Lecture Notes on Random Variables and Stochastic Processes

This lecture notes mainly follows Chapter 1-7 of the book *Foundations of Modern Probability* by Olav Kallenberg. We will omit some parts.

1 Elements of Measure Theory

We begin with elementary notation of set theory. We use union $A \cup B$ or $\bigcup_\alpha A_\alpha$, intersection $A \cap B$ or $\bigcap_\alpha A_\alpha$, difference $A \setminus B = \{x \in A : x \notin B\}$, and symmetric difference $A \Delta B = (A \setminus B) \cup (B \setminus A)$. A partition of a set A is a family $A_t \subset A$, $t \in T$, such that $A = \bigcup_t A_t$, and for any $t_1 \neq t_2$, $A_{t_1} \cap A_{t_2} = \emptyset$. If a whole space Ω is fixed and contains all relative sets, the complement A^c is $\Omega \setminus A$. Recall that

$$A \cap \Big(\bigcup_\alpha B_\alpha \Big) = \bigcup_\alpha (A \cap B_\alpha), \quad A \cup \Big(\bigcap_\alpha B_\alpha \Big) = \bigcap_\alpha (A \cup B_\alpha)$$

$$\Big(\bigcup_\alpha A_\alpha \Big)^c = \bigcap_\alpha A_\alpha^c, \quad \Big(\bigcap_\alpha A_\alpha \Big)^c = \bigcup_\alpha A_\alpha^c.$$

A σ-algebra or σ-field in a nonempty set Ω is defined as a collection of $\mathcal{A}$ of subsets of Ω such that

1. $\emptyset, \Omega \in \mathcal{A}$,
2. $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$,
3. $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$ implies that $\bigcup_n A_n \in \mathcal{A}$ and $\bigcap_n A_n \in \mathcal{A}$.

We may also say that a σ-algebra is a class of subsets, which contains the empty set and the whole space, and is closed under complement, countable union and countable intersection. There are two trivial examples of σ-algebras. First, $\{\emptyset, \Omega\}$ is the smallest σ-algebra. Second, the collection 2^Ω of all subsets of Ω is the biggest σ-algebra.

A measurable space is a pair $(\Omega, \mathcal{A})$, where Ω is a nonempty set and $\mathcal{A}$ is a σ-algebra in Ω. Every element of $\mathcal{A}$ is called a measurable set.

We observe that if $\mathcal{A}_\alpha$, $\alpha \in A$, is a family of σ-algebras in Ω, then $\bigcap_\alpha \mathcal{A}_\alpha$ is a σ-algebra in Ω. We use this fact to define the σ-algebra generated by a collection of sets. Let $\mathcal{C} \subset 2^\Omega$, i.e.,

$\mathcal{C}$ is a collection of subsets of Ω. Let $\mathcal{M}(\mathcal{C})$ be the set of all σ-algebra $\mathcal{A}$ in Ω such that $\mathcal{C} \subset \mathcal{A}$. We define

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \in \mathcal{M}(\mathcal{C})} \mathcal{A}.$$

Then

1. $\sigma(\mathcal{C}) \supset \mathcal{C}$ as $\mathcal{A} \supset \mathcal{C}$ for every $\mathcal{A} \in \mathcal{M}(\mathcal{C})$.

2. $\sigma(\mathcal{C})$ is a σ-algebra in Ω as it is the intersection of a collection of σ-algebras in Ω.

These two properties imply that $\sigma(\mathcal{C}) \in \mathcal{M}(\mathcal{C})$, and so is the smallest σ-algebra in Ω that contains $\mathcal{C}$. We call $\sigma(\mathcal{C})$ the σ-algebra generated by $\mathcal{C}$. There are no simple expressions of $\sigma(\mathcal{C})$ in terms of union, intersection, and complement of elements of $\mathcal{C}$.

If S is a topological space, then the Borel σ-algebra $\mathcal{B}(S)$ on S is the σ-algebra generated by the topology of S, i.e., the collection of open subsets of S. Thus, a topological space is also viewed as a measurable space. We write $\mathcal{B}$ for $\mathcal{B}(\mathbb{R})$.

Besides σ-algebras, the following notation will be useful for us.

1. A π-system $\mathcal{C}$ in Ω is a class of subsets of Ω, which is closed under finite intersection, i.e., $A, B \in \mathcal{C}$ implies that $A \cap B \in \mathcal{C}$.

2. A λ-system $\mathcal{D}$ in Ω is a class of subsets of Ω, which contains Ω, and is closed under proper difference and increasing limits. The former means that $A, B \in \mathcal{D}$ and $A \supset B$ implies that $A \setminus B \in \mathcal{D}$. The latter means that if $A_1 \subset A_2 \subset A_2 \subset \cdots \in \mathcal{D}$, then $\bigcup_n A_n \in \mathcal{D}$.

It is clear that $\mathcal{A}$ is a σ-algebra if and only if it is both a π-system and a λ-system. If $\mathcal{E} \subset 2^{\Omega}$, we may similarly define the π-system $\pi(\mathcal{E})$ and the λ-system $\lambda(\mathcal{E})$ generated by $\mathcal{E}$, respectively.

The following monotone class theorem is very useful. An application of this result is called a monotone class argument.

Theorem 1.1. *If $\mathcal{C}$ is a π-system, then $\sigma(\mathcal{C}) = \lambda(\mathcal{C})$.*

Proof. Since a σ-algebra containing $\mathcal{C}$ is also a λ-system containing $\mathcal{C}$, we have $\lambda(\mathcal{C}) \subset \sigma(\mathcal{C})$. We need to show that $\sigma(\mathcal{C}) \subset \lambda(\mathcal{C})$. It suffices to show that $\lambda(\mathcal{C})$ is a σ-algebra. Since it is already a λ-system, we only need to show that it is a π-system. This means we need to show that, if $A, B \in \lambda(\mathcal{C})$, then $A \cap B \in \lambda(\mathcal{C})$.

At the beginning, since $\mathcal{C}$ is a π-system, we know that if $A, B \in \mathcal{C}$, then $A \cap B \in \mathcal{C} \subset \lambda(\mathcal{C})$. Now we show that

$$A \in \mathcal{C} \text{ and } B \in \lambda(\mathcal{C}) \text{ implies that } A \cap B \in \lambda(\mathcal{C}). \tag{1.1}$$

We prove this statement in an indirect way. Fix $A \in \mathcal{C}$. Consider the set

$$\mathcal{S}_A := \{B \subset \Omega : A \cap B \in \lambda(\mathcal{C})\}.$$

Then

1. $\mathcal{C} \subset \mathcal{S}_A$,

2. $\mathcal{S}_A$ is a λ-system.

To check the second claim, we note that

1. $\Omega \in \mathcal{S}_A$ because $\Omega \cap A = A$;

2. If $B_1 \supset B_2 \in \mathcal{S}_A$, then $B_1 \cap A \supset B_2 \cap A$, and so $(B_1 \setminus B_2) \setminus A = (B_1 \cap A) \setminus (B_2 \cap A) \in \Lambda(\mathcal{C})$. Thus, $B_1 \setminus B_2 \in \mathcal{S}_A$;

3. If $B_1 \subset B_2 \subset B_3 \subset \cdots \in \mathcal{S}_A$, then $B_1 \cap A \subset B_2 \cap A \subset B_3 \cap A \subset \cdots \in \Lambda(\mathcal{C})$. So $\bigcup B_n \cap A = \bigcup (B_n \cap A) \in \Lambda(\mathcal{C})$, which implies that $\bigcup B_n \in \mathcal{S}_A$.

This means that $\mathcal{S}_A$ is a λ-system that contains $\mathcal{C}$. So $\mathcal{S}_A$ contains $\lambda(\mathcal{C})$. This finishes the proof of (1.1).

Next we show that

$$A \in \lambda(\mathcal{C}) \text{ and } B \in \lambda(\mathcal{C}) \text{ implies that } \mathcal{A} \cap \mathcal{B} \in \lambda(\mathcal{C}).$$

This is enough to conclude that $\lambda(\mathcal{C})$ is a π-system. For the proof, for any $A \in \lambda(\mathcal{C})$, we define $\mathcal{S}_A$ by the same way as before. By (1.1, $\mathcal{S}_A$ contains $\mathcal{C}$. The argument in the last paragraph shows that $\mathcal{S}_A$ is a λ-system. So $\mathcal{S}_A$ contains $\lambda(\mathcal{C})$, and the proof is complete. □

For any family of spaces Ω_t, $t \in T$, the Cartesian product $\prod_t \Omega_t$ is the class of all collections $(\omega_t : t \in T)$, where $\omega_t \in \Omega_t$ for all $t \in T$. When $T = \{1, \dots, n\}$ or $T = \mathbb{N} = \{1, 2, \dots\}$, we write the product space as $\Omega_1 \times \cdots \times \Omega_n$ and $\Omega_1 \times \Omega_2 \times \cdots$. If all $\Omega_t = \Omega$, we use the notation Ω^T, Ω^n, or Ω^∞.

If each Ω_t is equipped with a σ-algebra $\mathcal{A}_t$, then we introduce the product σ-algebra $\prod_t \mathcal{A}_t$ as the σ-algebra in $\prod_t \Omega_t$ generated by the class of cylinder sets

$$\{A_t \times \prod_{s \neq t} \Omega_s = \{(\omega_s : s \in T) : \omega_t \in A_t \text{ and } \omega_s \in \Omega_s \text{ for } s \neq t\} : t \in T, A \in \mathcal{A}_t\}. \tag{1.2}$$

We call $(\prod_t \Omega_t, \prod_t \mathcal{A}_t)$ the product of the measurable spaces $(\Omega_t, \mathcal{A}_t)$, $t \in T$. In special cases, we use the symbols $\mathcal{A}_1 \times \cdots \mathcal{A}_n$, $\mathcal{A}_1 \times \mathcal{A}_2 \times \cdots$, $\mathcal{A}^T$, $\mathcal{A}^n$, $\mathcal{A}^\infty$.

In Topology, one may define product of topological space, which is also a topological space. A natural question to ask is whether the Borel σ-algebra generated by the product topology agrees with the product of the Borel σ-algebra generated by each topology. The answer is Yes if we only consider a countable product and each space is a separable metric space. A topological space is called separable if it contains a countable dense set.

Lemma 1.2. *Let $S_1, S_2, \dots$ be separable metric spaces. Then*

$$\mathcal{B}(S_1 \times S_2 \times \cdots) = \mathcal{B}(S_1) \times \mathcal{B}(S_2) \times \cdots .$$

We remark that the product on the left is about topological spaces, and the product on the right is about measurable spaces. For example, since $\mathbb{R}$ is a separable metric space, $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}^n$.

Proof. Let $\mathcal{T}_n$ denote the topology in S_n. Then $\sigma(\mathcal{T}_n) = \mathcal{B}(S_n)$. Let

$$\mathcal{C}_\sigma^n = \{A_n \times \prod_{m \neq n} S_m : A_n \in \mathcal{B}(S_n)\}, \quad \mathcal{C}_\mathcal{T}^n = \{A_n \times \prod_{m \neq n} S_m : A_n \in \mathcal{T}_n\}, \quad n \in \mathbb{N};$$

$\mathcal{C}_\sigma = \bigcup_n \mathcal{C}_\sigma^n$ and $\mathcal{C}_\mathcal{T} = \bigcup_n \mathcal{C}_\mathcal{T}^n$. By definition of product σ-algebra,

$$\mathcal{B}(S_1) \times \mathcal{B}(S_2) \times \cdots = \sigma(\mathcal{C}_\sigma).$$

On the other hand, the product topology on $S_1 \times S_2 \times \cdots$ is the topology generated by $\mathcal{C}_\mathcal{T}$. We denote it by $\mathcal{T}(\mathcal{C}_\mathcal{T})$. Thus, the Borel σ-algebra on the product space is

$$\mathcal{B}(S_1 \times S_2 \times \cdots) = \sigma(\mathcal{T}(\mathcal{C}_\mathcal{T})).$$

It remains to show that $\sigma(\mathcal{C}_\sigma) = \sigma(\mathcal{T}(\mathcal{C}_\mathcal{T}))$. It is easy to show that $\mathcal{C}_\sigma^n = \sigma(\mathcal{C}_\mathcal{T}^n)$ for each n. So

$$\sigma(\mathcal{C}_\sigma) = \sigma(\bigcup_n \mathcal{C}_\sigma^n) \subset \sigma(\bigcup_n \sigma(\mathcal{C}_\mathcal{T}^n)) = \sigma(\bigcup_n \mathcal{C}_\mathcal{T}^n) = \sigma(\mathcal{C}_\mathcal{T}) \subset \sigma(\mathcal{T}(\mathcal{C}_\mathcal{T})).$$

For the other direction, we use the fact that each $\mathcal{T}_n$ has a countable base, i.e., there is a countable set $\mathcal{T}_n' \subset \mathcal{T}_n$ such that each element of $\mathcal{T}_n$ can be expressed as a union of some elements of $\mathcal{T}_n'$. To construct $\mathcal{T}_n'$, let A_n be a countable dense subset of S_n (because S_n is separable), and let

$$\mathcal{T}_n' = \{\{w \in S_n : \operatorname{dist}(w, z) < q\} : z \in A_n, q \in \mathbb{Q}_+\}.$$

It is easy to check that $\mathcal{T}_n'$ satisfies the desired property. We may use $\mathcal{T}_n'$ to construct a countable basis of the topology in $S_1 \times S_2 \times \cdots$, namely

$$A_1 \times A_2 \times \cdots \times A_m \times S_{m+1} \times S_{m+1} \times \cdots,$$

where $m \in \mathbb{N}$ and $A_j \in \mathcal{T}_j'$ for $1 \leq j \leq m$. Every element of the countable basis belongs to $\sigma(\mathcal{C}_\sigma)$. Since every open set in $S_1 \times S_2 \times \cdots$ is a countable union of elements in the basis, we have $\mathcal{T}(\mathcal{C}_\mathcal{T}) \subset \sigma(\mathcal{C}_\sigma)$. Thus, $\sigma(\mathcal{T}(\mathcal{C}_\mathcal{T})) \subset \sigma(\mathcal{C}_\sigma)$. The proof is then complete. □

Let S and T be two nonempty sets. A point mapping $f : S \to T$ induces two set mappings $f : 2^S \to 2^T$ and $f^{-1} : 2^T \to 2^S$ such that

$$fA = \{f(x) : x \in A\}, \quad f^{-1}B = \{x \in S : f(x) \in B\}$$

for $A \subset S$ and $B \subset T$. Note that for the definition of f^{-1} we do not need f to be surjective or injective. Then we have

$$f^{-1}B^c = (f^{-1}B)^c, \quad f^{-1}\bigcup_t B_t = \bigcup_t f^{-1}B_t, \quad f^{-1}\bigcap_t B_t = \bigcap_t f^{-1}B_t. \tag{1.3}$$

For a class $\mathcal{C} \subset 2^T$, we define

$$f^{-1}\mathcal{C} = \{f^{-1}B : B \in \mathcal{C}\}.$$

Lemma 1.3. *Let $\overline{S}$ and $\overline{T}$ be σ-algebras in S and T, respectively. Then $f^{-1}\overline{T}$ is a σ-algebra in S and $\{B \subset T : f^{-1}B \in \overline{S}\}$ is a σ-algebra in T.*

Proof. It follows directly from (1.3). $\square$

In the setup of Lemma 1.3, we call $f^{-1}\overline{T}$, denoted by $\sigma(f)$, the σ-algebra induced or generated by f; and if $f^{-1}\overline{T} \subset \overline{S}$, then we say that f is $\overline{S}/\overline{T}$-measurable or simply measurable if $\overline{S}$ and $\overline{T}$ are fixed. Note that $\sigma(f)$ is the smallest σ-algebra in S w.r.t. which f is measurable.

Lemma 1.4. *If $\mathcal{C} \subset 2^T$ satisfies that $\overline{T} = \sigma(\mathcal{C})$, then $f^{-1}\overline{T} \subset \overline{S}$ if and only if $f^{-1}(\mathcal{C}) \subset \overline{S}$.*

Proof. Clearly $f^{-1}\overline{T} \subset \overline{S}$ implies that $f^{-1}(\mathcal{C}) \subset \overline{S}$. On the other hand, if $f^{-1}(\mathcal{C}) \subset \overline{S}$ then by Lemma 1.3, the class of sets $B \subset T$ such that $f^{-1}(B) \in \overline{S}$ is a σ-algebra in T. Such class contains $\mathcal{C}$ by assumption, and so it contains $\sigma(\mathcal{C}) = \overline{T}$. Thus, we get $f^{-1}\overline{T} \subset \overline{S}$. $\square$

Lemma 1.5. *If $f : S \to T$ is a continuous mapping between two topological spaces, then f is measurable with respect to the Borel σ-algebras $\mathcal{B}(S)$ and $\mathcal{B}(T)$.*

Proof. Let $\mathcal{T}_S$ and $\mathcal{T}_T$ be the topologies in S and T, respectively. Then $\mathcal{B}(S) = \sigma(\mathcal{T}_S)$ and $\mathcal{B}(T) = \sigma(\mathcal{T}_T)$. By continuity of f, $f^{-1}\mathcal{T}_T \subset \mathcal{T}_S \subset \mathcal{B}(S)$. By Lemma 1.4, $f^{-1}\mathcal{B}(T) \subset \mathcal{B}(S)$. $\square$

Let $\mathcal{C} \subset 2^S$ and $A \subset S$. We define

$$A \cap \mathcal{C} = \{A \cap B : B \in \mathcal{C}\} \subset 2^A.$$

It is clear that if $\mathcal{C}$ is a σ-algebra in S, then $A \cap \mathcal{C}$ is a σ-algebra in A. We then call $(A, A \cap \mathcal{C})$ a (measurable) subspace of $(S, \mathcal{C})$. This definition mimics that of topological subspaces.

Lemma 1.6 (slight variation)**.** *If $A \subset S$ and $\mathcal{C} \subset 2^S$, then $\sigma_A(A \cap \mathcal{C}) = A \cap \sigma_S(\mathcal{C})$. Here we use $\sigma_A(\cdot)$ (resp. $\sigma_S(\cdot)$) to denote the σ-algebra in A (resp. S) generated by some class.*

Proof. Since $\mathcal{C} \subset \sigma_S(\mathcal{C})$, $A \cap \mathcal{C} \subset A \cap \sigma_S(\mathcal{C})$. Since the RHS is a σ-algebra in A, we get $\sigma_A(A \cap \mathcal{C}) \subset A \cap \sigma_S(\mathcal{C})$. To prove the other direction, let $\overline{S}$ denote the class of $B \subset S$ such that $A \cap B \in \sigma_A(A \cap \mathcal{C})$. Then $\overline{S}$ contains $\mathcal{C}$ and $A \cap \overline{S} \subset \sigma_A(A \cap \mathcal{C})$. Since $\sigma_A(A \cap \mathcal{C})$ is a σ-algebra in A, it is easy to see that $\overline{S}$ is a σ-algebra in S. Thus, $\overline{S} \supset \sigma_S(\mathcal{C})$, and so $A \cap \sigma_S(\mathcal{C}) \subset \sigma_A(A \cap \mathcal{C})$. $\square$

Suppose $(S, \mathcal{C})$ is a topological space, and $A \subset S$. Then A is a topological subspace with topology $A \cap \mathcal{C}$. By Lemma 1.6, $\mathcal{B}(A) = A \cap \mathcal{B}(S)$, and so A is also a measurable subspace of S.

Lemma 1.7 (composition)**.** *For three measurable spaces $(S, \overline{S})$, $(T, \overline{T})$, and $(U, \overline{U})$, and two measurable mappings $f : S \to T$ and $g : T \to U$, the composition $g \circ f : S \to U$ is measurable.*

Proof. We have $(g \circ f)^{-1}\overline{U} = f^{-1}g^{-1}\overline{U} \subset f^{-1}\overline{T} \subset \overline{S}$. $\square$

Lemma 1.8. *Let $(\Omega, \mathcal{A})$ and $(S_t, \overline{S}_t)$, $t \in T$. be measurable spaces. Let $U \subset \prod_t S_t$ and $f : \Omega \to U$. Then f is $U \cap \prod_t \overline{S}_t$-measurable if and only if for each $t \in T$, $f_t := \pi_t \circ f$ is $\overline{S}_t$-measurable, where $\pi_t : \prod_r S_r \to S_t$ is the t-th coordinate map.*

Proof. Suppose f is $U \cap \prod_t \overline{S}_t$-measurable. Fix $t \in T$ and $B \in \overline{S}_t$. We have

$$f_t^{-1} B = f^{-1}(B \times \prod_{s \neq t} S_s) = f^{-1}(U \cap (B \times \prod_{s \neq t} S_s)) \in \mathcal{A}.$$

So f_t is $\overline{S}_t$-measurable. Now suppose each f_t is $\overline{S}_t$-measurable. Then for each cylinder set in S^T of the form $B \times \prod_{s \neq t} S_s$, $B \in \overline{S}_t$, we have $f^{-1}(B \times \prod_{s \neq t} S_s) = f_t^{-1} B \in A$. Since the class of such cylinder sets generates the σ-algebra $\prod_t \overline{S}_t$, by Lemma 1.4, $f^{-1} \prod_t \overline{S}_t \subset \mathcal{A}$. Thus, f is $\prod_t \overline{S}_t$-measurable if we treat it as a function from Ω to $\prod_t S_t$. For any $A \in U \cap \prod_t \overline{S}_t$, there is $B \in \prod_t \overline{S}_t$ such that $A = U \cap B$. Then $f^{-1} A = f^{-1} B \in \mathcal{A}$. So f is $U \cap \prod_t \overline{S}_t$-measurable. □

Recall that $\sigma(f) = f^{-1} \prod_t \overline{S}_t$ and $\sigma(f_t) = f_t^{-1}$, $t \in T$, are the σ-algebras in Ω induced by f and f_t, respectively. Let

$$\sigma(f_t : t \in T) = \sigma(\bigcup_{t \in T} \sigma(f_t)),$$

and we call it the σ-algebra generated by f_t, $t \in T$.

Corollary . *$\sigma(f) = \sigma(f_t : t \in T)$.*

Proof. This follows immediately from Lemma 1.8. We leave it as an exercise. □

We use the following symbols:

$$\mathbb{R}_+ = [0, \infty), \quad \overline{\mathbb{R}} = [-\infty, \infty], \quad \overline{\mathbb{R}}_+ = [0, \infty].$$

The latter two spaces have Borel σ-algebras

$$\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{B}, \{\infty\}, \{-\infty\}), \quad \mathcal{B}(\overline{\mathbb{R}}_+) = \sigma(\mathcal{B}(\mathbb{R}_+), \{\infty\}).$$

We now fix a measurable space $(\Omega, \mathcal{A})$. A function f from Ω into an interval $I \subset \mathbb{R}$ is measurable if and only if for any $x \in I$, $\{\omega : f(\omega) \leq x\}$ is measurable. This follows from Lemma 1.4 and the fact that the class $(-\infty, x] \cap I$, $x \in I$, generates the σ-algebra $\mathcal{B}(I) = I \cap \mathcal{B}$. We will often write $\{f \leq x\}$ for $\{\omega : f(\omega) \leq x\}$. The inequality $\leq x$ may be replaced by $< x$, $\geq x$, or $> x$. The statements also hold for $I = \overline{\mathbb{R}}$ or $\overline{\mathbb{R}}_+$.

Lemma 1.9. *For any sequence of measurable functions $f_1, f_2, \ldots$ from $(\Omega, \mathcal{A})$ into $\overline{\mathbb{R}}$, $\sup_n f_n$, $\inf_n f_n$, $\limsup f_n$ and $\liminf f_n$ are also measurable.*

Proof. We use the equalities

$$\{\sup_n f_n \leq x\} = \bigcap_n \{f_n \leq x\}, \quad \{\inf_n f_n \geq x\} = \bigcap_n \{f_n \geq x\},$$

$$\limsup f_n = \inf_n \sup_{m \geq n} f_m, \quad \liminf f_n = \sup_n \inf_{m \geq n} f_m.$$

□

This lemma in particular implies that the limit of measurable functions (if it exists pointwise) is measurable. This statement also holds for a general metric space.

Lemma 1.10. *Let $f_1, f_2, \dots$ be measurable functions from $(\Omega, \mathcal{A})$ into some metric space (S, ρ). Then*

(i) If $f_n \to f$, then f is measurable.

(ii) If (S, ρ) is separable and complete, then $\{\omega : \lim f_n(\omega)$ converges$\}$ is measurable.

Proof. (i) If $f_n \to f$, then for any continuous function $g : S \to \mathbb{R}$, we have $g \circ f_n \to g \circ f$. So $g \circ f$ from Ω to $\mathbb{R}$ is measurable by Lemmas 1.5, 1.7 and 1.9. Fixing an open set $G \subset S$. We may choose some continuous functions $g_n : S \to \mathbb{R}_+$ such that $g_n \uparrow \mathbf{1}_G$. In fact, we may let

$$g_n(s) = \min\{1, n\rho(s, G^c)\},$$

where $\rho(s, G^c) = \inf\{\rho(s,t) : t \in G^c\}$ is the distance from s to G^c, which is continuous in s by the triangle inequality. Since each $g_n \circ f$ is measurable, $\mathbf{1}_G \circ f = \mathbf{1}_{f^{-1}G}$ is measurable. So $f^{-1}(G)$ is measurable for every open set G. By Lemma 1.4, f is measurable.

(ii) Since S is complete, $\lim f_n(\omega)$ converges if and only if $(f_n(\omega))$ is a Cauchy sequence in S. Now

$$\{\omega : (f_n(\omega)) \text{ is Cauchy in } S\} = \bigcap_m \bigcup_N \bigcap_{n_1 \ge N} \bigcap_{n_2 \ge N} \{\omega : \rho(f_{n_1}(\omega), f_{n_2}(\omega)) < \frac{1}{m}\}.$$

This formula is another way to state that $(f_n(\omega))$ is a Cauchy sequence if and only if for any $m \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for any $n_1, n_2 \ge N$, $\rho(f_{n_1}(\omega), f_{n_2}(\omega)) < \frac{1}{m}$. To prove that the set on the RHS is measurable it suffices to show that for any m, n_1, n_2, $\{\omega : \rho(f_{n_1}(\omega), f_{n_2}(\omega)) < \frac{1}{m}\}$ is measurable. For that purpose, we use the fact that

(i) by Lemma 1.8, $(f_{n_1}, f_{n_2}) : \Omega \to S^2$ is $\mathcal{A}/\mathcal{B}(S)^2$-measurable;

(ii) the map $S^2 \ni (s_1, s_2) \mapsto \rho(s_1, s_2) \in \mathbb{R}_+$ is continuous (follows easily from the triangle inequality), and so by Lemma 1.5 is measurable w.r.t. $\mathcal{B}(S^2)$;

(iii) by Lemma 1.2, $\mathcal{B}(S^2) = \mathcal{B}(S)^2$; (we use the separability of S here);

(iv) by Lemma 1.7, $\rho(f_{n_1}, f_{n_2}) : \Omega \to \mathbb{R}_+$ is $\mathcal{A}$-measurable.

□

Lemma 1.12. *For any measurable function $f, g : (\Omega, \mathcal{A}) \to \mathbb{R}$ and $a, b \in \mathbb{R}$, $af + bg$ and fg are measurable. If, in addition, g does not take value 0, then f/g is measurable.*

Proof. To prove the measurability of $af + bg$, we express $af + bg$ as the composition of the map $(f, g) : \Omega \to \mathbb{R}^2$ and the continuous function $\mathbb{R}^2 \ni (x, y) \mapsto ax + by \in \mathbb{R}$. The proof for fg is similar. For f/g, we express f/g as the composition of $(f, g) : \Omega \to \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ and the the continuous function $\mathbb{R} \times (\mathbb{R} \setminus \{0\}) \ni (x, y) \mapsto x/y \in \mathbb{R}$. □

For any $A \subset \Omega$, we define the associated indicator function $\mathbf{1}_A : \Omega \to \mathbb{R}$ to be equal to 1 on A and to 0 on A^c. Sometimes we write $\mathbf{1}A$ instead of $\mathbf{1}_A$. It is clear that $\mathbf{1}_A$ is measurable (w.r.t. $\mathcal{A}$) if and only if A is a measurable set (w.r.t. $\mathcal{A}$).

Linear combinations of indicator functions are called simple functions. Thus, a simple function $f : \Omega \to \mathbb{R}$ is of the form

$$f = c_1 \mathbf{1}_{A_1} + \cdots c_n \mathbf{1}_{A_n},$$

where $n \in \mathbb{N}$, $A_1, \dots, A_n \subset \Omega$ and $c_1, \dots, c_n \in \mathbb{R}$. Here we only allow finite sums. If $A_1, \dots, A_n \in \mathcal{A}$, then f is $\mathcal{A}$-measurable, and called a measurable simple function.

Lemma 1.11. *For any measurable function $f : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}_+$, there exist a sequence of measurable simple functions $f_n : (\Omega, \mathcal{A}) \to \mathbb{R}_+$ such that $f_n \uparrow f$.*

We use the following symbols from now on. For $a, b \in \overline{\mathbb{R}}$, we use $a \wedge b$ and $a \vee b$ to denote $\min\{a, b\}$ and $\max\{a, b\}$, respectively. The symbols also extend to $a_1 \wedge \cdots \wedge a_n$, $a_1 \vee \cdots \vee a_n$, $\wedge_t a_t$, and $\vee_t a_t$, where the latter two are alternative ways to write $\inf_t a_t$ and $\sup_t a_t$.

For $x \in \mathbb{R}$, we use $\lfloor x \rfloor$ to denote the biggest integer n with $n \le x$, and use $\lceil x \rceil$ to denote the smallest integer n with $n \ge x$. Then $\lfloor x \rfloor$ and $\lceil x \rceil$ are monotone increasing.

Proof. We let

$$f_n = \frac{\lfloor 2^n (f \wedge n) \rfloor}{2^n}, \quad n \in \mathbb{N}.$$

Then $0 \le f_n \le f \wedge n$. We se that f_n is a simple measurable function because it takes values in $\{\frac{k}{2^n} : 0 \le k \le n2^n\}$,

$$f_n^{-1}(\{\frac{k}{2^n}\}) = \{\omega : \frac{k}{2^n} \le f(\omega) < \frac{k+1}{2^n}\}, \quad 0 \le k < n2^n, \tag{1.4}$$

$$f_n^{-1}(\{\frac{n2^n}{2^n}\}) = \{\omega : n \le f(\omega)\},$$

and the sets on the RHS are all measurable. To see that (f_n) is increasing in n, we use the inequality

$$\frac{\lfloor 2^n (f \wedge n) \rfloor}{2^n} \le \frac{\lfloor 2^n (f \wedge (n+1)) \rfloor}{2^n} \le \frac{\lfloor 2^{n+1} (f \wedge (n+1)) \rfloor}{2^{n+1}},$$

where the second "$\le$" follows from $\lfloor 2x \rfloor \ge 2\lfloor x \rfloor$. Finally, we show that $f_n \to f$ pointwise. Fix $\omega \in \Omega$. If $f(\omega) = \infty$, then $f_n(\omega) = n \to f(\omega)$. Suppose $f(\omega) < \infty$. Let $\varepsilon > 0$. We may choose N such that $N > f(\omega)$ and $\frac{1}{2^N} < \omega$. For $n \ge N$, by (1.4), we get the inequality $|f_n(\omega) - f(\omega)| \le \frac{1}{2^n} < \varepsilon$. □

We say that two measurable spaces $(S, \overline{S})$ and $(T, \overline{T})$ are Borel isomorphic if there is a bijection $f : S \to T$ such that both f and f^{-1} are measurable. This means that $f^{-1}\overline{T} = \overline{S}$ and $f\overline{S} = \overline{T}$. A space S that is Borel isomorphic to a Borel subset I of $[0, 1]$, equipped with the Borel σ-algebra $\mathcal{B}(I) = I \cap \mathcal{B}([0, 1])$, is called a Borel space. By the following lemma, a Polish space is a Borel space.

Definition . A Polish space is a topological space, which admits a separable and complete metrization.

Lemma A1.6. *A Polish space S is a Borel space.*

Sketch of the proof. The first step is to construct a continuous and injective function $f : S \to [0,1]^\infty$. Let (s_n) be a dense sequence in S. Then we define $f(x) = (1 \wedge \rho(x, s_n))$. The second step is to use binary expansions to construct a measurable injective function $g : [0,1]^\infty \to [0,1]$. See Chapter 13 of Dudley, R.M.'s "Real Analysis and Probability" for details. □

For two functions $f : \Omega \to (S, \overline{S})$ and $g : \Omega \to (T, \overline{T})$, where $(S, \overline{S})$ and $(T, \overline{T})$ are measurable spaces, we say that f is g-measurable if $\sigma(f) \subset \sigma(g)$, or equivalently, $f^{-1}\overline{S} \subset g^{-1}\overline{T}$. If there is a $(\overline{T}/\overline{S}$-)measurable map $h : T \to S$ such that $f = h \circ g$, then

$$f^{-1}\overline{S} = g^{-1}h^{-1}\overline{S} \subset g^{-1}\overline{T}.$$

So f is g-measurable. Under some mild conditions, the converse is also true.

Lemma 1.13. *Under the above setup, if $(S, \overline{S})$ is a Borel space, then f is g-measurable if and only if there exists some measurable map $h : T \to S$ such that $f = h \circ g$.*

Proof. We only need to show the "only if" part. Since S is Borel, we may assume that $S \in \mathcal{B}([0,1])$. We may then view f as a map from Ω into $[0,1]$. This new viewpoint does not change $\sigma(f)$. So f is still g-measurable. If in this case, there exists a measurable map $\widetilde{h} : T \to [0,1]$ such that $f = \widetilde{h} \circ g$. Then we may define h such that $h = \widetilde{h}$ on $\widetilde{h}^{-1}(S)$, and $h = s_0$ on $\widetilde{h}^{-1}([0,1] \setminus S)$, where s_0 is a fixed point in S. Then $h : T \to S$ is measurable, and $f = h \circ g$. Thus, it suffices to assume that $S = [0,1]$.

If $f = \mathbf{1}_A$, and $A \in \sigma(g)$, then $A = g^{-1}B$ for some $B \in \overline{T}$. So $f = \mathbf{1}_B \circ g$ and we may choose $h = \mathbf{1}_B$. The result extends by linearity to any g-measurable simple functions. In the general case, by Lemma 1.11, there exists a sequence of g-measurable simple functions $f_n : \Omega \to [0,1]$ such that $f_n \uparrow f$. For each n, there exists an $\overline{T}$-measurable map $h_n : T \to [0,1]$ such that $f_n = h_n \circ g$. Then $h := \sup_n h_n : T \to [0,1]$ is also $\overline{T}$-measurable by Lemma 1.9. Finally, we note that

$$h \circ g = (\sup_n h_n) \circ g = \sup_n (h_n \circ g) = \sup_n f_n = f.$$

□

Definition . A measure on a measurable space $(\Omega, \mathcal{A})$ is a map $\mu : \mathcal{A} \to \overline{\mathbb{R}}_+$, which satisfies $\mu\emptyset = 0$ and

$$\mu \bigcup_n A_n = \sum_n \mu A_n, \quad \text{for all mutually disjoint } A_1, A_2, \cdots \in \mathcal{A}. \tag{1.5}$$

The triple $(\Omega, \mathcal{A}, \mu)$ is then called a measure space. The measure μ is called finite if $\mu\Omega < \infty$, and is called a probability measure if $\mu\Omega = 1$. In the latter case, $(\Omega, \mathcal{A}, \mu)$ is called a probability space. The μ is called a σ-finite measure if there is a sequence $A_1, A_2, \cdots \in \mathcal{A}$ such that $\Omega = \bigcup_n A_n$ and $\mu A_n < \infty$ for each n.

Remark . The property (1.5) is called *countably additivity*, which clearly implies *finitely additivity*:

$$\mu \bigcup_{n=1}^{N} A_n = \sum_{n=1}^{N} \mu A_n, \quad \text{for all mutually disjoint } A_1, A_2, \dots A_n \in \mathcal{A},$$

by setting $A_n = \emptyset$ for $n > N$, and *countably subadditivity*:

$$\mu \bigcup_n B_n \le \sum_n \mu B_n, \quad \text{for all } B_1, B_2, \dots \in \mathcal{A},$$

by defining $A_n = B_n \setminus \bigcup_{k<n} B_k$.

Lemma 1.14 (Continuity)**.** *Let μ be a measure on $(\Omega, \mathcal{A})$, and let $A_1, A_2, \dots \in \mathcal{A}$.*

(i) If $A_n \uparrow A$, then $\mu A_n \uparrow \mu A$.

(ii) If $A_n \downarrow A$, and $\mu A_1 < \infty$, then $\mu A_n \downarrow \mu A$.

Proof. (i) We apply (1.5) to $D_n = A_n \setminus A_{n-1}$ with $A_0 = \emptyset$. (ii) We apply (i) to $B_n = A_1 \setminus A_n$. Since $\mu A_1 < \infty$, we have $\mu A_n < \infty$ as well, and $\mu B_n = \mu A - \mu A_n \uparrow \mu A_1 - \mu A$. □

Exercise . Suppose $\mu : \mathcal{A} \to \mathbb{R}_+$ satisfies finitely additivity and the property that if $B_1 \supset B_2 \supset \dots \in \mathcal{A}$, and there is $\varepsilon > 0$ such that $\mu B_n \ge \varepsilon > 0$ for all n, then $\bigcap_n B_n \ne \emptyset$. Prove that μ is a measure.

Exercise . Prove that for two measures μ and ν on $(\Omega, \mathcal{A})$ with $\mu\Omega = \nu\Omega < \infty$, the class $\mathcal{D} = \{A \in \mathcal{A} : \mu A = \nu A\}$ is a λ-system.

By monotone class theorem and the above exercise, we conclude that if two probability measures on $(\Omega, \mathcal{A})$ agree on a π-system $\mathcal{C}$ with $\sigma(\mathcal{C}) = \mathcal{A}$, then the two measures must agree.

We may do the following operations on measures. If μ is a measure, and $c \in \mathbb{R}_+$, then $c\mu$ is also a measure. If μ is finite, then $\frac{1}{\mu\Omega}\mu$ is a probability measure. The sum of two measures is a measure. If (μ_n) is an increasing sequence of measures, then $\lim \mu_n$ is also a measure; if (μ_n) is a decreasing sequence of measures, and μ_1 is finite, then $\lim \mu_n$ is also a measure (Lemma 1.15). Thus, if $\mu_1, \mu_2, \dots$ are measures on the same space, then $\sum_n \mu_n$ is a measure.

If μ is a measure on $(\Omega, \mathcal{A})$ and $B \in \mathcal{A}$, then $\mu(\cdot \cap B) : \mathcal{A} \ni A \mapsto \mu(A \cap B)$ is also a measure on $(\Omega, \mathcal{A})$. It is called the restriction of μ to B. One may also view the restriction as a measure on the measurable subspace $(B, B \cap \mathcal{A})$.

The simplest measure is the zero measure, which takes value zero at all $A \in \mathcal{A}$. Another natural measure is the counting measure: $\mu A = \#(A)$ if A is finite; $\mu A = \infty$ if otherwise. For $s \in \Omega$, the *Dirac measure* (also called point mass) δ_s is defined by $\delta_s(A) = 1$ if $s \in A$, and $\delta_s(A) = 0$ if otherwise.

The most important nontrivial measure is the *Lebesgue measure* λ. It is the unique measure on $(\mathbb{R}, \mathcal{B})$ such that for any interval I, λI equals $|I|$, the length of I. It is σ-finite because $\mathbb{R} = \bigcup_{n\in\mathbb{Z}} [n, n+1)$. The proof uses the Carathéodory extension theorem stated below.

We call a class $\mathcal{R} \subset 2^\Omega$ a ring if it contains $\emptyset$ and is closed under finite union and difference, i.e., $A, B \in \mathcal{R}$ implies that $A \cup B, A \setminus B \in \mathcal{R}$. A map $\mu : \mathcal{R} \to \overline{\mathbb{R}}_+$ is called a pre-measure if $\mu\emptyset = 0$ and μ satisfies countably additivity, i.e., if $A_1, A_2, \dots \in \mathcal{R}$ is a partition of $A \in \mathcal{R}$, then $\mu A = \sum_n \mu A_n$. By considering the sets $B_n = A \setminus \bigcup_{k=1}^n B_k$, we find that countably additivity is equivalent to the combination of finitely countability and the statement that for any $B_1 \supset B_2 \supset \dots \in \mathcal{R}$, if there is $\varepsilon > 0$ such that $\mu B_n \ge \varepsilon$ for all n, then we have $\bigcap_n B_n \neq 0$. If $\mathcal{R}$ has a partition $A_1, A_2, \dots \in \mathcal{R}$ such that $\mu A_n < \infty$ for each n, then μ is called σ-finite.

Theorem (Carathéodory extension theorem)**.** *A pre-measure μ on a ring $\mathcal{R}$ extends to a measure on $\sigma(\mathcal{R})$. The extension is unique if μ is σ-finite.*

We will only give a sketch of the proof of Carathéodory extension theorem, but will provide details of the application of the theorem in constructing the Lebesgue measure because similar arguments will be used later.

Proof of Carathéodory extension theorem (Sketch). The uniqueness part follows from a monotone class argument. Note that for any n, the class $A_n \cap \mathcal{R}$ is a π-system in A_n, and if μ_1 and μ_2 are two extensions, then the set of $B \in A_n \cap \sigma(\mathcal{R})$ such that $\mu_1 B = \mu_2 B$ form a λ-system in A_n. The existence part uses *outer measures.* For every $A \subset \Omega$, we define the outer measure of A by

$$\mu^* A = \inf_{\mathcal{R} \ni I \supset A} \mu I.$$

It is clear that $\mu^* = \mu$ on $\mathcal{R}$. Then we consider the set $\mathcal{F}$ of all $A \subset \Omega$ such that for every $E \subset \Omega$,

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

Then one can prove the following statements:

(i) $\mathcal{F}$ is a σ-algebra containing $\mathcal{R}$;

(ii) μ^* restricted to $\mathcal{F}$ is a measure.

By (i), $\mathcal{F} \subset \sigma(\mathcal{R})$. By (ii), $\mu^*|_{\sigma(\mathcal{R})}$ is the extension that we want. □

To construct Lebesgue measure, we define a ring $\mathcal{R}$ in $\mathbb{R}$ to be the class of finite disjoint unions of intervals of the form $(a, b]$, where $a < b \in \mathbb{R}$. For an element $A \in \mathcal{R}$ expressed as disjoint union $\bigcup_{k=1}^m (a_k, b_k]$, we define $\mu A = \sum_{k=1}^m (b_k - a_k)$. It is easy to check that μ satisfies finitely additivity. Then we need to show that, if $A_1 \supset A_2 \dots \in \mathcal{R}$, and $\mu A_n \ge \varepsilon > 0$ for all n, then $\bigcap_n A_n \neq \emptyset$. For each n, we may pick $A_n' \in \mathcal{R}$ such that $\overline{A_n'} \subset A_n$ and $\mu(A_n \setminus A_n') < \varepsilon/2^n$ (if $A_n = \bigcup_{k=1}^m (a_k, b_k]$, we set $A_n' = \bigcup_{k=1}^m (a_k', b_k]$ such that $a_k < a_k' < b_k$ and $a_k' - a_k$ is small enough). Let $A_n'' = \bigcap_{k=1}^n A_n'$. Then $\overline{A_n''} \subset A_n$ for each n, and $A_1'' \supset A_2'' \supset \cdots$. Since $A_n \setminus A_n'' \subset \bigcup_{k=1}^n (A_k \setminus A_k')$, we get $\mu(A_n \setminus A_n'') \le \sum_{k=1}^n \mu(A_k \setminus A_k') < \sum_{k=1}^n \frac{\varepsilon}{2^k} < \varepsilon$. From $\mu A_n > \varepsilon$ we get $\mu A_n'' > 0$, and so $A_n'' \neq \emptyset$. Since each $\overline{A_n''}$ is compact and $\overline{A_1''} \supset \overline{A_2''} \supset \cdots$, we get $\bigcap_n \overline{A_n''} \neq \emptyset$, which together with $\overline{A_n''} \subset A_n$ implies that $\bigcap_n A_n \neq \emptyset$. So μ is a pre-measure on $\mathcal{R}$. We may then use Carathéodory extension theorem to extend μ to a measure on $\mathbb{R}$. It is easy to check that the extension is the Lebesgue measure.

Lemma 1.16 (Regularity)**.** *Let μ be a finite measure on some metric space S. Then for any $B \in \mathcal{B}(S)$,*

$$\mu B = \sup_{F \subset B} \mu F = \inf_{G \supset B} \mu G, \tag{1.6}$$

with F and G restricted to the classes of closed and open subsets of S, respectively.

Proof. Let $\mathcal{C}$ denote the set of B which satisfies (1.6). Then (i) $S \in \mathcal{C}$ because S is both closed and open; (ii) $B \in \mathcal{C}$ implies that $B^c \in \mathcal{C}$ since $F \subset B$ and F is closed if and only if $F^c \supset B^c$ and F^c is open; (iii) $B^1, B^2 \in \mathcal{C}$ implies that $B^1 \cup B^2 \in \mathcal{C}$ because if for $j = 1, 2$, closed sets $F_n^j \subset B^j$, $n \in \mathbb{N}$, satisfy $\mu F_n^j \to \mu B^j$ and open sets $G_n^j \supset B^j$, $n \in \mathbb{N}$, satisfy $\mu G_n^j \to \mu B^j$, then $\mu(F_n^1 \cup F_n^2) \to \mu(B^1 \cup B^2)$ and $\mu(G_n^1 \cup G_n^2) \to \mu(B^1 \cup B^2)$. The first follows from

$$(B^1 \cup B^2) \setminus (F_n^1 \cup F_n^2) \subset (B^1 \setminus F_n^1) \cup (B^2 \setminus F_n^2),$$

and the second is similar. The (ii) and (iii) together imply that $\mathcal{C}$ is closed under difference. Suppose (B_n) is an increasing sequence in $\mathcal{C}$, and $B = \bigcup_n B_n$. Fix any $\varepsilon > 0$. We may first choose n such that $\mu B_n > \mu B - \varepsilon/2$, and then choose closed $F \subset B_n$ such that $\mu F > \mu B_n - \varepsilon/2$. Since $F \subset B$ and $\mu F > \mu B - \varepsilon$, we get $\mu B = \sup_{F \subset B} \mu F$. On the other hand, for each $n \in \mathbb{N}$, we may choose open $G_n \supset B_n$ such that $\mu G_n < \mu B_n + \frac{\varepsilon}{2^n}$. Let $G = \bigcup_n G_n$. Then G is open, $G \supset B$, and $\mu(G \setminus B) < \sum_n \frac{\varepsilon}{2^n} = \varepsilon$. Thus, $\mu B = \inf_{G \supset B} \mu G$. So $B \in \mathcal{C}$. Hence $\mathcal{C}$ is a λ-system. We also know that $\mathcal{C}$ contains all open sets since every open set G can be written as a union of an increasing sequence of closed sets. By monotone class theorem, $\mathcal{C}$ contains the Borel σ-algebra $\mathcal{B}(S)$, i.e., (1.6) holds for any $B \in \mathcal{B}(S)$. □