pdf/2312.12440

Lemma C.9 [56]: Any countable set in $\mathbb{R}$ has zero Lebesgue measure.

Proof: Let $X = \{x_i\}_{i=1}^{\infty}$ be the countable set. For any $\epsilon > 0$, let $I_i = (x_i - \frac{\epsilon}{2^{i+1}}, x_i + \frac{\epsilon}{2^{i+1}})$ such that $x_i \in I_i$, and let $m(I_i) = \frac{\epsilon}{2^i}$ where m is the Lebesgue measure [57]. Since $\bigcup_{i=1}^{\infty} I_i$ is a Borel set, which is Lebesgue measurable, we have

$$m(X) \leq m(\bigcup_{i=1}^{\infty} I_i) \leq \sum_{i=1}^{\infty} m(I_i) = \epsilon$$

which implies $m(X) = 0$. $\square$

Since $\mathbb{Q}$ is a countable set in $\mathbb{R}$, by Lemma C.9, the Lebesgue measure of $\mathbb{Q}$ is zero.

[56] https://math.stackexchange.com/questions/3033316/

[57] We have $m([a,b]) = m([a,b)) = m((a,b]) = m((a,b)) = b - a$, because the end points of an interval have zero Lebesgue measure, i.e., $m(\{a\}) = m([a,a]) = 0$.