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

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

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 $1 \le j \le m$. 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). □

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 $f^{-1}\overline{T} \subset \overline{S}$. □

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, $f^{-1}\mathcal{B}(T) \subset \mathcal{B}(S)$. □

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 $A \cap \sigma_S(\mathcal{C}) \subset \sigma_A(A \cap \mathcal{C})$. □

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 $(g \circ f)^{-1}\overline{U} = f^{-1}g^{-1}\overline{U} \subset f^{-1}\overline{T} \subset \overline{S}$. □

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 $f_1, f_2, \ldots$ 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 $\{\omega : \lim f_n(\omega) \text{ converges}\}$ 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

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

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 $n_1, n_2 \geq N$, $\rho(f_{n_1}(\omega), f_{n_2}(\omega)) < \frac{1}{m}$. 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 $B \in \mathcal{B}(S)$. □

Let μ be a measure on $(S, \overline{S})$, and f is a measurable map from $(S, \overline{S})$ into $(T, \overline{T})$, then we get a measure $\mu \circ f^{-1}$ (also denoted by $f_* \mu$) on $(T, \overline{T})$ defined by

$$(\mu \circ f^{-1}) A = \mu f^{-1} A.$$

It is called the pushforward of μ under f.

Given a measure space $(\Omega, \mathcal{A}, \mu)$, we are going to define the integral

$$\mu f = \int f d\mu = \int f(\omega) \mu(d\omega)$$

for certain real valued measurable function f on $(\Omega, \mathcal{A})$. The construction is composed of several steps.

Step 1. If f is a nonnegative measurable simple function of the form

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

with $c_1, \ldots, c_n \in \mathbb{R}_+$ and $A_1, \ldots, A_n \in \mathcal{A}$, we define

$$\mu f = c_1 \mu A_1 + \cdots + c_n \mu A_n.$$

Throughout measure theory we follow the convention that $0\cdot\infty=0$. Using the finite additivity of μ, one can show that the definition is consistent, i.e., if f has another expression: $d_1\mathbf{1}_{B_1}+\cdots d_m\mathbf{1}_{B_m}$, then $d_1\mu B_1+\cdots+d_m\mu B_m$ equals the same number. We then get linearity and monotonicity: for nonnegative measurable simple functions f and g:

$$\mu(af+bg)=a\mu f+b\mu g,\quad \text{for } a,b\ge 0; \tag{1.7}$$

$$\mu f\ge \mu g\ge 0,\quad \text{if } f\ge g. \tag{1.8}$$

Exercise . Check the consistency and formulas (1.7) and (1.8).

Step 2. If $f:\Omega\to\overline{\mathbb{R}}_+$ is measurable, by Lemma 1.11 we may choose a sequence of nonnegative measurable simple functions (f_n) such that $f_n\uparrow f$. Then we define

$$\mu f=\lim \mu f_n.$$

We also need to prove the consistency, i.e., the definition does not depend on the choice of (f_n).

Lemma 1.18. *Let $f_1,f_2,\cdots$ and g be simple measurable functions on Ω such that $0\le f_1\le f_2\le\cdots$ and $0\le g\le \lim f_n$. Then $\lim\mu f_n\ge \mu g$.*

Proof. First suppose $g=c\mathbf{1}_A$ for $c\in\mathbb{R}_+$ and $A\in\mathcal{A}$. If $c=0$, it is trivial. For $c>0$, fix $\varepsilon\in(0,c)$ and let $A_n=A\cap\{f_n\ge c-\varepsilon\}$. Then $A_n\uparrow A$, and so

$$\mu f_n\ge \mu(c-\varepsilon)\mathbf{1}_{A_n}=(c-\varepsilon)\mu A_n\uparrow(c-\varepsilon)\mu A.$$

So $\lim\mu f_n\ge(c-\varepsilon)\mu A$. Letting $\varepsilon\to 0$, we get $\lim\mu f_n\ge c\mu A=\mu g$.

Now suppose $g=c_1\mathbf{1}_{A_1}+\cdots c_m\mathbf{1}_{A_m}$ with $c_1,\dots,c_m\in\mathbb{R}_+$ and $A_1,\dots,A_m\in\mathcal{A}$. We may assume that $A_1,\dots,A_m$ are mutually disjoint. Let $\mu_k=\mu(\cdot\cap A_k)$, $1\le k\le m$, and $\mu_0=\mu(\cdot\cap(\bigcup_k A_k)^c)$. Then $\mu=\sum_{k=0}^n\mu_k$. So $\mu f_n\ge\sum_{k=1}^m\mu_k f_n$. For $1\le k\le m$, since $\lim_n f_n\ge g\ge c_k\mathbf{1}_{A_k}$, by the above paragraph we get $\lim_n\mu_k f_n\ge c_k\mu A_k$. Thus,

$$\lim_n \mu f_n\ge\lim_n\sum_{k=1}^m\mu_k f_n=\sum_{k=1}^m\lim_n\mu_k f_n\ge\sum_{k=1}^m c_k\mu A_k=\mu g.$$

□

Applying this lemma, we see that if (f_n) and (g_m) are two sequences of measurable simple functions with $0\le f_n\uparrow f$ and $0\le g_m\uparrow f$, then for each m, $\lim_n\mu f_n\ge\mu g_m$. So $\lim_n\mu f_n\ge\lim_m\mu g_m$. By symmetry, we have $\lim_m\mu g_m\ge\lim_n\mu f_n$. So $\lim_n\mu f_n=\lim_m\mu g_m$, and we get the consistency in the definition of μf.

We can easily prove the linearity and monotonicity: for measurable functions f and g from Ω into $\overline{\mathbb{R}}_+$, (1.7) and (1.8) both hold.

Theorem 1.19 (Monotone Convergence Theorem)**.** *Let $f_1,f_2,\cdots:(\Omega,\mathcal{A})\to\overline{\mathbb{R}}_+$ be measurable. Suppose $f_n\uparrow f$. Then $\mu f_n\uparrow\mu f$.*

Proof. For each n, we choose a sequence of measurable simple functions (g_k^n) such that $g_k^n \uparrow f_n$ as $k \to \infty$. Then $\mu f_n = \lim_k \mu g_k^n$. Define

$$h_k = g_k^1 \vee g_k^2 \vee \cdots \vee g_k^k.$$

Then (h_k) is an increasing sequence of nonnegative simple measurable functions. Since for each $k \in \mathbb{N}$, $h_k \le f_1 \vee f_2 \vee \cdots f_k = f_k \le f$, we have $\lim h_k \le f$ and

$$\lim \mu h_k \le \lim \mu f_k \le \mu f. \tag{1.9}$$

For any fixed $n \in \mathbb{N}$, we have $h_k \ge g_k^n$ for $k \ge n$. So $\lim h_k \ge \lim_k g_k^n = f_n$. Thus, $\lim h_k \ge \sup f_n = f$. So we get $h_k \uparrow f$ and $\mu f = \lim \mu h_k$. By (1.9) we get $\lim \mu f_k = \mu f$. □

Lemma 1.20 (Fatou)**.** *For any measurable functions $f_1, f_2, \cdots : (\Omega, \mathcal{A}) \to \overline{\mathbb{R}}_+$, we have*

$$\liminf \mu f_n \ge \mu \liminf f_n.$$

Proof. Fix $n \in \mathbb{N}$. Since $f_k \ge \inf_{m \ge n} f_m$ for all $k \ge n$, by monotonicity,

$$\inf_{k \ge n} \mu f_k \ge \mu \inf_{m \ge n} f_m.$$

Letting $n \to \infty$ and using monotone convergence theorem, we get

$$\liminf \mu f_n = \lim_n \inf_{k \ge n} \mu f_k \ge \lim_n \mu \inf_{m \ge n} f_m = \mu \lim_n \inf_{m \ge n} f_m = \mu \liminf f_n.$$

□

Step 3. We define μf for integrable functions. A measurable function $f : (\Omega, \mathcal{A}, \mu) \to \mathbb{R}$ is called integrable if $\mu|f| < \infty$. Here since $|f|$ is a nonnegative measurable function, $\mu|f|$ was defined in Step 2. For the definition, we find two nonnegative measurable functions f_1 and f_2 such that $f = f_1 - f_2$ and $\mu f_1, \mu f_2 < \infty$, and then let

$$\mu f = \mu f_1 - \mu f_2.$$

For the existence of such f_1 and f_2, we may let $f_1 = f_+ := f \vee 0$ and $f_2 = f_- := (-f) \vee 0$. In fact, we have $f_+, f_- \ge 0$, $f = f_+ - f_-$, and $|f| = f_+ + f_-$. So $0 \le f_\pm \le |f|$, which implies that $\mu f_\pm \le \mu|f| < \infty$. For the consistency, suppose g_1 and g_2 satisfy the same properties as f_1 and f_2. Then from $f_1 - f_2 = g_1 - g_2$ we get $f_1 + g_2 = g_1 + f_2$, and so $\mu f_1 + \mu g_2 = \mu g_1 + \mu f_2$. Since every item is a real number, we get $\mu f_1 - \mu f_2 = \mu g_1 - \mu g_2$. Thus, μf is well defined. Finally, since $\mu f = \mu f_+ - \mu f_-$ and $\mu|f| = \mu f_+ + \mu f_-$, we get $|\mu f| \le \mu|f|$.

We then have the monotonicity and the linearity with real coefficient: if $f, g : \Omega \to \mathbb{R}$ are integrable, and $a, b \in \mathbb{R}$, then $af + bg$ is also integrable, and $\mu(af + bg) = a\mu f + b\mu g$.

In summary, the integral μf is defined for (i) all measurable functions $f : (\Omega, \mathcal{A}, \mu) \to \overline{\mathbb{R}}_+$; and (ii) all measurable functions $f : (\Omega, \mathcal{A}, \mu) \to \mathbb{R}$ such that $\mu|f| < \infty$. In the former case, μf takes values in $\overline{\mathbb{R}}_+$, and in the latter case, μf takes values in $\mathbb{R}$.

Theorem 1.21 (Dominated Convergence)**.** *Let $f, f_1, f_2, \dots$ and $g, g_1, g_2, \dots$ be $\mathbb{R}$-valued measurable functions on $(\Omega, \mathcal{A}, \mu)$ with $|f_n| \le g_n$ for all n, and such that $f_n \to f$, $g_n \to g$, and $\mu g_n \to \mu g < \infty$. Then $\mu f_n \to \mu f$.*

Proof. The sequence $(g_n \pm f_n)$ are nonnegative measurable functions and $g_n \pm f_n \to g \pm f$. Since $\mu g < \infty$ and $\mu g_n \to \mu g$, g and g_n are integrable for all but finitely many n. Since $|f_n| \le g_n$ and $|f| \le g$, the same statement holds for g and f. By Fatou's lemma and linearity of integral,

$$\mu g \pm \mu f = \mu(g \pm f) \le \liminf \mu(g_n \pm f_n) = \liminf(\mu g_n \pm \mu f_n) = \mu g + \liminf(\pm \mu f_n).$$

So we get $\mu f \le \liminf \mu f_n$ and $-\mu f \le \liminf(-\mu f_n) = -\limsup \mu f_n$, which implies that $\limsup \mu f_n \le \mu f \le \liminf \mu f_n$. So $\lim \mu f_n = \mu f$. □

Lemma 1.22 (Substitution)**.** *Let f from a measurable map from $(\Omega, \mathcal{A}, \mu)$ to $(S, \overline{S})$. Let $\mu \circ f^{-1}$ be the pushforward measure on $(S, \overline{S})$. Then for measurable function $g : S \to \overline{\mathbb{R}}$,*

$$(\mu \circ f^{-1})g = \mu(g \circ f). \tag{1.10}$$

Here the equality means that when one side is defined, then the other side is also defined, and the two sides agree.

Proof. We first show that if $g : S \to \overline{\mathbb{R}}_+$, and so $g \circ f : \Omega \to \overline{\mathbb{R}}_+$ and both sides are well defined, then (1.10) holds. The simplest case is $g = \mathbf{1}_A$. In this case

$$(\mu \circ f^{-1})g = (\mu \circ f^{-1})A = \mu f^{-1}A = \mu \mathbf{1}_{f^{-1}A} = \mu(g \circ f).$$

By linearity, (1.10) then holds for all nonnegative measurable simple functions. By monotone convergence, (1.10) also holds for all nonnegative measurable functions.

For measurable $g : S \to \mathbb{R}$, since $|g \circ f| = |g| \circ f$, by (1.10) g is integrable w.r.t. $\mu \circ f^{-1}$ if and only if $g \circ f$ is integrable w.r.t. μ. Moreover, if $g = g_1 - g_2$ such that $g_1, g_2 : S \to \mathbb{R}$ are measurable and $(\mu \circ f^{-1})g_j < \infty$, $j = 1, 2$, then by applying (1.10) to g_j we get (1.10) for g. □

Given a measurable function $f : (\Omega, \mathcal{A}, \mu) \to \overline{\mathbb{R}}_+$, we may define another measure $f \cdot \mu$ on $(\Omega, \mathcal{A})$ by

$$(f \cdot \mu)A = \int_A f d\mu = \int \mathbf{1}_A f.$$

The countably additivity of $f \cdot \mu$ follows from monotone convergence theorem. The f is referred as the μ-density of $f \cdot \mu$.

Lemma 1.23 (Chain Rule)**.** *For any measurable maps $f, g : (\Omega, \mathcal{A}, \mu) \to \overline{\mathbb{R}}$ with $f \ge 0$,*

$$(f \cdot \mu)g = \mu(fg).$$

The meaning of the equality should be explained in the same way as (1.10), i.e., when one side is define, the other side is also defined, and the two sides agree.

Proof. As in the last proof, we may begin with the case when g is an indicator function and then extend in steps to the general case. $\square$

This lemma implies that, if $f, g : \Omega \to \overline{\mathbb{R}}_+$ are measurable, then $f \cdot (g \cdot \mu) = (fg) \cdot \mu$.

Given a measure space $(\Omega, \mathcal{A}, \mu)$, a set $A \in \mathcal{A}$ is called μ-null if $\mu A = 0$. A relation depending on $\omega \in \Omega$ is said to hold μ-almost everywhere if there is a μ-null set A such that it holds for all $\omega \in A^c$. We often write μ-a.e. or simply a.e.

Lemma 1.24. *If $f, g : (\Omega, \mathcal{A}, \mu) \to \overline{\mathbb{R}}$ satisfy that μ-a.e. $f = g$, then $\mu f = \mu g$. Again the equality means that if any of μf and μg is defined, then the other is also defined, and the two values are equal.*

Proof. First, suppose $g = 0$ and $f \geq 0$. Let (f_n) be a sequence of measurable simple functions with $0 \leq f_n \uparrow f$. Then $\{f_n \neq 0\} \subset \{f \neq 0\}$, and so $\{f_n \neq 0\}$ is a null set. We may express each f_n as $c_1 \mathbf{1}_{A_1} + \cdots c_m \mathbf{1}_{A_m}$ with $c_1, \dots, c_m \in \overline{\mathbb{R}}_+$ and $A_1, \dots, A_m$ are null sets. Then $\mu f_n = \sum c_k \mu A_k = 0$. So $\mu f = \lim \mu f_n = 0 = \mu g$.

Second, suppose $f, g \geq 0$. Let $h = f \vee g$. Then $h \geq f$ and μ-a.e., $h = f$. We may write $h = f + \phi$, where $\phi : \Omega \to \overline{\mathbb{R}}_+$ is measurable and μ-a.e., $\phi = 0$. By the first paragraph, $\mu\phi = 0$. So $\mu h = \mu f + \mu \phi = \mu f$. Similarly, $\mu h = \mu g$. So $\mu f = \mu g$.

Now we consider integrable functions. Since μ-a.e., $|f| = |g|$, by the second paragraph, $\mu|f| = \mu|g|$. So f is integrable if and only if g is integrable. Now suppose f and g are integrable. Since $f_\pm = (\pm f) \vee 0 = (\pm g) \vee 0 = g_\pm$ a.e., by the previous result we have $\mu f_\pm = \mu g_\pm$. So $\mu f = \mu f_+ - \mu f_- = \mu g_+ - \mu g_- = \mu g$. $\square$

On the other hand, if $f : (\Omega, \mathcal{A}, \mu) \to \overline{\mathbb{R}}_+$ satisfies that $\mu f = 0$, then μ-a.e. $f = 0$. In fact, since $\{f \neq 0\} = \bigcup_n \{f \geq 1/n\}$, if $\mu\{f \neq 0\} > 0$, then there is $n \in \mathbb{N}$ such that $\mu\{f \geq 1/n\} > 0$. Then we get

$$\mu f \geq \mu \frac{1}{n} \mathbf{1}_{\{f \geq 1/n\}} = \frac{1}{n} \mu\{f \geq 1/n\} > 0.$$

Since two integrals agree when two integrands agree μ-a.e., we may allow the integrands to be undefined on some μ-null sets. Monotone Convergence Theorem, Fatou's Lemma, and Dominated Convergence Theorem remain valid if the hypothesis are only fulfilled outside some null sets. We also note that if $f : \Omega \to \overline{\mathbb{R}}_+$ satisfies $\mu f < \infty$, then a.e. $f \in \mathbb{R}_+$ because from $\infty > \mu f \geq \infty \cdot \mu f^{-1}\{\infty\}$ we get $\mu f^{-1}\{\infty\} = 0$.

Definition . Let μ and ν be two measures on a measurable space $(\Omega, \mathcal{A})$. We say that ν is absolutely continuous with respect to μ and write $\nu \ll \mu$ if every μ-null set is also a ν-null set. We say that μ and ν are mutually singular and write $\mu \perp \nu$ if there is $A \in \mathcal{A}$ such that $\mu A = 0$ and $\nu A^c = 0$.

If $\nu = f \cdot \mu$, then for any μ-null set A, $\nu A = \int \mathbf{1}_A f d\mu = 0$ since μ-a.e., $\mathbf{1}_A f = 0$. So A is also a ν-null set. Thus, we have $f \cdot \mu \ll \mu$. We focus on σ-finite measures.

Theorem A1.3 (Radon-Nikodym)**.** *Let μ and ν are two σ-finite measures on* $(\Omega, \mathcal{A})$, *(i) If $\nu \ll \mu$, there there is a μ-a.e. unique measurable function $f : \Omega \to \mathbb{R}_+$ such that $\nu = f \cdot \mu$.*

(ii) In the general case, there is a μ-a.e. unique measurable function $f : \Omega \to \mathbb{R}_+$ such that $\sigma := \nu - f \cdot \mu$ is a measure that is singular to μ.

In Part (i) of the theorem, we also call f the Radon-Nikodym derivative of ν against μ. For the proof of Radon-Nikodym Theorem, we introduce the notation of real measures, which is important on its own.

Definition . Let $(\Omega, \mathcal{A})$ be a measurable space. A function $\nu : \mathcal{A} \to \mathbb{R}$ is called a real measure or signed measure if it satisfies countably additivity with $\nu\emptyset = 0$, i.e., if $A_1, A_2, \dots \in \mathcal{A}$ are mutually disjoint, then $\nu \bigcup_n A_n = \sum_n \nu A_n$, where the series converges absolutely.

A finite measure is a real measure, and the space of all real measures on $(\Omega, \mathcal{A})$ is a linear space. Thus, the difference of two finite measures is a real measure. If μ is a measure, and $f : \Omega \to \mathbb{R}$ is integrable with respect to μ, then $(f \cdot \mu)(A) := \int_A f d\mu$ is a real measure. The countably additivity follows from the Dominated Convergence Theorem.

A real measure ν satisfies continuity: if $A_n \uparrow A$ or $A_n \downarrow A$, then $\nu A_n \to \nu A$. Actually, if $A_n \uparrow A$, we may write $A = \bigcup_n (A_n \setminus A_{n-1})$ with $A_0 = \emptyset$. Since $A_n \setminus A_{n-1}$ are mutually disjoint, $\nu A = \sum_n \nu(A_n \setminus A_{n-1}) = \sum_n (\nu A_n - \nu A_{n-1}) = \lim \nu A_n$. If $A_n \downarrow A$, then $A_n^c \uparrow A^c$ and $\nu A^c = \nu\Omega - \nu A$ and $\nu A_n^c = \nu\Omega - \nu A_n$.

Theorem (Hahn decomposition)**.** *Given a real measure ν on $(\Omega, \mathcal{A})$, there exists a partition $\{P, N\}$ of Ω such that $P, N \in \mathcal{A}$, $\nu E \ge 0$ for all $E \in P \cap \mathcal{A}$, and $\nu E \le 0$ for all $E \in N \cap \mathcal{A}$.*

Proof. Let $s = \sup\{\nu A : A \in \mathcal{A}\}$. Then $s \ge 0$ since $\nu\emptyset = 0$. We now exclude the possibility that $s = +\infty$. Suppose $s = +\infty$. Let

$$\mathcal{B} = \{A \in \mathcal{A} : \sup\{\nu B : B \in \mathcal{A}, B \subset A\} = +\infty\}.$$

Then $\Omega \in \mathcal{B}$. It is also easy to see that if $A_1, A_2 \in \mathcal{A} \setminus \mathcal{B}$ and $A_1 \cap A_2 = \emptyset$, then $A_1 \cup A_2 \in \mathcal{A} \setminus \mathcal{B}$. Thus, if $A_1 \in \mathcal{B}$, $A_2 \in \mathcal{A} \setminus \mathcal{B}$, and $A_2 \subset A_1$, then $A_1 \setminus A_2 \in \mathcal{B}$. First, suppose

$$\sup\{\nu B : B \in \mathcal{B}, B \subset A\} = +\infty, \quad \forall A \in \mathcal{B}. \tag{1.11}$$

Then we can inductively construct a sequence $A_0 \supset A_1 \supset A_2 \supset \cdots$ in $\mathcal{B}$ with $A_0 = \Omega$ and $\nu A_{n+1} > \nu A_n + 1$. Then (νA_n) does not converge, which contradicts the continuity of ν. Second, suppose (1.11) does not hold. Then there exist $A_0 \in \mathcal{B}$ and $M \in (0, \infty)$ such that for any $B \in \mathcal{B}$ with $B \subset A_0$, we have $\nu B \le M$. We inductively choose a sequence of mutually disjoint sets (A_n) in $A_0 \cap \mathcal{A}$ such that $\nu A_n > M$ for each n. First, since $A_0 \in \mathcal{B}$, we may choose $A_1 \in \mathcal{A}$ such that $\nu A_1 > M$. Since $\nu B \le M$ for any $B \in \mathcal{B}$ with $B \subset A_0$, we see that $A_1 \in \mathcal{A} \setminus \mathcal{B}$. So $A_0 \setminus A_1 \in \mathcal{B}$. Suppose we have found mutually disjoint sets $A_1, \dots, A_n \in A_0 \cap \mathcal{A}$ such that $A_0 \setminus \bigcup_{k=1}^n A_k \in \mathcal{B}$ (this is the case for $n = 1$). Then by the definition of $\mathcal{B}$, we can find $A_{n+1} \in \mathcal{A}$ with $A_{n+1} \subset A_0 \setminus \bigcup_{k=1}^n A_k$ and $\nu A_{n+1} \ge M$. Now $A_1, \dots, A_{n+1}$ are mutually disjoint. Since

$A_{n+1} \subset \mathcal{A}$, we get $A_{n+1} \in \mathcal{A} \setminus \mathcal{B}$. Thus, $A_0 \setminus \bigcup_{k=1}^{n+1} A_k = (A_0 \setminus \bigcup_{k=1}^{n} A_k) \setminus A_{n+1} \in \mathcal{B}$. So the sequence (A_n) is constructed. However, by the countably additivity of ν, we should have $\nu A_n \to 0$, which is a contradiction. Thus, $s < +\infty$.

For any $A, B \in \mathcal{A}$, we have by inclusion-exclusion,

$$\nu(A \cap B) = \nu A + \nu B - \nu(A \cup B) \geq \nu A + \nu B - s.$$

So $s - \nu A \cap B \leq (s - \nu A) + (s - \nu B)$. By induction, we have

$$s - \nu \bigcap_{k=1}^{n} A_k \leq \sum_{k=1}^{n} (s - \nu A_k), \quad A_1, \dots, A_n \in \mathcal{A}.$$

If $A_1, A_2, \dots$ is a sequence in $\mathcal{A}$, then by continuity $\nu \bigcap_n A_n = \lim_n \nu \bigcap_{k=1}^{n} A_k$. So

$$s - \nu\Big(\bigcap_n A_n\Big) \leq \sum_n (s - \nu A_n), \tag{1.12}$$

By the definition of s, there is a sequence $A_1, A_2, \dots \in \mathcal{A}$ such that $\nu A_n > s - \frac{1}{2^n}$ for each n. Define an increasing sequence (B_n) by $B_n = \bigcap_{m=n}^{\infty} A_m$. By (1.12),

$$\nu B_n \geq s - \sum_{k=n}^{\infty} \frac{1}{2^k} = s - \frac{1}{2^{n-1}}, \quad n \in \mathbb{N}. \tag{1.13}$$

Let $P = \bigcup_n B_n$ and $N = P^c$. Then $\{P, N\}$ is a measurable partition of Ω. By continuity of ν and (1.13), $\nu P = \lim \nu B_n \geq s$. By the definition of s, $\nu P \leq s$. So $\nu P = s$. If there is $E \in P \cap \mathcal{A}$ such that $\nu E < 0$, then $\nu(P \setminus E) = \nu P - \nu E > \nu P = s$, which contradicts the definition of s. So $\nu E \geq 0$ for any $E \in \mathcal{A}$ with $E \subset P$. If there is $E \in N \cap \mathcal{A}$ such that $\nu E > 0$, then $\nu(P \cup E) = \nu P + \nu E > \nu P = s$, which again contradicts the definition of s. So $\nu E \geq 0$ for any $E \in \mathcal{A}$ with $E \subset P$. □

If we set $\nu_+ = \nu(\cdot \cap P)$ and $\nu_- = -\nu(\cdot \cap N)$, then ν_+ and ν_- are two finite (nonnegative) measures, and $\nu = \nu_+ - \nu_-$. Since $\nu_+ P^c = \nu_- P = 0$, we have $\nu_+ \perp \nu_-$. We call $\nu = \nu_+ - \nu_-$ the Jordan decomposition of ν.

Lemma . *The Jordan decomposition of a real measure is unique.*

Proof. We leave this as an exercise. □

If $\nu_+ - \nu_-$ is the Jordan decomposition of a real measure ν, then we define the measure $|\nu| = \nu_+ + \nu_-$, and call it the total variation of ν.

Proof of Radon-Nikodym Theorem. (i) The uniqueness part is easy. If $\nu = f \cdot \mu = g \cdot \mu$, and $\mu\{f \neq g\} > 0$, then $\mu\{f > g\} > 0$ or $\mu\{g > f\} > 0$. By symmetry we assume that $\mu\{f > g\} > 0$. Then there is $n \in \mathbb{N}$ such that $\mu\{f > g + 1/n\} > 0$. Then $f \cdot \mu$ does not agree with $g \cdot \mu$ on $\{f > g + 1/n\}$, a contradiction.

For the existence, we may assume that μ and ν are finite. This is because we may find a measurable partition $\{A_n : n \in \mathbb{N}\}$ of Ω such that $\mu A_n, \nu A_n < \infty$ for each n. Then $\mu_n := \mu(\cdot \cap A_n)$ and $\nu_n := \nu(\cdot \cap A_n)$ are finite measures with $\nu_n \ll \mu_n$ for each n. If for each n, $\nu_n = f_n \cdot \mu_n$ for some $f_n : A_n \to \mathbb{R}_+$, then we may construct the μ-density f of ν with $f|_{A_n} = f_n$.

Now μ and ν are finite measures. Let F be the set of measurable functions $f : \Omega \to \mathbb{R}_+$ such that $f \cdot \mu \le \nu$, i.e., $\nu A \ge (f \cdot \mu)A$ for all $A \in \mathcal{A}$. Here F contains 0. For $f_1, f_2 \in F$, let $A_1 = \{f_1 > f_2\}$ and $A_2 = \{f_1 \le f_2\}$. For any $A \in \mathcal{A}$,

$$\int_A f_1 \vee f_2 d\mu = \int_{A\cap A_1} f_1 d\mu + \int_{A\cap A_2} f_2 d\mu \le \nu A \cap A_1 + \nu A \cap A_2 = \nu A.$$

So $f_1 \vee f_2 \in F$. Let $s = \sup\{\mu f : f \in F\}$. Then $0 \le s \le \nu\Omega < \infty$. We may find a sequence $g_1, g_2, \cdots \in F$ such that $\mu g_n \to s$. Let $f_n = g_1 \vee \cdots \vee g_n$, $n \in \mathbb{N}$. Then (f_n) is increasing, and for each n, $f_n \in F$, and $f_n \ge g_n$. So $\mu f_n \to s$. Let $f = \lim f_n$. By monotone convergence theorem, for any $A \in \mathcal{A}$, $\int_A f d\mu = \lim \int_A f_n d\mu \le \nu A$. So $f \in F$. Moreover, $\mu f = \lim \mu f_n = s$. We claim that $\nu = f \cdot \mu$. If it is not true, then $\nu_0 := \nu - f \cdot \mu$ is a none-zero measure. Since μ is finite, there is $\varepsilon > 0$ such that $\nu_0\Omega > \varepsilon\mu\Omega$. Now $\tau := \nu_0 - \varepsilon\mu$ is a real measure with $\tau\Omega > 0$. By Hahn decomposition theorem, there is a partition $\Omega = P \cup N$ such that $\tau(\cdot \cap P)$ and $-\tau(\cdot \cap N)$ are measures. For every $A \in \mathcal{A}$, from $\tau(A \cap P) \ge 0$, we get $\nu_0(A \cap P) \ge \varepsilon\mu(A \cap P)$, and so

$$\nu A = \int_A f d\mu + \nu_0 A \ge \int_A f d\mu + \nu_0 A \cap P \ge \int_A f d\mu + \varepsilon\mu A \cap P = \int_A (f + \varepsilon \mathbf{1}_P) d\mu.$$

Thus, $f + \varepsilon\mathbf{1}_P \in F$. From $s = \mu f \le \mu(f + \varepsilon\mathbf{1}_P) \le s$ we get $\mu P = 0$. So $\nu_P = \nu_0 P = \tau P = 0$. Then we see that $-\tau$ is a (positive) measure, which contradicts that $\tau\Omega > 0$. The contradiction shows that $\nu = f \cdot \mu$.

(ii) Let $\tau = \mu + \nu$. Then τ is also a σ-finite measure. Since $0 \le \nu \le \tau$, we have $\nu \ll \tau$. By (i) there is a measurable $g : \Omega \to \mathbb{R}_+$ such that $\nu = g \cdot \tau$. We have τ-a.e. $g \le 1$ because for any $A \in \mathcal{A}$, $\int_A 1 - g d\tau = \tau A - (g \cdot \tau)A = \tau A - \nu A = \mu A \ge 0$. By changing the values of g on a τ-null set, we may assume that $0 \le g \le 1$. From $\nu = g \cdot \tau$ we get $\mu = (1 - g) \cdot \tau$. Let $A = \{g < 1\}$. Then $\mu A^c = 0$. Define $f = \frac{g}{1-g}$ on A and $f = 0$ on A^c. Then $\nu(\cdot \cap A) = f \cdot \mu$. Let $\sigma = \nu - f \cdot \mu = \nu(\cdot \cap A^c)$. Then $\sigma A = 0$. So $\sigma \perp \mu$.

For the uniqueness, we still let $\tau = \mu + \nu$. Suppose $\nu = f \cdot \mu + \sigma$ for some measurable $f : \Omega \to \mathbb{R}_+$ and some measure σ with $\sigma \perp \mu$. Let $A \in \mathcal{A}$ be such that $\mu A^c = \sigma A = 0$. Then

$$\nu = \mathbf{1}_A f \cdot \mu + \mathbf{1}_{A^c} \cdot \sigma, \quad \tau = \mathbf{1}_A (f + 1) \cdot \mu + \mathbf{1}_{A^c} \cdot \sigma.$$

So $\nu = (\mathbf{1}_A \frac{f}{f+1} + \mathbf{1}_{A^c}) \cdot \tau$. By the uniqueness part of (i), if $\tau = g \cdot \mu + \rho$ and $\mu B^c = \rho B = 0$, then

$$\mathbf{1}_A \frac{f}{f+1} + \mathbf{1}_{A^c} = \mathbf{1}_B \frac{g}{g+1} + \mathbf{1}_{B^c}, \quad \tau - \text{a.e.}.$$

This implies that τ-a.e. $\mathbf{1}_A f = \mathbf{1}_B g$. Since $\mu A^c = \mu B^c = 0$ and $\mu \ll \tau$, we get μ-a.e. $f = g$. □

Radon-Nikodym theorem also extends to real measures.

Corollary . *Let μ be a σ-finite measure on $(\Omega, \mathcal{A})$. Let ν be a real measure on $(\Omega, \mathcal{A})$. Suppose $\nu \ll \mu$, i.e., for any $A \in \mathcal{A}$, $\mu A = 0$ implies $\nu A = 0$. Then there a μ-a.e. unique $f : \Omega \to \mathbb{R}$, which is integrable w.r.t. μ, such that $\nu = f \cdot \mu$.*

Proof. This follows from the Radon-Nikodym theorem and Jordan decomposition. □

Example (An important application)**.** Suppose μ is a probability measure on $(\Omega, \mathcal{A})$, $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$, and $f : \Omega \to \mathbb{R}$ is $\mathcal{A}$-measurable with $\mu|f| < \infty$. Let $\nu = f \cdot \mu$. Then ν is a signed measure on $(\Omega, \mathcal{A})$, and $\nu \ll \mu$. Let $\mu' = \mu|_{\mathcal{F}}$ and $\nu' = \nu|_{\mathcal{F}}$. Then μ' is a probability measure on $(\Omega, \mathcal{F})$, ν' is a signed measure on $(\Omega, \mathcal{F})$, and $\nu' \ll \mu'$. By the above corollary, there is an $\mathcal{F}$-measurable $f' : \Omega \to \mathbb{R}$ with $\mu'|f'| < \infty$ such that $\nu' = f' \cdot \mu$. Then for any $A \in \mathcal{F}$,

$$\int_A f' d\mu = \int_A f' d\mu' = \nu' A = \nu A = \int_A f d\mu.$$

Such f' is μ-a.e. unique, and is called the expectation of f conditionally on $\mathcal{F}$ with respect to μ.

A measure space $(\Omega, \mathcal{A}, \mu)$ is called complete if for every $B \subset A \subset \Omega$ with $A \in \mathcal{A}$ and $\mu A = 0$, we have $B \in \mathcal{A}$. Given a measure space $(\Omega, \mathcal{A}, \mu)$, a μ-completion of $\mathcal{A}$ is the σ-algebra

$$\mathcal{A}^\mu := \sigma(\mathcal{A}, \mathcal{N}_\mu),$$

where $\mathcal{N}_\mu$ is the class of all subsets of μ-null sets in $\mathcal{A}$. Note that $\mathcal{N}_\mu$ is closed under countable union because if $N_1, N_2, \dots \in \mathcal{N}_\mu$, there there are $A_1, A_2, \dots \in \mathcal{A}$ with $N_n \subset A_n$ and $\mu A_n = 0$ for each n. Then $\bigcup_n N_n \subset \bigcup_n A_n \in \mathcal{A}$, and $\mu \bigcup_n A_n = 0$. So $\bigcup_n N_n \in \mathcal{N}_\mu$.

Lemma 1.25. *(i) A set $A \subset \Omega$ is $\mathcal{A}^\mu$-measurable if and only if there exist $A', A'' \in \mathcal{A}$ with $A' \subset A \subset A''$ and $\mu(A'' \setminus A') = 0$. (ii) A function f from Ω to a Borel space $(S, \overline{S})$ is $\mathcal{A}^\mu$-measurable if and only if there is an $\mathcal{A}$-measurable map $g : \Omega \to (S, \overline{S})$ such that μ-a.e., $f = g$.*

Proof. (i) Let $\widetilde{\mathcal{A}}^\mu$ denote the set of $A \subset \Omega$ such that the A', A'' in the statement exist. We need to show that $\widetilde{\mathcal{A}}^\mu = \mathcal{A}^\mu$. Clearly, $\mathcal{A}, \mathcal{N}_\mu \subset \widetilde{\mathcal{A}}^\mu \subset \mathcal{A}^\mu$. It suffices to show that $\widetilde{\mathcal{A}}^\mu$ is a σ-algebra. We need to show that (a) if $A \in \widetilde{\mathcal{A}}^\mu$, then $A^c \in \widetilde{\mathcal{A}}^\mu$; and (b) if $A_1, A_2, \dots \in \widetilde{\mathcal{A}}^\mu$, then $\bigcup_n A_n \in \widetilde{\mathcal{A}}^\mu$. For (a), note that if $A' \subset A \subset A''$ with $A', A'' \in \mathcal{A}$ and $\mu(A'' \setminus A')$, then $(A'')^c \subset A^c \subset (A')^c$, and $\mu((A')^c \setminus (A'')^c) = 0$. For (b), note that if for each n, $A'_n \subset A_n \subset A''_n$, $A'_n, A''_n \in \mathcal{A}$ and $\mu(A''_n \setminus A'_n) = 0$, then $A' := \bigcup_n A'_n, A'' := \bigcup_n A''_n \in \mathcal{A}$ and satisfy that $A' \subset A \subset A''$ and $0 \le \mu(A'' \setminus A') \le \sum_n \mu(A''_n \setminus A'_n) = 0$.

(ii) If the g exists, then there is $N \in \mathcal{A}$ with $\mu N = 0$ such that $f = g$ on N^c. For any $B \in \overline{S}$, we have

$$f^{-1}B = ((f^{-1}B) \setminus N) \cup ((f^{-1}B) \cap N) = ((g^{-1}B) \setminus N) \cup ((f^{-1}B) \cap N).$$

So $(g^{-1}B)\setminus N \subset f^{-1}B \subset (g^{-1}B)\cup N$. Since $(g^{-1}B)\setminus N, (g^{-1}B)\cup N \in \mathcal{A}$ and $\mu N = 0$, by (i), $f^{-1}B \in \mathcal{A}^\mu$. So f is $\mathcal{A}^\mu$-measurable.

Now suppose f is $\mathcal{A}^\mu$-measurable. Since S is a Borel space, we may assume that it is a Borel subset of $[0,1]$. We first show that there is an $\mathbb{R}$-valued $\mathcal{A}$-measurable function g such that μ-a.e., $f = g$. If $f = \mathbf{1}_A$ for some $A \in \mathcal{A}^m u$, then by (i), there exist $A', A'' \in \mathcal{A}$ with $A' \subset A \subset A''$. Then μ-a.e., $f = \mathbf{1}_{A'} := g$. The statement then extends to simple measurable functions by linearity. Now suppose $f \ge 0$. There exists a sequence of $\mathcal{A}^\mu$-measurable simple functions (f_n) such that $0 \le f_n \uparrow f$. For each n, there exists an $\mathcal{A}$-measurable simple function g_n such that μ-a.e. $f_n = g_n$. The sequence (g_n) may not be nonnegative or increasing. However, we may choose $N_n \in \mathcal{A}$ such that $\mu N_n = 0$ and $f_n = g_n$ on N_n^c. Let $N = \bigcup_n N_n$. Then $N \in \mathcal{A}$ and $\mu N = 0$, and $0 \le g_n \uparrow f$ on N^c. Let $g = \lim g_n$ on N^c and $= 0$ on N. Then g is $\mathcal{A}$-measurable and μ-a.e., $f = g$. Finally, we may modify the value of g such that g takes values in S, and still satisfies other properties that we want. Let $N \in \mathcal{A}$ be such that $\mu N = 0$ and $f = g$ on N^c. Then $g \in S$ on N^c since f takes values in S. So $g^{-1}S \subset N^c$. We now choose $s_0 \in S$, and define $\widetilde{g}$ such that $\widetilde{g} = g$ on $g^{-1}S \in \mathcal{A}$ and $\widetilde{g} = s_0$ on $(g^{-1}S)^c$. Then $\widetilde{g} : \Omega \to S$ is $\mathcal{A}$-measurable, and μ-a.e., $\widetilde{g} = g$, so μ-a.e., $f = \widetilde{g}$. □

It is natural to extend μ to the completion $\mathcal{A}^\mu$ in the way such that if $A' \subset A \subset A''$ with $A', A'' \in \mathcal{A}$ and $\mu(A'' \setminus A') = 0$, then $\mu A = \mu A'$. The definition is consistent, and defines a measure on $(\Omega, \mathcal{A}^\mu)$.

Exercise . Prove the statements in the above paragraph.

We are going to construct product measures. Let $(S, \overline{S}, \mu)$ and $(T, \overline{T}, \nu)$ be two σ-finite measure spaces. We want the product measure $\mu \times \nu$ be a measure on $\overline{S} \times \overline{T}$ that satisfies

$$(\mu \times \nu)(A \times B) = \mu A \times \nu B, \quad \forall A \in \overline{S} \text{ and } B \in \overline{T}. \tag{1.14}$$

We will also show that such measure is unique. The $\mu \times \nu$ is called the product of μ and ν.

Lemma 1.26. *For any measurable function $f : S \times T \to \overline{\mathbb{R}}_+$, and any $t \in T$, the function $f(\cdot, t) : S \to \overline{\mathbb{R}}_+$ is $\overline{S}$-measurable. If we integrate $f(\cdot, t)$ against μ and get $\mu f(\cdot, t) \in \overline{\mathbb{R}}_+$ for each $t \in T$, then $t \mapsto \mu f(\cdot, t)$ is $\overline{T}$-measurable.*

Proof. First suppose μ is finite. Let $\mathcal{C}$ denote the set of $C \in \overline{S} \times \overline{T}$ such that the lemma holds for $f = \mathbf{1}_C$. Then $\mathcal{C}$ contains the π-system $\{A \times B : A \in \overline{S}, B \in \overline{T}\}$. In fact, if $f = \mathbf{1}_{A \times B}$, then for $t \in B$, $f(\cdot, t) = \mathbf{1}_A$, and for $t \in B^c$, $f(\cdot, t) \equiv 0$. In either case $f(\cdot, t)$ is $\overline{S}$-measurable. Moreover, $\mu f(\cdot, t) = \mu A \mathbf{1}_B(t)$ is $\overline{T}$-measurable. Using the linearity of integrals, we easily see that $\mathcal{C}$ is a λ-system. By monotone class theorem, $\mathcal{C} = \overline{\mathcal{S}} \times \overline{\mathcal{T}}$. Thus, the lemma holds for indicator functions. By linearity and monotone convergence, the statement extends to nonnegative measurable functions.

Now we do not assume that μ is finite. Since it is σ-finite, we may express $\mu = \sum_n \mu_n$, where each μ_n is a finite measure. The measurability of each $f(\cdot, t)$ does not rely on the finiteness of μ. Since $t \mapsto \mu_n f(\cdot, t)$ is $\overline{T}$-measurable for each n, the same is true for $t \mapsto \mu f(\cdot, t) = \sum_n \mu_n f(\cdot, t)$. □

Theorem 1.27 (Fubini)**.** *The product measure $\mu \times \nu$ exists uniquely, and for any measurable $f : S \times T \to \overline{\mathbb{R}}_+$ or $f : S \times T \to \mathbb{R}$ with $(\mu \times \nu)|f| < \infty$, we have*

$$(\mu \times \nu)f = \int \mu(ds) \int f(s,t)\nu(dt) = \int \nu(dt) \int f(s,t)\mu(ds). \tag{1.15}$$

Here the meaning of the second double integral is that we first fix $t \in T$, treat $f(s,t)$ as a function in $s \in S$, and integrate the function against the measure μ. The integral is a function of $t \in T$. We then integrate the function against the measure ν. The procedure is valid for measurable $f : S \times T \to \overline{\mathbb{R}}_+$ by Lemma 1.26. The meaning of the first double integral is similar.

Proof. By a monotone class argument involving partitions of S and T into finite measurable sets, it is easy to see that there exists at most one product measure.

By Lemma 1.26, we may define

$$(\mu \times \nu)C = \int \mu(ds) \int \mathbf{1}_C(s,t)\nu(dt), \quad C \in \overline{S} \times \overline{T}.$$

Then $\mu \times \nu$ is clearly a measure that satisfies (1.14). By uniqueness and symmetry, we also have

$$(\mu \times \nu)C = \int \nu(dt) \int \mathbf{1}_C(s,t)\mu(ds), \quad C \in \overline{S} \times \overline{T}.$$

Thus, (1.15) holds for indicator functions. By linearity and monotone convergence, the statement extends to measurable $\overline{\mathbb{R}}_+$-valued functions.

If $f : S \times T \to \mathbb{R}$ is integrable w.r.t. $\mu \times \nu$, then $(\mu \times \nu)|f| < \infty$. By (1.15),

$$\int \nu(dt) \int |f(s,t)|\mu(ds) < \infty. \tag{1.16}$$

So for ν-a.e. $t \in T$, $\int |f(s,t)|\mu(ds) < \infty$, i.e., $f(\cdot,t)$ is integrable w.r.t. μ. So we may define $\int f(s,t)\mu(ds)$ (as a function of t) outside a ν-null set. Since $|\int f(s,t)\mu(ds)| \le \int |f(s,t)|\mu(ds)$ whenever $f(\cdot,t)$ is μ-integrable, by (1.16), $t \mapsto \int f(s,t)\mu(ds)$ is ν-integrable. So the double integral $\int \nu(dt) \int f(s,t)\mu(ds)$ is well defined. Similarly, $\int \mu(ds) \int f(s,t)\nu(dt)$ is also well defined. We may prove (1.15) for such f by expressing $f = f_+ - f_-$. □

Note that the product $\mu \times \nu$ is also a σ-finite measure, and we may then define $(\mu \times \nu) \times \sigma$ for another σ-finite measures. If $(S_k, \overline{S}_k, \mu_k)$, $1 \le k \le n$, are σ-finite measure spaces, then we may use induction to construct the product measure $\mu_1 \times \cdots \times \mu_n$ on $\overline{S}_1 \times \cdots \times \overline{S}_n$, which is the unique measure that satisfies

$$(\mu_1 \times \cdots \times \mu_n)(A_1 \times \cdots \times A_n) = \prod_{k=1}^n \mu_k A_k, \quad \forall A_k \in \overline{S}_k, \quad 1 \le k \le n.$$

In the case all μ_n are the same μ, we write the product as μ^n. For the Lebesgue measure λ on $\mathbb{R}$, its power μ^n is called the Lebesgue measure on $\mathbb{R}^n$.

We may define the product of infinitely many measures, but need to assume that they are all probability measures.

Definition . Let $(S_t, \overline{S}_t, \mu_t)$, $t \in T$, be a family of probability spaces. A probability measure μ on the product measurable space $(\prod_t S_t, \prod_t \overline{S}_t)$ is called the product of μ_t, $t \in T$, denoted by $\prod_t \mu_t$, if for any finite $\Lambda \subset T$, and $A_\lambda \in \overline{S}_\lambda$, $\lambda \in \Lambda$, we have

$$\mu\Big(\prod_{\lambda\in\Lambda} A_\lambda \times \prod_{t\in T\setminus\Lambda} S_t\Big) = \prod_{\lambda\in\Lambda} \mu_\lambda A_\lambda.$$

By a monotone argument, we see that the product measure in the definition is unique, if it exists. The existence of the infinite product measure (assuming S_t are Borel spaces) will be proved in the next chapter.

Definition . A measurable group is a group G endowed with a σ-algebra $\overline{G}$ such that the group operations in G are measurable. This means

(i) the map $g \mapsto g^{-1}$ from G to G is $\overline{G}/\overline{G}$-measurable;

(ii) the map $(f, g) \mapsto fg$ from G^2 to G is $\overline{G}^2/\overline{G}$-measurable.

If G is a topological group, i.e., endowed with a topology such that the group operations are continuous, and has a countable basis, then it is a measurable group. We will mainly work with the Euclidean space $\mathbb{R}^n$ as a measurable group.

Definition . For two σ-finite measures μ and ν on a measurable group G, the convolution of μ and ν, denoted by $\mu * \nu$, is the pushforward of the product measure $\mu \times \nu$ under the map $(f, g) \mapsto fg$.

The convolution $\mu * \nu$ may not be σ-finite. If both μ and ν are finite, $\mu * \nu$ is also finite. If μ_1, μ_2, μ_3 are finite measures, then the associative law holds: $(\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$. If G is Abelian, then the commutative law holds: $\mu * \nu = \nu * \mu$.

Definition . A measure μ on a measurable group G is said to be right- or left invariant if $\mu \circ T_g^{-1} = \mu$ for any $g \in G$, where T_g denotes the right or left shift $x \mapsto xg$ or $x \mapsto gx$. If G is Abelian, right-invariance and left-invariance are equivalent.

Example . The Lebesgue measure λ^n is an invariant measure on $\mathbb{R}^n$, and any locally finite invariant measure on $\mathbb{R}^n$ is a scalar product of λ^n.