Notes on Measure and Probability Theory

Liu Zhizhou

First Created: September 6, 2022
Last Modified: September 6, 2022

Contents

Part I

Measure Theory

0.1 An Introduction: From Riemann Integral to Measure Theory

Riemann Integral A viewpoint of understanding Riemann Integral is considering the upper sum and lower sum, namely

$$U_p(R, f) \triangleq \sum_{j=1}^{n}(x_j - x_{j-1})M_i, \quad L_p(R, f) \triangleq \sum_{j=1}^{n}(x_j - x_{j-1})m_i,$$

where $M_i = \sup_{x\in[x_{j-1},x_j]} f(x)$, $m_i = \inf_{x\in[x_{j-1},x_j]} f(x)$, p is the partition and R means we are in the case of Riemann. Then define

$$\overline{\int_a^b} f(x)\mathrm{d}x \triangleq \inf_p U_p(R, f), \quad \underline{\int_a^b} f(x)\mathrm{d}x \triangleq \sup_p L_p(R, f).$$

If the two values above equal, then we say the $f(x)$ is Riemann integrable. There are two import results in Riemann integral.

Proposition 0.1.1. Every continuous real-valued function on each closed bounded interval is Riemann integrable.

And also the following theorem.

Theorem 0.1.2. Let $(f_n)_{n=1}^{\infty}$ a sequence of Riemann integrable functions on $[a, b]$. Suppose that it converges uniformly on $[a, b]$ to a function f. Then f is Riemann integrable.

This two results seems nice and make the integral apply to different areas. However, there are issues that occur in Riemann integral.

1. Many "simple" functions are not Riemann integrable. For example, the Dirichlet function defined below.

$$D(x) \triangleq \begin{cases} 1, & x \in \mathbb{Q} \\ 0, & x \notin \mathbb{Q} \end{cases}.$$

   It is not Riemann integrable since the upper integral equals $b-a$ yet the lower integral equals zero.

2. If $0 \leq f_1 \leq f_2 \leq \dots$ and f_i is Riemann integrable for each $i \in \mathbb{N}$. The pointwise convergent of $f_n \to f$ does not grantee that f is Riemann integrable. Indeed, let $\mathbb{Q} = \{r_1, r_2, \dots\}$. Define

$$f_n(x) = \begin{cases} 1, & x \in \{r_1, \dots, r_n\} \\ 0, & x \notin \{r_1, \dots, r_n\} \end{cases}.$$

   Then for each x, $f_n(x) \to D(x)$. Note that f_n is Riemann integrable for each n.

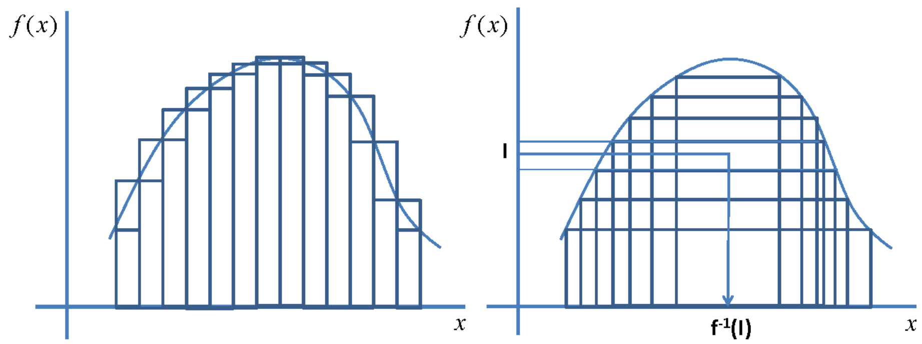

Figure 1: The figure one the left is an illustration of Riemann Integral; on the right is an illustration of Lebesgue Integral.

3. Suppose $f_n \geq 0$ for all n and Riemann integrable. The equation

$$\int_a^b \sum_{i=1}^{\infty} f_n(x) = \sum_{n=1}^{\infty} \int_a^b f_n(x) \tag{0.1.1}$$

is not true in general. Consider

$$f_n(x) = \begin{cases} 1, & x = r_n \\ 0, & x \neq r_n \end{cases}.$$

Then $\sum_{i=1}^{\infty} f_n(x) = D(x)$. The right hand side of the equation is even not integrable.

Here comes Lebesgue who defined Lebesgue Integral to fix the above problems.

Lebesgue Integral Consider an easy case: the domain of f is $[a, b]$. Rather than considering the partition on $[a, b]$ as in Riemann Integral, Lebesgue considered the partial on the range of f. The preimage would then be

$$E_i = \{x : y_{i-1} < f(x_i) \leq y_i\}.$$

The corresponding upper sum and lower sum is

$$U_p(L, f) \triangleq \sum_{i=1}^{n} y_i l(E_i), \quad L_p(L, f) \triangleq \sum_{i=1}^{n} y_{i-1} l(E_i),$$

where $l(E_i)$ is the length of E_i. Then

$$U_p(L, f) - L_p(L, f) = \sum_{i=1}^{n} (y_i - y_{i-1}) l(E_i) \leq \max_i \{y_i - y_{i-1}\}(b - a). \tag{0.1.2}$$

Therefore, $U_p(L, f) = L_p(L, f)$ when $\max_i\{y_i - y_{i-1}\} \to 0$. It seems that the integral is well-defined already and all functions are "Lebesgue Integrable". However, the issue is: what is $l(E_i)$? How to calculate it?

Since the only knowledge about length is the case of E_i is interval, we have no choice but to do the following definition, which is called *(Lebesgue) outer measure*,

$$\mu^*(E) \triangleq \inf\{\sum_{i=1}^{\infty} |I_i| : E \subset \cup_{i=1}^{\infty} I_i, \text{ where } I_i \text{ is open interval}\}.$$

However, this measure does not have the countable additive property when $\{E_i\}$ are disjoint. Instead, it only has sub-additive property:

$$\mu^*(\cup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu^*(E_i),$$

which does not guarantee Eq. (0.1.2) to be true.

The definition of Lebesgue measure fixed the problem. There are many equivalent definitions of Lebesgue measure:

Proposition 0.1.3. The following definitions are equivalent:

1. Suppose $\mu_*(E) = \sup\{\mu^*(A) : A \subseteq E, \text{ where } A \text{ is closed}\}$, which is called the *(Lebesgue) outer measure* of E. If $\mu^*(E) = \mu_*(E)$, then we say that E is measurable.

2. If for any $\epsilon > 0$, there exists an open set U with $E \subset U$ such that $\mu^*(U - E) \leq \epsilon$, then we say that E is measurable.

3. If for any subset A, we have

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c), \tag{0.1.3}$$

   then we say that E is measurable.

We can show that for every measurable set,

$$\mu^*(\cup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu^*(E_i).$$

Then define *measurable function* to be the functions that satisfies $f^{-1}([-\infty, a])$ is measurable for all $a \in \mathbb{R}$. Then the issue explained after Eq. (0.1.2) is completely solved.

Measure Theory The result in $\mathbb{R}$ can be easily generalize to the case of $\mathbb{R}^n$, however, how to define measure in abstract spaces, for example, the L^2 space. A Mathematician called Frechet used the same procedure defined the measure in abstract spaces: first define open set, then the measure on those sets, and the outer measure... However, have a second look at the Proposition 0.1.3, we find that item 3 does not need any information of open sets! Therefore, the definition of measurable set in abstract space follows.

Chapter 1

Fundamental Concepts

1.1 Algebra and σ-algebra

Definition 1.1.1 (algebra). A non-empty subset $\mathscr{A} \subset 2^X$ is said to be an *algebra* if

1. $\varnothing \in \mathscr{A}$,
2. if $A \in \mathscr{A}$, then $A^c \in \mathscr{A}$;
3. if $A_1, \ldots, A_n \in \mathscr{A}$, then $\cup_{i=1}^n A_i \in \mathscr{A}$.

Definition 1.1.2 (σ-algebra). A non-empty subset $\mathscr{A} \subset 2^X$ is said to be an *algebra* if

1. $\varnothing \in \mathscr{A}$,
2. if $A \in \mathscr{A}$, then $A^c \in \mathscr{A}$;
3. if $A_1, A_2 \cdots \in \mathscr{A}$, then $\cup_{i=1}^\infty A_i \in \mathscr{A}$.

In this case, we call $(X, \mathscr{A})$ a *measurable space* and the elements of $\mathscr{A}$ is called $\mathscr{A}$*-measurable sets* (or simply measurable sets).

Sometimes $\mathscr{C} \subset 2^X$ is not a σ-algebra.

Definition 1.1.3. Assume $\mathscr{C} \subset 2^X$. The σ-algebra generated by $\mathscr{C}$ is the smallest σ-algebra containing $\mathscr{C}$, denoted $\sigma(\mathscr{C})$.

Note that $\sigma(\mathscr{C})$ is all the intersection of all σ-algebra that contains $\mathscr{C}$.

Definition 1.1.4 (Borel σ-algebra). If X is a topological space, the σ-algebra generated by all open subsets of X is called the Borel σ-algebra of X and it is denoted by $\mathscr{B}(X)$.

Part II

Probability Theory

Chapter 2

Fundamental Concepts

2.1 Intuition of Probability

We have the following concepts illustrated in words. A *sample space* (or state space), denoted Ω, is the space of all possible outcomes of a *random experiment*, where the random experiment is a kind of experiment that you could not know the outcome before implementing. A *(random) event* is a subset of sample space.

Intuitively, the probability of an event A, denoted $\mathbb{P}(A)$, is the frequency that A will occur if we repeat the experiment many times, i.e.

$$\mathbb{P}(A) = \lim_{n\to\infty} f_n(A), \tag{2.1.1}$$

where $f_n(A)$ is the frequency of A after n trials of experiments.

The *probability model* is a triple-set $(\Omega, \mathscr{A}, \mathbb{P})$, where $\mathscr{A}$ is the class of events which we can define probability. For example, if $\Omega = \mathbb{R}$, then $\mathscr{A}$ could be all Lebesgue measurable set or all Borel sets.

A *random variable* $X(\omega)$ is a function which maps $\Omega \to E$, where E is a "simple" space, often $\mathbb{R}$ or $\mathbb{R}^n$ and even a Banach space or Hilbert space. And we can induce a probability on E by the probability on Ω. Define for $B \subseteq E$,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

and

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)). \tag{2.1.2}$$

Then we have a new probability model $(E, \mathscr{E}, \mathbb{P}_X)$. And $\mathbb{P}_X$ is called the *distribution* of X.