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. $\square$

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

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

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. $\square$

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 $\{\omega : \lim f_n(\omega) \text{ converges}\}$ 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.

$\square$

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 $\mathbb{R} \times (\mathbb{R} \setminus \{0\}) \ni (x, y) \mapsto x/y \in \mathbb{R}$. $\square$

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 $n \in \mathbb{N}$, $A_1, \ldots, A_n \subset \Omega$ and $c_1, \ldots, c_n \in \mathbb{R}$. Here we only allow finite sums. If $A_1, \ldots, A_n \in \mathcal{A}$, 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 $\{\frac{k}{2^n} : 0 \le k \le n 2^n\}$,

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

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

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 $|f_n(\omega) - f(\omega)| \le \frac{1}{2^n} < \varepsilon$. $\square$

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. $\square$

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

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

$\square$

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 $\bigcap_n B_n \ne 0$. 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 $\bigcap_n A_n \ne \emptyset$. 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 $A_n'' \ne \emptyset$. Since each $\overline{A_n''}$ is compact and $\overline{A_1''} \supset \overline{A_2''} \supset \cdots$, we get $\bigcap_n \overline{A_n''} \ne \emptyset$, which together with $\overline{A_n''} \subset A_n$ implies that $\bigcap_n A_n \ne \emptyset$. 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 $B \in \mathcal{B}(S)$. $\square$

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:

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

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

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 $0 \leq f_1 \leq f_2 \leq \cdots$ and $0 \leq g \leq \lim f_n$. Then $\lim \mu f_n \geq \mu g$.*

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 $A_n = A \cap \{f_n \geq c - \varepsilon\}$. Then $A_n \uparrow A$, and so

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

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

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

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

□

Applying this lemma, we see that if (f_n) and (g_m) are two sequences of measurable simple functions with $0 \leq f_n \uparrow f$ and $0 \leq g_m \uparrow f$, then for each m, $\lim_n \mu f_n \geq \mu g_m$. So $\lim_n \mu f_n \geq \lim_m \mu g_m$. By symmetry, we have $\lim_m \mu g_m \geq \lim_n \mu f_n$. 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 $f, f_1, f_2, \ldots$ and $g, g_1, g_2, \ldots$ 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 $\lim \mu f_n = \mu f$. $\square$

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. $\square$

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 $A_1, A_2, \cdots \in \mathcal{A}$ 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 $P, N \in \mathcal{A}$, $\nu E \geq 0$ for all $E \in P \cap \mathcal{A}$, and $\nu E \leq 0$ for all $E \in N \cap \mathcal{A}$.*

Proof. Let $s = \sup\{\nu A : A \in \mathcal{A}\}$. Then $s \geq 0$ 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 $\nu B \leq M$. 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 $\nu B \leq M$ 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 $\nu A_{n+1} \geq M$. 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 $E \subset P$. $\square$

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. $\square$

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. $f = g$. $\square$

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. $\square$

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 $f = f_+ - f_-$. $\square$

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.

Lemma 1.28. *Let $(G, +)$ be an Abelian measurable group with an invariant measure λ. Suppose μ and ν are σ-finite measures on G with λ-densities f and g. Then $\mu * \nu$ has a λ-density $f * g$ given by*

$$(f * g)(s) = \int f(s-t) g(t) \lambda(dt) = \int f(t) g(s-t) \lambda(dt), \quad s \in G. \tag{1.17}$$

Proof. Let $\pi : G \times G \to G$ be the map $(s, t) \mapsto s + t$. Let $A \in \mathcal{G}$. Then $(s, t) \in \pi^{-1} A$ if and only if $t \in A - s := \{x - s : x \in A\}$. So

$$(\mu * \nu)A = (\mu \times \nu)(\pi^{-1} A) = \int \mu(ds) \int \mathbf{1}_{\pi^{-1}A}(s, t) \nu(dt)$$

$$= \int \mu(ds) \int \mathbf{1}_{A-s}(t)\nu(dt) = \int \mu(ds) \int \mathbf{1}_{A-s}(t)g(t)\lambda(dt)$$
$$= \int \mu(ds) \int \mathbf{1}_A(t)g(t-s)\lambda(dt) = \int f(s)\lambda(ds) \int \mathbf{1}_A(t)g(t-s)\lambda(dt)$$
$$= \int \mathbf{1}_A(t)\lambda(dt) \int f(s)g(t-s)\lambda(ds) = \int \mathbf{1}_A(t)(f*g)(t)\lambda(dt).$$

Here in the third line we use the invariance of λ. Thus, $\mu * \nu$ has a λ-density $f * g$. $\square$

Note that when $G = \mathbb{R}^n$ and λ is the Lebesgue measure on $\mathbb{R}^n$, the $f * g$ defined by (1.17) agrees with the convolution of f and g.

We now define L^p-spaces for $p > 0$. Given a measure space $(\Omega, \mathcal{A}, \mu)$ and $p > 0$, we write $L^p = L^p(\Omega, \mathcal{A}, \mu)$ for the class of all measurable functions $f : \Omega \to \mathbb{R}$ with

$$\|f\|_p := (\mu|f|^p)^{1/p} < \infty.$$

In particular, L^1 is the space of all integrable functions. We have a scaling property $\|cf\|_p = |c|\|f\|_p$ for any $c \in \mathbb{R}$.

Lemma 1.30 (Hölder inequality and norm inequality)**.** *For any measurable functions f and g on Ω,*

(i) if $p, q > 1$ and $1 = p^{-1} + q^{-1}$, then $\|fg\|_1 \le \|f\|_p \|g\|_q$;

(ii) for all $p > 0$, $\|f + g\|_p^{p\wedge 1} \le \|f\|_p^{p\wedge 1} + \|g\|_p^{p\wedge 1}$.

Proof. (i) If $\|f\|_p$ or $\|g\|_q$ equals 0, then the inequality is trivial because $fg = 0$ a.e. If $\|f\|_p$ and $\|g\|_q$ are both positive, and one of them is ∞, the inequality is also trivial because the RHS is ∞. So we may assume that $\|f\|_p, \|g\|_q \in (0, \infty)$. By scaling we may assume that $\|f\|_p = \|g\|_q = 1$.

The relation $p^{-1} + q^{-1} = 1$ implies that $(p-1)(q-1) = 1$. So for $x, y \ge 0$, $y = x^{p-1}$ if and only if $x = y^{q-1}$. Consider two subsets of $\mathbb{R}_+^2$: $A_1 = \{(x, y) : 0 \le x \le x_0, 0 \le y \le x^{p-1}\}$ and $A_2 = \{(x, y) : 0 \le y \le y_0, 0 \le x \le y^{q-1}\}$. By Fubini theorem, $\lambda^2 A_1 = \int_0^{x_0} x^{p-1} dx$ and $\lambda^2 A_2 = \int_0^{y_0} y^{q-1} dy$. Suppose $(x, y) \in [0, x_0] \times [0, y_0]$. If $y \le x^{p-1}$, then $(x, y) \in A_1$; if $y \ge x^{p-1}$, then $x \le y^{q-1}$, and $(x, y) \in A_2$. So $[0, x_0] \times [0, y_0] \subset A_1 \cup A_2$. Thus,

$$x_0 y_0 = \lambda^2 [0, x_0] \times [0, y_0] \le \lambda^2 A_1 + \lambda^2 A_2 = \int_0^{x_0} x^{p-1} dx + \int_0^{y_0} y^{q-1} dy = x_0^p/p + y_0^q/q.$$

Applying the inequality to $x_0 = |f|$ and $y_0 = |g|$, we get

$$\|fg\|_1 = \mu|f||g| \le \mu(|f|^p/p + |g|^q/q) = 1/p + 1/q = 1 = \|f\|_p \|g\|_q.$$

(ii) If $p \in (0, 1]$, the inequality follows from the inequality $(x + y)^p \le x^p + y^p$ for any $x, y \ge 0$ (because $x \mapsto x^p$ is a concave function). Suppose $p > 1$. If $\|f\|_p$ or $\|g\|_p = \infty$, the inequality trivially holds. Suppose $\|f\|_p, \|g\|_q < \infty$. Since $|f+g|^p \le 2^p(|f| \vee |g|)^p \le 2^p(|f|^p + |g|^p)$, we get $\|f+g\|_p < \infty$. By applying (i) to $q := \frac{p}{p-1}$, we get

$$\begin{aligned}\|f+g\|_p^p = \int |f+g|^p d\mu &\le \int |f||f+g|^{p-1} d\mu + \int |g||f+g|^{p-1} d\mu \\ &\le \|f\|_p \||f+g|^{p-1}\|_q + \|g\|_p \||f+g|^{p-1}\|_q.\end{aligned}$$

Note that

$$\||f+g|^{p-1}\|_q = \Big(\int |f+g|^{(p-1)q} d\mu\Big)^{1/q} = \Big(\int |f+g|^p d\mu\Big)^{\frac{p-1}{p}} = \|f+g\|_p^{p-1}.$$

So $\|f+g\|_p^p \le \|f+g\|_p^{p-1}(\|f\|_p + \|g\|_p)$, which implies (ii) because $\|f+g\|_p < \infty$. $\square$

Since $\|f\|_p = 0$ if and only if a.e. $f = 0$. By the norm inequality, L^p becomes a metric space with distance $\rho(f,g) = \|f-g\|_p^{p\wedge 1}$ if we identify functions that agree μ-a.e. From now on, L^p will be a space of measurable functions with $\|f\|_p < \infty$ modulus the "equal almost everywhere" equivalence. We say that $f_n \to f$ in L^p if $\|f_n - f\|_p \to 0$. For $p \ge 1$, L^p is a normed space. We now show that L^p is complete for all $p > 0$. Then for $p \ge 1$, L^p is a Banach space.

Lemma 1.31. *Let (f_n) be a Cauchy sequence in L^p, where $p > 0$, then for some $f \in L^p$, $\|f_n - f\|_p \to 0$.*

Proof. First choose a subsequence (f_{n_k}) with $\sum_k \|f_{n_{k+1}} - f_{n_k}\|_p^{p\wedge 1} < \infty$. By Lemma 1.30 and monotone convergence, we get $\|\sum_k |f_{n_{k+1}} - f_{n_k}|\|_p^{p\wedge 1} < \infty$, and so $\sum_k |f_{n_{k+1}} - f_{n_k}| < \infty$ a.e. Hence (f_{n_k}) is Cauchy in $\mathbb{R}$ a.e. So there is a measurable function f such that $f_{n_k} \to f$ a.e. By Fatou's lemma,

$$\int |f_n - f|^p d\mu \le \liminf_k \int |f_n - f_{n_k}|^p d\mu \le \sup_{m \ge n} \int |f_n - f_m|^p d\mu \to 0, \quad n \to \infty.$$

Thus, $f \in L^p$ and $\|f_n - f\|_p \to 0$. $\square$

Lemma 1.32. *For any $p > 0$, let $f, f_1, f_2, \cdots \in L^p$ with $f_n \to f$ a.e. Then $f_n \to f$ in L^p if and only if $\|f_n\|_p \to \|f\|_p$.*

Proof. If $f_n \to f$ in L^p, by the norm inequality,

$$|\|f_n\|_p^{p\wedge 1} - \|f\|_p^{p\wedge 1}| \le \|f_n - f\|_p^{p\wedge 1} \to 0,$$

and so $\|f_n\|_p \to \|f\|_p$. If $\|f_n\|_p \to \|f\|_p$, then we define

$$g_n = 2^p(|f_n|^p + |f|^p), \quad g = 2^{p+1}|f|^p.$$

We have $g_n \to g$ a.e. and $\mu g_n \to \mu g = 2^{p+1}\|f\|_p^p < \infty$. Since $g_n \ge |f_n - f|^p \to 0$, by dominated convergence theorem, $\mu|f_n - f|^p \to 0$, i.e., $f_n \to f$ in L^p. $\square$

Lemma 1.33. *Given a metric space (S,ρ) and a finite measure μ on $(S,\mathcal{B}(S))$, for any $p>0$, the space $C_b(S,\mathbb{R})$ of bounded real valued continuous functions on S is dense in $L^p(S,\mathcal{B}(S),\mu)$.*

Proof. Since μ is finite, we have $C_b \subset L^p(\mu)$. We need to show that the closure $\overline{C_b}$ of C_b in L^p equals L^p. First, for every open set G, there is a sequence (f_n) in C_b such that $f_n \to \mathbf{1}_G$ pointwise. We may choose $f_n(s) = 1 \wedge n\rho(x, G^c)$. Since $0 \le f_n \le 1$, by dominated convergence theorem, $f_n \to \mathbf{1}_G$ in L^p. So $\mathbf{1}_G \in \overline{C_b}$. By Lemma 1.16, for every $B \in \mathcal{B}(S)$, $\mathbf{1}_B \in \overline{C_b}$. Since $\overline{C_b}$ is a linear space, it then contains all measurable simple functions. By monotone convergence, we see that $\overline{C_b}$ contains all nonnegative functions in L^p, and so equals L^p. $\square$

Because of Hölder's inequality, if $f,g\in L^2$, fg is integrable, and

$$|\int fgd\mu| \le \|f\|_2\|g\|_2.$$

So L^2 is a Hilbert space with inner product: $\langle f,g\rangle := \int fgd\mu$.

Another important space is $L^\infty(\mu)$: the space of bounded measurable functions modulo "equal almost everywhere"' equivalence. It is a Banach space with the norm

$$\|f\|_\infty := \inf\{a \ge 0 : |f| \le a\ \mu - \text{a.e.}\}.$$

Theorem . *Suppose μ is a σ-finite measure. Let $p\in[1,\infty)$. Let $q=\frac{p}{p-1}$ if $p>1$; and $q=\infty$ if $p=1$. Then every continuous linear function $T: L^p \to \mathbb{R}$ corresponds to a unique $g \in L^q$ such that for any $f \in L^p$, $T(f) = \int fgd\mu$. Conversely, every $g\in L^q$ determines a continuous linear function on L^p defined by $f \mapsto \int fgd\mu$. Moreover, for any $g \in L^q$,*

$$\sup_{f\in L^p\setminus\{0\}} \frac{|\int fgd\mu|}{\|f\|_p} = \|g\|_q.$$

This means that L^q can be identified as $(L^p)^*$, the dual of L^p.

Sketch of the proof. Let T be given. Let $\{A_n\}$ be a partition of Ω such that $\mu A_n < \infty$ for every n. For each n, we may define a real measure ν_n on A_n such that $\nu_n A = T(\mathbf{1}_A)$ for $A \in \mathcal{A}$ and $A \subset A_n$. If $\mu A = 0$, then $\mathbf{1}_A = 0$ a.e. and so $T(\mathbf{1}_A)=0$, which implies that $\nu_n A = 0$. So $\nu_n \ll A$. By Radon-Nikodym theorem, there is a measurable g_n on A_n such that $\nu_n A = \int_A g_n d\mu$. Define g on Ω such that $g|_{A_n} = g_n$ for each n. Then using Hölder inequality, one can check that such g satisfies the properties. $\square$

Exercise . Complete the above proof.

Fix a measurable space $(S,\overline{S})$. Let $\mathcal{M}(S)$ denote the spaces of σ-finite measures on $(S,\overline{S})$. For each $B\in\overline{S}$, we define a map $\pi_B : \mathcal{M} \to \overline{\mathbb{R}}_+$ such that $\pi_B(\mu) = \mu B$. We endow $\mathcal{M}(S)$ with the σ-algebra generated by the mappings π_B for $B \in \overline{S}$, i.e.,

$$\sigma(\pi_B^{-1}(\mathcal{B}(\overline{\mathbb{R}}_+)) : B \in \overline{S}).$$

Then $\mathcal{M}(S)$ becomes a measurable space. Let $\mathcal{P}(S)$ denote the space of all probability measures on $(S,\overline{S})$. Then $\mathcal{P}(S) = \pi_S^{-1}\{1\}$ is a measurable subset of $\mathcal{M}(S)$.

Lemma 1.35. *For any measurable spaces $(S,\overline{S})$ and $(T,\overline{T})$, the product mapping $(\mu,\nu)\mapsto\mu\times\nu$ is measurable from $\mathcal{P}(S)\times\mathcal{P}(T)$ to $\mathcal{P}(S\times T)$.*

Proof. It suffices to show that for any $C\in\overline{S}\times\overline{T}$, $\pi_C(\mu\times\nu)=(\mu\times\nu)C$ from $\mathcal{P}(S)\times\mathcal{P}(T)$ to $\mathbb{R}$ is measurable. Let $\mathcal{C}$ denote the class of all such C. Then $\mathcal{C}$ is a λ-system. On the other hand, it contains the π-system $\{A\times B: A\in\overline{S}, B\in\overline{T}\}$, which generates the σ-algebra $\overline{S}\times\overline{T}$. By monotone class theorem, $\mathcal{C}$ equals $\overline{S}\times\overline{T}$. $\square$

Definition . Given two measurable spaces $(S,\overline{S})$ and $(T,\overline{T})$, a mapping $\mu: S\times\overline{T}\to\overline{\mathbb{R}}_+$ is called a (probability) kernel from S to T if for every $s\in S$, $\mu_s:=\mu(s,\cdot)$ is a (probability) measure on $(T,\overline{T})$, and for every $B\in\overline{T}$, $s\mapsto\mu(s,B)$ is a measurable function on $(S,\overline{S})$.

A measure μ on T can be viewed as a kernel: $\mu_s=\mu$ for every $s\in S$. In general, a kernel from S to T can be understood as a $\overline{S}$-measurable measure on $(T,\overline{T})$. For a nonnegative measurable function $f:T\to\mathbb{R}$, we may define the integral $\mu f=\int\mu(s,dt)f(t)$. The value is a function on S.

Lemma 1.37. *Let $\mathcal{C}$ be a π-system in T with $\sigma(\mathcal{C})=\overline{T}$. Let $\{\mu_s: s\in S\}$ be a family of probability measures on $(T,\overline{T})$. The following are equivalent.*

(i) $\mu(s,B):=\mu_s(B)$ is a probability kernel from S to T;

(ii) the map $s\mapsto\mu_s$ from S to $\mathcal{P}(T)$ is measurable;

(iii) for any $B\in\mathcal{C}$, $s\mapsto\mu_s B$ from S to $[0,1]$ is measurable.

Proof. The equivalence between (i) and (iii) follows from monotone class theorem since the set of $B\in\overline{T}$ such that $s\mapsto\mu_s B$ is measurable form a λ-system. The equivalence between (i) and (ii) is also straightforward because by the definition of the σ-algebra on $\mathcal{P}(T)$, the map $s\mapsto\mu_s$ is measurable if and only if for any $B\in\overline{T}$, $s\mapsto\mu_s B$ is measurable. $\square$

Lemma 1.38. *Fix three measurable spaces $(S,\overline{S})$, $(T,\overline{T})$, and $(U,\overline{U})$. Let μ be a probability kernel from S to T, and ν be a probability kernel from $S\times T$ to U. Let $f:S\times T\to\mathbb{R}_+$ and $g:S\times T\to U$ be measurable. Then*

(i) $\mu_s f(s,\cdot)$ is a measurable function of $s\in S$;

(ii) $\mu_s\circ(g(s,\cdot))^{-1}$ is a kernel from S to U;

(iii) we may define a probability kernel $\mu\otimes\nu$ from S to $T\times U$ by

$$(\mu\otimes\nu)(s,C)=\int\mu(s,dt)\int\nu(s,t,du)\mathbf{1}_C(t,u),\quad C\in\overline{T}\times\overline{U}. \tag{1.18}$$

Proof. (i) By Lemma 1.26, for every $s\in S$, $f(s,\cdot)$ is measurable. So $\mu_s f(s,\cdot)$ is well defined. If $f=\mathbf{1}_{A\times B}$ for $A\in\overline{S}$ and $B\in\overline{T}$, then $\mu_s f(s,\cdot)=\mathbf{1}_A(s)\mu_s B$ is measurable in s. This then extends to all indicator functions by a monotone class argument, and to arbitrary f by linearity and monotone convergence. (ii) For every $s \in S$, $\mu_s \circ (g(s,\cdot))^{-1}$ is a probability measure on U. For any $B \in \overline{U}$, $(\mu_s \circ (g(s,\cdot))^{-1})B = \mu_s(\mathbf{1}_B \circ g(s,\cdot))$. Since $(s,t) \mapsto \mathbf{1}_B(t) \circ g(s,t)$ from $S \times T$ to $\mathbb{R}_+$ is measurable, applying (i) to the function $f(s,t) := \mathbf{1}_B(t) \circ g(s,t)$, we see that $s \mapsto (\mu_s \circ (g(s,\cdot))^{-1})B$ is measurable. (iii) Applying (i) to the function $f((s,t),u) := \mathbf{1}_C(t,u)$, we see that $\int \nu(s,t,du)\mathbf{1}_C(t,u)$ is a measurable function of $(s,t) \in S \times T$. Applying (i) again to the function $f(s,t) := \int \nu(s,t,du)\mathbf{1}_C(t,u)$, we see that the RHS of (1.18) is well defined and measurable in $s \in S$ for a fixed $C \in \overline{T} \times \overline{U}$. When s is fixed, by monotone convergence, $(\mu \otimes \nu)(s,\cdot)$ is a measure on $S \times T$. Since $\mu(s,\cdot)$ and $\nu(s,t,\cdot)$ are both probability measures, we get $(\mu \otimes \nu)(s, T \times U) = 1$. So $\mu \otimes \nu$ is a probability kernel from S to $T \times U$. □

Note that when μ and ν are probability measures, i.e., μ does not depend on s and ν does not depend on (s,t), then $\mu \otimes \nu$ is the product measure $\mu \times \nu$.

By linearity and monotone convergence, for any measurable $f : T \times U \to \mathbb{R}_+$,

$$(\mu \otimes \nu)_s f = \int \mu(s,dt) \int \nu(s,t,du) f(t,u).$$

We may simply write it as $(\mu \otimes \nu)f = \mu(\nu f)$.

Suppose we have kernels μ_k from $S_0 \times \cdots \times S_{k-1}$ to S_k, $k = 1, \dots, n$. By iteration we may combine them into a kernel $\mu_1 \otimes \cdots \otimes \mu_n$ from S_0 to $S_1 \times \cdots \times S_n$, given by

$$(\mu_1 \otimes \cdots \otimes \mu_n) f = \mu_1(\mu_2(\cdots(\mu_n f)\cdots))$$

for any measurable $f : S_1 \times \cdots S_n \to \mathbb{R}_+$. In the context of Markov chains, μ_k is often a kernel from S_{k-1} to S_k, $1 \le k \le n$, and we can get a kernel $\mu_1 \cdots \mu_n$ from S_0 to S_n given by

$$(\mu_1 \cdots \mu_n)_s B = (\mu_1 \otimes \cdots \otimes \mu_n)_s (S_1 \times \cdots \times S_{n-1} \times B)$$

$$= \int \mu_1(s, ds_1) \int \mu_2(s_1, ds_2) \cdots \int \mu_{n-1}(s_{n-2}, ds_{n-1}) \mu_n(s_{n-1}, B), \quad s \in S_0, \quad B \in \overline{S}_n.$$

Exercise . Problems 1, 6, 7, 15, 19 in Exercises of Chapter 1.

2 Processes, Distributions, and Independence

We now begin the study of probability theory. Throughout, fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In the probability context, the sets $A \in \mathcal{A}$ are called events, and $\mathbb{P}A = \mathbb{P}(A)$ is called the probability of A. Given a sequence of events, we may be interested in the events

$$\limsup A_n = \bigcap_n \bigcup_{m \ge n} A_m, \quad \liminf A_n = \bigcup_n \bigcap_{m \ge n} A_m.$$

Since $\omega \in \limsup A_n$ if and only if there are infinitely many n such that $\omega \in A_n$, we also call $\limsup A_n$ the event that A_n happens *infinitely often*, and denote it as $\{A_n \text{ i.o.}\}$. Since $\omega \in \liminf A_n$ if and only if there is N such that $\omega \in A_n$ for all $n > N$, we also call $\liminf A_n$

the event that A_n happens *ultimately*, and denote it as $\{A_n \text{ ult.}\}$. By basic set theory, we get $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ ult.}\}$. We may understand $\{A_n \text{ i.o.}\}$ and $\{A_n \text{ ult.}\}$ from another perspective. We view every $\omega \in \Omega$ as a universe. The space Ω is a collection of parallel universes. For a universe ω, we understand A_n as something that we know whether it happens at the time n. If $\omega \in A_n$, then in the universe ω, A_n happens at the time n. Then $\{A_n \text{ i.o.}\}$ is the collection of universes in which A_n happen infinitely many times; and $\{A_n \text{ ult.}\}$ is the collection of universes in which all A_n happen for n big enough.

By countably subadditivity of $\mathbb{P}$, for any $m \in \mathbb{N}$,

$$\mathbb{P}\{A_n \text{ i.o.}\} \le \mathbb{P}[\bigcup_{n=m}^{\infty} A_n] \le \sum_{n=m}^{\infty} \mathbb{P}A_n.$$

If $\sum_n \mathbb{P}A_n < \infty$, then $\sum_{n=m}^{\infty} \mathbb{P}A_n \to 0$ as $m \to \infty$. So we get $\mathbb{P}\{A_n \text{ i.o.}\} = 0$. This is the easy part of the *Borel-Cantelli lemma*.

A measurable mapping f from Ω to another measurable space $(S, \overline{S})$ is called a random element in S. It is called a random variable when $S = \mathbb{R}$, a random vector when $S = \mathbb{R}^n$, a random sequence when $S = \mathbb{R}^\infty$, a random or stochastic process when S is a function space, and a random measure (kernel) when S is a class of measures. The notation $\mathbb{P}$-almost everywhere will now be called almost surely (abbreviated as a.s.). Let $(S, \overline{S})$ be a measurable space and T be an abstract index set. Let $U \subset S^T$. A mapping X from Ω to U, which is $U \cap \overline{S}^T$-measurable, is called an S-valued (random) process on T with paths in U. By Lemma 1.8, X can be treated as a family of random elements X_t in the *state space* S.

Given a random element ζ in $(S, \overline{S})$, the pushforward $\mathbb{P} \circ \zeta^{-1}$ is a probability measure on $(S, \overline{S})$, and is called the *distribution* or *law* of ζ. We write it as $\text{Law}(\zeta)$. For two random elements ζ and η in the same measurable space, the equality $\zeta \overset{\text{d}}{=} \eta$ means that $\text{Law}(\zeta) = \text{Law}(\eta)$.

If for every $t \in T$, X_t is a random element in a measurable space $(S_t, \overline{S}_t)$. Then $X = (X_t : t \in T)$ is a random element in $(\prod_t S_t, \prod_t \overline{S}_t)$. For every finite subset $\Lambda \subset T$, the associated finite-dimensional distribution is given by

$$\mu_\Lambda = \text{Law}(X_t : t \in \Lambda).$$

For $\Lambda_1 \subset \Lambda_2 \subset T$, we use π_{Λ,Λ_1} to denote the natural projection from $\prod_{t \in \Lambda_2} S_t$ to $\prod_{t \in \Lambda_1} S_t$, which is measurable. We omit Λ_2 when it is equal to T. Since $(X_t : t \in \Lambda) = \pi_\Lambda(X)$, the finite dimensional distribution μ_Λ is the pushforwards of the law of X under π_Λ, i.e.,

$$\mu_\Lambda = \text{Law}(X_t : t \in \Lambda) = (\pi_\Lambda)_* \text{Law}(X).$$

Let $\mathcal{P}_*(T)$ to denote the class of all nonempty finite subset of T. Suppose $\Lambda_1 \subset \Lambda_2 \in \mathcal{P}_*(T)$. From $\pi_{\Lambda_1} = \pi_{\Lambda_2,\Lambda_1} \circ \pi_{\Lambda_2}$ we get

$$\mu_{\Lambda_1} = (\pi_{\Lambda_2,\Lambda_1})_* \mu_{\Lambda_2}, \quad \Lambda_1 \subset \Lambda_2 \in \mathcal{P}_*(T). \tag{2.1}$$

If we have a family of finite dimensional distributions μ_Λ, $\Lambda \in \mathcal{P}_*(T)$, on $\prod_{t \in \Lambda} S_t$, and the consistency condition (2.1) holds for every pair $\Lambda_1 \subset \Lambda_2 \in \mathcal{P}_*(T)$, then we call $(\mu_\Lambda \Lambda)_{\Lambda \in \mathcal{P}_*(T)}$ a consistent family.

Theorem 5.16 (Kolmogorov extension theorem)**.** *Suppose each S_t, $t \in T$, is a Borel space. Then for any consistent family $(\mu_\Lambda \Lambda)_{\Lambda \in \mathcal{P}_*(T)}$, there exists a unique probability measure μ on $\prod_{t\in T} S_t$ such that for every $\Lambda \in \mathcal{P}_*(T)$, $\mu_\Lambda = (\pi_\Lambda)_* \mu$.*

Remark . One important application of Kolmogorov extension theorem is the existence of infinite product measure. Suppose T is an infinite index set, and for each $t \in T$, μ_t is a probability measure on a Borel measurable space $(S_t, \overline{S}_t)$. We define the family

$$\mu_\Lambda = \prod_{t\in\Lambda} \mu_t, \quad \Lambda \in \mathcal{P}_*(T),$$

where $\mathcal{P}_*(T)$ is the class of nonempty subsets of T. We have known that the finite product measures are well defined. The consistency condition is easy to check. Since S_t are all Borel spaces, by Kolmogorov extension theorem, there is a unique probability measure μ on $\prod_t \overline{S}_t$ such that $\mu_\Lambda = (\pi_\Lambda)_*(\mu)$ for every $\Lambda \in \mathcal{P}_*(T)$. Such μ is the product $\prod_{t\in T} \mu_t$.

For a random variable ζ, the *expected value*, *expectation*, or *mean* of ζ is defined as

$$\mathbb{E}\zeta = \int \zeta d\mathbb{P} = \int x d\,\mathrm{Law}(\zeta)$$

whenever either integral exists. The last equality follows from Lemma 1.22. By that lemma, we also note that for any random element ζ in a measurable space S and a measurable map $f : S \to \mathbb{R}$,

$$\mathbb{E}f(\zeta) = \int_\Omega f(\zeta) d\mathbb{P} = \int_S f(s) d\,\mathrm{Law}(\zeta) = \int_\mathbb{R} x d\,\mathrm{Law}(f \circ \zeta),$$

if any integral exists. For a random variable ζ and an event A, we often write $\mathbb{E}[\zeta; A]$ for $\mathbb{E}[\mathbf{1}_A \zeta] = \int_A \zeta d\mathbb{P}$.

Proof of Kolmogorov extension theorem. The uniqueness part follows from the monotone class theorem.

We now consider the existence part. First assume that $T = \mathbb{N}$. Every Borel space S_t is Borel isomorphic to a Borel subset of $[0, 1]$. Since the theorem depends only on the σ-algebra structure of S_t, we may assume that each S_t is a Borel subset of $[0, 1]$. Then each μ_Λ can be also viewed as a probability measure on $[0, 1]^\Lambda$.

The proof uses Carathéodory extension theorem. For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ denote the σ-algebra on $\prod_{k\in\mathbb{N}} S_k$ generated by the projection $\pi_{\mathbb{N}_n}$, where $\mathbb{N}_n = \{1, \dots, n\}$. This means that $\mathcal{F}_n$ is the family of subsets $A \subset [0, 1]^\infty$ of the form $B \times [0, 1]^\infty$, where $B \in \mathcal{B}([0, 1])^n$. Then $\mathcal{F}_n$ is increasing in n. Let $\mathcal{R} = \bigcup_n \mathcal{F}_n$. Then $\mathcal{R}$ is a ring in $[0, 1]^\infty$, and $\mathcal{B}([0, 1])^\infty = \sigma(\mathcal{R})$. We define $\mu : \mathcal{R} \to [0, 1]$ such that if $A = B \times [0, 1]^\infty \in \mathcal{F}_n$ for some $B \in \mathcal{B}([0, 1])^n$, then $\mu A = \mu_{\mathbb{N}_n} B$. Such μ is well defined thanks to the consistency condition.

We now show that μ is a pre-measure. It is easy to see that μ satisfies the finitely additivity. It remains to show that if $A_1 \supset A_2 \supset \cdots \in \mathcal{R}$ with $\mu A_n \ge \varepsilon > 0$ for all n, then $\bigcap_n A_n \ne \emptyset$. Assume that $A_k \in \mathcal{F}_{n_k}$. Since $\mathcal{F}_n$ is increasing in n, we may assume that (n_k) is increasing in k. By inserting repeated sets (e.g., if $n_1 = 2$, $n_2 = 5$, $n_3 = 7$, then we use a new sequence $(A_1, A_1, A_2, A_2, A_2, A_3, A_3, \dots)$ to replace $(A_1, A_2, A_3, \dots)$), we may assume that $A_n \in \mathcal{F}_n$ for each n. Suppose $A_n = B_n \times [0,1]^\infty$ for some $B_n \in \mathcal{B}([0,1])^n$.

By Lemma 1.16, for each n, there is a closed set $K_n \subset B_n$ such that $\mu_{\mathbb{N}_n}(B_n \setminus K_n) < \frac{\varepsilon}{2^n}$. Let $A_n' = K_n \times [0,1]^\infty \subset A_n$. Then $\mu(A_n \setminus A_n') < \frac{\varepsilon}{2^n}$, and each A_n' is a compact subset of $[0,1]^\infty$. Let $A_n'' = \bigcap_{j=1}^n A_j'$, $n \in \mathbb{N}$. Then for every n, A_n'' is a compact subset of A_n, and $A_n \setminus A_n'' \subset \bigcap_{j=1}^n (A_j \setminus A_j')$. The latter implies that $\mu(A_n \setminus A_n'') \le \sum_{j=1}^n \frac{\varepsilon}{2^j} < \varepsilon$, which together with $\mu A_n > \varepsilon$ implies that $A_n'' \ne \emptyset$. Since $A_1'' \supset A_2'' \supset \cdots$ and each A_n'' is compact, we get $\bigcap_n A_n'' \ne \emptyset$, which together with $\overline{A_n''} \subset A_n$ implies that $\bigcap_n A_n \ne \emptyset$.

Thus, μ is a pre-measure on $\mathcal{R}$. By Carathéodory extension theorem, μ extends to a probability measure on $[0,1]^\infty$. By the definition of μ on $\mathcal{R}$, for every $n \in \mathbb{N}$, $\mu(\prod_{j=1}^n S_j \times [0,1]^\infty) = \mu_{\mathbb{N}_n} \prod_{j=1}^n S_j = 1$. So $\mu \prod_{n=1}^\infty S_n = \lim_n \mu(\prod_{j=1}^n S_j \times [0,1]^\infty) = 1$. Thus, μ is also a probability measure on $\prod_{n=1}^\infty S_n$. For every $A_n \in \prod_{j=1}^n \overline{S}_j \in \mathcal{B}([0,1])^n$, we have $\mu(A_n \times \prod_{j=n+1}^\infty S_j) = \mu(A_n \times [0,1]^\infty) = \mu_{\mathbb{N}_n} A_n$. So $\mu_{\mathbb{N}_n} = (\pi_{\mathbb{N}_n})_*(\mu)$ for every $n \in \mathbb{N}$. For every $\Lambda \in \mathcal{P}_*(\mathbb{N})$, there is $n \in \mathbb{N}$ such that $\Lambda \subset \mathbb{N}_n$. By (2.1) we have

$$\mu_\Lambda = (\pi_{\mathbb{N}_n,\Lambda})_*(\mu_{\mathbb{N}_n}) = (\pi_{\mathbb{N}_n,\Lambda})_* \circ (\pi_{\mathbb{N}_n})_*(\mu) = (\pi_\Lambda)_*(\mu).$$

So μ is what we need. We now know that the theorem holds if T is countable.

Finally, we consider a general T. Let $\mathcal{P}_\sigma(T)$ denote the class of all nonempty countable subsets of T. We have proved that for any $\Gamma \in \mathcal{P}_\sigma(T)$, there exists a unique probability measure μ_Γ on $\prod_{t\in\Gamma} S_t$ such that for any finite subset Λ of Γ, $\mu_\Lambda = (\pi_{\Gamma,\Lambda})_*(\mu_\Gamma)$. By the uniqueness, if $\Gamma_1 \subset \Gamma_2 \in \mathcal{P}_\sigma(T)$, then $\mu_{\Gamma_1} = (\pi_{\Gamma_2,\Gamma_1})_*(\mu_{\Gamma_2})$. For each $\Gamma \in \mathcal{P}_\sigma(T)$, let

$$\mathcal{F}_\Gamma = (\pi_\Gamma)^{-1} \prod_{t\in\Gamma} S_t = \prod_{t\in\Gamma} \overline{S}_t \times \prod_{t\in T\setminus\Gamma} S_t.$$

It is easy to check that $\bigcup_{\Gamma\in\mathcal{P}_\sigma(T)} \mathcal{F}_\Gamma$ is a σ-algebra, and so equals $\prod_{t\in T} \overline{S}_t$. We define $\mu : \bigcup_{\Gamma\in\mathcal{P}_\sigma(T)} \mathcal{F}_\Gamma \to [0,1]$ such that if A has an expression $\pi_\Gamma^{-1} B \in \mathcal{F}_\Gamma$ for some $\Gamma \in \mathcal{P}_\sigma(T)$ and $B \in \prod_{t\in\Gamma} \overline{S}_t$, then $\mu A = \mu_\Gamma B$. The value of μA does not depend on the choice of the expression of A thanks to the consistency condition $\mu_{\Gamma_1} = (\pi_{\Gamma_2,\Gamma_1})_*(\mu_{\Gamma_2})$. So μ is well defined. From the definition, $\mu_\Gamma = (\pi_\Gamma)_*\mu$ for every $\Gamma \in \mathcal{P}_\sigma(T)$. If $\Lambda \in \mathcal{P}_*(T)$, we may pick $\Gamma \in \mathcal{P}_\sigma(T)$ with $\Gamma \supset \Lambda$. Then we get the desired equality $\mu_\Lambda = (\pi_{\Gamma,\Lambda})_* \circ (\pi_\Gamma)_*\mu = (\pi_\Lambda)_*\mu$. □

Remark . For the existence of infinite product measure, we do not need to assume that the S_t are Borel spaces. The proof still uses Carathéodory extension theorem. Following the proof of Kolmogorov extension theorem and the construction of the infinite product measure, we need to show that, if $T = \mathbb{N}$, and $A_1 \supset A_2 \supset \cdots$ satisfy that for some $\varepsilon > 0$,

$$A_n = B_n \times \prod_{j>n} S_j, \quad \text{for some } B_n \in \prod_{j=1}^n \overline{S}_j \text{ with } (\prod_{j=1}^n \mu_j) B_n \ge \varepsilon,$$

for all $n \in \mathbb{N}$, then $\bigcap_n A_n \ne \emptyset$.

For $n > m \in \mathbb{N}$ and $(x_1, \dots, x_m) \in \prod_{j=1}^m S_j$, we define

$$B_n(x_1, \dots, x_m) = \{(x_{m+1}, \dots, x_n) \in \prod_{j=m+1}^n S_j : (x_1, x_2, \dots, x_n) \in B_n.$$

By Lemma 1.26, for each $(x_1, \dots, x_m) \in \prod_{j=1}^m S_j$, $B_n(x_1, \dots, x_m)$ is a measurable subset of $\prod_{j=m+1}^n S_j$, and $(x_1, \dots, x_m) \mapsto (\prod_{j=m+1}^n \mu_j) B_n(x_1, \dots, x_m)$ is a measurable function on $\prod_{j=1}^m S_j$. For $n \geq 2$, let

$$F_n^{(1)} = \{x_1 \in S_1 : (\prod_{j=2}^n \mu_j) B_n(x_1) > \varepsilon/2\}.$$

Then $F_2^{(1)} \supset F_3^{(1)} \supset \cdots$ are measurable subsets of S_1. By Fubini theorem,

$$\varepsilon \leq (\prod_{j=1}^n \mu_j) B_n = \int \mu_1(dx_1)(\prod_{j=2}^n \mu_j) B_n(x_1) \leq \frac{\varepsilon}{2} \mu_1(F_n^{(1)})^c + \mu_1 F_n^{(1)},$$

which implies that $\mu_1 F_n^{(1)} \geq \varepsilon/2$ for all $n \geq 2$. So $\mu_1 \bigcap_n F_n^{(1)} \geq \varepsilon/2$, and then we have $\bigcap_{n \geq 2} F_n^{(1)} \neq \emptyset$.

Pick $\overline{x_1} \in \bigcap_{n \geq 2} F_n^{(1)}$. Let $B_n^{(1)} = B_n(\overline{x}_1)$, $n \geq 2$. For every $n \geq 3$, and $x_2 \in S_2$, let

$$B_n^{(1)}(x_2) = B_n(\overline{x}_1, x_2) = \{(x_3, \dots, x_n) \in \prod_{j=3}^n S_j : (\overline{x}_1, x_2, x_3, \dots, x_n) \in B_n.$$

For $n \geq 3$, let

$$F_n^{(2)} = \{x_2 \in S_2 : (\prod_{j=3}^n \mu_j) B_n^{(1)}(x_2) > \varepsilon/4\}.$$

Using Fubini theorem and a similar argument as above, we get $\bigcap_{n \geq 3} F_n^{(2)} \neq \emptyset$. So we may pick $\overline{x_2} \in \bigcap_{n \geq 3}^{(2)} F_n^{(2)}$. Then $(\prod_{j=3}^n \mu_j) B_n(\overline{x}_1, \overline{x}_2) > \varepsilon/4$ for any $n \geq 3$.

Repeating the argument, we can find a sequence $\overline{x} := (\overline{x}_1, \overline{x}_2, \dots) \in \prod_k S_k$ such that $\overline{x}_k \in S_k$, $k \in \mathbb{N}$, and

$$(\prod_{j=m+1}^n \mu_j) B_m(\overline{x}_1, \dots, \overline{x}_n) > \varepsilon/2^n, \quad \forall m > n \in \mathbb{N}.$$

We now show that $\overline{x} \in \bigcap_n A_n$. Pick any $n \in \mathbb{N}$, since $A_n = B_n \times \prod_{j=n+1}^\infty S_j$, to prove that $\overline{x} \in A_n$, it suffices to show that $(\overline{x}_1, \dots, \overline{x}_n) \in B_n$. To prove this assertion, note that from $\mu_{n+1} B_{n+1}(\overline{x}_1, \dots, \overline{x}_n) > 0$ we get $B_{n+1}(\overline{x}_1, \dots, \overline{x}_n) \neq \emptyset$. So there is $x_{n+1} \in S_{n+1}$ such that $(\overline{x}_1, \dots, \overline{x}_n, x_{n+1}) \in B_{n+1}$. From $A_{n+1} \subset A_n$, we get $B_{n+1} \subset B_n \times S_{n+1}$, which then implies $(\overline{x}_1, \dots, \overline{x}_n) \in B_n$.

A random vector ζ in $\mathbb{R}^n$ is called integrable if every component ζ_j, $1 \le j \le n$, is integrable.

Lemma 2.5 (Jensen's inequality)**.** *Let ζ be an integrable random vector in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be convex, i.e.,*

$$f(px + (1-p)y) \le pf(x) + (1-p)f(y), \quad x, y \in \mathbb{R}^n, \quad 0 \le p \le 1.$$

Then $f(\mathbb{E}\zeta) \le \mathbb{E}[f(\zeta)]$.

Proof. We use a version of Hahn-Banach Theorem, which asserts that

$$f(x) = \sup_L L(x),$$

where the supremum is over all affine functions $L : \mathbb{R}^n \to \mathbb{R}$ with $L \le f$. Since for every affine function $L \le f$,

$$L(\mathbb{E}\zeta) = \mathbb{E}[L(\zeta)] \le \mathbb{E}[f(\zeta)],$$

taking the supremum over all affine functions $L \le f$, we get $f(\mathbb{E}\zeta) \le \mathbb{E}[f(\zeta)]$. □

For a random variable ζ and $p > 0$, the integral $\mathbb{E}|\zeta|^p = \|\zeta\|_p^p$ is called the *p-th absolute moment* of ζ.

Lemma 2.4. *For any random variable $\zeta \ge 0$ and $p > 0$,*

$$\mathbb{E}\zeta^p = p\int_0^\infty \mathbb{P}\{\zeta > t\}t^{p-1}dt = p\int_0^\infty \mathbb{P}\{\zeta \ge t\}t^{p-1}dt.$$

Proof. By Fubini's theorem and change of variables,

$$\mathbb{E}\zeta^p = \mathbb{E}\int_0^\infty \mathbf{1}\{\zeta^p > s\}ds = \int_0^\infty \mathbb{E}\mathbf{1}\{\zeta > s^{1/p}\}ds$$
$$= \int_0^\infty \mathbb{E}\mathbf{1}\{\zeta > t\}pt^{p-1}dt = p\int_0^\infty \mathbb{P}\{\zeta > t\}t^{p-1}dt.$$

Here in the third "=" we used $s = t^p$. The proof if the second expression is similar. □

Exercise . Show that $\|\zeta\|_p \le \|\zeta\|_q$ if $p \le q$. Here we use the fact that $\mathbb{P}\Omega = 1$. So the L^p-spaces are decreasing in p.

The *covariance* of two random variables $\zeta, \eta \in L^2$ is given by

$$\mathrm{cov}(\zeta, \eta) = \mathbb{E}(\zeta - \mathbb{E}\zeta)(\eta - \mathbb{E}\eta) = \mathbb{E}\zeta\eta - \mathbb{E}\zeta\mathbb{E}\eta.$$

It is clearly bilinear. The *variance* of $\zeta \in L^2$ is defined by

$$\mathrm{var}(\zeta) = \mathrm{cov}(\zeta, \zeta) = \mathbb{E}(\zeta - \mathbb{E}\zeta)^2 = \mathbb{E}\zeta^2 - (\mathbb{E}\zeta)^2.$$

By Cauchy inequality,

$$|\operatorname{cov}(\zeta,\eta)|^2 \le \operatorname{var}(\zeta)\operatorname{var}(\eta).$$

We say that ζ and η are *uncorrelated* if $\operatorname{cov}(\zeta,\eta)=0$.

For any collection $\zeta_t \in L^2$, $t\in T$, the associated *covariance function* $\rho_{s,t}=\operatorname{cov}(\zeta_s,\zeta_t)$, $s,t\in T$, is *nonnegative definite*, in the sense that for any $n\in\mathbb{N}$, $t_1,\dots,t_n\in T$, and $a_1,\dots,a_n\in\mathbb{R}$, $\sum_{i,j} a_i a_j \rho_{t_i,t_j}\ge 0$. This is because

$$\sum_{i,j} a_i a_j \rho_{t_i,t_j} = \sum_{i,j} a_i a_j \mathbb{E}(\zeta_{t_i}-\mathbb{E}\zeta_{t_i})(\zeta_{t_j}-\mathbb{E}\zeta_{t_j}) = \mathbb{E}(\sum_i a_i(\zeta_{t_i}-\mathbb{E}\zeta_{t_i}))^2 \ge 0.$$

Example . We now study the following distributions (i.e. probability measures) on $\mathbb{R}$. In each case below, we suppose ζ is a random variable with $\operatorname{Law}(\xi)=\mu$. Recall that $\mathbb{E}\zeta=\int x d\mu$ and $\operatorname{var}(\zeta)=\mathbb{E}\zeta^2-(\mathbb{E}\zeta)^2=\int x^2 d\mu-(\int x d\mu)^2$ are determined by μ. We first consider discrete distributions, which are combinations of Dirac measures.

(i) The degenerate distribution at x_0. This is the point mass $\mu=\delta_{x_0}$, $x_0\in\mathbb{R}$. We have $\mathbb{E}\zeta=\int x d\delta_{x_0}=x_0$ and $\mathbb{E}\zeta^2=\int x^2 d\delta_{x_0}=x_0^2$ and so $\operatorname{var}(\zeta)=0$.

(ii) The Bernoulli distribution with parameter $p\in[0,1]$. The measure, denoted by $\mathrm{B}(p)$, has the form $\mu=p\delta_1+(1-p)\delta_0$. We have $\mathbb{E}\zeta=p(1)+(1-p)(0)=p$ and $\mathbb{E}\zeta^2=p(1^2)+(1-p)(0^2)=p$. So $\operatorname{var}(\zeta)=p-p^2$.

(iii) The binomial distribution with parameter $n\in\mathbb{N}$ and $p\in[0,1]$. The measure, denoted by $\mathrm{B}(n,p)$, has the form $\mu=\sum_{k=0}^n p^k(1-p)^{n-k}\binom{n}{k}\delta_k$. It is a probability measure because $\sum_{k=0}^n p^k(1-p)^{n-k}\binom{n}{k}=(p+(1-p))^n=1$. We have

$$\mathbb{E}\zeta=\sum_{k=0}^n p^k(1-p)^{n-k}k\binom{n}{k}=\sum_{k=1}^n p^k(1-p)^{n-k}\frac{n!}{(k-1)!(n-k)!}$$

$$=np\sum_{k=1}^n p^{k-1}(1-p)^{n-k}\frac{(n-1)!}{(k-1)!(n-k)!}=np;$$

$$\mathbb{E}(\zeta^2-\zeta)=\sum_{k=0}^n p^k(1-p)^{n-k}k(k-1)\binom{n}{k}$$

$$=n(n-1)p^2\sum_{k=2}^n p^{k-2}(1-p)^{n-k}\frac{(n-2)!}{(k-1)!(n-k)!}=n(n-1)p^2.$$

So $\operatorname{var}(\zeta)=\mathbb{E}(\zeta^2-\zeta)+\mathbb{E}\zeta-(\mathbb{E}\zeta)^2=n(p-p^2)$.

(iv) The geometric distribution with parameter $p \in (0,1]$. The measure, denoted by $\mathrm{Geom}(p)$, has the form $\mu = \sum_{k=1}^{\infty}(1-p)^{k-1}p\delta_k$. It is a probability measure because $\sum_{k=1}^{\infty}(1-p)^{k-1}p = \frac{p}{1-(1-p)} = 1$, where we used $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ for $|x| < 1$. We have

$$\mathbb{E}\zeta = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{p}{(1-(1-p))^2} = \frac{1}{p};$$

$$\mathbb{E}(\zeta^2 - \zeta) = \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1}p$$

$$= p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = \frac{2p(1-p)}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}.$$

Here we used the equalities $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$ and $\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$ for $|x| < 1$, which can be proved by differentiating the equality $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$. Thus, $\mathrm{var}(\zeta) = \mathbb{E}(\zeta^2-\zeta) + \mathbb{E}\zeta - (\mathbb{E}\zeta)^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

(v) The Poisson distribution with parameter $\lambda > 0$. The measure, denoted by $\mathrm{Pois}(\lambda)$, has the form $\mu = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}\delta_k$. It is a probability measure because $\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda}$. We have

$$\mathbb{E}\zeta = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda;$$

$$\mathbb{E}(\zeta^2-\zeta) = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} = \lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.$$

So $\mathrm{var}(\zeta) = \mathbb{E}(\zeta^2-\zeta) + \mathbb{E}\zeta - (\mathbb{E}\zeta)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Below are continuous distributions on $\mathbb{R}$, which have density functions w.r.t. the Lebesgue measure λ. In each example below, f is the λ-density of $\mathrm{Law}(\zeta)$. Then $\mathbb{E}\zeta = \int_{\mathbb{R}} xf(x)dx$ and $\mathbb{E}\zeta^2 = \int_{\mathbb{R}} x^2 f(x)dx$.

(i) The uniform distribution $U[a,b]$ for $a < b \in \mathbb{R}$. The density is $f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}$. Then $\mathbb{E}\zeta = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\frac{x^2}{2}|_a^b = \frac{a+b}{2}$ and $\mathbb{E}\zeta^2 = \frac{1}{b-a}\int_a^b x^2dx = \frac{1}{b-a}\frac{x^3}{3}|_a^b = \frac{1}{3}(a^2+ab+b^2)$. So $\mathrm{var}(\zeta) = \frac{1}{3}(a^2+ab+b^2) - (\frac{a+b}{2})^2 = \frac{(a-b)^2}{12}$.

(ii) The exponential distribution $\mathrm{Exp}(\lambda)$ with parameter $\lambda > 0$. The density is $\mathbf{1}_{[0,\infty)}\lambda e^{-\lambda x}$. It is a probability measure because $\int_0^{\infty}\lambda e^{-\lambda x}dx = 1$. We have

$$\mathbb{E}\zeta = \int_0^{\infty} x\lambda e^{-\lambda x}dx = -\int_0^{\infty}(-e^{-\lambda x})dx = \frac{1}{\lambda};$$

$$\mathbb{E}\zeta^2 = \int_0^{\infty} x^2\lambda e^{-\lambda x}dx = -\int_0^{\infty} 2x(-e^{-\lambda x})dx = \frac{2}{\lambda^2}.$$

Here we use integration by parts. So $\mathrm{var}(\zeta) = \mathbb{E}\zeta^2 - (\mathbb{E}\zeta)^2 = \frac{1}{\lambda^2}$.

(iii) The normal distribution $\mathrm{N}(\mu, \sigma^2)$ with parameter $\mu \in \mathbb{R}$ and $\sigma > 0$. The density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

It is a probability measure because using a change of variable $x = \mu + \sqrt{\sigma} y$, we get

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} dy,$$

and by Fubini's theorem and using polar coordinate,

$$\Big(\int_{\mathbb{R}} e^{-y^2/2} dy \Big)^2 = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x^2/2} e^{-y^2/2} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta$$

$$= 2\pi \int_0^{\infty} e^{-r^2/2} r dr = 2\pi (-e^{r^2/2})|_0^{\infty} = 2\pi.$$

Using the same change of variable $x = \mu + \sigma y$, we get

$$\mathbb{E}\zeta = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\mu + \sigma y) e^{-y^2/2} dy = \mu;$$

$$\mathbb{E}\zeta^2 = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\mu + \sigma y)^2 e^{-y^2/2} dy$$

$$= \mu + \sigma^2 \frac{1}{\sqrt{2\pi}} \int y^2 e^{-y^2/2} dy.$$

Here we used that $\int_{\mathbb{R}} y e^{-y^2/2} dy = 0$ because the integrand is odd. Thus,

$$\mathrm{var}(\zeta) = \sigma^2 \frac{1}{\sqrt{2\pi}} \int y^2 e^{-y^2/2} dy = \sigma^2 \frac{1}{\sqrt{2\pi}} \int e^{-y^2/2} dy = \sigma^2,$$

where we used integration by parts: differentiating y and integrating $y e^{-y^2/2}$.

We understand the degenerate distribution δ_μ as a normal distribution $\mathrm{N}(\mu, 0)$, which does not have a λ-density. In this case it trivially holds that $\mathbb{E}\zeta = \mu$ and $\mathrm{var}(\zeta) = 0$. If $\mathrm{Law}(\zeta) = \mathrm{N}(\mu, \sigma^2)$, then for any $a, b \in \mathbb{R}$, $\mathrm{Law}(a\zeta + b) = \mathrm{N}(a\mu + b, a^2\sigma^2)$.

Exercise . Prove the following

(i) The binomial distribution $B(n, p)$ is the n-th convolution power of the Bernoulli distribution $B(p)$, i.e.,

$$\underbrace{\mathrm{B}(p) * \cdots * \mathrm{B}(p)}_{n \text{ copies}} = \mathrm{B}(n, p).$$

(ii) The Poisson distributions satisfy that for any $\lambda_1, \lambda_2 > 0$,

$$\mathrm{Pois}(\lambda_1) * \mathrm{Pois}(\lambda_2) = \mathrm{Pois}(\lambda_1 + \lambda_2).$$

(iii) The normal distributions satisfy that for any $\mu_1, \mu_2 \in \mathbb{R}$ and $v_1, v_2 \geq 0$,

$$\mathrm{N}(\mu_1, v_1) * \mathrm{N}(\mu_2, v_2) = \mathrm{N}(\mu_1 + \mu_2, v_1 + v_2).$$

Example . There exists a probability measure on $\mathbb{R}$, which is not a combination of a discrete distribution and a continuous distribution. Consider the Cantor 1/3 set:

$$C = \{\sum_{n=1}^{\infty} \frac{a_n}{3^n} : a_n \in \{0, 2\}, n \in \mathbb{N}\}.$$

It is Borel isomorphic to the product space $\{0, 2\}^\infty$. Let $f : \{0, 2\}^\infty \to C$ be the bijective measurable map

$$f((a_n)_{n\in\mathbb{N}}) = \sum_{n=1}^{\infty} \frac{a_n}{3^n}.$$

Let $\mu = \frac{1}{2}(\delta_0 + \delta_2)$ be a probability measure on $\{0, 2\}$. We have known that the product measure μ^∞ exists on $\{0, 2\}^\infty$. The pushforward measure $f_*\mu^\infty$ is a probability measure on C. Then $f_*\mu^\infty(C^c) = 0$. We know that $\lambda(C) = 0$. So $f_*\mu^\infty \perp \lambda$. We also see that $f_*\mu^\infty$ has no point mass, i.e., there does not exist $x \in C$ such that $f_*\mu^\infty(\{x\}) > 0$, because μ^∞ has no point mass.

Exercise . Let $\mu = \frac{1}{2}(\delta_0 + \delta_1)$ be a probability measure on $\{0, 1\}$. Let $f : \{0, 1\}^\infty \to [0, 1]$ be defined by

$$f((a_n)_{n\in\mathbb{N}}) = \sum_{n=1}^{\infty} \frac{a_n}{2^n}.$$

Prove that f is measurable, and $f_*\mu^\infty = \lambda(\cdot \cap [0, 1])$.

We now define and study the notation of independence. The events A_t, $t \in T$, are said to be *(mutually) independent* (w.r.t. $\mathbb{P}$) if for any distinct indices $t_1, \dots, t_n \in T$,

$$\mathbb{P}[\bigcap_{k=1}^{n} A_{t_k}] = \prod_{k=1}^{n} \mathbb{P}A_{t_k}. \tag{2.2}$$

We say that a class of families $\mathcal{C}_t$, $t \in T$, are independent, if when we pick an A_t in every $\mathcal{C}_t$, then A_t, $t \in T$, are independent. We do not require the independence between events in the same family $\mathcal{C}_t$. The random elements ζ_t, $t \in T$, are said to be independent if the independence holds for the generated σ-algebras $\sigma(\zeta_t)$, $t \in T$.

Lemma 2.10 (Strengthened version)**.** *For each $t \in T$, let ζ_t be a random element in a measurable space $(S_t, \overline{S}_t)$. Let $\zeta = (\zeta_t : t \in T)$ be a random element in $\prod_{t\in T} S_t$. Then ζ_t, $t \in T$, are independent iff*

$$\mathrm{Law}(\zeta) = \prod_{t\in T} \mathrm{Law}(\zeta_t).$$

Proof. This is a strengthened version of Lemma 2.10 of the textbook, which assumes that T is finite. We leave the proof as an exercise. $\square$

Corollary . *Let T be an arbitrary index set. Suppose for each $t \in T$, μ_t is a probability measure on a Borel space $(S_t, \overline{S}_t)$. Then there is a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and an independent family of random elements ζ_t, $t \in T$, defined on it such that $\mathrm{Law}(\zeta_t) = \mu_t$ for each t.*

Proof. We have shown that the product measure $\prod_{t\in T} \mu_t$ on $(\prod_{t\in T} S_t, \prod_{t\in T} \overline{S}_t)$ exists. Let $(\Omega, \mathcal{A}, \mathbb{P}) = (\prod_{t\in T} S_t, \prod_{t\in T} \overline{S}_t, \prod_{t\in T} \mu_t)$. For each $t \in T$, let the random element $\zeta_t : \Omega \to S_t$ be the projection map $\pi_{\{t\}}$. Then the random element $\zeta = (\zeta_t : t \in T)$ from Ω to $\prod_{t\in T} S_t = \Omega$ is just the identity map. So $\mathrm{Law}(\zeta_t) = (\pi_{\{t\}})_* \prod_{s\in T} \mu_s = \mu_t$, $t \in T$, and $\mathrm{Law}(\zeta) = \prod_{t\in T} \mu_t$. By Lemma 2.10, ζ_t, $t \in T$, are independent. $\square$

Lemma 2.6. *If the π-systems $\mathcal{C}_t$, $t \in T$, are independent, then so are the σ-fields $\mathcal{F}_t := \sigma(\mathcal{C}_t)$, $t \in T$.*

Proof. We need to show that for any distinct indices $t_1, \dots, t_n \in T$, and any $A_{t_k} \in \mathcal{F}_{t_k}$, $1 \le k \le n$, (2.2) holds. By assumption, it is true if $A_{t_k} \in \mathcal{C}_{t_k}$, $1 \le k \le n$. By a monotone class argument, we may first weaken the assumption on A_{t_1} from $A_{t_1} \in \mathcal{C}_{t_1}$ to $A_{t_1} \in \mathcal{F}_{t_1}$, and then weaken the assumption on A_{t_2} from $A_{t_2} \in \mathcal{C}_{t_2}$ to $A_{t_2} \in \mathcal{F}_{t_2}$. Repeating the argument until we weaken the assumptions of all A_{t_k} from $A_{t_k} \in \mathcal{C}_{t_k}$ to $A_{t_k} \in \mathcal{F}_{t_k}$. Then we get the desired equality. $\square$

Corollary 2.7. *Let $\mathcal{F}_t$, $t \in T$, be independent σ-algebras. Let R_s, $s \in S$, be a partition of T. Then the σ-algebras $\mathcal{F}'_s = \vee_{t\in R_s} \mathcal{F}_t := \sigma(\bigcup_{t\in R_s} \mathcal{F}_t)$, $s \in S$, are independent.*

Proof. For each $s \in S$, let $\mathcal{C}_s$ denote the set of all finite intersections of sets in $\bigcup_{t\in R_s} \mathcal{F}_t$. Then each $\mathcal{C}_s$ is a π-system, and it is straightforward to check that $\mathcal{C}_s$, $s \in S$, are independent. By Lemma 2.6, we have $\mathcal{F}'_s = \sigma(\mathcal{C}_s)$, $s \in S$, are independent. $\square$

Pairwise independence between two objects A and B will be denoted by $A \perp\!\!\!\perp B$. In general, pairwise independence between all pairs of A_t, $t \in T$, say, does not imply the (total) independence of the group A_t, $t \in T$.

Lemma 2.8. *The σ-algebras $\mathcal{F}_1, \mathcal{F}_2, \dots$ are independent iff $\vee_{k\le n} \mathcal{F}_k \perp\!\!\!\perp \mathcal{F}_{n+1}$ for all n.*

Proof. The "only if" part follows from Corollary 2.7. For the "if" part, it suffices to show that for any $n \in \mathbb{N}$ and $A_k \in \mathcal{F}_k$, $1 \le k \le n$, we have $\mathbb{P} \bigcap_{k=1}^n A_k = \prod_{k=1}^n \mathbb{P} A_k$. This follows from induction and the fact that $\mathbb{P} \bigcap_{k=1}^n A_k = \mathbb{P} A_n \cdot \mathbb{P} \bigcap_{k=1}^{n-1} A_k$ because $\mathcal{F}_n \perp\!\!\!\perp \vee_{k\le n-1} \mathcal{F}_k$, and $\bigcap_{k=1}^{n-1} A_k \in \vee_{k\le n-1} \mathcal{F}_k$. $\square$

A σ-algebra $\mathcal{F} \subset \mathcal{A}$ is called ($\mathbb{P}$-)trivial if for any $A \in \mathcal{F}$, $\mathbb{P} A \in \{0, 1\}$.

Lemma 2.9. *(i) A σ-algebra $\mathcal{F} \subset \mathcal{A}$ is trivial iff $\mathcal{F} \perp\!\!\!\perp \mathcal{F}$. (ii) If $\mathcal{F}$ is trivial, and ζ is an $\mathcal{F}$-measurable random element in a separable metric space S, then ζ is a.s. constant.*

Proof. (i) First suppose $\mathcal{F}$ is trivial. Let $A, B \in \mathcal{F}$. Then $\mathbb{P}A$ and $\mathbb{P}B$ equal to 0 or 1. If $\mathbb{P}A = 0$, then since $A \cap B \subset A$, we have $\mathbb{P}[A \cap B] = 0 = \mathbb{P}A \cdot \mathbb{P}B$. Similarly, if $\mathbb{P}B = 0$, then $\mathbb{P}[A \cap B] = \mathbb{P}A \cdot \mathbb{P}B$. Now suppose $\mathbb{P}A = \mathbb{P}B = 1$. Then $\mathbb{P}A^c = \mathbb{P}B^c = 0$. Thus, $\mathbb{P}[A^c \cup B^c] = 0$. So $\mathbb{P}[A \cap B] = 1 - \mathbb{P}[(A \cap B)^2] = 1 - \mathbb{P}[A^c \cup B^c] = 1$. If $\mathcal{F} \perp\!\!\!\perp \mathcal{F}$, then for any $A \in \mathcal{F}$, $\mathbb{P}A = \mathbb{P}(A \cap A) = (\mathbb{P}A)^2$, which implies that $\mathbb{P}A \in \{0, 1\}$, and so $\mathcal{F}$ is trivial.

(ii) Suppose $\mathcal{F}$ is trivial. For each $n \in \mathbb{N}$, we may partition S into mutually disjoint countably many Borel sets $B_{n,j}$ of diameter $< 1/n$. Fix $n \in \mathbb{N}$. Since $\mathbb{P}[\zeta \in B_{n,j}] \in \{0, 1\}$ for each j, and $(B_{n,j})$ is a partition of S, there is j_n such that $\mathbb{P}[\zeta \in B_{n,j_n}] = 1$. So there is a null event N_n such that $\zeta \in B_{n,j_n}$ on N_n^c. Let $N = \bigcup_n N_n$. Then N is a null set, and $\zeta \in \bigcap_n B_{n,j_n}$ on N^c. Since $\operatorname{diam}(B_{n,j_n}) < 1/n$ for all n, ζ is a constant on N^c. □

Lemma 2.11. *Let ζ and η be independent random elements in measurable spaces S and T, and let $f : S \times T \to \mathbb{R}$ be measurable. If $f \ge 0$, then $\mathbb{E}f(\zeta, \eta) = \mathbb{E}[\mathbb{E}[f(s, \eta)]|_{s=\zeta}]$. Here the RHS means that we first fix $s \in S$ and integrate the random variable $f(s, \eta)$, which is a measurable function in $s \in S$ by Lemma 1.38; then we compose it with ζ to get a random variable, and integrate it. If we do not assume that $f \ge 0$, but assume that either $\mathbb{E}|f(\zeta, \eta)| < \infty$ or $\mathbb{E}[\mathbb{E}[f(s, \eta)]|_{s=\zeta}] < \infty$, then the equality also holds.*

Proof. This lemma essentially follows from Fubini's theorem. We now only work on the case that $f \ge 0$. Let μ and ν be the laws of ζ and η, respectively. Since $\zeta \perp\!\!\!\perp \eta$, by Lemma 2.10, $\operatorname{Law}(\zeta, \eta) = \mu \times \nu$. By Fubini's theorem,

$$\mathbb{E}f(\zeta, \eta) = \int f(s, t) \mu \times \nu(ds, dt) = \int \mu(ds) \int f(s, t) \nu(dt)$$

$$= \mathbb{E}\Big[\int f(s, t) \nu(dt)|_{s=\zeta} \Big] = \mathbb{E}[\mathbb{E}[f(s, \eta)]|_{s=\zeta}].$$

The case without assuming $f \ge 0$ follows from linearity. □

Corollary . *For independent random variables $\zeta_1, \dots, \zeta_n$,*

1. *(i) if $\zeta_1, \dots, \zeta_n \in L^1$, then $\mathbb{E} \prod_{k=1}^n \zeta_k = \prod_{k=1}^n \mathbb{E}\zeta_k$;*
2. *(ii) if $\zeta_1, \dots, \zeta_n \in L^2$, then $\operatorname{var}(\sum_{k=1}^n \zeta_k) = \sum_{k=1}^n \operatorname{var}(\zeta_k)$.*

Proof. By induction and Corollary 2.7, it suffices to prove the case $n = 2$. Suppose $\zeta \perp\!\!\!\perp \eta$. To prove $\mathbb{E}\zeta\eta = \mathbb{E}\zeta\mathbb{E}\eta$, we apply Lemma 2.11 with $f(x, y) = xy$. For the variance, we note that

$$\operatorname{var}(\zeta + \eta) - (\operatorname{var}(\zeta) + \operatorname{var}(\eta)) = 2\operatorname{cov}(\zeta, \eta) = 2\mathbb{E}(\zeta - \mathbb{E}\zeta)(\eta - \mathbb{E}\eta) = 2\mathbb{E}(\zeta - \mathbb{E}\zeta)\mathbb{E}(\eta - \mathbb{E}\eta) = 0,$$

where the second equality holds because $\zeta - \mathbb{E}\zeta \perp\!\!\!\perp \eta - \mathbb{E}\eta$. So $\operatorname{var}(\zeta + \eta) = \operatorname{var}(\zeta) + \operatorname{var}(\eta)$. □

Corollary 2.12. *Let ζ, η be independent random elements in a measurable group. Then $\operatorname{Law}(\zeta + \eta) = \operatorname{Law}(\zeta) * \operatorname{Law}(\eta)$.*

Proof. By Lemma 2.10, $\mathrm{Law}(\zeta,\eta)=\mathrm{Law}(\zeta)\times\mathrm{Law}(\eta)$. So $\mathrm{Law}(\zeta+\eta)$ equals the pushforward of $\mathrm{Law}(\zeta)\times\mathrm{Law}(\eta)$ under the map $(x,y)\mapsto xy$, which is the convolution of $\mathrm{Law}(\zeta)$ and $\mathrm{Law}(\eta)$. $\square$

By an exercise, if $\zeta_1,\dots,\zeta_n$ are independent random variables with Bernoulli distribution $\mathrm{B}(p)$, then $\zeta_1+\cdots+\zeta_n$ has the binomial distribution $\mathrm{B}(n,p)$. Suppose ζ_1 and ζ_2 are independent random variables. If they have Poisson distributions $\mathrm{Pois}(\lambda_1)$ and $\mathrm{Pois}(\lambda_2)$, respectively, then $\zeta_1+\zeta_2$ has Poisson distributions $\mathrm{Pois}(\lambda_1+\lambda_2)$. If they have Normal distributions $\mathrm{N}(\mu_1,v_1)$ and $\mathrm{N}(\mu_2,v_2)$, respectively, then $\zeta_1+\zeta_2$ has Normal distributions $\mathrm{N}(\mu_1+\mu_2,v_1+v_2)$.

We now study some zero-one laws. Given a sequence of σ-algebras $\mathcal{F}_1,\mathcal{F}_2,\dots$, the associated *tail* σ-algebra is defined by

$$\mathcal{T}=\bigcap_n\bigvee_{k\ge n}\mathcal{F}_k=\bigcap_n\sigma(\bigcup_{k\ge n}\mathcal{F}_k).$$

Example . Suppose $\zeta_1,\zeta_2\dots$ is a sequence of random variables, and $\mathcal{F}_n=\sigma(\zeta_n)$ for each n. Let $\mathcal{T}$ be the tail σ-algebra. Then

(i) The set A_1 of $\omega\in\Omega$ such that $\lim_n\zeta_n(\omega)$ converges is measurable w.r.t. $\mathcal{T}$.

(ii) The set A_2 of $\omega\in\Omega$ such that $\sum_n\zeta_n(\omega)$ converges is measurable w.r.t. $\mathcal{T}$.

(iii) The set of $\omega\in\Omega$ such that $\frac{1}{n}\sum_{k=1}^n\zeta_k(\omega)$ converges is measurable w.r.t. $\mathcal{T}$.

(iv) If we define $\eta_1=\lim_n\zeta_n$ on A_1, then η_1 is $A_1\cap\mathcal{T}$-measurable.

(v) If we define $\eta_2=\sum_n\zeta_n$ on A_2, then η_2 may not be $A_2\cap\mathcal{T}$-measurable.

(vi) If we define $\eta_3=\lim_n\frac{1}{n}\sum_{k=1}^n\zeta_k$ on A_3, then η_3 is $A_3\cap\mathcal{T}$-measurable.

Theorem 2.13 (Kolmogorov's zero-one law). *Let $\mathcal{F}_1,\mathcal{F}_2,\dots$ be independent σ-algebras in $\mathcal{A}$. Then the associated tail σ-algebra is trivial.*

Proof. For $n\in\mathbb{N}$, define $\mathcal{T}_n=\bigvee_{k>n}\mathcal{F}_k$. Then $\mathcal{T}=\bigcap_n\mathcal{T}_n$. By Corollary 2.7, for any n, $\mathcal{F}_1,\dots,\mathcal{F}_n,\mathcal{T}_n$ are independent. Since $\mathcal{T}\subset\mathcal{T}_n$, $\mathcal{T},\mathcal{F}_1,\dots,\mathcal{F}_n$ are independent for all n. Then we conclude that, $\mathcal{T},\mathcal{F}_1,\mathcal{F}_2,\dots$ are independent. By Corollary 2.7 again, we get $\mathcal{T}\perp\!\!\!\perp\bigvee_{n=1}^\infty\mathcal{F}_n$. Since $\mathcal{T}\subset\bigvee_{n=1}^\infty\mathcal{F}_n$, we get $\mathcal{T}\perp\!\!\!\perp\mathcal{T}$. By Lemma 2.9 (i), $\mathcal{T}$ is trivial. $\square$

Corollary 2.14. *Let $\zeta_1,\zeta_2,\dots$ be independent random variables. Let $S_n=\sum_{k=1}^n\zeta_k$, $n\in\mathbb{N}$. Then each of the sequences (ζ_n), (S_n) and $(\frac{1}{n}S_n)$ is either a.s. convergent or a.s. divergent. If (ζ_n) or $(\frac{1}{n}S_n)$ a.s. converges, then the limit is a.s. constant.*

There is another zero-one law, which works best for the sum of independent and identically distributed (i.i.d.) sequences of random vectors.

A bijective map $p:\mathbb{N}\to\mathbb{N}$ is called a *finite permutation* of $\mathbb{N}$ if there is N such that $p_n=n$ for $n>N$. A finite permutation p of $\mathbb{N}$ induces a bijective map $T_p:S^\infty\to S^\infty$ given by $T_p(s_1,s_2,\dots)=(s_{p_1},s_{p_2},\dots)$. A set $I\subset S^\infty$ is called symmetric if $T_p^{-1}I=I$ for all finite permutation p of $\mathbb{N}$. Let $(S,\overline{S})$ be a measurable space. Then for every p, $\mathcal{I}_p:=\{I\in\overline{S}^\infty: T_p^{-1}I=I\}$ is a σ-algebra. So the set of symmetric $I\in\overline{S}^\infty$ form a σ-algebra $\mathcal{I}=\bigcap_p\mathcal{I}_p$, which is called the *permutation invariant* σ-algebra in $\overline{S}^\infty$.

Example . Suppose G is an Abelian measurable group (e.g. $\mathbb{R}^d$). Let $B \subset G$ be measurable. Then the set

$$E_B = \{(v_1, v_2, \dots) \in G : \sum_{k=1}^{n} v_k \in B \text{ for infinitely many } n\}$$

belongs to the permutation invariant σ-algebra.

Theorem 2.15 (Hewitt-Savage zero-one law)**.** *Let $\zeta_1, \zeta_2, \dots$ be an i.i.d. sequence of random elements in a measurable space $(S, \overline{S})$, and let $\zeta = (\zeta_1, \dots, \zeta_n)$. Let $\mathcal{I}$ be the permutation invariant σ-algebra in $\overline{S}^\infty$. Then $\zeta^{-1}\mathcal{I}$ is trivial.*

Lemma 2.16. *Given any σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ in S, a probability measure μ on $\vee_n \mathcal{F}_n$, and a set $A \in \vee_n \mathcal{F}_n$, there exist a sequence $A_1, A_2, \dots \in \bigcup \mathcal{F}_n$ with $\mu(A_n \Delta A) \to 0$.*

Proof. Let $\mathcal{D}$ denote the set of $A \in \vee_n \mathcal{F}_n$ with the stated property. Then $\mathcal{D}$ is a λ-system containing the π-system $\mathcal{C} := \bigcup \mathcal{F}_n$. Here we use the fact that $\mu(A \Delta B) = \|\mathbf{1}_A - \mathbf{1}_B\|_1$. By monotone class theorem, $\mathcal{D}$ contains $\sigma(\mathcal{C}) = \vee_n \mathcal{F}_n$. □

Proof of Theorem 2.15. Let $\mu = \mathbb{P} \circ \zeta^{-1}$. Set $\mathcal{F}_n = \overline{S}^n \times S^\infty$, $n \in \mathbb{N}$. Note that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$, and $\vee_n \mathcal{F}_n = \overline{S}^\infty \supset \mathcal{I}$. For any $I \in \mathcal{I}$, by Lemma 2.16 there is a sequence I_n of the form $B_n \times S^\infty$ with $B_n \in \overline{S}^n$ such that $\mu(I_n \Delta I) \to 0$, and so $\mu I_n \to \mu I$. Writing $\widetilde{I}_n = S^n \times B_n \times S^\infty$, then by the symmetry of μ and I, we have $\mu \widetilde{I}_n = \mu I_n$ and $\mu(\widetilde{I}_n \Delta I) = \mu(I_n \Delta I) \to 0$. Hence

$$\mu((I_n \cap \widetilde{I}_n) \Delta I) \le \mu(I_n \Delta I) + \mu(\widetilde{I}_n \Delta I) \to 0$$

because $(A \cap B) \Delta C \subset (A \Delta C) \cup (B \Delta C)$. So $\mu(I_n \cap \widetilde{I}_n) \to \mu I$. By independence of ζ_k, we have

$$\mu(I_n \cap \widetilde{I}_n) = \mathbb{P}[(\zeta_1, \dots, \zeta_n) \in B_n, (\zeta_{n+1}, \dots, \zeta_{2n}) \in B_n] = \mathbb{P}[(\zeta_1, \dots, \zeta_n) \in B_n]^2 = \mu(I_n)^2.$$

So $\mu(I_n \cap \widetilde{I}_n) \to \mu(I)^2$. Then we get $\mu I = (\mu I)^2$ and so $\mu I \in \{0, 1\}$. □

Corollary 2.17. *Let $\zeta_1, \zeta_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^d$, and put $S_n = \zeta_1 + \cdots + \zeta_n$. Then for any $B \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}\{S_n \in B \text{ i.o.}\} = 0$ or 1.*

Note that Kolmogorov's zero-one law does not apply here because $\{S_n \in B \text{ i.o.}\}$ is not a tail event.

The sequence (S_n) is called a random walk on $\mathbb{R}^d$. For a more specific example, we may consider the case that every ζ_k has the distribution

$$\frac{1}{2d} \sum_{\sigma \in \{+,-\}} \sum_{j=1}^{d} \delta_{\sigma e_j},$$

where e_j is the vector in $\mathbb{R}^d$ whose j-th component is 1 and all other components are 0. In this case (S_n) is called a simple random walk on $\mathbb{Z}^d$. By Corollary 2.17, for every $v_0 \in \mathbb{Z}^d$, $\mathbb{P}\{S_n = v_0 \text{ i.o.}\} = 0$ or 1. By translation invariance of $\mathbb{Z}^d$, one easily see that the value of $\mathbb{P}\{S_n = v_0 \text{ i.o.}\}$ depends only on d. If the value is 1, the random walk is called *recurrent*; if the value is 0, the random walk is called *transient*. It turns out (not easy!) that, when $d \le 2$, the random walk is recurrent, and when $d \ge 3$, the random walk is transient.

Theorem 2.18 (Borel-Cantelli lemma). *Let $A_1, A_2, \dots \in \mathcal{A}$. Then $\sum_n \mathbb{P}A_n < \infty$ implies that $\mathbb{P}[A_n \text{ i.o.}] = 0$, and when the A_n are independent, $\mathbb{P}[A_n \text{ i.o.}] = 0$ implies that $\sum_n \mathbb{P}A_n < \infty$.*

Proof. We have proved the first assertion. Now suppose A_n are independent. Then A_n^c are also independent. For any $n < N \in \mathbb{N}$,

$$1 - \mathbb{P}\bigcup_{m=n}^{N} A_m = \mathbb{P}\bigcap_{m=n}^{N} A_m^c = \prod_{m=n}^{N}(1 - \mathbb{P}A_m).$$

Letting $N \to \infty$, we get

$$1 - \mathbb{P}\bigcup_{m=n}^{\infty} A_m = \prod_{m=n}^{\infty}(1 - \mathbb{P}A_m).$$

If $\mathbb{P}[A_n \text{ i.o.}] = 0$, then there is n such that $1 - \mathbb{P}\bigcup_{m=n}^{\infty} A_m > 0$, which implies by calculus that $\sum_{m=n}^{\infty} \mathbb{P}A_m < \infty$, and so $\sum_n \mathbb{P}A_n < \infty$. $\square$

For $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ in $\mathbb{R}^d$, we write $x \le y$ (resp. $x < y$) if $x_k \le y_k$ (resp. $x_k < y_k$) for all $1 \le k \le d$. For $x < y \in \mathbb{R}^d$, we define

$$(-\infty, y] = \{z \in \mathbb{R}^d : z \le y\} = \prod_{k=1}^{d}(-\infty, y_k], \quad (x, y] = \{z \in \mathbb{R}^d : x < z \le y\} = \prod_{k=1}^{d}(x_k, y_k].$$

For a random vector ζ in $\mathbb{R}^d$, we define the associated distribution function F by

$$F(x) = \mathbb{P}[\zeta_j \le x_j, 1 \le j \le d] = \mathrm{Law}(\zeta)(-\infty, x].$$

By a monotone argument, we get

Lemma 2.3. *Two random vectors in $\mathbb{R}^d$ have the same distribution iff they have the same distribution function.*

We may use F to calculate $\mu(x, y]$. For $d = 1$, $\mu(x, y] = F(y) - F(x)$. For $d \ge 2$, we need an inclusion-exclusion argument.

Exercise . Prove that for any $x < y \in \mathbb{R}^d$,

$$\mu(x, y] = \sum_{S \subset \{1, \dots, d\}} (-1)^{|S|} F(z^S), \tag{2.3}$$

where $z^S \in \mathbb{R}^d$ such that $z_k^S = x_k$ if $k \in S$ and $z_k^S = y_k$ if $k \notin S$.

Then F satisfies the following properties.

(i) $F(x, y] \ge 0$ for every $x < y \in \mathbb{R}^d$, where we define $F(x, y]$ to be the RHS of (2.3).

(ii) F is right-continuous in the sense that $\lim_{x \downarrow y} F(x) = F(y)$ for any $y \in \mathbb{R}^d$, where $x \downarrow y$ means that $x_k > y_k$ and $x_k \to y_k$ for all $1 \le k \le d$.

(iii) $\lim_{\min x_k \to -\infty} F(x) = 0$.

(iv) $\lim_{\min x_k \to \infty} F(x) = 1$.

Here (ii)-(iv) follow from the continuity of μ and the fact that $\mu(\mathbb{R}^d) = 1$.

Theorem 2.25-2.26. *If F satisfies (i-iii), then it is the distribution function of some σ-finite measure μ on $\mathbb{R}^d$. If F also satisfies (iv), then μ is a probability measure.*

Proof. We define a ring $\mathcal{R}$ on $\mathbb{R}^d$ to be the class of disjoint unions of sets of the form $(x, y]$ for $x < y \in \mathbb{R}^d$. Define $\mu : \mathcal{R} \to \mathbb{R}_+$ such that if A has a disjoint union expression $\bigcup_{j=1}^m (x^j, y^j]$, then

$$\mu A = \sum_{j=1}^m F(x^j, y^j].$$

Such μ is well defined and satisfies finitely additivity. We then show that μ is a pre-measure. Suppose $A_1 \supset A_2 \supset \cdots \in \mathcal{R}$ with $\mu A_n \geq \varepsilon > 0$ for all n. We need to show that $\bigcap_n A_n \neq \emptyset$. For every $n \in \mathbb{N}$, we may choose $A_n' \in \mathcal{R}$ such that $\overline{A_n'} \subset A_n$ and $\mu(A_n \setminus A_n') < \frac{\varepsilon}{2^n}$. Here we use the fact that if $x^n \downarrow x < y$, then $F(x^n, y] \to F(x, y]$, which follows from the right-continuity of F.

Let $A_n'' = A_1' \cap \cdots \cap A_n'$. 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 $\mu(A_n \setminus A_n'') \leq \sum_{k=1}^n \mu(A_k \setminus A_k') < \sum_{k=1}^n \frac{\varepsilon}{2^k} < \varepsilon$. 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 $\mathbb{R}^d$. It is σ-finite because $\mu(x, x + \underline{1}] < \infty$ for every $x \in \mathbb{Z}^d$, where $\underline{1} = (1, \ldots, 1)$.

By (iii) we have, for every $y \in \mathbb{R}^d$,

$$F(y) = \lim_{\min x_k \to -\infty} F(x, y] = \lim_{\min x_k \to -\infty} \mu(x, y] = \mu(-\infty, y].$$

So F is the distribution function of μ. If (iv) holds, then

$$\mu\mathbb{R}^d = \lim_{n \to \infty} \mu(-\infty, (n, \ldots, n)] = \lim_{n \to \infty} F(n, \ldots, n) = 1,$$

which implies that μ is a probability measure. □

Exercise . Problems 4, 5, 8, 12 of Exercises of Chapter 2.

3 Random Sequences, Series, and Averages

We still fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and assume that all random elements are defined on this space. We will study several different concepts of convergence of random variables: almost sure convergence, $\zeta_n \to \zeta$ a.s., convergence in probability, $\zeta_n \xrightarrow{\text{P}} \zeta$, convergence in distribution, $\zeta_n \xrightarrow{\text{d}} \zeta$, and convergence in L^p.

Definition . Let $\zeta, \zeta_1, \zeta_2, \dots$ be random elements in a metric space (S, ρ).

(i) We say that ζ_n converges almost surely to ζ, and write $\zeta_n \to \zeta$ a.s., if there is a null event N such that $\rho(\zeta_n(\omega), \zeta(\omega)) \to 0$ for every $\omega \in \Omega \setminus N$.

(ii) We say that ζ_n converges in probability to ζ, and write $\zeta_n \xrightarrow{\mathrm{P}} \zeta$, if for every $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\{\rho(\zeta_n, \zeta) > \varepsilon\} = 0$.

(iii) We say that ζ_n converges in distribution to ζ, and write $\zeta_n \xrightarrow{\mathrm{d}} \zeta$, if for every $f \in C_b(S, \mathbb{R})$, the space of bounded real-valued continuous functions on S, we have $\mathbb{E}f(\zeta_n) \to \mathbb{E}f(\zeta)$.

(iv) In the case that $S = \mathbb{R}$, we say that ζ_n converges to ζ in L^p for some $p > 0$, if $\zeta, \zeta_1, \zeta_2, \dots \in L^p$ and $\|\zeta_n - \zeta\|_p = (\mathbb{E}|\zeta_n - \zeta|^p)^{1/p} \to 0$.

Lemma 3.1 (Chebyshev inequality)**.** *For any measurable $\zeta : \Omega \to \overline{\mathbb{R}}_+$ and $r > 0$,*

$$\mathbb{P}\{\zeta \geq r\} \leq \frac{1}{r}\mathbb{E}\zeta.$$

Proof. Since $\zeta \geq r\mathbf{1}_{\{\zeta \geq r\}}$, we get $\mathbb{E}\zeta \geq \mathbb{E}(r\mathbf{1}_{\{\zeta\geq r\}}) = r\mathbb{P}\{\zeta \geq r\}$. □

Exercise . Prove that $\zeta_n \to \zeta$ in L^p for some $p > 0$ implies that $\zeta_n \xrightarrow{\mathrm{P}} \zeta$.

Lemma . *For $\zeta, \zeta_1, \zeta_2, \dots$ in the above definition, $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ iff $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \to 0$.*

Proof. For every $\varepsilon \in (0, 1)$, from $\varepsilon\mathbf{1}\{\rho(\zeta_n, \zeta) > \varepsilon\} \leq 1 \wedge \rho(\zeta_n, \zeta) \leq \mathbf{1}\{\rho(\zeta_n, \zeta) > \varepsilon\} + \varepsilon$, we get

$$\varepsilon\mathbb{P}\{\rho(\zeta_n, \zeta) > \varepsilon\} \leq \mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \leq \mathbb{P}\{\rho(\zeta_n, \zeta) > \varepsilon\} + \varepsilon.$$

These inequalities imply the equivalence. □

Remark . The lemma means that the convergence in probability is determined by a metric

$$\rho_V(\zeta, \eta) = \mathbb{E}[1 \wedge \rho(\zeta, \eta)].$$

This is in general not true for almost surely convergence

Lemma 3.2 (subsequence criterion)**.** *Let $\zeta, \zeta_1, \zeta_2, \dots$ be as before. Then $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ iff every subsequence $N' \subset \mathbb{N}$ has a further subsequence $N'' \subset N'$ such that $\zeta_n \to \zeta$ a.s. along N''. In particular, the almost sure convergence implies the convergence in probability.*

Proof. Suppose $\zeta_n \xrightarrow{\mathrm{P}} \zeta$. Then $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \to 0$ by the above lemma. Suppose $N' \subset \mathbb{N}$. Then $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \to 0$ along N'. We may then choose a subsequence $N'' \subset N'$ such that $\sum_{n\in N''} \mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] < \infty$. By monotone convergence theorem, we get

$$\mathbb{E}[\sum_{n\in N''} 1 \wedge \rho(\zeta_n, \zeta)] < \infty,$$

which implies that a.s. $\sum_{n\in N''} 1 \wedge \rho(\zeta_n, \zeta) < \infty$. So a.s. $\zeta_n \to \zeta$ along N''. On the other hand, suppose $\zeta_n \overset{\mathrm{P}}{\not\to} \zeta$. Then $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \not\to 0$. So there is $\varepsilon > 0$ and a subsequence $N' \subset \mathbb{N}$ such that $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] > \varepsilon$ for any $n \in N'$. It there is a further subsequence $N'' \subset N'$ such that $\zeta_n \to \zeta$ a.s. along N'', then since $1 \wedge \rho(\zeta_n, \zeta) \to 0$ a.s. along N'', by dominated convergence theorem, $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \to 0$ along N'', which is a contradiction.

Finally, if $\zeta_n \to \zeta$ a.s. then for any $N' \subset \mathbb{N}$, $\zeta_n \to \zeta$ a.s. along N'. So we get $\zeta_n \xrightarrow{\mathrm{P}} \zeta$. □

Remark . From Lemma 3.2, we see that the condition that $\zeta_n \to \zeta$ a.s. in dominated convergence theorem can be further weakened to $\zeta_n \xrightarrow{\mathrm{P}} \zeta$. This means that if $\zeta_n \to P\zeta$, $|\zeta_n| \le \eta$ for all n, and $\mathbb{E}\eta < \infty$, then $\mathbb{E}\zeta_n \to \mathbb{E}\zeta$.

Example . We may find a sequence of random variables ζ_n on $([0,1], \lambda)$ such that $\zeta_n \xrightarrow{\mathrm{P}} 0$ but ζ_n does not a.s. converge to 0. In fact, we may choose $\zeta_n = \mathbf{1}_{A_n}$, where

$$A_1 = [0,1], \quad A_2 = [0,1/2], \quad A_3 = [1/2,1],$$

$$A_4 = [0,1/4], \quad A_5 = [1/4,2/4], \quad A_6 = [2/4,3/4], \quad A_7 = [3/4,1], \dots$$

The general formula is: for $2^k \le n \le 2^{k+1}-1$, $\zeta_k = \mathbf{1}_{[\frac{n}{2^k}-1, \frac{n+1}{2^k}-1]}$. We observe that $\|\zeta_n\|_1 = 2^{-k}$ if $2^k \le n \le 2^{k+1}-1$. So $\zeta_n \to 0$ in L^1, which implies that $\zeta_n \xrightarrow{\mathrm{P}} 0$. However, for every $t \in [0,1]$, there are infinitely many n such that $\zeta_n(t) \to 1$. So ζ_n does not a.s. tend to 0.

Lemma 3.3. *Let S and T be two metric spaces. Suppose $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ in S, and $f : S \to T$ be continuous. If $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ in S, then $f(\zeta_n) \xrightarrow{\mathrm{P}} f(\zeta)$ in T.*

Proof. By Lemma 3.2, every subsequence $N' \subset \mathbb{N}$ contains a further subsequence $N'' \subset N'$ such that $\zeta_n \to \zeta$ a.s. in S along N''. By the continuity of f, we see that $f(\zeta_n) \to f(\zeta)$ a.s. in T along N''. Thus, by Lemma 3.2 $f(\zeta_n) \xrightarrow{\mathrm{P}} f(\zeta)$ in T. □

Corollary 3.5. *Let $\zeta, \zeta_1, \zeta_2, \dots$ and $\eta, \eta_1, \eta_2, \dots$ be random variables with $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ and $\eta_n \xrightarrow{\mathrm{P}} \eta$. Then $a\zeta_n + b\eta_n \to a\zeta + b\eta$ for any $a, b \in \mathbb{R}$ and $\zeta_n\eta_n \to \zeta\eta$. Furthermore, $\zeta_n/\eta_n \xrightarrow{\mathrm{P}} \zeta/\eta$ whenever η_n and η do not take value zero.*

Proof. From $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ and $\eta_n \xrightarrow{\mathrm{P}} \eta$ we get $(\zeta_n, \eta_n) \xrightarrow{\mathrm{P}} (\zeta, \eta)$. We may then apply Lemma 3.3 to continuous functions $\mathbb{R}^2 \ni (x,y) \mapsto ax+by \in \mathbb{R}$, $\mathbb{R}^2 \ni (x,y) \mapsto xy$, and $\mathbb{R} \times (\mathbb{R} \setminus \{0\}) \ni (x,y) \mapsto x/y$, respectively. □

Definition . For random elements $\zeta_1, \zeta_2, \dots$ in a metric space (S, ρ), we say that (ζ_n) is a *Cauchy sequence in probability* if for any $\varepsilon > 0$, $\mathbb{P}\{\rho(\zeta_n, \zeta_m) > \varepsilon\} \to 0$ as $n, m \to \infty$. Using a similar argument as before, we can show that this is equivalent to that $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta_m)] \to 0$ as $n, m \to \infty$.

If $\zeta_n \xrightarrow{\text{P}} \zeta$, then $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta)] \to 0$ as $n \to \infty$. By triangle inequality, we get $\mathbb{E}[1 \wedge \rho(\zeta_n, \zeta_m)] \to 0$ as $n, m \to \infty$, which implies that (ζ_n) is a Cauchy sequence in probability. The converse is true if (S, ρ) is complete. This is the lemma below.

Lemma 3.6. *If (S, ρ) is complete, then (ζ_n) is a Cauchy sequence in probability iff $\zeta_n \xrightarrow{\text{P}} \zeta$ for some random element ζ in S.*

Proof. We have proved the "if" part. Now we prove the "only if" part. Assume that (ζ_n) is a Cauchy sequence in probability. We may choose a subsequence (n_k) of $\mathbb{N}$ such that $\mathbb{E}[1 \wedge \rho(\zeta_{n_k}, \zeta_{n_{k+1}})] \le 2^{-k}$ for all $k \in \mathbb{N}$. Then we have

$$\mathbb{E}[\sum_k 1 \wedge \rho(\zeta_{n_k}, \zeta_{n_{k+1}})] \le \sum_k 2^{-k} < \infty,$$

which implies that a.s. $\sum_k 1 \wedge \rho(\zeta_{n_k}, \zeta_{n_{k+1}}) < \infty$, and so $\sum_k \rho(\zeta_{n_k}, \zeta_{n_{k+1}}) < \infty$. So almost surely (ζ_{n_k}) is a Cauchy sequence in S. By the completeness of S, there is a random element ζ in S such that a.s. $\zeta_{n_k} \to \zeta$. Thus, $\mathbb{E}[1 \wedge \rho(\zeta_{n_k}, \zeta)] \to 0$ as $k \to \infty$. To see that $\zeta_n \xrightarrow{\text{P}} \zeta$, write

$$\mathbb{E}[1 \wedge \rho(\zeta_m, \zeta)] \le \mathbb{E}[1 \wedge \rho(\zeta_{n_k}, \zeta)] + \mathbb{E}[1 \wedge \rho(\zeta_m, \zeta_{n_k})],$$

and use the convergence of the RHS to 0 as $m, k \to \infty$. □

This lemma shows that the space of random elements on S with metric $\rho_V(\zeta, \eta) = \mathbb{E}[1 \wedge \rho(\zeta, \eta)]$ is complete when S is complete.

Lemma 3.7. *The convergence in probability implies the convergence in distribution.*

Proof. Suppose $\zeta_n \xrightarrow{\text{P}} \zeta$ in S, and $f \in C_b(S)$. Then $f(\zeta_n) \xrightarrow{\text{P}} f(\zeta)$ by Lemma 3.3. By monotone convergence theorem (for convergence in probability), we have $\mathbb{E}f(\zeta_n) \to \mathbb{E}f(\zeta)$. So $\zeta_n \xrightarrow{\text{d}} \zeta$. □

Definition . Let $\mu, \mu_1, \mu_2, \dots$ be probability measures on a metric space (S, ρ). We say that μ_n converges weakly to μ, and write $\mu_n \xrightarrow{\text{w}} \mu$, if for any $f \in C_b(S, \mathbb{R})$, $\mu_n f \to \mu f$.

Remark . By Lemma 1.22, $\mathbb{E}f(\zeta) = \text{Law}(\zeta)f$. So $\zeta_n \xrightarrow{\text{d}} \zeta$ iff $\text{Law}(\zeta_n) \xrightarrow{\text{w}} \text{Law}(\zeta)$. This means that the convergence in distribution depends only on the distributions of ζ and ζ_n (and not on the exact value of $\zeta_n(\omega)$ and $\zeta(\omega)$).

Lemma 3.25 (Portmanteau)**.** *For any probability measures $\mu, \mu_1, \dots, \mu_n$ on a metric space (S, ρ), these conditions are equivalent:*

(i) $\mu_n \xrightarrow{\text{w}} \mu$;

(ii) $\liminf_n \mu_n G \ge \mu G$ *for any open set* $G \subset S$;

(iii) $\limsup_n \mu_n F \le \mu F$ *for any closed set* $F \subset S$;

(iv) $\lim_n \mu_n B = \mu B$ *for any* $B \in \mathcal{B}(S)$ *with* $\mu\partial B = 0$.

A set B *satisfying the condition in (iv) is called a* μ*-continuity set.*

Example . Suppose (x_n) is a sequence in S and $x_n \to x_0 \in S$. Then we have $\delta_{x_n} \xrightarrow{\text{w}} \delta_{x_0}$ because for any $f \in C_b$,

$$\delta_{x_n} = f(x_n) \to f(x_0) = \delta_{x_0} f.$$

Suppose $G \subset S$ is open, and $x_0 \in \partial G$, then we can find a sequence (x_n) in G such that $x_n \to x_0$. Then $\delta_{x_0} G = 0$ but $\delta_{x_n} G = 1$ for each n. So we do not get a strict inequality in (ii).

Proof. Assume (i), and fix an open set $G \subset S$. Let $f_m(x) = 1 \wedge (m\rho(x, G^c))$, $m \in \mathbb{N}$. Then $f_m \in C_b(S)$ and $f_m \uparrow \mathbf{1}_G$. For each m, by $\mu_n \xrightarrow{\text{w}} \mu$, we have $\mu f_m = \lim_n \mu_n f_m \le \liminf_n \mu_n G$. Sending $m \to \infty$ and using monotone convergence, we then get (ii). The equivalence between (ii) and (iii) are clear from taking complements. Now assume (ii) and (iii). For any $B \in \mathcal{B}$,

$$\mu B^\circ \le \liminf_n \mu_n B^\circ \le \liminf_n \mu_n B \le \limsup_n B \le \limsup_n \overline{B} \le \mu\overline{B}.$$

If $\mu\partial B = 0$, then $\mu\overline{B} = \mu B^\circ = \mu B$, and (iv) follows.

Assume (iv), and fix a closed set $F \subset S$. Write $F^\varepsilon = \{s \in S : \rho(s, F) < \varepsilon\}$. Then the sets $\partial F^\varepsilon \subset \{s \in S : \rho(s, F) = \varepsilon\}$, $\varepsilon > 0$, are disjoint. So there are at most countably many $\varepsilon > 0$ such that $\mu\partial F^\varepsilon = 0$. We can find a positive sequence $\varepsilon_m \to 0$ such that for every m, $\mu\partial F^{\varepsilon_m} = 0$. So $\mu F^{\varepsilon_m} = \lim_n \mu_n F^{\varepsilon_m} \ge \limsup_n \mu_n F$. Sending $m \to \infty$, we get (iii). Finally, assume (ii) and let $f : S \to \mathbb{R}_+$ be continuous. By Lemma 2.4 and Fatou's lemma,

$$\mu f = \int_0^\infty \mu\{f > t\}dt \le \int_0^\infty \liminf_n \mu_n\{f > t\}dt \le \liminf_n \int_0^\infty \mu_n\{f > t\}dt = \liminf_n \mu_n f.$$

Suppose now $f \in C_b(S)$ and $|f| \le c$. Applying the above formula to $c \pm f$, we get $c \pm \mu f \le \liminf_n (c \pm \mu_n f)$, which implies $\lim_n \mu_n f = \mu f$, i.e., (i) holds. $\square$

Exercise . Let $\mu, \mu_1, \mu_2, \dots$ be probability measures on $\mathbb{R}^d$. Let $F, F_1, F_2, \dots$ be their distribution functions. Prove that $\mu_n \xrightarrow{\text{w}} \mu$ iff for any continuity point x of F, $F_n(x) \to F(x)$.

Definition . A family of probability measures μ_t, $t \in T$, on a topological space S is called tight, if for any $\varepsilon > 0$, there is a compact set $K \subset S$ such that $\mu_t(S \setminus K) < \varepsilon$ for any $t \in T$.

Suppose (S, ρ) is a metric space. For $x \in S$ and $\varepsilon > 0$, let $B(x, \varepsilon) = \{y \in S : \rho(x, y) < \varepsilon\}$. For $A \subset S$ and $\varepsilon > 0$, let

$$A^\varepsilon = \bigcup_{x \in A} B(x, \varepsilon) = \{y \in S : \rho(y, A) < \varepsilon\}.$$

We now state some results about weak convergence without proofs.

Theorem 14.3 (Prokhorov's theorem)**.** *Let* (S, ρ) *be a separable metric space. Then* *(i) The Prokhorov metric ρ_* on the space $\mathcal{P}(S)$ defined by*

$$\rho_*(\mu,\nu) = \inf\{\varepsilon > 0 : \mu A \le \nu A^\varepsilon + \varepsilon \text{ and } \nu A \le \mu A^\varepsilon + \varepsilon \text{ for any } A \in \mathcal{B}(S)\}$$

is a metric such that the weak convergence of probability measures on S is equivalent to the convergence w.r.t. the Prokhorov metric.

(ii) A tight family is relatively sequential compact w.r.t the weak convergence, i.e., every sequence in the family contains a weak convergent subsequence.

(iii) If S is complete, then $(\mathcal{P}(S), \rho_*)$ is complete and every relatively compact subset of $\mathcal{P}(S)$ is a tight family.

This lemma tells us that the weak convergence is induced by some explicitly defined metric, and if S is complete, then the a tight family is equivalent to a relatively compact set w.r.t. weak convergence.

In the case that $S = \mathbb{R}^d$, we sketch a proof of (ii) as follows. Suppose $\mu_1, \mu_2, \dots$ is a sequence of probability measures on $\mathbb{R}^d$. Let $F_1, F_2, \dots$ be the distribution functions. Since $0 \le F_n \le 1$, for every $x \in \mathbb{Q}^d$, $(F_n(x))$ contains a convergent subsequence. By a diagonal argument and passing to a subsequence, we may assume that $(F_n(x))$ converges for each $x \in \mathbb{Q}^d$. Let $\widetilde{F}(x)$, $x \in \mathbb{Q}^d$, be the limit function. Such $\widetilde{F}$ is non-decreasing on $\mathbb{Q}^d$. We use $\widetilde{F}$ to define a function F on $\mathbb{R}^d$ such that $F(x) = \lim_{\mathbb{Q}^d \ni y \downarrow x} \widetilde{F}(y)$, $x \in \mathbb{R}^d$. Then F is non-decreasing and right-continuous, and $F_n(x) \to F(x)$ for each continuity point x of F. If $\{\mu_n\}$ is tight, then F is the distribution function of some probability measure μ, which is the weak limit of μ_n.

To understand the Prokhorov metric, suppose X and Y are two random elements in S defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$\mathbb{P}\{\rho(X,Y) > \varepsilon\} < \varepsilon. \tag{3.1}$$

Then it is straightforward to check that $\rho_*(\mathrm{Law}(X), \mathrm{Law}(Y)) < \varepsilon$. The converse is not true, but we have the following coupling theorem, whose proof is omitted.

Theorem (coupling theorem)**.** *If $\rho_*(\mu,\nu) < \varepsilon$, then there are a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and two random elements X, Y in S defined on Ω such that $\mathrm{Law}(X) = \mu$, $\mathrm{Law}(Y) = \nu$, and (3.1) holds.*

From Lemma 3.7, $\zeta_n \xrightarrow{\mathrm{P}} \zeta$ implies that $\mathrm{Law}(\zeta_n) \xrightarrow{\mathrm{w}} \mathrm{Law}(\zeta)$ and $\zeta \xrightarrow{\mathrm{d}} \zeta$. We have a converse statement in the following sense. We omit its proof.

Theorem 3.30 (Skorokhod's representation theorem)**.** *Let $\mu, \mu_1, \mu_2, \dots$ be probability measures on a separable metric space (S, ρ). Then there exist a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random elements $\zeta, \zeta_1, \zeta_2, \dots$ in S defined on Ω such that $\mathrm{Law}(\zeta) = \mu$, $\mathrm{Law}(\zeta_n) = \mu_n$, and $\zeta_n \to \zeta$ pointwise.*

Exercise . Suppose $\zeta_n \xrightarrow{\mathrm{d}} \zeta$ and $\mathrm{Law}(\zeta)$ is a point mass. Prove that $\zeta_n \xrightarrow{\mathrm{P}} \zeta$.

There are other types of convergence of measures, such as the strong convergence: $\mu_n A \to \mu A$ for every $A \in \mathcal{A}$, and an even stronger convergence: the total variation convergence:

$$\|\mu_n - \mu\|_{\mathrm{TV}} := 2 \sup_{A \in \mathcal{A}} |\mu A - \nu A| \to 0.$$

They are stronger than the weak convergence, but do not rely on the topology of S.

Example . Let S be a metric space. Let (x_n) be a sequence in S that converges to x_0. Suppose $x_n \neq x_0$ for all n. Then δ_{x_n} converges to δ_{x_0} weakly but not strongly. If we take $A = \{x_0\}$, then $\delta_{x_n} A = 0$ for all n but $\delta_{x_0} A = 1$.

Exercise . Let $\mu, \mu_1, \mu_2, \dots$ be probability measures on a measurable space S. Let ν be a finite measure on S such that $\mu \ll \nu$ and $\mu_n \ll \nu$ for all n. Such ν always exists, e.g., let $\nu = \mu + \sum_n \frac{\mu_n}{2^n}$. Let $f = d\mu/d\nu$ and $f_n = d\mu_n/d\nu$. Then $f, f_n \in L^1(\nu)$; $\mu_n \to \mu$ in total variation iff $f_n \to f$ in $L^1(\nu)$; and $\mu_n \to \mu$ strongly iff $f_n \to f$ weakly in $L^1(\nu)$, i.e., for any $g \in L^\infty$, $\int f_n g d\nu \to \int f g d\nu$.

We now introduce a new concept: *uniformly integrability*, which plays an important role in the theory of martingales. To motivate the definition, we observe that if $\zeta \in L^1$, then by dominated convergence theorem, $\mathbb{E}[\mathbf{1}_{|\zeta| \ge R} \zeta] \to 0$ as $R \to \infty$.

Definition . A family of random variables ζ_t, $t \in T$, is called uniformly integrable, if

$$\lim_{R \to \infty} \sup_{t \in T} \mathbb{E}[\mathbf{1}_{|\zeta_t| \ge R} \zeta] = 0.$$

The previous observation shows that any finite set of integrable random variables is uniformly integrable. The uniformly integrability depends only on the distributions of the random variables, and is stronger than the tightness of the distributions.

Exercise . For $t \in T$, let ζ_t be a random variable with distribution μ_t, and let $p_{t,n} = \mathbb{P}[|\zeta_t| \ge n]$. Prove that ζ_t, $t \in T$, is uniformly integrable iff $\sum_n p_{t,n}$ converges uniformly in $t \in T$, which then implies that the family μ_t, $t \in T$, is tight.

Exercise . Prove that a sequence $\zeta_1, \zeta_2, \dots \in L^1$ is uniformly integrable iff

$$\lim_{R \to \infty} \limsup_{n \to \infty} \int_{\{|\zeta_n| \ge R\}} |\zeta_n| d\mathbb{P} = 0.$$

Lemma . *If for some $p > 1$, $\{\zeta_t : t \in T\}$ is L^p-bounded, i.e., there is $C < \infty$ such that $\|\zeta_t\|_p \le C$ for all $t \in T$, then ζ_t, $t \in T$, is uniformly integrable.*

Proof. To see this, note that

$$\int_{\{|\zeta_t| \ge R\}} |\zeta_t| d\mathbb{P} \le \int_{\{|\zeta_t| \ge R\}} (|\zeta_t|/R)^{p-1} |\zeta_t| d\mathbb{P} \le R^{1-p} \mathbb{E}|\zeta_t|^p = R^{1-p} \|\zeta_t\|_p^p \le R^{1-p} C^p.$$

□

The lemma does not hold for $p=1$. For example, if $\zeta_n = n\mathbf{1}_{[0,1/n]}$, $n\in\mathbb{N}$, are defined on $([0,1],\lambda)$, then $\|\zeta_n\|_1 = 1$ for all n, but for any $R>0$, $\mathbb{E}[\mathbf{1}_{|\zeta_n|\ge R}\zeta_n] = 1$ if $n\ge R$.

Lemma 3.10. *The random variables ζ_t, $t\in T$, are uniformly integrable iff they are L^1-bounded, and*

$$\lim_{\mathbb{P}A\to 0}\sup_{t\in T}\mathbb{E}[\mathbf{1}_A|\zeta_t|]\to 0. \tag{3.2}$$

Proof. Suppose ζ_t, $t\in T$, are uniformly integrable. Then

$$\mathbb{E}[\mathbf{1}_A|\zeta_t|]\le R\mathbb{P}A + \mathbb{E}[\mathbf{1}_{|\zeta_t|\ge R}|\zeta_t|].$$

For any $\varepsilon>0$, we may choose $R>0$ such that $\mathbb{E}[\mathbf{1}_{|\zeta_t|\ge R}|\zeta_t|]<\varepsilon/2$ for all $t\in T$. Thus, if $\mathbb{P}A<\varepsilon/(2R)$, then $\mathbb{E}[\mathbf{1}_A|\zeta_t|]<\varepsilon$ for all $t\in T$. To get the L^1-boundedness, we take $A=\Omega$ and take R to be sufficiently big in the displayed formula.

Suppose now ζ_t, $t\in T$, are L^1-bounded, and (3.2) holds. By Chebyshev's inequality we get

$$\mathbb{P}\{|\zeta_t|\ge R\}\le \frac{1}{R}\sup_{t\in T}\|\zeta_t\|_1\to 0,\quad R\to\infty,$$

which together with (3.2) implies the uniformly integrability. $\square$

Exercise . Let ζ_s, $s\in S$, and η_t, $t\in T$, be two uniformly integrable families of random variables. Then $|\zeta_s|+|\eta_t|$, $(s,t)\in S\times T$, are also uniformly integrable.

Proposition 3.12. *Fix $p>0$. Suppose $\zeta_1,\zeta_2,\dots\in L^p$ are such that $|\zeta_n|^p$, $n\in\mathbb{N}$, are uniformly integrable. Suppose $\zeta_n\xrightarrow{\mathrm{P}}\zeta$. Then $\zeta_n\to\zeta$ in L^p.*

Proof. By Fatou's lemma and the L^1-boundedness of $|\zeta_n|^p$ (by Lemma 3.10), we have

$$\mathbb{E}|\zeta|^p\le\liminf_n \mathbb{E}|\zeta_n|^p<\infty.$$

So $\zeta\in L^p$. Since $|\zeta_n-\zeta|^p\le 2^p(|\zeta_n|^p+|\zeta|^p)$, by the exercise above, $|\zeta_n-\zeta|^p$, $n\in\mathbb{N}$, are also uniformly integrable. Fix $\varepsilon>0$. Then

$$\mathbb{E}[|\zeta_n-\zeta|^p]\le\varepsilon^p+\mathbb{E}[\mathbf{1}_{\{|\zeta_n-\zeta|\ge\varepsilon\}}|\zeta_n-\zeta|^p].$$

Since $\zeta_n\xrightarrow{\mathrm{P}}\zeta$, as $n\to\infty$, $\mathbb{P}\{|\zeta_n-\zeta|\ge\varepsilon\}\to 0$, which implies $\mathbb{E}[\mathbf{1}_{\{|\zeta_n-\zeta|\ge\varepsilon\}}|\zeta_n-\zeta|^p]\to 0$ by Lemma 3.10. Sending $n\to\infty$, we get $\limsup_n\mathbb{E}[|\zeta_n-\zeta|^p]\le\varepsilon^p$. Since this holds for any $\varepsilon>0$, we get $\mathbb{E}[|\zeta_n-\zeta|^p]\to 0$. So $\zeta_n\to\zeta$ in L^p. $\square$

Theorem 3.23 (strong law of large numbers)**.** *Let $\zeta,\zeta_1,\zeta_2,\dots$ be i.i.d. random variables with $\mathbb{E}|\zeta|<\infty$. Let $S_n=\sum_{k=1}^n\zeta_k$. Then a.s. $\frac{1}{n}S_n\to\mathbb{E}\zeta$.*

We are not going to prove the theorem following the approach of the textbook (Proposition 3.14, Lemma 3.15, Lemma 3.16, Theorem 3.17, Theorem 3.18, Lemma 3.19, Lemma 3.20, Lemma 3.21, Corollary 3.22). Instead, we give elementary proofs of some weaker results, and postpone the proof of Theorem 3.23 to the chapter of martingales.

Theorem (weak law of large numbers for L^2). *In the setup of Theorem 3.23, if $\zeta \in L^2$, then $\frac{1}{n}S_n \xrightarrow{\mathrm{P}} \mathbb{E}\zeta$.*

Proof. By subtracting $\mathbb{E}\zeta$ from ζ_n, $n \in \mathbb{N}$, we may assume that $\mathbb{E}\zeta = 0$. Since $\zeta_1, \zeta_2, \dots$ are independent,

$$\mathbb{E}\Big[\Big|\frac{1}{n}\sum_{j=1}^n \zeta_j\Big|^2\Big] = \frac{1}{n^2}\operatorname{var}\Big(\sum_{j=1}^n \zeta_j\Big) = \frac{1}{n^2}\sum_{j=1}^n \operatorname{var}(\zeta_j) = \frac{1}{n}\operatorname{var}(\zeta).$$

By Chebyshev inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\Big[\Big|\frac{1}{n}\sum_{j=1}^n \zeta_j\Big| \ge \varepsilon\Big] \le \frac{1}{\varepsilon^2}\mathbb{E}\Big[\Big|\frac{1}{n}\sum_{j=1}^n \zeta_j\Big|^2\Big] \le \frac{1}{n}\frac{\operatorname{var}(\zeta)}{\varepsilon^2} \to 0,$$

as $n \to \infty$. So $\frac{1}{n}\sum_{j=1}^n \zeta_j \xrightarrow{\mathrm{P}} 0$. □

Theorem (strong law of large numbers for L^4). *Theorem 3.23 holds if $\zeta \in L^4$.*

Proof. We again assume that $\mathbb{E}\zeta = 0$. We have

$$\mathbb{E}\Big[(\frac{1}{n}S_n)^4\Big] = \frac{1}{n^4}\sum_{1\le j_1,j_2,j_3,j_4\le n} \mathbb{E}[\zeta_{j_1}\zeta_{j_2}\zeta_{j_3}\zeta_{j_4}].$$

If for some $s \in \{1,2,3,4\}$, $j_s \notin \{j_t : t \ne s\}$, then by independence of $\zeta_1, \zeta_2, \dots$ and that $\mathbb{E}\zeta_{j_s} = 0$, we get

$$\mathbb{E}[\zeta_{j_1}\zeta_{j_2}\zeta_{j_3}\zeta_{j_4}] = \mathbb{E}\zeta_{j_s}\mathbb{E}[\prod_{t\ne s}\zeta_{j_t}] = 0.$$

Thus,

$$\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{j=1}^n \zeta_j\Big)^4\Big] = \frac{1}{n^4}\sum_{j=1}^n \mathbb{E}\zeta_j^4 + \frac{12}{n^4}\sum_{1\le j<k\le n}\mathbb{E}\zeta_j^2\zeta_k^2 = \frac{1}{n^3}\mathbb{E}\zeta^4 + \frac{6(n-1)}{n^3}(\mathbb{E}\zeta^2)^2 \le \frac{6}{n^2}\mathbb{E}\zeta^4.$$

In the last inequality, we used $(\mathbb{E}\zeta^2)^2 \le \mathbb{E}\zeta^4$. So for any $\varepsilon > 0$, by Chebyshev inequality,

$$\mathbb{P}\Big[\Big|\frac{1}{n}S_n\Big| \ge \varepsilon\Big] \le \frac{1}{\varepsilon^4}\mathbb{E}\Big[\Big(\frac{1}{n}S_n\Big)^4\Big] \le \frac{6}{n^2}\frac{\mathbb{E}\zeta^4}{\varepsilon^4}.$$

Since $\sum_n \frac{6}{n^2}\frac{\mathbb{E}\zeta^4}{\varepsilon^4} < \infty$, by Borel-Cantelli lemma, a.s. there is a random N such that for $n > N$, $|\frac{1}{n}S_n| < \varepsilon$. This implies that $\frac{1}{n}\sum_{j=1}^n \zeta_j \to 0$ a.s. □

Exercise . Problems 4, 5, 6, 8, 11, of Exercises of Chapter 3.

4 Characteristic Functions and Classical Limit Theorems

Suppose ζ is a random vector in $\mathbb{R}^d$ with distribution μ, the associated *characteristic function* $\widehat{\mu}$ is given by

$$\widehat{\mu}(t) = \int e^{itx}\mu(dx) = \mathbb{E}e^{it\zeta}, \quad t \in \mathbb{R}^d, \tag{4.1}$$

where tx denotes the inner product $t_1x_1+\cdots+t_dx_d$. The function $x \mapsto e^{itx}$ is integrable because $|e^{itx}| = 1$. In the language of Analysis, $\widehat{\mu}$ is the Fourier transform of μ. If μ is a distribution on $\mathbb{R}^d_+$, i.e., $\mu\mathbb{R}^d_+ = 1$, then the *Laplace transform* $\widetilde{\mu}$ is defined by

$$\widetilde{\mu}(t) = \int e^{-tx}\mu(dx) = \mathbb{E}e^{-t\zeta}, \quad t \in \mathbb{R}^d_+.$$

The function $x \mapsto e^{-tx}$ is integrable because $0 < e^{-tx} \le 1$ as $tx \ge 0$. Finally, for a distribution μ on $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$, the generating function ψ is defined by

$$\psi(s) = \sum_{n=0}^{\infty} s^n \mathbb{P}\{\zeta = n\} = \mathbb{E}s^{\zeta}, \quad s \in [0, 1].$$

Formally, $\widetilde{\mu}(u) = \widehat{\mu}(iu)$ and $\widehat{\mu}(t) = \widetilde{\mu}(-it)$, $\widetilde{\mu}(u) = \psi(e^{-u})$ and $\psi(s) = \widetilde{\mu}(-\log s)$. We will focus on characteristic functions. Many results also apply to Laplace transforms and generating functions with similar proofs.

We first list some simple properties of characteristic functions.

(i) If ϕ is the characteristic function for ζ, then for any $a \in \mathbb{R}$ and $b \in \mathbb{R}^d$, the characteristic function for $a\zeta + b$ is $e^{itb}\phi(at)$.

(ii) If $\phi_1, \dots, \phi_n$ are characteristic functions for independent $\zeta_1, \dots, \zeta_n$, then the characteristic function for $\zeta_1 + \cdots + \zeta_n$ is $\prod_{j=1}^n \phi_j(t)$. We used the fact that $e^{it\zeta_1}, \dots, e^{it\zeta_n}$ are independent. Thus, if $\zeta_1, \dots, \zeta_n$ are i.i.d. with characteristic function ϕ, and $S_n = \sum_{k=1}^n \zeta_k$, then the characteristic function for $\frac{1}{n}S_n$ is $\phi(t/n)^n$.

(iii) For any characteristic function ϕ, $\phi(0) = 1$ and for any $t \in \mathbb{R}^d$, $|\phi(t)| \le 1$ and $\phi(-t) = \overline{\phi(t)}$, where the bar stands for the complex conjugate. Here we use the inequality $|\mathbb{E}f| \le \mathbb{E}|f|$ for complex random variable f and the equality $\overline{e^{itx}} = e^{-itx}$.

(iv) ϕ is uniformly integrable. Here we use that $|e^{it_1x} - e^{it_2x}| = |e^{i(t_1-t_2)x} - 1| \le 2 \wedge (|t-s||x|)$ and dominated convergence theorem.

(v) If μ_n converges weakly to μ with associated characteristic functions ϕ_n and ϕ, then for any $t \in \mathbb{R}^d$, $\phi_n(t) \to \phi(t)$. Here we used the fact that, for any $t \in \mathbb{R}^d$, $x \mapsto e^{itx}$ is a bounded and continuous.

(vi) In the case $d = 1$, if $\mathbb{E}|\zeta|^n < \infty$ for some $n \in \mathbb{N}$, then ϕ is n-times continuously differentiable, $\phi^{(n)}$ is bounded and uniformly continuous, and $\phi^{(n)}(0) = i^n \mathbb{E}\zeta^n$. To see this is true, we may formally differentiate (4.1) w.r.t. t and get $\phi'(t) = \mathbb{E}[ixe^{itx}]$. If we continue differentiation, then we get $\phi^{(k)}(t) = \mathbb{E}[(ix)^k e^{itx}]$ for all $k \in \mathbb{N}$. In general, this equalities may not hold. In fact, $(ix)^k e^{itx}$ may not be integrable. However, if $\mathbb{E}|\zeta|^n < \infty$ for some $n \in \mathbb{N}$, then for any $0 \le k \le n$ and $t \in \mathbb{R}^d$, $(i\zeta)^k e^{it\zeta}$ is integrable, and we may define $\phi^{[k]}(t) = \mathbb{E}[(i\zeta)^k e^{it\zeta}]$, $0 \le k \le n$. Here $\phi^{[0]} = \phi$. Since $|(ix)^k e^{itx} - (ix)^k e^{isx}| \le |x|^k (2 \wedge |s-t||x|)$, by DCT, we see that $\phi^{[k]}$ is uniformly continuous for each $0 \le k \le n$. By Fubini Theorem, for $1 \le k \le n$ and $a < b \in \mathbb{R}$, $\int_a^b \phi^{[k]}(t) = \phi^{[n-1]}(b) - \phi^{[k-1]}(a)$. Thus, $\phi^{[k]}$ is the derivative of $\phi^{[k-1]}$. So $\phi^{(n)}(t) = \phi^{[n]}(t) = \mathbb{E}[(i\zeta)^k e^{it\zeta}]$. Taking $t = 0$, we get $\phi^{(n)}(0) = \mathbb{E}[(i\zeta)^k] = i^k \mathbb{E}[\zeta^k]$.

The following theorem is important for us.

Theorem 4.3. *For probability measures $\mu, \mu_1, \mu_2, \dots$ on $\mathbb{R}^d$, $\mu_n \xrightarrow{\text{w}} \mu$ iff $\widehat{\mu}_n \to \widehat{\mu}$ pointwise iff $\widehat{\mu}_n \to \widehat{\mu}$ uniformly on every bounded set.*

That $\mu_n \xrightarrow{\text{w}} \mu$ implies that $\widehat{\mu}_n \to \widehat{\mu}$ pointwise is Property (v) above. We postpone the proof to the end of this chapter. This theorem in particular implies that $\widehat{\mu}$ determines μ.

Example . (i) If μ is the degeneracy distribution δ_{x_0}, then $\widehat{\mu}(t) = e^{itx_0}$.

(ii) If μ is the Bernoulli distribution $\mathrm{B}(p)$, then $\widehat{\mu}(t) = pe^{it\cdot 1} + (1-p)e^{it\cdot 0} = 1 - p + pe^{it}$.

(iii) If μ is the binomial distribution $\mathrm{B}(n,p)$, since it is the n-th convolution power of the Bernoulli distribution $\mathrm{B}(p)$, we get $\widehat{\mu}(t) = (1-p+pe^{it})^n$.

(iv) If μ is the geometric distribution $\mathrm{Geom}(p)$, then

$$\widehat{\mu}(t) = \sum_{k=1}^{\infty} (1-p)^{k-1} p e^{itk} = \frac{pe^{it}}{1-(1-p)e^{it}}.$$

(v) If μ is the Poisson distribution $\mathrm{Pois}(\lambda)$, then

$$\widehat{\mu}(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} = e^{\lambda e^{it} - \lambda}.$$

(vi) If μ is the uniform distribution $\mathrm{U}[a,b]$, then

$$\widehat{\mu}(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{itb} - e^{ita}}{itb - ita}.$$

(vii) If μ is the exponential distribution $\mathrm{Exp}(\lambda)$, then

$$\widehat{\mu}(t) = \int_0^{\infty} \lambda e^{-\lambda x} e^{itx} dx = \frac{\lambda}{\lambda - it}.$$

(viii) If μ is the normal distribution $\mathrm{N}(a, \sigma^2)$, then

$$\widehat{\mu}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} e^{itx} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} e^{it(a+\sigma y)} dy$$
$$= e^{-\frac{\sigma^2}{2}t^2+iat} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-it\sigma)^2} dy = e^{-\frac{\sigma^2}{2}t^2+iat}.$$

Here the last equality follows from contour integral in complex analysis. The statement holds true even if $\sigma = 0$. When μ is $\mathrm{N}(0,1)$, the characteristic function is $e^{-\frac{t^2}{2}}$.

We now study some applications of Theorem 4.3.

Theorem (weak law of large numbers for L^1)**.** *Let $\zeta, \zeta_1, \zeta_2, \dots$ be an i.i.d. sequence random variables in L^1. Let $S_n = \sum_{j=1}^n \zeta_j$. Then $\frac{1}{n}S_n \xrightarrow{\mathrm{P}} \mathbb{E}\zeta$.*

Proof. Let ϕ be the characteristic function for ζ. Since $\zeta \in L^1$, $\phi \in C^1$, $\phi(0) = 1$ and $\phi'(0) = i\mathbb{E}\zeta$. The characteristic function for $\frac{1}{n}S_n$ is $\phi(t/n)^n = \exp(n \log \phi(t/n))$, which tends to $\exp(t\frac{d}{dt}(\log\phi)|_0) = e^{it\phi'(0)/\phi(0)} = e^{it\mathbb{E}\zeta}$ as $n \to \infty$. Since $e^{it\mathbb{E}\zeta}$ is the characteristic function for $\delta_{\mathbb{E}\zeta}$, by Theorem 3.4, $\mathrm{Law}(\frac{1}{n}S_n) \to \delta_{\mathbb{E}\zeta}$. So $\frac{1}{n}S_n \xrightarrow{\mathrm{P}} \mathbb{E}\zeta$. ☐

Proposition 4.9 (central limit theorem)**.** *Let $\zeta, \zeta_1, \zeta_2, \dots$ be i.i.d. random variables in L^2 with $\mathbb{E}\zeta = 0$ and $\mathbb{E}\zeta^2 = 1$. Let $S_n = \sum_{j=1}^n \zeta_j$. Then $\mathrm{Law}(n^{-1/2}S_n) \xrightarrow{\mathrm{w}} N(0,1)$.*

Proof. Let ϕ be the characteristic function for ζ. Since $\zeta \in L^2$, we have $\phi \in C^2$, $\phi(0) = 1$, $\phi'(0) = i\mathbb{E}\zeta = 0$, and $\phi''(0) = -\mathbb{E}\zeta^2 = -1$. The characteristic function for $n^{-1/2}S_n$ is $\phi(n^{-1/2}t)^n = \exp(n \log \phi(t/n^{1/2}))$. By Taylor theorem, as $n \to \infty$,

$$\phi(\frac{t}{\sqrt{n}}) = 1 - \frac{t^2}{2n} + o(\frac{1}{n}),$$

which implies that

$$\log \phi(\frac{t}{\sqrt{n}}) = -\frac{t^2}{2n} + o(\frac{1}{n}).$$

So we have

$$\phi(n^{-1/2}t)^n = \exp(n \log \phi(t/n^{1/2})) \to e^{-\frac{t^2}{2}}, \quad \text{as } n \to \infty.$$

Since $e^{-\frac{t^2}{2}}$ is the characteristic function for $\mathrm{N}(0,1)$. The proof is complete by Theorem 4.3. ☐

Theorem (Poisson limit theorem)**.** *For any $\lambda > 0$, as $n \to \infty$, the binomial distributions $B(n, \lambda/n)$ tend weakly to the Poisson distribution Pois(λ).*

Proof. The characteristic function for $\mathrm{B}(n, \lambda/n)$ is

$$\phi_n = (1 - \lambda/n + \lambda/ne^{it})^n \to e^{\lambda e^{it} - \lambda}, \quad \text{as } n \to \infty.$$

Since $e^{\lambda e^{it}-\lambda}$ is the characteristic function for $\mathrm{Pois}(\lambda)$, the proof is done. ☐

The rest of this chapter is devoted to the proof of Theorem 4.3. Recall the definition of tightness: a family of probability measures μ_t, $t \in T$, on a topological space S is tight if for any $\varepsilon > 0$, there is a compact set $K \subset S$ such that $\mu_t K^c < \varepsilon$ for all $t \in T$. When $S = \mathbb{R}^d$, this is equivalent to $\lim_{r\to\infty} \sup_{t\in T} \mu_t\{x : |x| \ge r\} = 0$. If, in addition, $T = \mathbb{N}$, this is further equivalent to $\lim_{r\to\infty} \limsup_n \mu_n\{x : |x| \ge r\} = 0$.

Lemma 3.8. *A weakly convergent sequence of probability measures on $\mathbb{R}^d$ is tight.*

This is a special case of Prokhorov Theorem. But we can now give a direct proof.

Proof. For any $r > 1$, we define a bounded continuous function f_r on $\mathbb{R}^d$ by $f_r(x) = 0$ if $|x| \le r-1$, $f_r(x) = 1$ if $|x| \ge r$, and $f_r(x) = |x| - (r-1)$ if $r-1 \le |x| \le r$. Then

$$\limsup_n \mu_n\{x : |x| \ge r\} \le \lim_n \mu_n f_r = \mu f_r \le \mu\{x : |x| \ge r-1\}.$$

Here the RHS tends to 0 as $r \to \infty$. So $\lim_{r\to\infty} \limsup_n \mu_n\{x : |x| \ge r\} = 0$. □

Lemma 3.9. *Let $\zeta_1, \zeta_2, \dots$ be random vectors in $\mathbb{R}^d$ with laws $\mu_1, \mu_2, \dots$. Then $\{\mu_n\}$ is tight iff $c_n\zeta_n \xrightarrow{\mathrm{P}} 0$ for any constants $c_1, c_2, \dots \ge 0$ with $c_n \to 0$.*

Proof. First assume that $\{\mu_n\}$ is tight. Let $c_n \to 0$. Fix any $r, \varepsilon > 0$. We note that $|c_n r| \le \varepsilon$ for all but finitely many n. So $|c_n\zeta_n| > \varepsilon$ implies $|\zeta_n| > r$ for all but finitely many n. So we get

$$\limsup \mathbb{P}\{|c_n\zeta_n| > \varepsilon\} \le \limsup \mathbb{P}\{|\zeta_n| > r\}.$$

Here the RHS tends to 0 as $r \to \infty$, and the LHS does not depend on r. So $\limsup \mathbb{P}\{|c_n\zeta_n| > \varepsilon\} \le 0$, which implies that $\lim \mathbb{P}\{|c_n\zeta_n| > \varepsilon\} = 0$. Since this holds for any $\varepsilon > 0$, we get $c_n\zeta_n \xrightarrow{\mathrm{P}} 0$.

If $\{\mu_n\}$ is not tight. Then we can find $\varepsilon_0 > 0$ and a subsequence (μ_{n_k}) such that $\mathbb{P}[|\zeta_{n_k}| \ge k] \ge \varepsilon_0$ for all k. We may then find $c_1, c_2, \dots \ge 0$ with $c_n \to 0$ such that $c_{n_k} = \frac{1}{k}$. Then $\mathbb{P}[|c_{n_k}\zeta_{n_k}| \ge 1] \ge \varepsilon_0$ for all k, which implies that $c_n\zeta_n$ does not converge to 0 in probability. □

Lemma 4.1. *For any probability measure μ on $\mathbb{R}$ and $r > 0$, we have*

$$\mu\{x : |x| \ge r\} \le \frac{r}{2}\int_{-2/r}^{2/r} (1 - \widehat{\mu}(t))dt; \tag{4.2}$$

Proof. Let $c > 0$. By Fubini Theorem and straightforward calculation,

$$\int_{-c}^{c} (1 - \widehat{\mu}(t))dt = \int_{-c}^{c}\int_{\mathbb{R}} (1 - e^{itx})\mu(dx)dt = \int_{\mathbb{R}}\int_{-c}^{c} (1 - e^{itx})dt\mu(dx)$$

$$= \int_{\mathbb{R}} \Big(t - \frac{e^{itx}}{ix}\Big)\Big|_{t=-c}^{t=c} \mu(dx) = 2c\int_{\mathbb{R}} \Big(1 - \frac{\sin(cx)}{cx}\Big)\mu(dx) \ge c\mu\{x : |cx| \ge 2\},$$

where the last step follows from $\sin x \le 1 \le x/2$ for $x \ge 2$. Letting $c = \frac{2}{r}$, we get (4.2). □

Remark . For $1 \le k \le d$, let $e_k \in \mathbb{R}^d$ be the vector whose k-th coordinate is 1 and other coordinates are 0; let π_k be the projection $x \mapsto x_k = e_k x$ from $\mathbb{R}^d$ to $\mathbb{R}$. For a probability measure μ on $\mathbb{R}^d$, and $1 \le k \le d$, let $\mu^k = (\pi_k)_* \mu$. Then we get

$$\widehat{\mu}^k(t) = \int_{\mathbb{R}} e^{itx} \mu^k(dx) = \int_{\mathbb{R}^d} e^{itx_k} \mu(dx) = \int_{\mathbb{R}^d} e^{i(te_k)x} \mu(dx) = \widehat{\mu}(te_k), \quad 1 \le k \le d.$$

By Lemma 4.1, we have

$$\mu(\mathbb{R}^d \setminus [-\delta, \delta]^d) \le \sum_{k=1}^{n} \mu\{x \in \mathbb{R}^d : |x_k| \ge \delta\} = \sum_{k=1}^{d} \mu_k\{x \in \mathbb{R} : |x| \ge \delta\}$$

$$\le \sum_{k=1}^{d} \frac{\delta}{2} \int_{-2/\delta}^{2/\delta} (1 - \widehat{\mu}^k(t)) dt = \sum_{k=1}^{d} \frac{\delta}{2} \int_{-2/\delta}^{2/\delta} (1 - \widehat{\mu}(te_k)) dt. \tag{4.3}$$

Lemma 4.2. *A family $\{\mu_\alpha\}$ of probability measures on $\mathbb{R}^d$ is tight iff $\{\widehat{\mu}_\alpha\}$ is equicontinuous at 0, and then $\{\widehat{\mu}_\alpha\}$ is uniformly equicontinuous on $\mathbb{R}^d$.*

Proof. Note that $\{\mu_\alpha\}$ is tight iff $\mu_\alpha(\mathbb{R}^d \setminus [-r, r]^d) \to 0$ as $r \to \infty$, uniformly in α. First, suppose $\{\widehat{\mu}_\alpha\}$ is equicontinuous at 0. Then for each $1 \le k \le d$, $\frac{r}{2} \int_{-2/r}^{2/r} (1 - \widehat{\mu}_\alpha(te_k)) dt \to 0$ as $r \to \infty$, uniformly in α. By (4.3) we see that $\{\mu_\alpha\}$ is tight.

Next, suppose $\{\mu_\alpha\}$ is tight. Let ζ_α be a random vector with law μ_α. We compute that for $s, t \in \mathbb{R}^d$,

$$|\widehat{\mu}_\alpha(s) - \widehat{\mu}_\alpha(t)| = \mathbb{E}|e^{i(s-t)\zeta_\alpha} - 1| \le \mathbb{E}[|(s-t)\zeta_\alpha| \wedge 2].$$

By Lemma 3.9, for any sequence (α_n) of indices and any two sequences (s_n) and (t_n) in $\mathbb{R}^d$ with $|s_n - t_n| \to 0$, we get $(s_n - t_n)\zeta_n \xrightarrow{\mathrm{P}} 0$, which implies by DCT that $\mathbb{E}[|(s_n - t_n)\zeta_{\alpha_n}| \wedge 2] \to 0$, and so by the above formula, $|\widehat{\mu}_{\alpha_n}(s_n) - \widehat{\mu}_{\alpha_m}(t_n)| \to 0$. This shows that $\{\widehat{\mu}_\alpha\}$ is uniformly equicontinuous on $\mathbb{R}^d$, and in particular is equicontinuous at 0. $\square$

We also need the following approximation lemma from Analysis.

Lemma 4.4 (Stone-Weierstrass approximation)**.** *Every continuous function $f : \mathbb{R}^d \to \mathbb{R}$ with period 2π in each coordinate admits a uniform approximation by linear combinations of e^{ikx}, $k \in \mathbb{Z}^d$.*

Proof. We first consider the case $d = 1$. In this case f has a Fourier series $\sum_{n \in \mathbb{Z}} a_n e^{inx}$, where $a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$. The truncated series $\sum_{n=-N}^{N} a_n e^{inx}$ is a linear combinations of e^{inx}, $n \in \mathbb{Z}$, but may not converge uniformly to f.

The approximation sequence is constructed as follows. Let $N \in \mathbb{N}$. Let $h_N(x)$ be the sum of the finite geometric series

$$h_N(x) = e^{i(1-N)x/2} + e^{i(3-N)x/2} + \cdots + e^{i(N-3)x/2} + e^{i(N-1)x/2} = e^{i(1-N)x/2} \sum_{k=0}^{N-1} e^{ikx}.$$

It has ratio e^{ix}, the leading term $e^{i(1-N)x/2}$ and the ending term $e^{i(N-1)x/2}$. We observe that $e^{ix/2}h_N(x) - e^{-ix/2}h_N(x) = e^{iNx/2} - e^{-iNx/2}$, and so $h_N(x) = \frac{e^{iNx/2}-e^{-iNx/2}}{e^{ix/2}-e^{-ix/2}} = \frac{\sin(Nx/2)}{\sin(x/2)}$. Calculating $h_N(x)^2$ using the series expression, we get

$$\frac{\sin^2(Nx/2)}{\sin^2(x/2)} = e^{i(1-N)x}\Big(\sum_{n=0}^{N-1} e^{ikx}\Big)\Big(\sum_{m=0}^{N-1} e^{imx}\Big) = e^{i(1-N)x}\sum_{k=0}^{2N-2}\sum_{0\le n,m\le N-1:n+m=k} e^{ikx}$$

$$= e^{i(1-N)x}\sum_{k=0}^{2N-2}(N-|N-1-k|)e^{ikx} = \sum_{j=1-N}^{N-1}(N-|j|)e^{ijx}.$$

Let $b_n^{(N)} = (1-\frac{|n|}{N})$. Then $\sum_{n=1-N}^{N-1} b_n^{(N)} e^{inx} = \frac{\sin^2(Nx/2)}{N\sin^2(x/2)}$. We define $g_N(x) = \frac{\sin^2(Nx/2)}{N\sin^2(x/2)}$. Then $g_N \ge 0$, and $\frac{1}{2\pi}\int_{-\pi}^{\pi} g_N(x)dx = b_0^{(N)} = 1$.

Let

$$f_N(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)g_N(y)dy = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)g_N(x-y)dy$$
$$= \sum_n \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)b_n^{(N)}e^{in(x-y)}dy = \sum_n a_n b_n^{(N)} e^{inx}.$$

So f_N is a linear combination of e^{ikx}, $k\in\mathbb{Z}$. To see that $f_N \to f$ uniformly, we compute

$$|f_N(x)-f(x)| = \Big|\frac{1}{2\pi}\int_{-\pi}^{\pi}(f(x-y)-f(x))g_N(y)dy\Big| \le \frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x-y)-f(x)|g_N(y)dy$$

$$= \frac{1}{2\pi}\int_{-\delta}^{\delta}|f(x-y)-f(x)|g_N(y)dy + \frac{1}{2\pi}\int_{[-\pi,\pi]\setminus[-\delta,\delta]}|f(x-y)-f(x)|g_N(y)dy$$
$$\le \sup_{|x-z|\le\delta}|f(x)-f(z)| + 2\|f\|\cdot\frac{1}{2\pi}\int_{[-\pi,\pi]\setminus[-\delta,\delta]} g_N(y)dy. \tag{4.4}$$

We note that for any fixed $\delta\in(0,\pi)$, $\frac{1}{2\pi}\int_{[-\pi,\pi]\setminus[-\delta,\delta]} g_N(y)dy \to 0$ as $N\to\infty$ because

$$\sup_{y\in[-\pi,\pi]\setminus[-\delta,\delta]} g_N(y) \le \frac{1}{N\sin^2(\delta/2)}.$$

Given any $\varepsilon>0$, we may first choose $\delta\in(0,\pi)$ such that $\sup_{|x-z|\le\delta}|f(x)-f(z)| < \frac{\varepsilon}{2}$, and then choose N_0 such that for $N>N_0$, the second term of (4.4) is also less than $\frac{\varepsilon}{2}$.

For general dimension d, we let $g_N^{(d)}(x) = \prod_{k=1}^d g_N(x_k)$. Then $g_N^{(d)} \ge 0$, is a linear combination of e^{ikx}, $k\in\mathbb{Z}^d$, and satisfies $\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} g_N^{(d)}(y)dy = 1$. Let

$$f_N(x) = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} f(x-y)g_N^{(d)}(y)dy = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} f(y)g_N^{(d)}(x-y)dy.$$

Then each f_N is a linear combination of e^{ikx}, $k \in \mathbb{Z}^d$. A similar computation with $[-\delta,\delta]^d$ and $[-\pi,\pi]^d$ in place of $[-\delta,\delta]$ and $[-\pi,\pi]$ shows that

$$|f_N(x) - f(x)| \le \sup_{\max_k |x_k - z_k| \le \delta} |f(x) - f(z)| + \frac{2\|f\|}{(2\pi)^d} \int_{[-\pi,\pi]^d \setminus [-\delta,\delta]^d} g_N^{(d)}(y)dy.$$

To conclude that $f_n \to f$ uniformly, we need to show that for any $\delta \in (0,\pi)$,

$$\int_{[-\pi,\pi]^d \setminus [-\delta,\delta]^d} g_N^{(d)}(y)dy \to 0, \quad \text{as } N \to \infty. \tag{4.5}$$

Now we do not have $\sup_{y \in [-\pi,\pi]^d \setminus [-\delta,\delta]^d} g_N^{(d)}(y) \to 0$ as $N \to \infty$. However, if we let $U_k = \{x \in [-\pi,\pi]^d : |x_k| \ge \delta\}$, $1 \le k \le d$, then the LHS of (4.5) is

$$\le \sum_{k=1}^d \int_{U_k} g_N^{(d)}(y)dy = d\Big(\int_{[-\pi,\pi]} g_N(y)dy\Big)^{d-1} \cdot \int_{[-\pi,\pi]\setminus[-\delta,\delta]} g_N(y)dy \le \frac{d(2\pi)^d}{N\sin^2(\delta/2)}.$$

So we get (4.5) and conclude the proof. $\square$

Proof of Theorem 4.3. If $\mu_n \xrightarrow{w} \mu$, then for each $t \to \mathbb{R}^d$, since $x \mapsto e^{itx}$ is bounded and continuous on $\mathbb{R}^d$, we get $\widehat{\mu}_n(t) \to \widehat{\mu}(t)$. By Lemma 3.8, $\{\mu_n\}$ is tight. By Lemma 4.2, $\{\widehat{\mu}_n\}$ is uniformly equicontinuous on $\mathbb{R}^d$. So $\widehat{\mu}_n \to \widehat{\mu}$ uniformly on every bounded set.

Suppose now $\widehat{\mu}_n \to \widehat{\mu}$ pointwise. By (4.3) we have

$$\limsup_n \mu_n(\mathbb{R}^d \setminus [-r,r]^d) \le \limsup_n \sum_{k=1}^d \frac{r}{2}\int_{-2/r}^{2/r} (1 - \widehat{\mu}_n(te_k))dt = \sum_{k=1}^d \frac{r}{2}\int_{-2/r}^{2/r}(1-\widehat{\mu}(te_k))dt,$$

where the equality follows from DCT. Since $\widehat{\mu}$ is continuous at 0, the RHS tends to 0 as $r \to \infty$, which shows that $\{\mu_n\}$ is tight.

Given any $\varepsilon > 0$, we may then choose $r > 0$ so large such that $\mu_n\{|x| \ge r\} < \varepsilon$ for each n and $\mu\{|x| \ge r\} < \varepsilon$. Now fix $f \in C_b(\mathbb{R}^d)$. We need to show that $\mu_n f \to \mu f$. By the definition of $\widehat{\mu}_n$ and $\widehat{\mu}$, we know this is true if f is of the form $x \mapsto e^{itx}$ for some $t \in \mathbb{R}^d$, or is a linear combination of such functions. Let $m = \|f\|$, the supernorm of f. Let $h \in C(\mathbb{R}^d)$ be such that $0 \le h \le 1$, $h \equiv 1$ on $\{|x| \le r\}$, and $h \equiv 0$ on $\mathbb{R}^d \setminus (-\pi r, \pi r)^d$. Then $\|hf\| \le m$, hf agrees with f in $\{|x| \le r\}$, and vanishes outside $(-\pi r,\pi r)^d$. So we may extend hf from $(-\pi r,\pi r)^d$ to $\widetilde{f} \in C(\mathbb{R}^d)$, which has period $2\pi r$ in each coordinate. Then $\widetilde{f}$ agrees with f on $\{|x| \le r\}$, and $\|\widetilde{f}\| = \|hf\| \le m$. By Lemma 4.4 there exists some linear combination g of $e^{ikx/r}$, $k \in \mathbb{Z}^d$, such that $\|\widetilde{f} - g\| < \varepsilon$. By earlier discussion, $\mu_n g \to \mu g$. For any $n \in \mathbb{N}$,

$$|\mu_n f - \mu_n g| \le \mu_n\{|x| \ge r\}\|f - \widetilde{f}\| + \|\widetilde{f} - g\| \le 2m\varepsilon + \varepsilon,$$

and similarly for μ. Thus,

$$|\mu_n f - \mu f| \le |\mu_n g - \mu g| + 2(2m+1)\varepsilon, \quad n \in \mathbb{N}.$$

Letting $n \to \infty$ and then $\varepsilon \to 0$, we get $\mu_n f \to \mu f$. Since $f \in C_b$ is arbitrary, we get $\mu_n \xrightarrow{w} \mu$. $\square$

Exercise . Problems 6, 14 of Exercises of Chapter 4.

5 Conditioning and Disintegration

We now study conditioning. We still fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Suppose $B \in \mathcal{A}$ is such that $\mathbb{P}B > 0$. We may then define a conditional probability

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, \quad A \in \mathcal{A}.$$

It is easy to see that $\mathbb{P}[\cdot|B]$ is a probability measure on $(\Omega, \mathcal{A})$. The expectation w.r.t. this probability measure is then given by

$$\mathbb{E}[\zeta|B] = \frac{\mathbb{E}[\mathbf{1}_B \zeta]}{\mathbb{P}[B]}.$$

We want to extend the above concept and define conditional expectation $\mathbb{E}[\cdot|\mathcal{F}]$, where $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$. To motivate the definition, we suppose $B_1, \dots, B_n$ is a measurable partition of Ω such that $\mathbb{P}[B_k] > 0$ for each $1 \le k \le n$. They together generate a sub-σ-algebra $\mathcal{F}_B$, each element is a union of some B_k's. Given an integrable random variable ζ, consider its conditional expectation given B_k, we get n real values $\mathbb{E}[\zeta|B_1], \dots, \mathbb{E}[\zeta|B_n]$. We now define a new random variable ζ_B on Ω by

$$\zeta_B = \sum_{k=1}^{n} \mathbb{E}[\zeta|B_k] \mathbf{1}_{B_k}. \tag{5.1}$$

Then ζ_B is $\mathcal{F}_B$-measurable, and for any B_k,

$$\mathbb{E}[\mathbf{1}_{B_k} \zeta_B] = \mathbb{E}[\zeta|B_k]\mathbb{P}[B_k] = \mathbb{E}[\mathbf{1}_{B_k} \zeta].$$

Since every $A \in \mathcal{F}_B$ is a disjoint union of some B_k's, we get

$$\mathbb{E}[\mathbf{1}_A \zeta_B] = \mathbb{E}[\mathbf{1}_A \zeta], \quad \forall A \in \mathcal{F}_B \tag{5.2}$$

On the other hand, suppose ζ_B is an $\mathcal{F}_B$-measurable random variable and satisfies (5.2). Then ζ_B takes constant value on each B_k, and so can be expressed as $\sum_k c_k \mathbf{1}_{B_k}$ for some $c_1, \dots, c_n \in \mathbb{R}$. Taking $A = B_k$ in (5.2), we get $c_k \mathbb{P}[B_k] = \mathbb{E}[\mathbf{1}_{B_k} \zeta]$, which implies that $c_k = \mathbb{E}[\zeta|B_k]$. So ζ_B is given by (5.1), and we can reveal $\mathbb{E}[\zeta|B_k]$ for each k from ζ_B.

Definition . For a sub-σ-algebra $\mathcal{F}$ of $\mathcal{A}$ and $\zeta \in L^1(\mathcal{A}, \mathbb{P})$, we use $\mathbb{E}[\zeta|\mathcal{F}]$ or $\mathbb{E}^{\mathcal{F}} \zeta$ to denote an element $\eta \in L^1(\mathcal{F}, \mathbb{P})$, which satisfies that

$$\mathbb{E}[\mathbf{1}_A \zeta] = \mathbb{E}[\mathbf{1}_A \eta], \quad \forall A \in \mathcal{F}. \tag{5.3}$$

For $A \in \mathcal{A}$, we define $\mathbb{P}^{\mathcal{F}} A = \mathbb{P}[A|\mathcal{F}]$ as $\mathbb{E}[\mathbf{1}_A|\mathcal{F}]$. If η is a random element, then we define $\mathbb{E}[\zeta|\eta] = \mathbb{E}^{\eta} \zeta$ as $\mathbb{E}[\zeta|\sigma(\eta)]$, and define $\mathbb{P}[A|\eta] = \mathbb{P}^{\eta} A$ as $\mathbb{E}[\mathbf{1}_A|\sigma(\eta)]$.

Theorem 5.1, Part I. *The $\mathbb{E}^{\mathcal{F}}\zeta$ as in the definition always exists and is a.s. unique. Moreover, the map $\zeta \mapsto \mathbb{E}^{\mathcal{F}}\zeta$ is a bounded linear map from $L^1(\mathcal{A})$ to $L^1(\mathcal{F})$ with $\|\mathbb{E}^{\mathcal{F}}\zeta\|_1 \le \|\zeta\|_1$, and if $\zeta \ge 0$ then a.s. $\mathbb{E}^{\mathcal{F}}\zeta \ge 0$.*

Proof. We may define a signed measure ν on $(\Omega, \mathcal{A})$ by $d\nu = \zeta d\mathbb{P}$. Then $\nu \ll \mathbb{P}$ on $\mathcal{A}$, and so we also have $\nu \ll \mathbb{P}$ on $\mathcal{F}$. Applying Radon-Nikodym Theorem to $\mathbb{P}$ and ν on $(\Omega, \mathcal{F})$, we get an $\mathcal{F}$-measurable random variable η, which is integrable w.r.t. $\mathbb{P}$, such that $d\nu = \eta d\mathbb{P}$ on $\mathcal{F}$. Let $A \in \mathcal{F}$. From $d\nu = \eta d\mathbb{P}$ on $\mathcal{F}$, we get $\mathbb{E}[\mathbf{1}_A\eta] = \nu(A)$. From $d\nu = \zeta d\mathbb{P}$ on $\mathcal{A}$, we get $\mathbb{E}[\mathbf{1}_A\zeta] = \nu(A)$. Thus, $\mathbb{E}[\mathbf{1}_A\eta] = \mathbb{E}[\mathbf{1}_A\zeta]$. So we get the existence of $\mathbb{E}^{\mathcal{F}}\zeta$.

Now suppose another $\mathcal{F}$-measurable random variable η' satisfies $\mathbb{E}[\mathbf{1}_A\eta'] = \mathbb{E}[\mathbf{1}_A\zeta]$ for any $A \in \mathcal{F}$. Then η and η' are both $\mathcal{F}$-measurable, and for any $A \in \mathcal{F}$, $\mathbb{E}[\mathbf{1}_A\eta'] = \mathbb{E}[\mathbf{1}_A\zeta] = \mathbb{E}[\mathbf{1}_A\eta]$. So $\eta' = \eta$ a.s., and we get the a.s. uniqueness of $\mathbb{E}^{\mathcal{F}}\zeta$. In particular, we see that $\mathbb{E}^{\mathcal{F}}\zeta$ is a uniquely defined element in $L^1(\mathcal{F})$.

If $\zeta \ge 0$, then the above ν is a positive measure, which implies that the Radon-Nikodym derivative $d\nu/d\mathbb{P} = \mathbb{E}^{\mathcal{F}}\zeta$ on $\mathcal{F}$ is a.s. nonnegative.

To see that the map $\zeta \mapsto \mathbb{E}^{\mathcal{F}}\zeta$ is linear, let $\zeta, \eta \in L^1(\mathcal{A})$ and $a, b \in \mathbb{R}$. Let $\zeta' = \mathbb{E}^{\mathcal{F}}\zeta$ and $\eta' = \mathbb{E}^{\mathcal{F}}\eta$. Then for any $A \in \mathcal{F}$, we have

$$\mathbb{E}[\mathbf{1}_A(a\zeta + b\eta)] = a\mathbb{E}[\mathbf{1}_A\zeta] + b\mathbb{E}[\mathbf{1}_A\eta] = a\mathbb{E}[\mathbf{1}_A\zeta'] + b\mathbb{E}[\mathbf{1}_A\eta'] = \mathbb{E}[\mathbf{1}_A(a\zeta' + b\eta')].$$

So we get $\mathbb{E}[a\zeta + b\eta|\mathcal{F}] = a\mathbb{E}^{\mathcal{F}}\zeta + b\mathbb{E}^{\mathcal{F}}\eta$.

To see that $\|\mathbb{E}^{\mathcal{F}}\zeta\|_1 \le \|\zeta\|_1$, we write $\zeta = \zeta_+ - \zeta_-$ such that $\zeta_\pm \ge 0$ and $\|\zeta\|_1 = \|\zeta_+\|_1 + \|\zeta_-\|_1$. Let $\zeta'_\pm = \mathbb{E}^{\mathcal{F}}\zeta_\pm \ge 0$. Then $\mathbb{E}^{\mathcal{F}}\zeta = \zeta'_+ - \zeta'_-$, and so

$$\|\mathbb{E}^{\mathcal{F}}\zeta\|_1 \le \|\zeta'_+\|_1 + \|\zeta'_-\|_1 = \mathbb{E}[\zeta'_+] + \mathbb{E}[\zeta'_-] = \mathbb{E}[\zeta_+] + \mathbb{E}[\zeta_-] = \|\zeta_+\|_1 + \|\zeta_-\|_1 = \|\zeta\|_1.$$

□

We refer to the property (5.3) as the *averaging property*; to the property that $\zeta \ge 0$ implies $\mathbb{E}[\zeta|\mathcal{F}] \ge 0$ as *positivity*; to the property that $\zeta \mapsto \mathbb{E}^{\mathcal{F}}\zeta$ is a linear map as linearity; and to the property that $\|\mathbb{E}^{\mathcal{F}}\zeta\|_1 \le \|\zeta\|_1$ as L^1-contractivity. Note that if we take $A = \Omega$ in (5.3), then we get $\mathbb{E}[\zeta] = \mathbb{E}[\mathbb{E}[\zeta|\mathcal{F}]]$. When $\mathcal{F}$ is generated by a partition $\{B_1, \dots, B_n\}$, we get the well-known formula $\mathbb{E}[\zeta] = \sum_{k=1}^n \mathbb{P}[B_k]\mathbb{E}[\zeta|B_k]$.

Since $\mathbb{E}^{\mathcal{F}}\zeta$ is only a.s. unique, any formula involves conditional expectation only holds almost surely no matter whether or not we use the phrase "a.s.".

For any $A \in \mathcal{A}$, since $0 \le \mathbf{1}_A \le 1$, we have $0 \le \mathbb{P}^{\mathcal{F}}A \le 1$. Since $\mathbf{1}_\Omega \equiv 1$ and $\mathbf{1}_\emptyset \equiv 0$, we get a.s. $\mathbb{P}^{\mathcal{F}}\Omega = 1$ and $\mathbb{P}^{\mathcal{F}}\emptyset = 0$.

Remark . There are two trivial cases. If $\mathcal{F} = \mathcal{A}$, since ζ is $\mathcal{A}$-measurable, we get $\mathbb{E}^{\mathcal{A}}\zeta = \zeta$. If $\mathcal{F} = \{\Omega, \emptyset\}$, then $\mathbb{E}[\zeta|\{\Omega, \emptyset\}]$ is a constant, which equals $\mathbb{E}[\zeta]$.

Remark . If $\zeta \perp\!\!\!\perp \mathcal{F}$, then $\mathbb{E}^{\mathcal{F}}\zeta = \mathbb{E}\zeta$. In fact, for any $A \in \mathcal{F}$, since $\mathbf{1}_A \perp\!\!\!\perp \zeta$, we have $\mathbb{E}[\mathbf{1}_A\zeta] = \mathbb{E}[\mathbf{1}_A]\mathbb{E}\zeta = \mathbb{E}[\mathbf{1}_A\mathbb{E}\zeta]$. Since $\mathbb{E}\zeta$ is $\mathcal{F}$-measurable, we get $\mathbb{E}^{\mathcal{F}}\zeta = \mathbb{E}\zeta$.

Remark . We have a.s. $\mathbb{E}^{\mathcal{F}}\zeta = \zeta$ iff ζ is $\mathcal{F}^{\mathbb{P}}$-measurable, where $\mathcal{F}^{\mathbb{P}}$ is the $\mathbb{P}$-completion of $\mathcal{F}$. If ζ is $\mathcal{F}^{\mathbb{P}}$-measurable, then there is an $\mathcal{F}$-measurable random variable ζ' such that a.s. $\zeta' = \zeta$. So for any $A \in \mathcal{F}$, $\int_A \zeta' d\mathbb{P} = \int_A \zeta d\mathbb{P}$, which implies that a.s. $\mathbb{E}^{\mathcal{F}}\zeta = \zeta' = \zeta$. On the other hand, if a.s. $\mathbb{E}^{\mathcal{F}}\zeta = \zeta$, we take $\zeta' = \mathbb{E}^{\mathcal{F}}\zeta$. Then ζ' is $\mathcal{F}$-measurable and a.s. $\zeta' = \zeta$. So ζ is $\mathcal{F}^{\mathbb{P}}$-measurable.

Example . Suppose $\mathcal{F}_B$ is generated by a measurable partition $\{B_1, \dots, B_n\}$ of Ω. Now we do not assume that $\mathbb{P}[B_k] > 0$ for every k. Since $\mathbb{E}^{\mathcal{F}_B}\zeta$ is $\mathcal{F}_B$-measurable, it is constant, say c_k, on each B_k. From $\mathbb{E}[\mathbf{1}_{B_k}\mathbb{E}^{\mathcal{F}_B}\zeta] = \mathbb{E}[\mathbf{1}_{B_k}\zeta]$ we get $c_k\mathbb{P}[B_k] = \mathbb{E}[\mathbf{1}_{B_k}\zeta]$. So if $\mathbb{P}[B_k] > 0$, then $c_k = \mathbb{E}[\zeta|B_k]$; if $\mathbb{P}[B_k] = 0$, then c_k can be any number. The choice of c_k does not affect the a.s. uniqueness of $\mathbb{E}[\zeta|\mathcal{F}_B]$.

Theorem 5.1, Part II. *We use the setup as before.*

(i) If $\zeta \in L^\infty$, then $\mathbb{E}^{\mathcal{F}}\zeta \in L^\infty$, and $\|\mathbb{E}^{\mathcal{F}}\zeta\|_\infty \le \|\zeta\|_\infty$.

(ii) For $p \in (1, \infty)$, if $\zeta \in L^p$, then $\mathbb{E}^{\mathcal{F}}\zeta \in L^p$ and $\|\mathbb{E}^{\mathcal{F}}\zeta\|_p \le \|\zeta\|_p$.

(iii) If $\mathcal{G} \subset \mathcal{F}$ is another σ-algebra, then $\mathbb{E}^{\mathcal{G}}\mathbb{E}^{\mathcal{F}}\zeta = \mathbb{E}^{\mathcal{G}}\zeta$.

(iv) If $0 \le \zeta_n \uparrow \zeta \in L^1$, then $\mathbb{E}[\zeta_n|\mathcal{F}] \uparrow \mathbb{E}[\zeta|\mathcal{F}]$.

Proof. (i) Let $M = \|\zeta\|_\infty$, then a.s. $M \pm \zeta \ge 0$, which implies that a.s.

$$0 \le \mathbb{E}^{\mathcal{F}}[M \pm \zeta] = M \pm \mathbb{E}^{\mathcal{F}}\zeta.$$

So a.s. $-M \le \mathbb{E}^{\mathcal{F}}\zeta \le M$, i.e., $\|\mathbb{E}^{\mathcal{F}}\zeta\|_\infty \le M = \|\zeta\|_\infty$.

(ii) Since the map $\mathbb{E}^{\mathcal{F}}$ is a contraction from $L^1(\mathcal{A})$ to $L^1(\mathcal{F})$, and a contraction from $L^\infty(\mathcal{A})$ to $L^\infty(\mathcal{F})$, by Marcinkiewicz interpolation theorem, it is also a contraction from $L^p(\mathcal{A})$ to $L^p(\mathcal{F})$ for any $p \in [1, \infty]$. This result also follows from Jensen's inequality below.

(iii) Let $\zeta' = \mathbb{E}^{\mathcal{F}}\zeta$ and $\zeta'' = \mathbb{E}^{\mathcal{G}}\zeta'$. Then ζ'' is $\mathcal{G}$-measurable, and for any $A \in \mathcal{G}$,

$$\mathbb{E}[\mathbf{1}_A\zeta''] = \mathbb{E}[\mathbf{1}_A\zeta'] = \mathbb{E}[\mathbf{1}_A\zeta.$$

So we get $\zeta'' = \mathbb{E}^{\mathcal{G}}\zeta$.

(iv) From $0 \le \zeta_1 \le \zeta_2 \le \cdots \le \zeta$ we get a.s.

$$0 \le \mathbb{E}[\zeta_1|\mathcal{F}] \le \mathbb{E}[\zeta_2|\mathcal{F}] \le \cdots \le \mathbb{E}[\zeta|\mathcal{F}].$$

Let $\zeta' = \lim_{n\to\infty} \mathbb{E}[\zeta_n|\mathcal{F}]$. Then ζ' is $\mathcal{F}$-measurable and a.s. $\zeta' \le \mathbb{E}[\zeta|\mathcal{F}]$. By Monotone convergence theorem and the averaging property,

$$\mathbb{E}[\zeta'] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[\zeta_n|\mathcal{F}]] = \lim_{n\to\infty} \mathbb{E}[\zeta_n] = \mathbb{E}[\zeta] = \mathbb{E}[\mathbb{E}[\zeta|\mathcal{F}]].$$

This equality together with a.s. $\zeta' \le \mathbb{E}[\zeta|\mathcal{F}]$ implies that a.s. $\mathbb{E}[\zeta|\mathcal{F}] = \zeta' = \lim_{n\to\infty} \mathbb{E}[\zeta_n|\mathcal{F}]$. $\square$

We refer to (i) and (ii) as the L^∞-*contractivity* and L^p-*contractivity*, to (iii) as the *chain rule*, and to (iv) as the *monotone convergence property*.

Theorem 5.2, Part III. *(i) Let $\zeta \in L^1(\mathcal{A})$ and let η be an $\mathcal{F}$-measurable random variable such that $\eta\zeta \in L^1(\mathcal{A})$. Then $\eta\mathbb{E}[\zeta|\mathcal{F}] \in L^1(\mathcal{F})$ and*

$$\mathbb{E}[\eta\zeta|\mathcal{F}] = \eta\mathbb{E}^{\mathcal{F}}\zeta. \tag{5.4}$$

(ii) Let $p, q \in [1,\infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $\eta \in L^p(\mathcal{A})$ and $\zeta \in L^q(\mathcal{A})$. Then $\zeta\mathbb{E}^{\mathcal{F}}\eta$, $\eta\mathbb{E}^{\mathcal{F}}\zeta$, and $\mathbb{E}^{\mathcal{F}}\zeta\mathbb{E}^{\mathcal{F}}\eta$ are all integrable, and have the same expectation, i.e.,

$$\mathbb{E}[\zeta\mathbb{E}^{\mathcal{F}}\eta] = \mathbb{E}[\eta\mathbb{E}^{\mathcal{F}}\zeta] = \mathbb{E}[\mathbb{E}^{\mathcal{F}}\zeta\mathbb{E}^{\mathcal{F}}\eta]. \tag{5.5}$$

Proof. (i) We first assume that η is an $\mathcal{F}$-measurable simple random variable. Then there are $A_1, \dots, A_n \in \mathcal{F}$ and $c_1, \dots, c_n \in \mathbb{R}$ such that $\eta = \sum_{k=1}^n c_k \mathbf{1}_{A_k}$. Then $\eta\mathbb{E}^{\mathcal{F}}\zeta \in L^1(\mathcal{F})$ because η is bounded and $\mathbb{E}^{\mathcal{F}}\zeta \in L^1(\mathcal{F})$. Moreover, for any $A \in \mathcal{F}$,

$$\mathbb{E}[\mathbf{1}_A\eta\mathbb{E}^{\mathcal{F}}\zeta] = \mathbb{E}[\sum_{k=1}^n c_k\mathbf{1}_{A\cap A_k}\mathbb{E}^{\mathcal{F}}\zeta] = \sum_{k=1}^n c_k\mathbb{E}[\mathbf{1}_{A\cap A_k}\zeta] = \mathbb{E}[\mathbf{1}_A\eta\zeta].$$

So we get (5.4). Next, we assume that $\zeta, \eta \geq 0$, but do not assume that η is simple. Then we can find a sequence of nonnegative $\mathcal{F}$-measurable simple random variables (η_n) with $\eta_n \uparrow \eta$. For each n, we have $\eta_n\mathbb{E}^{\mathcal{F}}\zeta \in L^1(\mathcal{F})$ and $\mathbb{E}[\eta_n\zeta|\mathcal{F}] = \eta_n\mathbb{E}^{\mathcal{F}}\zeta$. Since $\eta_n\zeta \uparrow \eta\zeta$, and $\eta\zeta \in L^1(\mathcal{A})$, we get

$$\mathbb{E}[\eta\zeta|\mathcal{F}] = \lim_n \mathbb{E}[\eta_n\zeta|\mathcal{F}] = \lim_n \eta_n\mathbb{E}[\zeta|\mathcal{F}] = \eta\mathbb{E}[\zeta|\mathcal{F}].$$

So we again get (5.4). Finally, we do not assume that $\zeta, \eta \geq 0$. We may write $\zeta = \zeta_+ - \zeta_-$ and $\eta = \eta_+ - \eta_-$. Then for any $\sigma_1, \sigma_2 \in \{+,-\}$, from $|\eta_{\sigma_1}| \leq |\eta|$ and $|\zeta_{\sigma_2}| \leq |\zeta|$ we get $\eta_{\sigma_1}\zeta_{\sigma_2} \in L^1(\mathcal{A})$. The previous result implies that (5.4) holds for η_{σ_1} and ζ_{σ_2}. Using the linearity, we get (5.4) for η and ζ.

(ii) Since $\eta \in L^p(\mathcal{A})$ and $\zeta \in L^q(\mathcal{A})$, we get $\mathbb{E}^{\mathcal{F}}\eta \in L^p(\mathcal{F})$ by L^p-contractivity of $\mathbb{E}^{\mathcal{F}}$ and then $\zeta\mathbb{E}^{\mathcal{F}}\eta \in L^1(\mathcal{A})$ by Hölder's inequality. Applying (i) with $\mathbb{E}[\eta|\mathcal{F}]$ in place of η, we get $\mathbb{E}[\zeta\mathbb{E}^{\mathcal{F}}\eta|\mathcal{F}] = \mathbb{E}^{\mathcal{F}}\zeta\mathbb{E}^{\mathcal{F}}\eta$. Symmetrically, we get $\mathbb{E}[\eta\mathbb{E}^{\mathcal{F}}\zeta|\mathcal{F}] = \mathbb{E}[\eta|\mathcal{F}]\mathbb{E}[\zeta|\mathcal{F}]$. Taking expectation, we get (5.5). $\square$

We refer to (i) as the *pull-out property*, and to (ii) as the *self-adjointness*. From (ii) we see that, for $1 \leq p < \infty$, the adjoint operator of the conditional expectation $\mathbb{E}^{\mathcal{F}} : L^p(\mathcal{A}) \to L^p(\mathcal{F})$ is the conditional expectation $\mathbb{E}^{\mathcal{F}} : L^q(\mathcal{A}) \to L^q(\mathcal{F})$. When $p = 2$, $\mathbb{E}^{\mathcal{F}} : L^2(\mathcal{A}) \to L^2(\mathcal{F})$ is in fact the orthogonal projection onto $L^2(\mathcal{F})$.

Lemma 5.2 (local property)**.** *Let $\mathcal{F}$ and $\mathcal{G}$ be two sub-σ-algebras of $\mathcal{A}$. Let ζ, η be two integrable random variables. Suppose there is $A \in \mathcal{F}\cap\mathcal{G}$ such that $A\cap\mathcal{F} = A\cap\mathcal{G}$ and a.s. $\zeta = \eta$ on A. Then a.s. $\mathbb{E}^{\mathcal{F}}\zeta = \mathbb{E}^{\mathcal{G}}\eta$ on A.*

Proof. Since $A \in \mathcal{F} \cap \mathcal{G}$ and $A \cap \mathcal{F} = A \cap \mathcal{G}$, both $\mathbf{1}_A \mathbb{E}^{\mathcal{F}} \zeta$ and $\mathbf{1}_A \mathbb{E}^{\mathcal{G}} \eta$ are $\mathcal{F} \cap \mathcal{G}$-measurable. For any $B \in \mathcal{F} \cap \mathcal{G}$, by the averaging property and that a.s. $\zeta = \eta$ on A,

$$\mathbb{E}[\mathbf{1}_B \mathbf{1}_A \mathbb{E}^{\mathcal{F}} \zeta] = \mathbb{E}[\mathbf{1}_{A\cap B} \mathbb{E}^{\mathcal{F}} \zeta] = \mathbb{E}[\mathbf{1}_{A\cap B} \zeta] = \mathbb{E}[\mathbf{1}_{A\cap B} \eta] = \mathbb{E}[\mathbf{1}_B \mathbf{1}_A \mathbb{E}^{\mathcal{G}} \eta].$$

Since this holds for any $B \in \mathcal{F} \cap \mathcal{G}$, we get a.s. $\mathbf{1}_A \mathbb{E}^{\mathcal{F}} \zeta = \mathbf{1}_A \mathbb{E}^{\mathcal{G}} \eta$. So we get the conclusion. $\square$

Lemma 5.5 (uniformly integrability, Doob)**.** *For any $\zeta \in L^1$, the family $\mathbb{E}^{\mathcal{F}} \zeta$, where $\mathcal{F}$ is any sub-σ-algebra of $\mathcal{A}$, are uniformly integrable.*

Proof. By the L^1-contractivity, $\{\mathbb{E}^{\mathcal{F}} \zeta\}$ is L^1-bounded. In order to show that it is uniformly integrable, by Lemma 3.10, it suffices to show that $\mathbb{E}[\mathbf{1}_A \mathbb{E}^{\mathcal{F}} \zeta] \to 0$ as $A \in \mathcal{A}$ and $\mathbb{P}A \to 0$, uniformly in $\mathcal{F}$. So we need to show that if $\mathcal{F}_1, \mathcal{F}_2, \dots$ are sub-σ-algebras of $\mathcal{A}$, and $A_1, A_2, \dots \in \mathcal{A}$ satisfy $\mathbb{P}A_n \to 0$, then $\mathbb{E}[\mathbf{1}_{A_n}\ EE^{\mathcal{F}_n} \zeta] \to 0$. By the self-adjointness, Since $\zeta \in L^1$ and $\mathbf{1}_{A_n} \in L^\infty$,

$$\mathbb{E}[\mathbf{1}_{A_n} \mathbb{E}^{\mathcal{F}_n} \zeta] = \mathbb{E}[\zeta \mathbb{E}^{\mathcal{F}_n}[\mathbf{1}_{A_n}]].$$

Since $\mathbb{E}[\mathbb{E}^{\mathcal{F}_n}[\mathbf{1}_{A_n}]] = \mathbb{P}A_n \to 0$, we know that $\mathbb{E}^{\mathcal{F}_n}[\mathbf{1}_{A_n}] \to 0$ in L^1. So $\mathbb{E}^{\mathcal{F}_n}[\mathbf{1}_{A_n}] \xrightarrow{\mathrm{P}} 0$. Thus, $\zeta \mathbb{E}^{\mathcal{F}_n}[\mathbf{1}_{A_n}] \xrightarrow{\mathrm{P}} 0$. By dominated convergence theorem (for convergence in probability), we get $\mathbb{E}[\zeta \mathbb{E}^{\mathcal{F}_n}[\mathbf{1}_{A_n}]] \to 0$, as desired. $\square$

We are going to use conditional expectation to define conditional distribution (or law). Suppose $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$, and ζ is a random element in a measurable space $(S, \overline{S})$. For every $A \in \overline{S}$, $\mathbb{P}^{\mathcal{F}}[\zeta \in A]$ is an element in $L^1(\mathcal{F})$, which satisfies a.s. $0 \le \mathbb{P}^{\mathcal{F}}[\zeta \in A] \le 1$, $\mathbb{P}^{\mathcal{F}}[\zeta \in S] = 1$, and $\mathbb{P}^{\mathcal{F}}[\zeta \in \emptyset] = 0$. Suppose for each $A \in \overline{S}$, we choose a representative, say ζ_A, of $\mathbb{P}^{\mathcal{F}}[\zeta \in A]$. Such ζ_A is an $\mathcal{F}$-measurable random variable. We may choose ζ_A such that $0 \le \zeta_A \le 1$, $\zeta_S \equiv 1$, and $\zeta_\emptyset \equiv 0$. Consider the map $\nu : \Omega \times \overline{S} \to [0,1]$ defined by

$$\nu(\omega, A) = \zeta_A(\omega).$$

We find that, for any $A \in \overline{S}$, $\nu(\cdot, A)$ is an $\mathcal{F}$-measurable random variable. On the other hand, by the linearity and monotone convergence property of conditional expectation, we have

$$\text{a.s. } \nu(\cdot, A) = \sum_{n=1}^\infty \nu(\cdot, A_n), \quad \text{if } A \text{ is a disjoint union of } A_1, A_2, \dots \in \overline{S}.$$

This means that there is an exceptional event N depending on $A_1, A_2, \dots$ with $\mathbb{P}N = 0$ such that

$$\mu(\omega, A) = \sum_{n=1}^\infty \nu(\omega, A_n), \quad \forall \omega \in \Omega \setminus N.. \tag{5.6}$$

Since there are uncountably many such sequences, in general, we may not be able to find a common exceptional null set, which is an $\mathcal{F}$-measurable set N with $\mathbb{P}N = 0$, such that (5.6) holds for any $A, A_1, A_2, \dots \in \mathcal{A}$ such that A is a disjoint union of $A_1, A_2, \dots$. However if such N does exist, we may modify the value of each ζ_A as follows. Pick $s_0 \in S$. For every $A \in \overline{S}$, we do not change the value of ζ_A on $\Omega \setminus N$, but for $\omega \in N$, we now define $\zeta_A(\omega) = \mathbf{1}_A(s_0) = \delta_{s_0} A$. Then the new ζ_A are still representatives of $\mathbb{P}^{\mathcal{F}}[\zeta \in A]$, and (5.6) holds true for all $\omega \in \Omega$. So we find that ν is a probability kernel from $(\Omega, \mathcal{F})$ to $(S, \overline{S})$.

Definition . Suppose ν is a probability kernel from $(\Omega, \mathcal{F})$ to $(S, \overline{S})$ and satisfies that for any $A \in \overline{S}$, a.s.

$$\mathbb{P}^{\mathcal{F}}[\zeta \in A] = \nu(\cdot, A).$$

Then ν is called a (regular) conditional distribution (or law) of ζ, given $\mathcal{F}$. When such ν exists, we write it as $\mathrm{Law}(\zeta|\mathcal{F})$ or $\mathrm{Law}^{\mathcal{F}}(\zeta)$.

A conditional law is convenient for us. Suppose further that η is a random element in another measurable space $(T, \overline{T})$. We may then consider conditional law of ζ given $\sigma(\eta)$. If such a conditional law ν exits, then it is a probability kernel from $(\Omega, \sigma(\eta))$ to $(S, \overline{S})$. Recall that, for any probability kernel μ from $(T, \overline{T})$ to $(S, \overline{S})$, $(\omega, A) \mapsto \mu(\eta(\omega), A)$ is a probability kernel from $(\Omega, \sigma(\eta))$ to $(S, \overline{S})$. It is desirable to have a probability kernel μ such that $\nu(\omega, A) = \mu(\eta(\omega), A)$. Then for any $A \in \overline{S}$, we have

$$\mathbb{P}^{\eta}[\zeta \in A] = \mu(\eta, A), \quad \text{a.s.} \tag{5.7}$$

When such μ exists, we then have the existence of $\mathrm{Law}(\zeta|\eta)$, which equals $\mu(\eta, \cdot)$. The following theorem concerns the existence of such kernel.

Theorem 5.3. *Let $(S, \overline{S})$ and $(T, \overline{T})$ be two measurable spaces, where S is a Borel space. Let ζ and η be two random elements in S and T, respectively. Then there is a probability kernel μ from $(T, \overline{T})$ to $(S, \overline{S})$ such that for any $A \in \overline{S}$, (5.7) holds. Moreover, such μ is* $\mathrm{Law}(\eta)$*-a.s. unique, which means that if another μ' satisfies the same property, then there exists $N \in \overline{T}$ with $\mathbb{P} \circ \eta^{-1} N = 0$ such that $\mu' \equiv \mu$ on $(T \setminus N) \times \overline{S}$.*

Corollary . *If ζ is a random element in a Borel space $(S, \overline{S})$, then for any sub-σ-algebra $\mathcal{F}$ of $\mathcal{A}$, the conditional law* $\mathrm{Law}(\zeta|\mathcal{F})$ *exists and is a.s. unique.*

Proof of the corollary. Take $(T, \overline{T}) = (\Omega, \mathcal{F})$. Let $\eta : \Omega \to \Omega$ be the identity. Since $\mathcal{F} \subset \mathcal{A}$, η is $\mathcal{A}/\mathcal{F}$-measurable. We have $\mathrm{Law}(\eta) = \mathbb{P}$ and $\sigma(\eta) = \mathcal{F}$. By Theorem 5.3, there is a probability kernel μ from $(\Omega, \mathcal{F})$ to $(S, \overline{S})$ such that for any $A \in \overline{S}$, a.s. $\mathbb{P}^{\mathcal{F}}[\zeta \in A] = \mathbb{P}^{\eta}[\zeta \in A] = \mu(\eta, A) = \mu(\cdot, A)$. So $\mu = \mathrm{Law}(\zeta|\mathcal{F})$. By Theorem 5.3, such μ is $\mathrm{Law}(\eta)$-a.s. unique. Since $\mathrm{Law}(\eta) = \mathbb{P}$, such μ is a.s. unique. □

Proof of Theorem 5.3. We may assume that $S \in \mathcal{B}(\mathbb{R})$. Then ζ is a random variable. For every $r \in \mathbb{Q}$, we consider a representative of $\mathbb{P}[\zeta \le r|\eta]$, which is an η-measurable random variable taking values in $[0, 1]$. By Lemma 1.13, for each $r \in \mathbb{Q}$, there is a random variable f_r defined on T such that

$$\text{a.s. } \mathbb{P}^{\eta}[\zeta \le r] = f_r(\eta). \tag{5.8}$$

For any $r_1, r_2 \in \mathbb{Q}$ with $r_1 < r_2$, by positivity we have a.s. $f_{r_1}(\eta) \le f_{r_2}(\eta)$, which implies that $\mathbb{P} \circ \eta^{-1}$-a.s. $f_{r_1} \le f_{r_2}$. By monotone convergence property, we have a.s. $\lim_{n \to +\infty} f_n(\eta) = 1$ and $\lim_{n \to +\infty} f_{-n}(\eta) = 0$. Thus, $\mathbb{P} \circ \eta^{-1}$-a.s. $\lim_{\mathbb{Z} \ni n \to +\infty} f_n = 1$ and $\lim_{\mathbb{Z} \ni n \to -\infty} f_n = 0$. Since there are at most countably many pairs (r_1, r_2) with $r_1, r_2 \in \mathbb{Q}$, we may find $N \in \overline{T}$ with $\mathbb{P}\eta^{-1}N = 0$ such that for $t \in T \setminus N$, $\mathbb{Q} \ni r \mapsto f_r(t)$ is increasing, and $\lim_{\mathbb{Z}\ni n\to+\infty} f_n(t) = 1$ and $\lim_{\mathbb{Z}\ni n\to-\infty} f_n(t) = 0$. Then we get $\lim_{\mathbb{Q}\ni r\to+\infty} f_r(t) = 1$ and $\lim_{\mathbb{Q}\ni r\to-\infty} f_r(t) = 0$ for $t \in T \setminus N$. We define a measurable function $F : T \times \mathbb{R} \to [0, 1]$ such that

$$F(t, x) = \begin{cases} \inf_{\mathbb{Q}\ni r>x} f_r(t), & t \in T \setminus N; \\ \mathbf{1}_{[0,\infty)}(x), & t \in N. \end{cases}$$

Then for every $t \in T$, $F(t, \cdot)$ is increasing and right continuous and satisfies $\lim_{x\to+\infty} F(t, x) = 1$ and $\lim_{x\to-\infty} F(t, x) = 0$, and so is a distribution of some probability measure $m(t, \cdot)$ on $\mathbb{R}$ (when $t \in N$, $m(t, \cdot) = \delta_0$ by the construction). From the measurability of F, we see that for any $x \in \mathbb{R}$, $t \mapsto m(t, (-\infty, x])$ is $\overline{T}$-measurable. Using a monotone class argument, we conclude that m is a probability kernel from T to $\mathbb{R}$.

By (5.8) and the monotone convergence property of conditional expectation, for any $x \in \mathbb{R}$, a.s.

$$m(\eta, (-\infty, x]) = F(\eta, x) = \inf_{\mathbb{Q}\ni r>x} \mathbb{P}^{\eta}[\zeta \le r] = \mathbb{P}^{\eta}[\zeta \in (-\infty, x]].$$

Using a monotone class argument based on the a.s. monotone convergence property, we may extend the last relation to

$$m(\eta, B) = \mathbb{P}^{\eta}[\zeta \in B] \text{ a.s.} \quad \forall B \in \mathcal{B}(\mathbb{R}). \tag{5.9}$$

In particular, we have a.s. $m(\eta, S^c) = 0$, i.e., $\mathbb{P} \circ \eta^{-1}$-a.s. $m(\cdot, S^c) = 0$. Taking $s_0 \in S$, (5.9) remains true if m is replaced by the kernel μ from T to S defined by

$$\mu(t, \cdot) = \begin{cases} m(t, \cdot), & \text{if } m(t, S^c) = 0; \\ \delta_{s_0}, & \text{if } m(t, S^c) > 0. \end{cases}$$

Such μ is what we need. If there is another probability kernel μ' from T to S with the stated property, then for any $r \in \mathbb{Q}$, a.s.

$$\mu(\eta, (-\infty, r]) = \mathbb{P}^{\eta}[\zeta \le r] = \mu'(\eta, (-\infty, r]).$$

Since $\mathbb{Q}$ is countable, we can exchange "for any $r \in \mathbb{Q}$" with "a.s.". A monotone class argument yields a.s. $\mu(\eta, \cdot) = \mu'(\eta, \cdot)$, and so $\mathbb{P} \circ \eta^{-1}$-a.s. $\mu = \mu'$. □

There are two trivial cases. If $\mathcal{F} = \{\Omega, \emptyset\}$, then a probability kernel from $(\Omega, \mathcal{F})$ to $(S, \overline{S})$ is just a probability measure on $(S, \overline{S})$. In this case, the conditional law $\mathrm{Law}(\zeta|\mathcal{F})$ agrees with $\mathrm{Law}(\zeta)$, which is often referred as the unconditional law of ζ. Another trivial case is $\mathcal{F} = \mathcal{A}$.

Exercise . Find the conditional law $\mathrm{Law}(\zeta|\mathcal{A})$.

Recall that if $f : S \to \mathbb{R}$ is measurable such that $\mathbb{E}|f(\zeta)| < \infty$, then

$$\mathbb{E}f(\zeta) = \int_S f(s) \,\mathrm{Law}(\zeta)(ds).$$

The following theorem extends this equality to conditional laws.

Theorem 5.4 (disintegration)**.** *Let ζ and η be random elements in measurable spaces $(S, \overline{S})$ and $(T, \overline{T})$, respectively. Let $\mathcal{F} \subset \mathcal{A}$ be a σ-field such that η is $\mathcal{F}$-measurable. Suppose* $\mathrm{Law}(\zeta|\mathcal{F})$ *exists. Let f be a measurable function on $S \times T$ such that $\mathbb{E}|f(\zeta, \eta)| < \infty$. Then for a.s. $\omega \in \Omega$, $s \mapsto f(s, \eta(\omega))$ is integrable w.r.t.* $\mathrm{Law}(\zeta|\mathcal{F})(\omega, \cdot)$*, $\omega \mapsto \int f(s, \eta(\omega)) \,\mathrm{Law}(\zeta|\mathcal{F})(\omega, ds)$ is $\mathcal{F}$-measurable, and equals $\mathbb{E}[f(\zeta, \eta)|\mathcal{F}](\omega)$. In short, this means a.s.*

$$\mathbb{E}[f(\zeta, \eta)|\mathcal{F}] = \int f(s, \eta) \,\mathrm{Law}(\zeta|\mathcal{F})(ds). \tag{5.10}$$

Integrating (5.10), we get the commonly used formula

$$\mathbb{E}[f(\zeta, \eta)] = \mathbb{E} \int f(s, \eta) \,\mathrm{Law}(\zeta|\mathcal{F})(ds). \tag{5.11}$$

When η disappears, (5.10 becomes $\mathbb{E}[f(\zeta)|\mathcal{F}] = \int f(s) \,\mathrm{Law}(\zeta|F)(ds)$.

Proof. We first prove (5.11). We write ν for $\mathrm{Law}(\zeta|\mathcal{F})$. First, suppose $f = \mathbf{1}_{B \times C}$, where $B \in \overline{S}$ and $C \in \overline{T}$. Then $\int f(s, \eta)\nu(ds) = \mathbf{1}_{\eta \in C} \nu B$ is $\mathcal{F}$-measurable because $\eta \in \mathcal{F}$. By $\eta \in \mathcal{F}$ and the pull-out property of conditional expectation,

$$\mathbb{E}[f(\zeta, \eta)] = \mathbb{E}[\mathbb{E}^{\mathcal{F}}[\mathbf{1}_{\zeta \in B} \mathbf{1}_{\eta \in C}]] = \mathbb{E}[\mathbf{1}_{\eta \in C} \mathbb{P}^{\mathcal{F}}[\zeta \in B]] = \mathbb{E}[\mathbf{1}_{\eta \in C} \nu B] = \mathbb{E} \int f(s, \eta)\nu(ds).$$

So we proved (5.11) for $f = \mathbf{1}_{B \times C}$. By a monotone class argument, we then conclude that, if f is an indicator function, then $\int f(s, \eta)\nu(ds)$ is $\mathcal{F}$-measurable and (5.11) holds. Using linearity and monotone convergence, we see that the measurability and (5.11) holds for any measurable function $f \geq 0$. In particular, if $\mathbb{E} f(\zeta, \eta) < \infty$, we find that a.s. $\int f(s, \eta)\nu(ds) < \infty$. So $s \mapsto f(s, \eta)$ is a.s. integrable w.r.t ν, and the measurability holds outside a null set on which $\int f(s, \eta)\nu(ds) = \infty$.

We now return to (5.10). Fix a measurable $f : S \times T \to \mathbb{R}_+$ with $\mathbb{E} f(\zeta, \eta) < \infty$, and let $A \in \mathcal{F}$. Then $\eta_A := (\eta, \mathbf{1}_A)$ is an $\mathcal{F}$-measurable random element in $T \times \{0, 1\}$. Note that $\mathbf{1}_A f(\zeta, \eta)$ can be expressed as $\widetilde{f}(\zeta, \eta_A)$ such that $\widetilde{f}(s, (t, 1)) = f(s, t)$ and $\widetilde{f}(s, (t, 0)) = 0$. Such $\widetilde{f}$ is $S \times (T \times \{0, 1\})$-measurable. Applying (5.11) with $T \times \{0, 1\}$ in place of T, η_A in place of η, and $\widetilde{f}$ in place of f, we get

$$\mathbb{E}[\mathbf{1}_A f(\zeta, \eta)] = \mathbb{E}[\widetilde{f}(\zeta, \eta_A)] = \mathbb{E} \int \widetilde{f}(s, \eta_A)\nu(ds) = \mathbb{E}[\mathbf{1}_A \int f(s, \eta)\nu(ds)].$$

Since $\int f(s, \eta)\nu(ds)$ is $\mathcal{F}$-measurable, we get (5.10) for $f \geq 0$. The general result follows by taking differences. $\square$

Remark . For two random elements ζ and η in T and S, respectively, if $\mathrm{Law}(\zeta|\eta)$ exists and is expressed by $\mu(\eta, \cdot)$ for a probability kernel μ from T to S, then we may recover the $\mathrm{Law}(\zeta, \eta)$ using μ and $\nu := \mathrm{Law}(\eta)$. For any $A \in \overline{S} \times \overline{T}$, applying (5.11) to $\mathcal{F} = \sigma(\eta)$ and $f = \mathbf{1}_A$, we get

$$\mathbb{P}[(\zeta, \eta) \in A] = \mathbb{E} \int_S \mathbf{1}_A(s, \eta)\mu(\eta, ds) = \int_T \nu(dt) \int_S \mathbf{1}_A(s, t)\mu(t, ds).$$

Thus, $\nu\otimes\mu$ as a probability measure on $T\times S$ is the law of (η,ζ). When $\zeta\perp\!\!\!\perp\eta$, μ is the constant $\mathrm{Law}(\zeta)$, and $\nu\otimes\mu$ is just the product measure $\mathrm{Law}(\eta)\times\mathrm{Law}(\zeta)$.

Example . Suppose ζ and η are two random variables such that the law of (ζ,η) is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^2$, and the Radon-Nikodym derivative is f. Define f_η on $\mathbb{R}$ by $f_\eta(y)=\int_{\mathbb{R}}f(x,y)dx\in[0,\infty]$. Then f_η is the density of the law of η against the Lebesgue measure on $\mathbb{R}$ because for any $B\in\mathcal{B}$,

$$\mathbb{P}[\eta\in B]=\mathbb{P}[(\zeta,\eta)\in\mathbb{R}\times B]=\int_B dy\int_{\mathbb{R}}dxf(x,y)=\int_B f_\eta(y)dy.$$

So $\mathrm{Law}(\eta)$-a.s. $f_\eta\in(0,\infty)$. Now we define a probability kernel μ from $\mathbb{R}$ to $\mathbb{R}$ such that for $y\in\mathbb{R}$ and $A\in\mathcal{B}$, if $f_\eta(y)\in(0,\infty)$, then

$$\mu(y,A)=\frac{1}{f_\eta(y)}\int_A f(x,y)dx;$$

and otherwise, $\mu(y,A)=\delta_0(A)$. This means that for $\mathrm{Law}(\eta)$-a.s. all y, $\mu(y,\cdot)$ has a density, which is $\frac{f(x,y)}{f_\eta(y)}$, w.r.t. the Lebesgue measure. The choice of $\mu(y,\cdot)$ when $f_\eta(y)\in\{0,\infty\}$ is not important. We claim that $\mathrm{Law}(\zeta|\eta)=\mu(\eta,\cdot)$. To see this, note that for any $A,B\in\mathcal{B}$, letting $B'=B\cap f_\eta^{-1}((0,\infty))$, we get

$$\mathbb{E}[\mathbf{1}_{\eta\in B}\mathbf{1}_{\zeta\in A}]=\mathbb{E}[\mathbf{1}_{\eta\in B'}\mathbf{1}_{\zeta\in A}]=\int_{B'}\int_A f(x,y)dxdy=\int_{B'}f_\eta(y)\int_A\frac{f(x,y)}{f_\eta(y)}dxdy$$

$$=\int_{B'}f_\eta(y)\mu(y,A)dy=\mathbb{E}[\mathbf{1}_{B'}\mu(\eta,A)]=\mathbb{E}[\mathbf{1}_B\mu(\eta,A)].$$

For a fixed $A\in\mathcal{B}$, since the above formula holds for any $B\in\mathcal{B}$, we get a.s. $\mathbb{P}[\zeta\in A|\eta]=\mu(\eta,A)$. Since this holds for any $A\in\mathcal{B}$, we get $\mathrm{Law}(\zeta|\eta)=\mu(\eta,\cdot)$.

Corollary (Jensen's inequality for conditional expectation)**.** *Let ζ be an integrable random variable. Let $\mathcal{F}\subset\mathcal{A}$ be a σ-algebra. Let $f:\mathbb{R}\to\mathbb{R}$ be convex such that $f(\zeta)$ is integrable. Then*

$$\mathbb{E}[f(\zeta)|\mathcal{F}]\ge f(\mathbb{E}[\zeta|\mathcal{F}]).$$

Proof. Applying (5.10) and using the unconditional Jensen's inequality, we get

$$\mathbb{E}[f(\zeta)|\mathcal{F}]=\int f(s)\,\mathrm{Law}(\zeta|\mathcal{F})(ds)\ge f\Big(\int s\,\mathrm{Law}(\zeta|\mathcal{F})(ds)\Big)=f(\mathbb{E}[\zeta|\mathcal{F}]).$$

□

Applying this Jensen's inequality to $f(x)=|x|^p$, $p\in(1,\infty)$, we see that for $\zeta\in L^p$, $\mathbb{E}[|\zeta|^p]\ge\mathbb{E}[|\mathbb{E}[\zeta|\mathcal{F}]|^p]$, and so we again get the L^p-contractivity $\|\mathbb{E}[\zeta|\mathcal{F}]\|_p\le\|\zeta\|_p$.

We now define conditional independence. For sub-σ-algebras $\mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n$ of $\mathcal{A}$, we say that $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are conditionally independent, given $\mathcal{G}$, if

$$\mathbb{P}^{\mathcal{G}}\Big[\bigcap_{k=1}^{n} B_k\Big] = \prod_{k=1}^{n} \mathbb{P}^{\mathcal{G}}[B_k] \text{ a.s.}, \quad B_k \in \mathcal{F}_k, \quad 1 \le k \le n.$$

If $(\mathcal{F}_t)_{t\in T}$ is an infinite family of sub-σ-algebras of $\mathcal{A}$, we say that they are conditionally independent, given $\mathcal{G}$, if the same property holds for every finite subcollection $\mathcal{F}_{t_1}, \ldots, \mathcal{F}_{t_n}$. Conditional independence involving events A_t or random elements ζ_t, $t \in T$, is defined as before in terms of the induced σ-algebras $\sigma(A_t)$ and $\sigma(\zeta_t)$. We use $\perp\!\!\!\perp_{\mathcal{G}}$ to denote pairwise conditional independence, given $\mathcal{G}$.

If ζ is $\mathcal{G}$-measurable, then for any $\mathbb{P}^{\mathcal{G}}[\zeta \in A] = \mathbf{1}_{\zeta\in A}$, and so ζ is conditionally independent of any $\mathcal{F} \subset \mathcal{A}$, given $\mathcal{G}$. If $\mathcal{F}_t$, $t \in T$, are all independent of $\mathcal{G}$, then for any $B \in \bigvee_{t\in T} \mathcal{F}_t$, $\mathbb{P}^{\mathcal{G}}[B] = \mathbb{P}[B]$, and so $\mathcal{F}_t$, $t \in T$, are conditionally independent, given $\mathcal{G}$, iff $\mathcal{F}_t$, $t \in T$, are unconditionally independent.

Proposition 5.6 (conditional independence, Doob)**.** *Let $\mathcal{F}, \mathcal{G}, \mathcal{H}$ be sub-σ-algebras of $\mathcal{A}$. Then $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$ iff*

$$\mathbb{P}[H|\mathcal{F}, \mathcal{G}] = \mathbb{P}[H|\mathcal{G}] \ \textit{a.s.}, \quad \forall H \in \mathcal{H}. \tag{5.12}$$

Proof. Assuming (5.12) and using the chain rule and pull-out properties, we get for any $F \in \mathcal{F}$ and $H \in \mathcal{H}$,

$$\begin{aligned}\mathbb{P}^{\mathcal{G}}[F \cap H] &= \mathbb{E}^{\mathcal{G}}[\mathbb{E}^{\mathcal{F}\vee\mathcal{G}}[\mathbf{1}_F\mathbf{1}_H]] = \mathbb{E}^{\mathcal{G}}[\mathbf{1}_F\mathbb{E}^{\mathcal{F}\vee\mathcal{G}}[\mathbf{1}_H]] \\ &= \mathbb{E}^{\mathcal{G}}[\mathbf{1}_F\mathbb{P}[H|\mathcal{F}, \mathcal{G}]] = \mathbb{E}^{\mathcal{G}}[\mathbf{1}_F\mathbb{P}[H|\mathcal{G}]] = \mathbb{P}^{\mathcal{G}}[H]\mathbb{P}^{\mathcal{G}}[F],\end{aligned}$$

which shows that $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$. Conversely, if $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$, then for any $F \in \mathcal{F}$, $G \in \mathcal{G}$, and $H \in \mathcal{H}$,

$$\mathbb{E}[\mathbf{1}_{F\cap G}\mathbb{P}^{\mathcal{G}}H] = \mathbb{E}[\mathbb{E}^{\mathcal{G}}[\mathbf{1}_F\mathbf{1}_G\mathbb{E}^{\mathcal{G}}[\mathbf{1}_H]]] = \mathbb{E}[\mathbf{1}_G\mathbb{E}^{\mathcal{G}}[\mathbf{1}_H]\mathbb{E}^{\mathcal{G}}[\mathbf{1}_F]]$$

$$= \mathbb{E}[\mathbf{1}_G\mathbb{P}^{\mathcal{G}}[H]\mathbb{P}^{\mathcal{G}}[F]] = \mathbb{E}[\mathbf{1}_G\mathbb{P}^{\mathcal{G}}[F \cap H]] = \mathbb{E}[\mathbb{P}^{\mathcal{G}}[G \cap F \cap H]] = \mathbb{P}[F \cap \mathcal{G} \cap H].$$

By a monotone class argument, we get that for any $A \in \mathcal{F} \vee \mathcal{G}$,

$$\mathbb{E}[\mathbf{1}_A\mathbb{P}^{\mathcal{G}}H] = \mathbb{P}[A \cap H].$$

Since $\mathbb{P}^{\mathcal{G}}H$ is $\mathcal{F} \vee \mathcal{G}$-measurable, we get (5.12). $\square$

From now on, for every sub-σ-algebra $\mathcal{F}$ of $\mathcal{A}$, we use $\overline{\mathcal{F}}$ to denote the completion of $\mathcal{F}$ w.r.t. $\mathbb{P}$.

Corollary 5.7. *Let $\mathcal{F}, \mathcal{G}, \mathcal{H}$ be sub-σ-algebras of $\mathcal{A}$. Then*

(i) $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$ iff $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} (\mathcal{G}, \mathcal{H})$;

(ii) $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}$ iff $\mathcal{F} \subset \overline{\mathcal{G}}$.

Proof. (i) By Proposition 5.6, both relations are equivalent to

$$\mathbb{P}[F|\mathcal{G},\mathcal{H}] = \mathbb{P}[F|\mathcal{G}] \text{ a.s.}, \quad \forall F \in \mathcal{F}.$$

(ii) If $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}$, then by Proposition 5.6, for any $F \in \mathcal{F}$,

$$\text{a.s. } \mathbf{1}_F = \mathbb{P}[F|\mathcal{F},\mathcal{G}] = \mathbb{P}[F|\mathcal{G}],$$

which implies that $F \in \overline{\mathcal{G}}$. So $\mathcal{F} \subset \overline{\mathcal{G}}$. On the other hand, if $\mathcal{F} \subset \overline{\mathcal{G}}$, then for any $F \in \mathcal{F}$,

$$\text{a.s. } \mathbb{P}[F|\mathcal{G}] = \mathbb{P}[F|\overline{\mathcal{G}}] = \mathbf{1}_F = \mathbb{P}[F|\mathcal{F},\overline{\mathcal{G}}].$$

Using Proposition 5.6 again, we get $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}$. □

Proposition 5.8 (chain rule)**.** *Let $\mathcal{G},\mathcal{H},\mathcal{F}_1,\mathcal{F}_2,\dots$ be sub-σ-algebras of $\mathcal{A}$. Then the following conditions are equivalent.*

(i) $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} (\mathcal{F}_1,\mathcal{F}_2,\dots)$*;*

(ii) $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G},\mathcal{F}_1,\dots,\mathcal{F}_n} \mathcal{F}_{n+1}$ *for all* $n \ge 0$*.*

Proof. If (i) holds, then for any $n \ge 0$, $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} (\mathcal{F}_1,\dots,\mathcal{F}_n)$. By Proposition 5.6, for any $H \in \mathcal{H}$ and $n \ge 0$, a.s.

$$\mathbb{P}[H|\mathcal{G},\mathcal{F}_1,\dots,\mathcal{F}_n] = \mathbb{P}[H|\mathcal{G}] = \mathbb{P}[H|\mathcal{G},\mathcal{F}_1,\dots,\mathcal{F}_n,\mathcal{F}_{n+1}],$$

which implies (ii) by Proposition 5.6.

Suppose (ii) holds. By Proposition 5.6, for any $H \in \mathbb{H}$ and $n \ge 0$, a.s.

$$\mathbb{P}[H|\mathcal{G},\mathcal{F}_1,\dots,\mathcal{F}_n] = \mathbb{P}[H|\mathcal{G},\mathcal{F}_1,\dots,\mathcal{F}_n,\mathcal{F}_{n+1}].$$

When $n = 0$, this means a.s. $\mathbb{P}[H|\mathcal{G}] = \mathbb{P}[H|\mathcal{G},\mathcal{F}_1]$. Thus, for any $m \ge 1$, a.s.

$$\mathbb{P}[H|\mathcal{G}] = \mathbb{P}[H|\mathcal{G},\mathcal{F}_1,\dots,\mathcal{F}_m].$$

So by Proposition 5.6,

$$\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} (\mathcal{F}_1,\dots,\mathcal{F}_m), \quad m \ge 1.$$

By a monotone class argument, we get (i). □

Remark . Taking $\mathcal{G} = \{\Omega,\emptyset\}$, we find that $\mathcal{H} \perp\!\!\!\perp (\mathcal{F}_1,\mathcal{F}_2,\dots)$ iff $\mathcal{H} \perp\!\!\!\perp_{\mathcal{F}_1,\dots,\mathcal{F}_n} \mathcal{F}_{n+1}$ for all $n \ge 0$.

Exercise . Do Problems 1, 2, 4, 5, 7, 8 in Chapter 5. Note that Problems 4,5 define $\mathbb{E}^{\mathcal{F}}\zeta$ for any $\overline{\mathbb{R}}_+$-valued random variable (may not be integrable); Problems 7,8 extend Fatou's lemma and dominated convergence theorem to conditional expectation.

6 Filtration and Stopping Times

Consider a measurable space $(\Omega, \mathcal{A})$. Let $T \subset \mathbb{R}$ be an index set. A filtration on T is an increasing family of σ-algebras $\mathcal{F}_t \subset \mathcal{A}$, $t \in T$. This means that $s < t \in T$ implies that $\mathcal{F}_s \subset \mathcal{F}_t$. We understand $\mathcal{F}_t$ as the knowledge at the time t with the memory of the past being kept. The increasingness of $\mathcal{F}_t$ reflects the arrow of time. From now on, we use $\mathcal{F}$ to denote a filtration rather than a σ-algebra. Let $(S, \overline{S})$ be a measurable space. An S-valued stochastic process X with index T is a family of measurable mappings X_t, $t \in T$, from Ω to S. It is called $\mathcal{F}$-adapted if X_t is $\mathcal{F}_t$-measurable for every $t \in T$. If we start with $X = (X_t)_{t\in T}$, and define $\mathcal{F}_t = \sigma(X_s : s \le t)$, $t \in T$, then $\mathcal{F} = (\mathcal{F}_t)$ is called the filtration induced by X. It is the smallest filtration to which X is adapted.

Given a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in T}$, a random variable τ taking values in $T \cup \{\sup T\}$ is called an $\mathcal{F}$-stopping time or $\mathcal{F}$-optional time if for any $t \in T$, $\{\tau \le t\} = \{\omega \in \Omega : \tau(\omega) \le t\} \in \mathcal{F}_t$. Intuitively, τ is a stopping time means that we can determine whether τ happens using only the knowledge of the past.

Exercise . Show that if T is countable, then τ is an $\mathcal{F}$-stopping time iff $\{\tau = t\} \in \mathcal{F}_t$, $\forall t \in T$.

Exercise . Show that the supremum of a sequence of $\mathcal{F}$-stopping times is an $\mathcal{F}$-stopping time, and the minimum of finitely many $\mathcal{F}$-stopping times is an $\mathcal{F}$-stopping time. We will see that the infimum of a sequence of $\mathcal{F}$-stopping times may not be an $\mathcal{F}$-stopping time.

Example . Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $\zeta_1, \zeta_2, \dots$ is an i.i.d. sequence of random variables with Bernoulli distribution $B(1/2)$. Let $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be the filtration generated by $\zeta = (\zeta_n)$. Let $X_n = \sum_{k=1}^n \zeta_k$, $n \in \mathbb{N}$. Then $X = (X_n)_{n\in\mathbb{N}}$ is an $\mathcal{F}$-adapted process. Let $N \in \mathbb{N}$. Let τ_N be the first n such that $X_n = N$; if such time does not exist, we set $\tau_N = \infty$. Then τ_N is an $\mathcal{F}$-stopping time because for any $n \in \mathbb{N}$,

$$\{\tau_N \le n\} = \bigcup_{k=1}^n \{X_k = N\} \in \mathcal{F}_n.$$

On the other hand, let σ_N be the last n such that $X_n = N$; and when such time does not exist, let $\sigma_N = \infty$. Then σ_N is not a stopping time because

$$\{\sigma_N = n\} = \{X_n = N\} \cap \{\zeta_{n+1} = 1\} \in \mathcal{F}_{n+1} \setminus \mathcal{F}_n.$$

Intuitively, σ_N is not a stopping time because we need future information to determine whether it happens.

For an $\mathcal{F}$-stopping time τ, we define

$$\mathcal{F}_\tau := \{A \in \mathcal{A} : A \cap \{\tau \le t\} \in \mathcal{F}_t, \forall t \in T\}.$$

It is easy to see that $\mathcal{F}_\tau$ is a σ-algebra. We understand $\mathcal{F}_\tau$ as the knowledge at the random time τ.

Exercise . Show that for any fixed $t_0 \in T$, the constant time $\tau \equiv t_0$ is an $\mathcal{F}$-stopping time, and the σ-algebra $\mathcal{F}_\tau$ associated with such τ agrees with the $\mathcal{F}_{t_0}$. Thus, τ and $\mathcal{F}_\tau$ naturally extend t and $\mathcal{F}_t$ for deterministic times $t \in T$.

Lemma 6.1. *For any $\mathcal{F}$-stopping times σ and τ, we have*

(i) τ *is* $\mathcal{F}_\tau$*-measurable;*

(ii) $\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau \subset \mathcal{F}_\tau$.

(iii) $\mathcal{F}_\sigma \cap \{\sigma = \tau\} = \mathcal{F}_\tau \cap \{\sigma = \tau\}$.

(iv) $\{\sigma < \tau\}, \{\sigma \le \tau\}, \{\sigma = \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.

(v) If $\sigma \le \tau$*, then* $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

Proof. Let $A \in \mathcal{F}_\sigma$. Then for any $t \in T$,

$$(A \cap \{\sigma \le \tau\}) \cap \{\tau \le t\} = (A \cap \{\sigma \le t\}) \cap \{\tau \le t\} \cap \{\sigma \wedge t \le \tau \wedge t\}.$$

Since $A \in \mathcal{F}_\sigma$, $A \cap \{\sigma \le t\} \in \mathcal{F}_t$. Since τ is an $\mathcal{F}$-stopping time, $\{\tau \le t\} \in \mathcal{F}_t$. Since $\sigma \wedge t$ takes values in $T \cap (-\infty, t]$, and for $s \in T \cap (-\infty, t]$, if $s < t$, $\{\sigma \wedge t \le s\} = \{\sigma \le s\} \in \mathcal{F}_s \subset \mathcal{F}_t$; and if $s = t$, $\{\sigma \wedge t \le s\} = \Omega \in \mathcal{F}_t$. So $\sigma \wedge t$ is $\mathcal{F}_t$-measurable. Similarly, $\tau \wedge t$ is $\mathcal{F}_t$-measurable. So $\{\sigma \wedge t \le \tau \wedge t\} \in \mathcal{F}_t$. Thus, $(A \cap \{\sigma \le \tau\}) \cap \{\tau \le t\} \in \mathcal{F}_t$. Since this holds for any $t \in T$, we get $A \cap \{\sigma \le \tau\} \in \mathcal{F}_\tau$. Thus,

$$\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_\tau. \tag{6.1}$$

If $\sigma \le \tau$, then $\{\sigma \le \tau\} = \Omega$, and we get $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$ from (6.1). So we proved (v). Since $\sigma \wedge \tau \le \sigma, \tau$, by (v) we get $\mathcal{F}_{\sigma\wedge\tau} \subset \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. On the other hand, if $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, then from

$$A \cap \{\sigma \wedge \tau \le t\} = (A \cap \{\sigma \le t\}) \cup (A \cap \{\tau \le t\}) \in \mathcal{F}_t, \quad t \in T,$$

we get $A \in \mathcal{F}_{\sigma\wedge\tau}$. So $\mathcal{F}_\sigma \cap \mathcal{F}_\tau \subset \mathcal{F}_{\sigma\wedge\tau}$. Thus, $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. From (6.1) we get $\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_\sigma \cap \mathcal{F}_\tau = \mathcal{F}_{\sigma\wedge\tau}$, which is (ii).

Taking $A = \Omega \in \mathcal{F}_\sigma$ in (6.1), we get $\{\sigma \le \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Swapping σ and τ, we get $\{\sigma < \tau\} = \{\tau \le \sigma\}^c \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Thus, $\{\sigma = \tau\} = \{\sigma \le \tau\} \setminus \{\sigma < \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. So we get (iv).

Since by (ii) $\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_\tau$, and by (iv) $\{\sigma = \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, we get $\mathcal{F}_\sigma \cap \{\sigma = \tau\} \subset \mathcal{F}_\tau \cap \{\sigma = \tau\}$. Swapping σ and τ, we get $\mathcal{F}_\tau \cap \{\sigma = \tau\} \subset \mathcal{F}_\sigma \cap \{\sigma = \tau\}$. So (iii) holds.

Finally, since τ takes values in $T \cup \{\sup T\}$, to prove (i) that τ is $\mathcal{F}_\tau$-measurable, it suffices to show that for any $t \in T$, $\{\tau \le t\} \in \mathcal{F}_\tau$. This follows from (iv) since any deterministic time t is an $\mathcal{F}$-stopping time. $\square$

Suppose now $T = \mathbb{R}_+ = [0, \infty)$. For a filtration $\mathcal{F}$ on $\mathbb{R}_+$, we define a new filtration $\mathcal{F}^+$ by $\mathcal{F}^+_t = \bigcap_{u>t} \mathcal{F}_u$, $t \ge 0$. We understand $\mathcal{F}^+_t$ as the knowledge at an infinitesimal time after t. It is clear that for $0 \le t < u$, $\mathcal{F}_t \subset \mathcal{F}^+_t \subset \mathcal{F}_u$. We may not have $\mathcal{F}_t = \mathcal{F}^+_t$.

Example . Let Ω be the space left-continuous $\mathbb{Z}_+$-valued increasing functions defined on $\mathbb{R}_+$ with initial value 0. For $t \ge 0$, let $\pi_t : \Omega \to \mathbb{Z}_+$ be the map $\omega \mapsto \omega(t)$. Let $\mathcal{F} = (\mathcal{F}_t)_{t \ge 0}$ be the filtration such that $\mathcal{F}_t = \sigma(\pi_s : 0 \le s \le t)$. Fix $t_0 \ge 0$. Let A_{t_0} denote the set of $\omega \in \Omega$ which are continuous at t_0. Then $A_{t_0} \in \mathcal{F}^+_{t_0} \setminus \mathcal{F}_{t_0}$. In fact, $A_{t_0} = \{\pi_{t_0^+} = \pi_{t_0}\}$, where $\pi_{t_0^+}(\omega) := \lim_{t \downarrow t_0} \omega(t)$. For any $t_0 \ge 0$, $\pi_{t_0} \in \mathcal{F}_{t_0} \subset \mathcal{F}^+_{t_0}$. For any $u > t_0$, we may pick a sequence (t_n) in $(t_0, u]$ with $t_n \downarrow t_0$. Then $\pi^+_{t_0} = \lim_n \pi_{t_n} \in \mathcal{F}_u$. Since this holds for any $u > t_0$, $\pi^+_{t_0} \in \bigcap_{u > t_0} \mathcal{F}_u = \mathcal{F}^+_{t_0}$. Thus, $A_{t_0} = \{\pi^+_{t_0} = \pi_{t_0}\} \in \mathcal{F}^+_{t_0}$.

Next, we show that $A_{t_0} \notin \mathcal{F}_{t_0}$. We define an equivalence relation "$\cong_{t_0}$" on Ω such that $\omega_1 \cong_{t_0} \omega_2$ iff ω_1 and ω_2 agree on $[0, t_0]$. Let $\mathcal{G}_{t_0}$ be the family of all subsets of Ω which are unions of the equivalence classes w.r.t. $\cong_{t_0}$. Then $\mathcal{G}_{t_0}$ is a σ-algebra, and $\pi_t \in \mathcal{G}_{t_0}$ for $0 \le t \le t_0$,. Thus, $\mathcal{F}_{t_0} \subset \mathcal{G}_{t_0}$. We see that $A_{t_0} \notin \mathcal{G}_{t_0}$ because for any $\omega_1 \in A_{t_0}$, we may define $\omega_2 \in \Omega \setminus A_{t_0}$ by $\omega_2(t) = \omega_1(t)$ for $0 \le t \le t_0$ and $\omega_2(t) = \omega_1(t) + 1$ for $t > t_0$. Thus, $A_{t_0} \notin \mathcal{F}_{t_0}$.

We say that $\mathcal{F}$ is right-continuous if $\mathcal{F}^+ = \mathcal{F}$. This means that the knowledge at time t is the same as the knowledge at the time $t + o(1)$. In particular, $\mathcal{F}^+$ is right-continuous because

$$(\mathcal{F}^+)^+_t = \bigcap_{u>t} \mathcal{F}^+_u = \bigcap_{u>t} \bigcap_{v>u} \mathcal{F}_v = \bigcap_{v>t} \mathcal{F}_v = \mathcal{F}^+_t.$$

We call $\mathcal{F}^+$ the right-continuation of $\mathcal{F}$. A random time $\tau : \Omega \to [0, \infty]$ is called a weak $\mathcal{F}$-stopping time if for any $t > 0$, $\{\tau < t\} \in \mathcal{F}_t$. In this case, for any $h > 0$, $\tau + h$ is an $\mathcal{F}$-stopping time because for any $t \ge 0$, when $t < h$, $\{\tau + h \le t\} = \emptyset \in \mathcal{F}_t$, and when $t \ge h$, we may take a sequence (t_n) in $(t - h, t)$ with $t_n \downarrow t - h$, and get

$$\{\tau + h \le t\} = \{\tau \le t - h\} = \bigcap_n \{\tau < t_n\} \in \mathcal{F}_t,$$

where the last relation holds because $\{\tau < t_n\} \in \mathcal{F}_{t_n} \subset \mathcal{F}_t$ as $t_n < t$. So for each $h > 0$, we may define a σ-algebra $\mathcal{F}_{\tau+h}$. If $0 < h_1 < h_2$, from $\tau + h_1 < \tau + h_2$ we get $\mathcal{F}_{\tau+h_1} \subset \mathcal{F}_{\tau+h_2}$. We now define $\mathcal{F}_{\tau+} = \bigcap_{h>0} \mathcal{F}_{\tau+h}$, which is also a sub-$\sigma$-algebra of $\mathcal{A}$.

Lemma 6.2. *A random time τ is a weak $\mathcal{F}$-stopping time iff it is an $\mathcal{F}^+$-stopping time, in which case*

$$\mathcal{F}_{\tau+} = \mathcal{F}^+_\tau = \{A \in \mathcal{A} : A \cap \{\tau < t\} \in \mathcal{F}_t, \forall t > 0\}.$$

Proof. For any $t \ge 0$ and $u > t$, we note that

$$\{\tau \le t\} = \bigcap_{r \in \mathbb{Q}_+ \cap (t,u]} \{\tau < r\}, \quad \{\tau < t\} = \bigcup_{r \in \mathbb{Q}_+ \cap (0,t)} \{\tau \le r\}.$$

If $A \cap \{\tau \le r\} \in \mathcal{F}^+_r$ for all $r \ge 0$, then for $t > 0$,

$$A \cap \{\tau < t\} = \bigcup_{r \in \mathbb{Q}_+ \cap (0,t)} (A \cap \{\tau \le r\}) \in \mathcal{F}_t$$

because for $r < t$, $\mathcal{F}_r^+ \subset \mathcal{F}_t$. On the other hand, if $A \cap \{\tau < r\} \in \mathcal{F}_r$ for all $r > 0$, then for $t \geq 0$ and $u > t$,

$$A \cap \{\tau \leq t\} = \bigcap_{r \in \mathbb{Q}_+ \cap (t,u]} (A \cap \{\tau < r\}) \in \mathcal{F}_u.$$

Since this holds for any $u > t$, we get $A \cap \{\tau \leq t\} \in \mathcal{F}_t^+$. So we have proved the first assertion by taking $A = \Omega$. For general $A \in \mathcal{A}$, this shows that $\mathcal{F}_\tau^+ = \{A \in \mathcal{A} : A \cap \{\tau < t\} \in \mathcal{F}_t, \forall t > 0\}$.

By the definition of $\mathcal{F}_{\tau+}$, $A \in \mathcal{F}_{\tau+}$ iff $A \in \mathcal{F}_{\tau+h}$ for each $h > 0$, i.e., $A \cap \{\tau + h \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$ and $h > 0$. Since $\{\tau + h \leq t\} = \emptyset$ when $h > t$, the above relation is further equivalent to that $A \cap \{\tau \leq t - h\} \in \mathcal{F}_t$ for any $t \geq h > 0$, which by a change of variable ($s = t - h$) is equivalent to $A \cap \{\tau \leq s\} \in \mathcal{F}_{s+h}$ for any $s \geq 0$ and $h > 0$, and hence to $A \cap \{\tau \leq s\} \in \mathcal{F}_s^+$ for all $s \geq 0$, i.e., $A \in \mathcal{F}_\tau^+$. Thus, $\mathcal{F}_{\tau+} = \mathcal{F}_\tau^+$. □

Note that if $\mathcal{F}$ is right-continuous, then a weak $\mathcal{F}$-stopping time is an $\mathcal{F}$-stopping time, and there is no difference between $\mathcal{F}_{\tau+}$ and $\mathcal{F}_\tau$. Intuitively, τ is a weak $\mathcal{F}$-stopping time means that we can determine that τ happens using the information of the past and a tiny bit of future.

Lemma 6.3. *Let $\tau_1, \tau_2, \dots$ be weak $\mathcal{F}$-stopping times. Then $\tau := \inf\{\tau_n\}$ is also a weak $\mathcal{F}$-stopping time, and $\mathcal{F}_{\tau+} = \bigcap_n \mathcal{F}_{\tau_n+}$.*

Proof. We see that for any $t > 0$ and $A \in \mathcal{A}$,

$$A \cap \{\tau < t\} = A \cap \bigcup_n \{\tau_n < t\} = \bigcup_n (A \cap \{\tau_n < t\}). \tag{6.2}$$

Taking $A = \Omega$, we see that τ is a weak $\mathcal{F}$-stopping time. By Lemma 6.1, $\mathcal{F}_{\tau+} \subset \bigcap_n \mathcal{F}_{\tau_n+}$. If $A \in \bigcap_n \mathcal{F}_{\tau_n+}$, then by Lemma 6.2, $A \cap \{\tau_n < t\} \in \mathcal{F}_t$ for each n, and so by (6.2), $A \cap \{\tau < t\} \in \mathcal{F}_t$, which implies that $A \in \mathcal{F}_{\tau+}$. So we get $\mathcal{F}_{\tau+} = \bigcap_n \mathcal{F}_{\tau_n+}$. □

Note that if $\mathcal{F}$ is right-continuous, this lemma tells us that the infimum of a sequence of $\mathcal{F}$-stopping times is an $\mathcal{F}$-stopping time. This is not true in general. The lemma below shows another reason that a right-continuous filtration is useful.

If $T = \mathbb{R}_+$ or $\mathbb{Z}_+$, for a set $B \subset S$, we may define the hitting time

$$\tau_B = \inf\{t \in T, t > 0 : X_t \in B\}.$$

As usual, we set $\inf \emptyset = \infty$ by convention. The following result helps us to decide whether τ_B is a stopping time.

Lemma 6.6. *Fix a filtration $\mathcal{F}$ on $T = \mathbb{R}_+$ or $\mathbb{Z}_+$, let X be an $\mathcal{F}$-adapted process on T with values in a measurable space S, and let $B \subset S$. Then we have the following*

(i) If $T = \mathbb{Z}_+$ and B is measurable, τ_B is an $\mathcal{F}$-stopping time.

(ii) If $T = \mathbb{R}_+$, S is a metric space, B is closed, and X is continuous, then τ_B is a weak $\mathcal{F}$-stopping time.

(iii) If $T = \mathbb{R}_+$, S is a topological space, B is open, and X is right- or left- continuous, then τ_B is a weak $\mathcal{F}$-stopping time.

In particular, in (ii) and (iii), if $\mathcal{F}$ is right-continuous, then τ_B is an $\mathcal{F}$-stopping time.

Proof. (i) For any $n \in \mathbb{Z}_+$,

$$\{\tau_B \le n\} = \bigcup_{k=1}^{n} \{X_k \in B\} \in \mathcal{F}_n$$

since for every $k \le n$, $\{X_k \in B\} \in \mathcal{F}_k \subset \mathcal{F}_n$. So τ_B is an $\mathcal{F}$-stopping time.

Suppose now $T = \mathbb{R}_+$. Let $t_0 > 0$. By the definition of τ_B,

$$\{\tau_B < t_0\} = \bigcup_{0<t<t_0} \{X_t \in B\} = \bigcup_{n\in\mathbb{N}} \bigcup_{t\in[\frac{t_0}{n},(1-\frac{1}{n})t_0]} \{X_t \in B\}.$$

(ii) If S is a metric space, B is closed and X is continuous, then for any $n \in \mathbb{N}$,

$$\bigcup_{t\in[\frac{t_0}{n},(1-\frac{1}{n})t_0]} \{X_t \in B\} = \bigcap_{m\in\mathbb{N}} \bigcup_{t\in[\frac{t_0}{n},(1-\frac{1}{n})t_0]} \{\rho(X_t, B) < \frac{1}{m}\}$$

$$= \bigcap_{m\in\mathbb{N}} \bigcup_{r\in\mathbb{Q}\cap[\frac{t_0}{n},(1-\frac{1}{n})t_0]} \{\rho(X_r, B) < \frac{1}{m}\}.$$

Thus, for any $t_0 > 0$,

$$\{\tau_B < t_0\} = \bigcup_{n\in\mathbb{N}} \bigcap_{m\in\mathbb{N}} \bigcup_{r\in\mathbb{Q}\cap[\frac{t_0}{n},(1-\frac{1}{n})t_0]} \{\rho(X_r, B) < \frac{1}{m}\}.$$

Since for any $n, m \in \mathbb{N}$ and $r \in \mathbb{Q} \cap [\frac{t_0}{n}, (1-\frac{1}{n})t_0]$, $\{\rho(X_r, B) < \frac{1}{m}\} \in \mathcal{F}_r \subset \mathcal{F}_{t_0}$, and the above formula involves only countable union and countable intersection, we set $\{\tau_B < t_0\} \in \mathcal{F}_{t_0}$.

(iii) If B is open and X is right-continuous or left-continuous, then for any $t_0 > 0$,

$$\{\tau_B < t_0\} = \bigcup_{t\in(0,t_0)} \{X_t \in B\} = \bigcup_{r\in\mathbb{Q}\cap(0,t_0)} \{X_r \in B\} \in \mathcal{F}_{t_0}.$$

So we again conclude that τ_B is a weak $\mathcal{F}$-stopping time. □

Remark . If we now define $\tau_B = \inf\{t \ge 0 : X_t \in B\}$, the above theorem still holds. If $T = \mathbb{Z}_+$ and σ is an $\mathcal{F}$-stopping time, then $\tau := \inf\{t \ge \sigma : X_t \in B\}$ is a stopping time. For the latter statement, we note that for any $u \in \mathbb{Z}_+$,

$$\{\tau \le u\} = \bigcup_{0\le t\le u} \{\sigma \le t\} \cap \{X_t \in B\} \in \mathcal{F}_u.$$

Lemma 6.4 (discrete approximation)**.** *For any weak $\mathcal{F}$ stopping time τ, there exists a sequence of countably valued $\mathcal{F}$-stopping times (τ_n) with $\tau_n \downarrow \tau$.*

Proof. Let $\tau_n = 2^{-n}\lceil 2^n\tau + 1\rceil$. This means that if $\frac{k}{2^n} \le \tau < \frac{k+1}{2^n}$ for some $k \in \mathbb{Z}_{\ge 0}$, then $\tau_n = \frac{k+1}{2^n}$. Then τ_n takes values in $2^{-n}\mathbb{N}$ and $\tau_n \downarrow \tau$. To see that each τ_n is an $\mathcal{F}$-stopping time, we note that for any $t \ge 0$, there is $k_0 \in \mathbb{Z}_{\ge 0}$ such that $\frac{k_0}{2^n} \le t < \frac{k_0+1}{2^n}$. Then $\tau_n \le t$ iff $\tau_n \le \frac{k_0}{2^n}$ iff $\tau \in [\frac{k}{2^n}, \frac{k+1}{2^n})$ for some $k \in \mathbb{Z}$ with $k+1 \le k_0$, which is equivalent to that $\tau < \frac{k_0}{2^n}$. Since τ is a weak $\mathcal{F}$-stopping time, we get $\{\tau_n \le t\} = \{\tau < \frac{k_0}{2^n}\} \in \mathcal{F}_{\frac{k_0}{2^n}} \subset \mathcal{F}_t$. □

The definition of $\mathcal{F}$-adaptedness does not imply the joint measurability $(t, \omega) \mapsto X_t(\omega)$. Now we introduce a stronger concept.

Definition . Let $\mathcal{F}$ be a filtration on $\mathbb{R}_+$. An S-valued process X on $\mathbb{R}_+$ is called $\mathcal{F}$-progressively measurable or simply progressive if for any $t_0 \in \mathbb{R}_+$, the map

$$\Omega \times [0, t_0] \ni (\omega, t) \mapsto X_t(\omega) \in S$$

is $\mathcal{F}_{t_0} \times \mathcal{B}[0, t_0]$-measurable. A set $A \in \Omega \times \mathbb{R}_+$ is called $\mathcal{F}$-progressive if $\mathbf{1}_A$ is $\mathcal{F}$-progressive.

Exercise . Show that (i) an $\mathcal{F}$-progressive process is $\mathcal{F}$-adapted; (ii) the class of all $\mathcal{F}$-progressive sets form a σ-algebra, denoted by $\mathcal{P}$; and (iii) a stochastic process X on $\mathbb{R}_+$ is $\mathcal{F}$-progressive iff it is measurable w.r.t. $\mathcal{P}$.

Lemma . *A left- or right-continuous adapted process is progressive.*

Proof. Let X be a left- or right-continuous adapted process. We need to show that for any $t_0 \ge 0$, $(\omega, t) \mapsto X_t(\omega)$ is $\mathcal{F}_{t_0} \times \mathcal{B}[0, t_0]$-measurable. Let $t_0 \ge 0$. It suffices to construct a sequence of functions $X^n : \Omega \times [0, t_0] \to S$ such that each X^n is $\mathcal{F}_{t_0} \times \mathcal{B}[0, t_0]$-measurable, and $X^n \to X$ pointwise on $\Omega \times [0, t_0]$. If X is left-continuous, we define $X^n(\omega, t) = X(\omega, \frac{k}{2^n} t_0)$ if $\frac{k}{2^n} t_0 \le t < \frac{k+1}{2^n} t_0$ for some $k \in \mathbb{Z}$. If X is right-continuous, we define $X^n(\omega, t) = X(\omega, \frac{k+1}{2^n} t_0)$ if $\frac{k}{2^n} t_0 \le t < \frac{k+1}{2^n} t_0$ for some $k \in \mathbb{Z}$ with $k < 2^n$; and $X^n(\omega, t_0) = X(\omega, t_0)$. From the adaptedness of X, we see that in both cases, X^n is $\mathcal{F}_{t_0} \times \mathcal{B}[0, t_0]$-measurable. The pointwise convergence of $X^n \to X$ follows from the left- or right-continuity of X. □

It is useful to have a progressive process for the following reasons.

Lemma 6.5. *Fix a filtration with index T. Let τ be a T-valued $\mathcal{F}$-stopping time. Let X be an $\mathcal{F}$-adapted process on T with values in a measurable space $(S, \overline{S})$. Then $X_\tau : \omega \mapsto X_{\tau(\omega)}(\omega)$ is $\mathcal{F}_\tau$-measurable in the following two cases.*

(i) T is countable;

(ii) $T = \mathbb{R}_+$ and X is progressive.

Proof. To prove that X_τ is $\mathcal{F}_\tau$-measurable, we need to show that for any $B \in \mathcal{B}(S)$ and $t_0 \in T$, $\{X_\tau \in B\} \cap \{\tau \le t_0\} \in \mathcal{F}_{t_0}$. Note that

$$\{X_\tau \in B\} \cap \{\tau \le t_0\} = \{X_{\tau\wedge t_0} \in B\} \cap \{\tau \le t_0\};$$

$$\{X_{\tau\wedge t_0} \in B\} = (\{X_\tau \in B\} \cap \{\tau \le t_0\}) \cup (\{X_{t_0} \in B\} \cap \{t_0 \le \tau\}).$$

Since X is $\mathcal{F}$-adapted, $\{X_\tau \in B\} \cap \{\tau \le t_0\} \in \mathcal{F}_{t_0}$ iff $\{X_{\tau\wedge t_0} \in B\} \in \mathcal{F}_{t_0}$. Note that $\tau \wedge t_0$ is an $\mathcal{F}$-stopping time bounded above by t_0. So it suffices to show that for any $\mathcal{F}$-stopping time σ bounded above by t_0, X_σ is $\mathcal{F}_{t_0}$-measurable.

(i) Since σ takes values in $\{t \in T : t \le t_0\}$, we have

$$\{X_\sigma \in B\} = \bigcup_{t\in T: t\le t_0} (\{X_t \in B\} \cap \{\sigma = t\}) \in \mathcal{F}_{t_0}$$

because T is countable, and for $t \in T$ with $t \le t_0$, $\{X_t \in B\}, \{\sigma = t\} \in \mathcal{F}_t \subset \mathcal{F}_{t_0}$.

(ii) Now σ takes values in $[0, t_0]$. The we write $X_\sigma = X^{t_0} \circ \psi$, where X^{t_0} is the restriction of X to $\Omega \times [0, t_0]$, and $\psi : \Omega \to \Omega \times [0, t_0]$ is given by $\omega \mapsto (\omega, \sigma(\omega))$. Since X is progressive, X^{t_0} is $\mathcal{F}_{t_0} \times \mathcal{B}[0, t_0]$-measurable. In order to show that X_σ is $\mathcal{F}_{t_0}$-measurable, it suffices to show that ψ is $\mathcal{F}_{t_0}/(\mathcal{F}_{t_0} \times \mathcal{B}[0, t_0])$-measurable. This holds because for any $B \in \mathcal{F}_{t_0}$ and $t \in [0, t_0]$, $\psi^{-1}(B \times [0, t]) = B \cap \{\sigma \le t\} \in \mathcal{F}_{t_0}$. □

Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{A})$ and we work on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For any σ-algebra $\mathcal{G} \subset \mathcal{A}$, we use $\overline{\mathcal{G}}$ to denote the completion of $\mathcal{G}$, and say that $\mathcal{G}$ is complete if $\mathcal{G} = \overline{\mathcal{G}}$. A filtration $(\mathcal{F}_t)$ is called complete if every $\mathcal{F}_t$ is complete. Given any filtration $\mathcal{F} = (\mathcal{F}_t)$, its completion is the filtration $(\overline{\mathcal{F}}_t)$. Suppose now $T = \mathbb{R}_+$ and $\mathcal{F}$ is a filtration on $\mathbb{R}_+$. We get two filtration extensions of $\mathcal{F}$: one is its completion $(\overline{\mathcal{F}}_t)$, the other is its right-continuation $(\mathcal{F}_t^+)$. The following lemma tells us that the right-continuation of the completion agrees with the completion of the right-continuation.

Lemma 6.8. *For any filtration $\mathcal{F}$ on $\mathbb{R}_+$, we have*

$$\overline{\mathcal{F}_t^+} = \overline{\mathcal{F}}_t^+, \quad \forall t \ge 0.$$

Proof. Since $\mathcal{F}_t \subset \overline{\mathcal{F}}_t$ for all $t \ge 0$, we have $\mathcal{F}_t^+ = \bigcap_{u>t} \mathcal{F}_u \subset \bigcap_{u>t} \overline{\mathcal{F}}_u = \overline{\mathcal{F}}_t^+$ for all $t \ge 0$. Since every $\overline{\mathcal{F}}_t$ is complete, every $\overline{\mathcal{F}}_t^+$ is also complete. So $\overline{\mathcal{F}_t^+} \subset \overline{\mathcal{F}}_t^+$, $t \ge 0$.

We now prove the opposite direction. Let $A \in \overline{\mathcal{F}}_t^+$ for some $t \ge 0$. Then $A \in \overline{\mathcal{F}_u}$ for every $u > t$. By Lemma 1.25, for each $u > t$, there is $A_u \in \mathcal{F}_u$ such that $\mathbb{P}[A \Delta A_u] = 0$. Choose $u_n \downarrow t$ and define $A' = \limsup A_{u_n} \in \mathcal{F}_t^+$. Then $\mathbb{P}[A \Delta A'] \le \sum_n \mathbb{P}[A \Delta A_{u_n}] = 0$. So $A \in \overline{\mathcal{F}_t^+}$. Thus, $\overline{\mathcal{F}}_t^+ \subset \overline{\mathcal{F}_t^+}$. □

The common filtration $(\overline{\mathcal{F}_t^+}) = (\overline{\mathcal{F}}_t^+)$ is both complete and right-continuous, and is called the (usual) augmentation of $\mathcal{F}$.

Exercise . Do problems 2 and 4 in Chapter 6.

7 Martingales

Definition . Let $\mathcal{F}$ be a filtration with index set $T \subset \mathbb{R}$. Let $X = (X_t)_{t\in T}$ be an $\mathcal{F}$-adapted process of integrable random variables. If for any $s, t \in T$ with $s \le t$, we have

$$X_s = \mathbb{E}[X_t|\mathcal{F}_s] \quad a.s., \tag{7.1}$$

then we say that X is an $\mathcal{F}$-martingale. If (7.1) holds with "$\le$" (resp. "$\ge$") in place of "$=$" for all $s \le t \in T$, then X is called an $\mathcal{F}$-submartingale (resp. $\mathcal{F}$-supermartingale). If $X = (X^1, \dots, X^d)$ is a process on T in $\mathbb{R}^d$, we say that X is an $\mathcal{F}$-vector martingale if for every $1 \le k \le d$, X^k is an $\mathcal{F}$-martingale.

Facts: X is an $\mathcal{F}$-martingale iff it is both an $\mathcal{F}$-submartingale and an $\mathcal{F}$-supermartingale; X is an $\mathcal{F}$-supermartingale iff $-X$ is an $\mathcal{F}$-submartingale; and a linear combination of $\mathcal{F}$-martingales is also an $\mathcal{F}$-martingale. We have some freedom to choose the filtration.

Exercise . Prove that if X is an $\mathcal{F}$-martingale (resp. supermartingale or submartingale), then it is also a martingale (resp. supermartingale or submartingale) w.r.t. (i) the completion of $\mathcal{F}$; (ii) the filtration induced by X.

Example . Let the filtration $\mathcal{F}$ be given. Let ζ be an integrable random variable. Let $X_t = \mathbb{E}[\zeta|\mathcal{F}_t]$, $t \in T$. Then X is an $\mathcal{F}$-martingale because for any $s \le t \in T$, by chain rule,

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[\zeta|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[\zeta|\mathcal{F}_s] = X_s.$$

By Lemma 5.5, X is uniformly integrable, and so is L^1-bounded.

For a process X on $\mathbb{Z}_+$, we define $\Delta X_n = X_n - X_{n-1}$, $n \in \mathbb{N}$.

Exercise . For an $\mathcal{F}$-adapted process X on $\mathbb{Z}_+$, prove that X is an $\mathcal{F}$-martingale (resp. supermartingale or submartingale) iff a.s. $\mathbb{E}[\Delta X_n|\mathcal{F}_{n-1}] = 0$ (resp. ≥ 0 or ≤ 0) for all $n \in \mathbb{N}$.

Example . Let $\zeta_1, \zeta_2, \dots$ be a sequence of independent integrable random variables. For $n \in \mathbb{Z}_+$, let $X_n = \sum_{k=1}^n \zeta_k$ and $\mathcal{F}_n = \sigma(\zeta_k : 1 \le k \le n)$. Then $\mathcal{F} = (\mathcal{F}_n)$ is a filtration, and $X = (X_n)$ is $\mathcal{F}$-adapted. For $n \in \mathbb{N}$, since $\Delta X_n = \zeta_n \perp\!\!\!\perp \mathcal{F}_{n-1}$, we get a,s. $\mathbb{E}[\Delta X_n|\mathcal{F}_{n-1}] = \mathbb{E}[\zeta_n]$. Thus, X is a martingale (resp. submartingale or supermartingale) if $\mathbb{E}\zeta_n = 0$ (resp. ≥ 0 or ≤ 0) for all $n \in \mathbb{N}$. If $\mathrm{Law}(\zeta_n) = \frac{1}{2}(\delta_1 + \delta_{-1})$, X is a random walk on $\mathbb{Z}$.

A martingale on $\mathbb{Z}_+$ may be thought of as a gambler's balance history, who always plays fair games.

Definition . For a filtration $\mathcal{F}$ on $\mathbb{Z}_+$, a process $A = (A_n)_{n\ge 0}$ is called $\mathcal{F}$-predictable if $A_0 \equiv 0$, and for $n \in \mathbb{N}$, $A_n \in \mathcal{F}_{n-1}$.

We use this name because we know the value of A_n at the time $n-1$. Note that a predictable process must be adapted.

Lemma 6.10. *For a filtration $\mathcal{F}$ on $\mathbb{Z}_+$, every $\mathcal{F}$-predictable process X can be expressed as the sum $M+A$, where M is an $\mathcal{F}$-martingale and A is $\mathcal{F}$-predictable, and such decomposition is a.s. unique. Moreover X is a submartingale (resp. supermartingale) iff the A in the decomposition is a.s. increasing (resp. decreasing).*

The decomposition $X = M + A$ is called the Doob's decomposition.

Proof. Define the process A by

$$A_n = \sum_{k=1}^{n} \mathbb{E}[\Delta X_k | \mathcal{F}_{k-1}], \quad n \geq 0.$$

Then $A_0 = 0$ and for $n \geq 1$, A_n is $\mathcal{F}_{n-1}$-adapted and $\Delta A_n = \mathbb{E}[\Delta X_n|\mathcal{F}_{n-1}]$. So A is $\mathcal{F}$-predictable. Let $M = X - A$. Then M is also an $\mathcal{F}$-adapted process, and for $n \geq 1$, a.s.

$$\mathbb{E}[\Delta M_n|\mathcal{F}_{n-1}] = \mathbb{E}[\Delta X_n - \Delta A_n|\mathcal{F}_{n-1}] = \mathbb{E}[\Delta X_n|\mathcal{F}_{n-1}] - \Delta A_n = 0.$$

So M is an $\mathcal{F}$-martingale. So we get the existence of Doob's decomposition. Suppose there is another such decomposition $M' + A'$, then $Y := M - M' = A' - A$ is both $\mathcal{F}$-martingale and $\mathcal{F}$-predictable, and has the initial value $Y_0 = 0$. So for any $n \in \mathbb{N}$, a.s. $Y_n = \mathbb{E}[Y_n|\mathcal{F}_{n-1}] = Y_{n-1}$. We then get a.s. $Y_n = 0$ for all $n \in \mathbb{N}$. So we get the a.s. uniqueness of Doob's decomposition. Moreover, X is a submartingale iff a.s. $\mathbb{E}[\Delta X_n|\mathcal{F}_{n-1}] = \Delta A_n \geq 0$ for each $n \geq 1$, which is equivalent to that a.s. A_n is increasing. Similarly, X is a supermartingale iff a.s. A_n is decreasing. $\square$

Lemma 6.11. *Let M be a martingale in $\mathbb{R}^d$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Suppose $X_t = f(M_t)$ is integrable for every t. Then X is a submartingale. The statement remains true if M is a submartingale, and $f : \mathbb{R} \to \mathbb{R}$ is convex and increasing.*

Proof. The statements follows from Jensen's inequality for conditional expectation. The first one holds because

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}[f(M_t)|\mathcal{F}_s] \geq f(\mathbb{E}[M_t|\mathcal{F}_s]) = f(M_s) = X_s, \quad s \leq t \in T.$$

The second one holds because

$$\mathbb{E}[X_1|\mathcal{F}_s] = \mathbb{E}[f(M_t)|\mathcal{F}_s] \geq f(\mathbb{E}[M_t|\mathcal{F}_s]) \geq f(M_s) = X_s, \quad s \leq t \in T.$$

$\square$

We say that X is an L^p-process if $X_t \in L^p$ for each $t \in T$. We say X is L^p-bounded if $\|X_t\|_p$, $t \in T$, is bounded. If M is an L^p-martingale, $p \in [1, \infty)$, applying Lemma 6.11 to $f(x) = |x|^p$, we see that $|M|^p$ is a submartingale.

Applying Lemma 6.11 to $f(x) = x \vee 0$, we see that if X is a submartingale, then the process $X_t^+ := X_t \vee 0$, $t \in T$, is also a submartingale.

We say that an $\mathcal{F}$-stopping time τ is bounded if there is a deterministic time $u \in T$ such that a.s. $\tau \leq u$. The following theorem generalizes the equality $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ to stopping times.

Theorem 6.12 (Optional Stopping Theorem). *Let M be a martingale on some index set T with filtration $\mathcal{F}$. Let σ and τ be two $\mathcal{F}$-stopping times taking countably many values. Suppose τ is bounded. Then M_τ and $M_{\sigma\wedge\tau}$ are integrable, and a.s.*

$$\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = M_{\sigma\wedge\tau}.$$

In particular, if a.s. $\sigma \le \tau$ are both bounded, then a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = M_\sigma$ and so $\mathbb{E}[M_\tau] = \mathbb{E}[M_\sigma]$.

Proof. Suppose $u \in T$ satisfies that a.s. $\tau \le u$. By Lemmas 5.2 (local property) and 6.1 ($\mathcal{F}_\tau$ agrees with $\mathcal{F}_t$ on $\{\tau = t\}$), for any $t \in T$ with $t \le u$, a.s.

$$\mathbb{E}[M_u|\mathcal{F}_\tau] = \mathbb{E}[M_u|\mathcal{F}_t] = M_t \quad \text{on } \{\tau = t\}.$$

Since τ takes countably many values, we get a.s. $\mathbb{E}[M_u|\mathcal{F}_\tau] = M_\tau$ and so M_τ is integrable. Since $\sigma \wedge \tau$ is also an $\mathcal{F}$-stopping time bounded by u taking countably many values, $M_{\sigma\wedge\tau}$ is also integrable, and a.s.

$$M_{\sigma\wedge\tau} = \mathbb{E}[M_u|\mathcal{F}_{\sigma\wedge\tau}] = \mathbb{E}[\mathbb{E}[M_u|\mathcal{F}_\tau]|\mathcal{F}_{\sigma\wedge\tau}] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge\tau}].$$

It remains to show that a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge\tau}]$. Since $\mathcal{F}_\sigma$ agrees with $\mathcal{F}_{\sigma\wedge\tau}$ on $\{\sigma = \sigma\wedge\tau\} = \{\sigma \le \tau\}$, by Lemma 5.2, a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge\tau}]$ on $\{\sigma \le \tau\}$. Since $\mathcal{F}_\tau \cap \{\tau \le \sigma\} \subset \mathcal{F}_\sigma$ and M_τ is $\mathcal{F}_\tau$-measurable, by Lemma 5.2, a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = M_\tau$ on $\{\tau \le \sigma\}$. Since $\mathcal{F}_\tau \cap \{\tau \le \sigma\} = \mathcal{F}_\tau \cap \{\tau \le \sigma\wedge\tau\} \subset \mathcal{F}_{\sigma\wedge\tau}$, by Lemma 5.2, a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = M_\tau$ on $\{\tau \le \sigma\}$. thus, a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge\tau}]$ on $\{\tau \le \sigma\}$. Combining this with that a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge\tau}]$ on $\{\sigma \le \tau\}$, we get a.s. $\mathbb{E}[M_\tau|\mathcal{F}_\sigma] = \mathbb{E}[M_\tau|\mathcal{F}_{\sigma\wedge\tau}]$, as desired. $\square$

Exercise . Prove that if $T = \mathbb{Z}_+$ or finite, and X is a submartingale (resp. supermartingale), then for σ, τ in the theorem, we have a.s. $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] - X_{\sigma\wedge\tau} \ge 0$ (resp. ≤ 0). Hint: Use Doob's decomposition.

Example . The condition on τ can not be removed. Suppose $\zeta_1, \zeta_2, \dots$ is a sequence of i.i.d. random variables with common distribution $\frac{1}{2}(\delta_1 + \delta_{-1})$. For $n \in \mathbb{Z}_+$, let $X_n = \sum_{k=1}^n 2^{k-1}\zeta_k$. Then X is a martingale. Let $\tau = \inf\{n \in \mathbb{N} : \zeta_n = 1\}$. Then τ is a stopping time, and a.s. takes values in $\mathbb{N}$. In fact,

$$\mathbb{P}[\tau = \infty] = \mathbb{P}[\bigcap_{N\in\mathbb{N}} \{\tau > N\}] = \lim_{N\to\infty} \mathbb{P}\{\zeta_n = -1, 1 \le n \le N\} = \lim_{N\to\infty} 2^{-N} = 0.$$

We observe that for any $N \in \mathbb{N}$, when $\tau = N$, $X_\tau = \sum_{k=1}^{N-1}(-1)2^{n-1} + 2^{N-1} = 1$. Thus, $\mathbb{E}[X_\tau] = 1$. But since $X_0 = 0$, $\mathbb{E}[X_0] = 0 \ne \mathbb{E}[X_\tau]$.

This example describes the balance history of a gambler, who bids one dollar on the first day, doubles his bid on every next day, and stops whenever he wins. In reality, a gambler can not win money with this game because he does not have infinite amount of money to bid.

Lemma 6.13 (Martingale Criterion). *Let M be an integrable adapted process on some index set T w.r.t. a filtration $\mathcal{F}$. Then M is an $\mathcal{F}$-martingale iff for any two T-valued $\mathcal{F}$-stopping times σ and τ taking at most two values, we have $\mathbb{E}[M_\sigma] = \mathbb{E}[M_\tau]$.*

Proof. The only if part follows from Theorem 6.12. For the if part, let $s < t \in T$. Let $A \in \mathcal{F}_s$. Then $\tau := s\mathbf{1}_A + t\mathbf{1}_{A^c}$ is an $\mathcal{F}$-stopping time because for any $u \in T$, if $u \ge t$, $\{\tau \le u\} = \Omega \in \mathcal{F}_u$; if $s \le u < t$, $\{\tau \le u\} = A \in \mathcal{F}_s \subset \mathcal{F}_u$; and if $u < s$, $\{\tau \le u\} = \emptyset \in \mathcal{F}_u$. By the assumption, we have

$$0 = \mathbb{E}M_t - \mathbb{E}M_\tau = \mathbb{E}M_t - \mathbb{E}[\mathbf{1}_A M_s] - \mathbb{E}[\mathbf{1}_{A^c} M_t] = \mathbb{E}[\mathbf{1}_A(M_t - M_s)].$$

Since this holds for any $A \in \mathcal{F}_s$, we get a.s. $\mathbb{E}[M_t - M_s | \mathcal{F}_s] = 0$. So M is an $\mathcal{F}$-martingale. □

Corollary 6.14 (Martingale Transforms)**.** *Let M be a martingale on some index set T with filtration $\mathcal{F}$. Fix a stopping time τ that takes countably many values, and let η be a bounded, $\mathcal{F}_\tau$-measurable random variable. Then the process $N_t = \eta(M_t - M_{t\wedge\tau})$ is again an $\mathcal{F}$-martingale.*

Taking $\eta \equiv 1$, from Corollary 6.14 we see that if M is an $\mathcal{F}$-martingale, and if τ is a bounded $\mathcal{F}$-stopping time taking countably many values, then the stopped process

$$M_t^\tau := M_{\tau\wedge t}, \quad t \in T,$$

is also an $\mathcal{F}$-martingale.