PROBABILITY THEORY

Zenghu Li

—

School of Mathematical Sciences
Beijing Normal University

—

December 12, 2005

Contents

Chapter 1

Measurable Spaces

1.1 Measurable spaces

In this section, we discuss some properties of σ-algebras and measurable transformations. Let Ω and E be non-empty sets.

Definition 1.1.1 A family $\mathscr{F}$ of subsets of Ω is called a *σ-algebra* on Ω if

(i) $\emptyset \in \mathscr{F}$ and $\Omega \in \mathscr{F}$;

(ii) $A \in \mathscr{F}$ implies $A^c \in \mathscr{F}$;

(iii) $\{A_1, A_2, \cdots\} \subseteq \mathscr{F}$ implies $\bigcup_{k=1}^{\infty} A_k \in \mathscr{F}$.

If $\mathscr{F}$ is a σ-algebra on Ω, we call $(\Omega, \mathscr{F})$ a *measurable space*. The sets in $\mathscr{F}$ are called *measurable sets*.

Proposition 1.1.1 *Suppose that I is an arbitrary index set and $\mathscr{F}_\alpha$ is a σ-algebra on Ω for each $\alpha \in I$. Then $\mathscr{F} := \bigcap_{\alpha \in I} \mathscr{F}_\alpha$ is a σ-algebra.*

Proof. By the above definition, $\emptyset \in \mathscr{F}_\alpha$ and $\Omega \in \mathscr{F}_\alpha$ for each $\alpha \in I$. Then $\emptyset \in \mathscr{F}$ and $\Omega \in \mathscr{F}$. If $A \in \mathscr{F}$, for each $\alpha \in I$ we have $A \in \mathscr{F}_\alpha$ and hence $A^c \in \mathscr{F}_\alpha$, so $A^c \in \mathscr{F}$. Suppose that $\{A_1, A_2, \cdots\} \subseteq \mathscr{F}$. For each $\alpha \in I$ we have $\{A_1, A_2, \cdots\} \subseteq \mathscr{F}_\alpha$ so that $\bigcup_{k=1}^{\infty} A_k \in \mathscr{F}_\alpha$, implying $\bigcup_{k=1}^{\infty} A_k \in \mathscr{F}$. □

Let f be a map from Ω to E. For $A \subseteq \Omega$ we write $f(A) = \{f(\omega) \in B : \omega \in \Omega\}$ and for $B \subseteq E$ we write $f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$. If $\mathscr{C}$ is a family of subsets of Ω, then

$$\sigma(\mathscr{C}) := \bigcap \{\mathscr{F} : \mathscr{F} \supseteq \mathscr{C} \text{ is a } \sigma\text{-algebra}\} \tag{1.1.1}$$

defines a σ-algebra, which is called the *σ-algebra generated by $\mathscr{C}$*.

Proposition 1.1.2 *Let $f : \Omega \to E$ be a map and $\mathscr{G}$ a σ-algebra on E. Then $f^{-1}(\mathscr{G}) := \{f^{-1}(B) \subseteq \Omega : B \in \mathscr{G}\}$ is a σ-algebra on Ω.*

Proof. (Homework.) □

In the situation of the above proposition, we call $f^{-1}(\mathscr{G})$ the *σ-algebra generated by f* and is denoted by $\sigma(f)$. For a class $\mathscr{H}$ of mappings from Ω to E, the *σ-algebra generated by $\mathscr{H}$* is defined as

$$\sigma(\mathscr{H}) := \sigma(\{f^{-1}(A) : A \in \mathscr{E}, f \in \mathscr{H}\}). \tag{1.1.2}$$

Proposition 1.1.3 *Let $\mathscr{U}$ be a family of subsets of E. Then we have $\sigma(f^{-1}(\mathscr{U})) = f^{-1}(\sigma(\mathscr{U}))$ for any mapping $f : \Omega \to E$.*

Proof. By Proposition 1.1.2 we find that $f^{-1}(\sigma(\mathscr{U}))$ is a σ-algebra. Since $f^{-1}(\sigma(\mathscr{U})) \supseteq f^{-1}(\mathscr{U})$, we have $f^{-1}(\sigma(\mathscr{U})) \supseteq \sigma(f^{-1}(\mathscr{U}))$. On the other hand, let $\mathscr{G} = \{B \subseteq E : f^{-1}(B) \in \sigma(f^{-1}(\mathscr{U}))\}$. It is easily seen that $f^{-1}(\mathscr{G}) \subseteq \sigma(f^{-1}(\mathscr{U}))$. If $B \in \mathscr{U}$, we have $f^{-1}(B) \in f^{-1}(\mathscr{U}) \subseteq \sigma(f^{-1}(\mathscr{U}))$ and hence $B \in \mathscr{G}$. That shows $\mathscr{U} \subseteq \mathscr{G}$. It is not hard to check that $\mathscr{G}$ is a σ-algebra on E so that $\sigma(\mathscr{U}) \subseteq \mathscr{G}$. Consequently, we have $f^{-1}(\sigma(\mathscr{U})) \subseteq f^{-1}(\mathscr{G}) \subseteq \sigma(f^{-1}(\mathscr{U}))$. □

Given two measurable spaces $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$, we say a mapping $f : \Omega \to E$ is *$\mathscr{F}/\mathscr{E}$-measurable*, or simply *$\mathscr{F}$-measurable*, provided $f^{-1}(\mathscr{E}) \subseteq \mathscr{F}$. A one-to-one correspondence $\phi : \Omega \to E$ is called an *isomorphism* if ϕ is $\mathscr{F}/\mathscr{E}$-measurable and ϕ^{-1} is $\mathscr{E}/\mathscr{F}$-measurable. We say $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$ and *isomorphic* if there is an isomorphism between them.

Proposition 1.1.4 *If every $f \in \mathscr{H}$ is $\mathscr{F}/\mathscr{E}$-measurable, then $\sigma(\mathscr{H}) \subseteq \mathscr{F}$.*

Proof. Under the assumption, we have $\{f^{-1}(A) : A \in \mathscr{E}, f \in \mathscr{H}\} \subseteq \mathscr{F}$ and hence $\sigma(\mathscr{H}) \subseteq \mathscr{F}$. □

Proposition 1.1.5 *Let $(\Omega, \mathscr{F})$, $(E, \mathscr{E})$ and $(G, \mathscr{G})$ be measurable spaces. If $f : \Omega \to E$ is $\mathscr{F}/\mathscr{E}$-measurable and $h : E \to G$ is $\mathscr{E}/\mathscr{G}$-measurable, the composition $h \circ f : \Omega \to G$ is $\mathscr{F}/\mathscr{G}$-measurable.*

Proof. Since h is $\mathscr{E}/\mathscr{G}$-measurable, we have $h^{-1}(\mathscr{G}) \subseteq \mathscr{E}$ and hence

$$(h \circ f)^{-1}(\mathscr{G}) = f^{-1}(h^{-1}(\mathscr{G})) \subseteq f^{-1}(\mathscr{E}).$$

From the measurability of f we get $f^{-1}(\mathscr{E}) \subseteq \mathscr{F}$. Consequently, $h \circ f$ is $\mathscr{F}/\mathscr{G}$-measurable. □

Let $\mathscr{O}(\mathbb{R}^d)$ denote the family of all open sets on the Euclidean space $\mathbb{R}^d$. We call $\mathscr{B}(\mathbb{R}^d) := \sigma(\mathscr{O}(\mathbb{R}^d))$ the *Borel σ-algebra* of on $\mathbb{R}^d$. Clearly, $\mathscr{B}(\mathbb{R}^d)$ includes all open sets, closed sets, intervals, singletons, finite sets and countable sets.

Proposition 1.1.6 *Let $\mathscr{L}(\mathbb{R}^d) = \{(-\infty, b] : b \in \mathbb{R}^d\}$ and $\mathscr{R}(\mathbb{R}^d) = \{[a, \infty) : a \in \mathbb{R}^d\}$. Then $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathscr{L}(\mathbb{R}^d)) = \sigma(\mathscr{R}(\mathbb{R}^d))$.*

Proof. As observed above, we have $\mathscr{L}(\mathbb{R}^d) \subseteq \mathscr{B}(\mathbb{R}^d)$. Since $\mathscr{B}(\mathbb{R}^d)$ is a σ-algebra, we conclude that $\sigma(\mathscr{L}(\mathbb{R}^d)) \subseteq \mathscr{B}(\mathbb{R}^d)$. On the other hand, for any $a \le b \in \mathbb{R}^d$ we have $(a, b] = (-\infty, b] \setminus (-\infty, a] \in \sigma(\mathscr{L}(\mathbb{R}^d))$. Since each element of $\mathscr{O}(\mathbb{R}^d)$ is the union of a countable number of intervals of the form $(a, b]$, we get $\mathscr{O}(\mathbb{R}^d) \subseteq \sigma(\mathscr{L}(\mathbb{R}^d))$ and hence $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathscr{O}(\mathbb{R}^d)) \subseteq \sigma(\mathscr{L}(\mathbb{R}^d))$. That shows $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathscr{L}(\mathbb{R}^d))$. The equality $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathscr{R}(\mathbb{R}^d))$ follows by similar arguments. (Homework.) □

Corollary 1.1.1 *Let* $\mathscr{S}(\mathbb{R}^d) = \{(a, b] : a \le b \in \mathbb{R}^d\}$*. Then* $\mathscr{B}(\mathbb{R}^d) = \sigma(\mathscr{S}(\mathbb{R}^d))$*.*

Proposition 1.1.7 *Let* f *be a real function on* $(\Omega, \mathscr{F})$*. Then following properties are equivalent:*

(i) f *is* $\mathscr{F}/\mathscr{B}(\mathbb{R})$*-measurable;*

(ii) $f^{-1}((-\infty, b]) \in \mathscr{F}$ *for every* $b \in \mathbb{R}$*;*

(iii) $f^{-1}((-\infty, b)) \in \mathscr{F}$ *for every* $b \in \mathbb{R}$*;*

(iv) $f^{-1}([a, -\infty)) \in \mathscr{F}$ *for every* $a \in \mathbb{R}$*;*

(v) $f^{-1}((a, -\infty)) \in \mathscr{F}$ *for every* $a \in \mathbb{R}$*.*

Proof. Clearly, (i) implies (ii) – (v). Now suppose (ii) holds. In other words, $f^{-1}(\mathscr{L}(\mathbb{R})) \subseteq \mathscr{F}$. Then Propositions 1.1.6 and 1.1.3 imply

$$f^{-1}(\mathscr{B}(\mathbb{R})) = f^{-1}(\sigma(\mathscr{L}(\mathbb{R}))) = \sigma(f^{-1}(\mathscr{L}(\mathbb{R}))) \subseteq \mathscr{F}.$$

Then (ii) holds. Clearly, (iii) implies (ii), so it also implies (i). The remaining assertions are immediate. □

Corollary 1.1.2 *Let* $\{f_n\}$ *be a bounded sequence of real functions on* $(\Omega, \mathscr{F})$*. If each* f_n *is* $\mathscr{F}$*-measurable, then the following real functions are* $\mathscr{F}$*-measurable:* $\inf_n f$, $\sup_n f$, $\liminf_n f$, $\limsup_n f$.

Example 1.1.1 A finite or countable family $\{U_i : i \in I\}$ of disjoint subsets of Ω satisfying $\bigcup_{i\in I} U_i = \Omega$ is called a *partition of* Ω. If $\mathscr{C} = \{U_i : i \in I\}$ is a partition of Ω, then $\sigma(\mathscr{C}) = \big\{\bigcup_{j\in J} U_j : J \subseteq I\big\}$ with $\bigcup_{j\in\emptyset} U_j = \emptyset$ by convention. (Homework.)

Example 1.1.2 Let $\mathscr{C} = \{U_i : i \in I\}$ be a finite or countable partition of Ω. We equip $\mathbb{R}$ with the σ-algebra $\mathscr{B}(\mathbb{R})$. Then a function $X : \Omega \to \mathbb{R}$ is $\sigma(\mathscr{C})$-measurable if and only if

$$X(\omega) = \sum_{i\in I} c_i 1_{U_i}(\omega), \qquad \omega \in \Omega, \tag{1.1.3}$$

for a family of real constants $\{c_i : i \in I\}$. (Homework.)

Example 1.1.3 Let $\{c_i : i \in I\}$ be distinct real numbers and let $\mathscr{C} = \{U_i : i \in I\}$ be a partition of Ω. Let $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ be given as in the last example. If the function $X : \Omega \to \mathbb{R}$ has representation (1.1.3), then $\sigma(X) = \sigma(\mathscr{C})$. (Homework.)

1.2 Metric spaces and Borel functions

Definition 1.2.1 Let E be a non-empty set. A function $\rho : E \times E \to \mathbb{R}$ is called a *metric* if it satisfies:

(i) $\rho(x, y) \geq 0$ for all $x, y \in E$;

(ii) $\rho(x, y) = 0$ if and only if $x = y$ for all $x, y \in E$;

(iii) $\rho(x, y) = \rho(y, x)$ for all $x, y \in E$;

(iv) $\rho(x, y) \leq \rho(x, z) + \rho(y, z)$ for all $x, y, z \in E$.

In this case, we call (E, ρ) a *metric space*.

Suppose that ρ is a metric on E. For $x, y \in E$ let $r(x, y) = 1 \wedge \rho(x, y)$. It is easy to show that $r(\cdot, \cdot)$ is also a metric on E. In this sequel, we assume (E, ρ) is a fixed metric space.

Definition 1.2.2 Let $a \in E$ and $r > 0$. We call $B(a, r) := \{x : \rho(a, x) < r\}$ a *ball* centered at a with radius r. A set $U \subseteq E$ is called an *open set*, if for each $a \in U$ there is some $r = r(a) > 0$, such that $B(a, r) \in U$. A set F is called a *closed set* if $E \setminus F$ is an open set.

Clearly, the sets E and $\emptyset$ are simultaneously open and closed. A typical metric space is the space $\mathbb{R}^d$ equipped with the Euclidean metric determined by

$$|x - y| = \Big(\sum_{j=1}^{d} |x_j - y_j|^2 \Big)^{1/2}.$$

Definition 1.2.3 Let $x \in A \subseteq E$. If there is some $r > 0$ such that $B(x, r) \subseteq A$, we call x an *interior point* of A. We call $A^\circ := \{\text{interior points of } A\}$ the *interior* of A and call $\bar{A} := ((A^c)^\circ)^c$ the *closure* of A. It is not hard to show that A° is the largest open subset of A and $\bar{A}$ is the smallest closed superset of A. The set $\partial A := \bar{A} \setminus A^\circ$ is called the *boundary* of A. We say $A \subseteq E$ is *dense* in $F \subseteq E$ if $\bar{A} \supseteq F$. The space E is said to be *separable* is it has a countable dense subset.

Proposition 1.2.1 *Let $A \subseteq E$. Then $x \in \bar{A}$ if and only if there is a sequence $\{x_n\} \subseteq A$, such that $x_n \to x$ as $n \to \infty$.*

Proof. "$\Rightarrow$" Suppose that $x \in \bar{A} = ((A^c)^\circ)^c$, so $x \notin (A^c)^\circ$. According to the definition of the interio, for every $n \geq 1$ the inclusion $B(x, 1/n) \subseteq A^c$ does not hold. In other word, for every $n \geq 1$ there is some $x_n \in B(x, 1/n)$ such that $x_n \notin A^c$. That is, $x_n \in B(x, 1/n) \cap A$. It follows that $x_n \in A$ and $\rho(x_n, x) < 1/n$, so we have $x_n \to x$ as $n \to \infty$.

"$\Leftarrow$" Suppose there is $\{x_n\} \subseteq A$ such that $x_n \to x$ as $n \to \infty$. We shall prove $x \in \bar{A} = ((A^c)^\circ)^c$. If this is not true, we have $x \in (A^c)^\circ$, that is, there is some $r > 0$ such that $B(x, r) \subseteq A^c$. Then $x_n \notin B(x, r)$ for every $n \geq 1$, which is in contradiction to the fact $x_n \to x$. $\square$

Definition 1.2.4 Let (E, ρ) and (F, r) be two metric spaces. We say $f : E \to F$ is *continuous at* $x \in E$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $f(B_\rho(x, \delta)) \subseteq B_r(f(x), \varepsilon)$. We say f is *continuous on* E if it is continuous at every $x \in E$. We say f is *uniformly continuous on* E if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $f(B_\rho(x, \delta)) \subseteq B_r(f(x), \varepsilon)$ for every $x \in E$.

Theorem 1.2.1 *A mapping $f:(E,\rho)\to(F,r)$ is continuous if and only if $f^{-1}(U)\subseteq E$ is open whenever $U\subseteq F$ is open.*

Proof. "⇒" Suppose that f is continuous and $U\subseteq F$ is an open set. For each $x\in f^{-1}(U)$ we have $y:=f(x)\in U$. Since U is open, there is some $\varepsilon>0$ such that $B_r(y,\varepsilon)\subseteq U$. By the continuity of f, there is $\delta>0$ so that

$$f(B_\rho(x,\delta))\subseteq B_r(f(x),\varepsilon))=B_r(y,\varepsilon).$$

It follows that

$$B_\rho(x,\delta)\subseteq f^{-1}(B_r(y,\varepsilon))\subseteq f^{-1}(U).$$

That shows that $f^{-1}(U)$ is an open set.

"⇐" Suppose that $f^{-1}(U)\subseteq E$ is open whenever $U\subseteq F$ is an open set. Let $x\in E$ and $\varepsilon>0$. Then $f^{-1}(B_r(f(x),\varepsilon))\subseteq E$ is an open set and $x\in f^{-1}(B_r(f(x),\varepsilon))$. It follows that $B_\rho(x,\delta)\subseteq f^{-1}(B_r(f(x),\varepsilon))$ for some $\delta>0$, which implies that $f(B_\rho(x,\delta))\subseteq B_r(f(x),\varepsilon)$. Then f is continuous. □

Corollary 1.2.1 *A mapping $f:(E,\rho)\to(F,r)$ is continuous if and only if $f^{-1}(A)\subseteq E$ is closed whenever $A\subseteq F$ is a closed set.*

Let $\mathscr{O}(E)$ denote the family of open subsets of E. We call $\mathscr{B}(E):=\sigma(\mathscr{O}(E))$ the *Borel σ-algebra* of E. A set $B\in\mathscr{B}(E)$ is called a *Borel set* and a $\mathscr{B}(E)$-measurable real function $f:E\to\mathbb{R}$ is called a *Borel function*.

Proposition 1.2.2 *A continuous function $f:E\to\mathbb{R}$ is a Borel function.*

Proof. Since f is continuous, by Theorem 1.2.1 we have $f^{-1}(\mathscr{O}(\mathbb{R}))\subseteq\mathscr{O}(E)$. It follows that

$$f^{-1}(\mathscr{B}(\mathbb{R}))=f^{-1}(\sigma(\mathscr{O}(\mathbb{R}))=\sigma(f^{-1}(\mathscr{O}(\mathbb{R})))\subseteq\sigma(\mathscr{O}(E))=\mathscr{B}(E).$$

Then f is $\mathscr{B}(E)/\mathscr{B}(\mathbb{R})$-measurable. □

Proposition 1.2.3 *For any non-empty set $A\subseteq E$, let*

$$\rho(x,A)=\inf\{\rho(x,y):y\in A\},\qquad x\in E. \tag{1.2.1}$$

Then $\rho(\cdot,A)$ is a uniformly continuous function on E.

Proof. Let $x,y\in E$. For any $z\in A$ we have

$$\rho(x,A)-\rho(y,z)\le\rho(x,z)-\rho(y,z)\le\rho(x,y).$$

Then we can take $\sup_{z\in A}$ in both sides to get

$$\rho(x,A)-\rho(y,A)\le\rho(x,y).$$

A similar argument shows that

$$\rho(y, A) - \rho(x, A) \le \rho(x, y).$$

Combining those two inequalities, we obtain

$$|\rho(x, A) - \rho(y, A)| \le \rho(x, y).$$

Then $\rho(\cdot, A)$ is uniformly continuous on E. □

Proposition 1.2.4 *Let $C(E)^+$ denote the class of bounded, non-negative and continuous functions on E. Then $\sigma(C(E)^+) = \mathscr{B}(E)$.*

Proof. By Proposition 1.2.2, each $f \in C(E)^+$ is $\mathscr{B}(E)$-measurable. Then Propositions 1.1.4 implies $\sigma(C(E)^+) \subseteq \mathscr{B}(E)$. For fixed $U \in \mathscr{O}(E)$ let $h(x) = 1 \wedge \rho(x, U^c)$. By Proposition 1.2.3, we have $h_n := h^{1/n} \in C(E)^+$ so h_n is $\sigma(C(E)^+)$-measurable. Clearly, $h(x) = 0$ for $x \in U^c$ and $0 < h(x) \le 1$ for $x \in U$. It follows that $h_n \to 1_U$, so 1_U is $\sigma(C(E)^+)$-measurable and consequently $U \in \sigma(C(E)^+)$. That shows that $\mathscr{O}(E) \subseteq \sigma(C(E)^+)$. Thus $\mathscr{B}(E) = \sigma(\mathscr{O}(E)) \subseteq \sigma(C(E)^+)$. □

We shall often need to consider functions taking values in the *extended real line* $\bar{\mathbb{R}} := [-\infty, \infty]$ denote the. A metric ρ_1 on $\bar{\mathbb{R}}$ is defined by

$$\rho_1(x, y) = \Big|\frac{x}{1 + |x|} - \frac{y}{1 + |y|}\Big|.$$

Proposition 1.2.5 *Let $(\Omega, \mathscr{F})$ be a measurable space and $f : \Omega \to \bar{\mathbb{R}}$ an extended real-valued function on Ω. Then following properties are equivalent:*

(i) f *is* $\mathscr{F}/\mathscr{B}(\bar{\mathbb{R}})$*-measurable;*

(ii) $f^{-1}([-\infty, b]) \in \mathscr{F}$ *for every* $b \in \mathbb{R}$*;*

(iii) $f^{-1}([-\infty, b)) \in \mathscr{F}$ *for every* $b \in \mathbb{R}$*;*

(iv) $f^{-1}([a, -\infty]) \in \mathscr{F}$ *for every* $a \in \mathbb{R}$*;*

(v) $f^{-1}((a, -\infty]) \in \mathscr{F}$ *for every* $a \in \mathbb{R}$*.*

Proof. (Homework.) □

Corollary 1.2.2 *Let $\{f_n\}$ be a sequence of extended real-valued functions on $(\Omega, \mathscr{F})$. If each f_n is $\mathscr{F}$-measurable, then the following extended real-valued functions are $\mathscr{F}$-measurable:* $\inf_n f$, $\sup_n f$, $\liminf_n f$, $\limsup_n f$.

Definition 1.2.5 We say a sequence $\{x_n\} \subseteq E$ is *Cauchy* if $\rho(x_m, x_n) \to 0$ as $m, n \to \infty$. We say x_n *converges* to $x \in E$ as $n \to \infty$, if $\rho(x_n, x) \to 0$ as $n \to \infty$. In this case, we write $x_n \to x$. The space (E, ρ) is said to be *complete* if every Cauchy sequence in it converges.

Proposition 1.2.6 *Let Ω be a Borel subset of some complete separable metric space with $\mathscr{F} = \mathscr{B}(\Omega)$. Then there is a closed subset F of $[0, 1]$ such that $(\Omega, \mathscr{F})$ is isomorphic to $(F, \mathscr{B}(F))$.*

Proof. This follows immediately from Parthasarathy (1967, p.14, Theorem 2.12). (Homework: Read and understand the proof.) □

1.3 Monotone classes of sets

We simply write $A_n\uparrow$ if $\{A_n\}$ is a non-decreasing sequence of sets and write $A_n\downarrow$ if $\{A_n\}$ is a non-increasing sequence of sets. If $A_n\uparrow$ and $A = \bigcup_{n=1}^{\infty} A_n$, we write $A_n \uparrow A$. Similarly, if $A_n\downarrow$ and $A = \bigcap_{n=1}^{\infty} A_n$, we write $A_n \downarrow A$.

Definition 1.3.1 A class $\mathscr{A}$ of subsets of Ω is called an *algebra* on Ω if

(i) $\emptyset \in \mathscr{A}$ and $\Omega \in \mathscr{A}$;

(ii) $A \in \mathscr{A}$ implies $A^c \in \mathscr{A}$;

(iii) $\{A_1, A_2, \cdots, A_n\} \subseteq \mathscr{F}$ implies $\bigcup_{k=1}^{n} A_k \in \mathscr{F}$.

Definition 1.3.2 A class $\mathscr{D}$ of subsets of the non-empty set Ω is called a *monotone class* if it has the following properties:

(i) If $\{A_n\} \subseteq \mathscr{D}$ and $A_n\uparrow$, then $\bigcup_{n=1}^{\infty} A_n \in \mathscr{D}$;

(ii) If $\{A_n\} \subseteq \mathscr{D}$ and $A_n\downarrow$, then $\bigcap_{n=1}^{\infty} A_n \in \mathscr{D}$.

Let $\mathscr{C}$ be a class of subsets of Ω. Then

$$\mu(\mathscr{C}) := \bigcap\{\mathscr{D} : \mathscr{D} \supseteq \mathscr{C} \text{ is a monotone class}\}. \tag{1.3.1}$$

is a monotone class, which is called the *monotone class generated by* $\mathscr{C}$.

Lemma 1.3.1 *If $\mathscr{D}$ is simultaneously an algebra and a monotone class, it is a σ-algebra.*

Proof. It is sufficient to show that $\mathscr{D}$ is closed under the operation of countable unions. Suppose that $\{A_n\} \subseteq \mathscr{D}$. Since $\mathscr{D}$ is an algebra, we have $B_n := \bigcup_{k=1}^{n} A_k \in \mathscr{D}$ for each $n \geq 1$. By the definition of the monotone class we have $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathscr{D}$. □

Theorem 1.3.1 *If $\mathscr{C}$ is an algebra, then $\mu(\mathscr{C}) = \sigma(\mathscr{C})$.*

Proof. Since a σ-algebra is a monotone class, we have $\sigma(\mathscr{C}) \supseteq \mu(\mathscr{C})$. To prove $\mu(\mathscr{C}) \supseteq \sigma(\mathscr{C})$, it is sufficient to show $\mu(\mathscr{C})$ is a σ-algebra. By Lemma 1.3.1 we only need to show $\mu(\mathscr{C})$ is an algebra. Since $\Omega \in \mathscr{C} \subset \mu(\mathscr{C})$, it suffices to show $A, B \in \mu(\mathscr{C})$ implies $B \cup A$ and $A \setminus B \in \mu(\mathscr{C})$. For $A \subseteq \Omega$, let

$$\mathscr{D}_A = \{B \subseteq \Omega : B \cup A, B \setminus A \text{ and } A \setminus B \in \mu(\mathscr{C})\}.$$

The desired result will follow if we can prove $\mathscr{D}_A \supseteq \mu(\mathscr{C})$ for all $A \in \mu(\mathscr{C})$. We show this in three steps as follows.

Step 1) We prove $\mathscr{D}_A$ is a monotone class. If $\{B_n\} \subseteq \mathscr{D}_A$ and $B_n \uparrow B$, we have $B_n \cup A$, $B_n \setminus A$ and $A \setminus B_n \in \mu(\mathscr{C})$. Since $\mu(\mathscr{C})$ is a monotone class, it is easily seen that $B \cup A$, $B \setminus A$ and $A \setminus B \in \mu(\mathscr{C})$, and so $B \in \mathscr{D}_A$. Similarly, if $\{B_n\} \subseteq \mathscr{D}_A$ and $B_n \downarrow B$, we have $B \in \mathscr{D}_A$. Thus $\mathscr{D}_A$ is a monotone class.

Step 2) For $A \in \mathscr{C}$, we prove $\mathscr{D}_A \supseteq \mu(\mathscr{C})$. Suppose that $B \in \mathscr{C}$. Since $\mathscr{C}$ is an algebra, we have $B \cup A$, $B \setminus A$ and $A \setminus B \in \mathscr{C} \subseteq \mu(\mathscr{C})$. It follows that $B \in \mathscr{D}_A$. That shows $\mathscr{D}_A \supseteq \mathscr{C}$. Since $\mathscr{D}_A$ is a monotone class we get $\mathscr{D}_A \supseteq \mu(\mathscr{C})$.

Step 3) For $A \in \mu(\mathscr{C})$, we prove $\mathscr{D}_A \supseteq \mu(\mathscr{C})$. Suppose that $B \in \mathscr{C}$. By the last step, we have $\mathscr{D}_B \supseteq \mu(\mathscr{C})$. In particular, we get $A \in \mathscr{D}_B$ which implies $B \in \mathscr{D}_A$. It follows that $\mathscr{D}_A \supseteq \mathscr{C}$ and hence $\mathscr{D}_A \supseteq \mu(\mathscr{C})$. □

Corollary 1.3.1 *If $\mathscr{C}$ is an algebra and $\mathscr{D} \supseteq \mathscr{C}$ is a monotone class, then $\mathscr{D} \supseteq \mu(\mathscr{C}) = \sigma(\mathscr{C})$.*

Definition 1.3.3 A class $\mathscr{C}$ of subsets of Ω is called a π-*class* if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$. A class $\mathscr{D}$ of subsets of Ω is called a λ-*class* if

(i) $\Omega \in \mathscr{D}$ and $\Omega \in \mathscr{D}$;

(ii) $A, B \in \mathscr{D}$ and $A \subseteq B$ imply $B \setminus A \in \mathscr{D}$;

(iii) $\{A_n\} \subseteq \mathscr{D}$ and $A_n\uparrow$ imply $\bigcup_{n=1}^{\infty} A_n \in \mathscr{D}$.

For a class $\mathscr{C}$ of subsets of Ω, let

$$\lambda(\mathscr{C}) := \bigcap\{\mathscr{D} : \mathscr{D} \supseteq \mathscr{C} \text{ is a } \lambda\text{-class}\}. \tag{1.3.2}$$

Then $\lambda(\mathscr{C})$ is a λ-class, which is called the λ-*class generated by* $\mathscr{C}$.

Lemma 1.3.2 *If $\mathscr{D}$ is simultaneously a π-class and a λ-class, it is a σ-algebra.*

Proof. By the definition of the λ-class, we have $\Omega \in \mathscr{D}$. Moreover, if $A \in \mathscr{D}$, then $A^c = \Omega \setminus A \in \mathscr{D}$. For any sequence $\{A_n\} \subseteq \mathscr{D}$ let $B_n = \bigcup_{j=1}^{n} A_j$. Then $A_n^c \in \mathscr{D}$ for each $n \geq 1$ and hence $B_n^c = \bigcap_{j=1}^{n} A_j^c \in \mathscr{D}$. It follows that $B_n = (B_n^c)^c \in \mathscr{D}$. Using the definition of the λ-class again we conclude that $B_n \uparrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{D}$. Then $\mathscr{D}$ is a σ-algebra. □

Theorem 1.3.2 *If $\mathscr{C}$ is a π-class, then $\lambda(\mathscr{C}) = \sigma(\mathscr{C})$.*

Proof. Since a σ-algebra is a λ-class, we have $\lambda(\mathscr{C}) \subseteq \sigma(\mathscr{C})$. To show $\sigma(\mathscr{C}) \subseteq \lambda(\mathscr{C})$ it suffices to prove $\lambda(\mathscr{C})$ is a σ-algebra. By Lemma 1.3.2, we only need to show that $\lambda(\mathscr{C})$ is a π-class. For $A \subseteq \Omega$ let $\mathscr{D}_A = \{B \subseteq \Omega : A \cap B \in \lambda(\mathscr{C})\}$. The desired result follows once it is proved that $\mathscr{D}_A \supseteq \lambda(\mathscr{C})$ for all $A \in \lambda(\mathscr{C})$. This property can be in three steps: Step 1) Prove $\mathscr{D}_A$ is a λ-class for $A \in \lambda(\mathscr{C})$; Step 2) Prove $\mathscr{D}_A \supseteq \lambda(\mathscr{C})$ for $A \in \mathscr{C}$; Step 3) Prove $\mathscr{D}_A \supseteq \lambda(\mathscr{C})$ for $A \in \lambda(\mathscr{C})$. We omit the details. (Homework: Gives the details of the last part of the proof.) □

Corollary 1.3.2 *If $\mathscr{C}$ is a π-class and $\mathscr{D} \supseteq \mathscr{C}$ is a λ-class, then $\mathscr{D} \supseteq \lambda(\mathscr{C}) = \sigma(\mathscr{C})$.*

1.4 Monotone systems of functions

Definition 1.4.1 A non-empty family $\mathscr{M}$ of non-negative extended real-valued functions on Ω is called a *monotone system* if the following conditions are satisfied:

(i) For any $f_1, f_2 \in \mathscr{M}$ and $a_1, a_2 \in \mathbb{R}$ with $a_1 f_1 + a_2 f_2 \geq 0$ we have $a_1 f_1 + a_2 f_2 \in \mathscr{M}$;

(ii) If $\{f_n\} \subseteq \mathscr{M}$ and $f_n \uparrow f$, then $f \in \mathscr{M}$.

Theorem 1.4.1 *Let $\mathscr{C}$ be an algebra and $\mathscr{M}$ a monotone system on Ω. If $\mathscr{M}$ contains the indicators of all sets in $\mathscr{C}$, it contains all non-negative $\sigma(\mathscr{C})$-measurable functions.*

Proof. Let $\mathscr{D} = \{A \subseteq \Omega : 1_A \in \mathscr{M}\}$. By the assumption, we have $\mathscr{D} \supseteq \mathscr{C}$. We claim that $\mathscr{D}$ is a monotone class. Indeed, suppose that $\{B_n\} \subseteq \mathscr{D}$ and $B_n \uparrow B$. Then $1_{B_n} \uparrow 1_B \in \mathscr{M}$ by the definition of the monotone system. It follows that $B \in \mathscr{D}$. If $\{B_n\} \subseteq \mathscr{D}$ and $B_n \downarrow B$, we have $1_{B_n} \in \mathscr{M}$. By the definition again we have $1_{B_1} - 1_{B_n} \in \mathscr{M}$ and $(1_{B_1} - 1_{B_n}) \uparrow 1_{B_1} - 1_B \in \mathscr{M}$. It follows that $1_B = 1_{B_1} - (1_{B_1} - 1_B) \in \mathscr{M}$ and hence $B \in \mathscr{D}$. Now we have $\mathscr{D} \supseteq \mu(\mathscr{C}) = \sigma(\mathscr{C})$ by Corollary 1.3.1. In other words, $1_A \in \mathscr{M}$ for each $A \in \sigma(\mathscr{C})$. For a non-negative and $\sigma(\mathscr{C})$-measurable function f and $n \geq 1$, let

$$f_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{\{(k-1)/2^n \leq f < k/2^n\}} + n 1_{\{f \geq n\}}.$$

It is easy to see that $f_n \in \mathscr{M}$ and $f_n \uparrow f$. Then $f \in \mathscr{M}$ by the definition of the monotone class. □

Definition 1.4.2 A family $\mathscr{L}$ of non-negative extended real-valued functions on Ω is called a λ-*system* if

(i) $1 \in \mathscr{L}$;

(ii) For any $f_1, f_2 \in \mathscr{L}$ and $a_1, a_2 \in \mathbb{R}$ with $a_1 f_1 + a_2 f_2 \geq 0$ we have $a_1 f_1 + a_2 f_2 \in \mathscr{L}$;

(iii) If $\{f_n\} \subseteq \mathscr{L}$ and $f_n \uparrow f$, then $f \in \mathscr{L}$.

Theorem 1.4.2 *Let $\mathscr{C}$ be a π-class and $\mathscr{L}$ a λ-system. If $\mathscr{L}$ contains the indicators of all sets in $\mathscr{C}$, it contains all non-negative $\sigma(\mathscr{C})$-measurable function.*

Proof. This follows by similar arguments as in the proof of Theorem 1.4.1. (Homework.) □

Given a family $\mathscr{U}$ of non-negative functions on Ω, we may define the following λ-system:

$$\Lambda(\mathscr{U}) = \bigcap \{\mathscr{L} : \mathscr{L} \supseteq \mathscr{U} \text{ is a } \lambda\text{-system}\}, \tag{1.4.1}$$

which is called the λ-*system generated by* $\mathscr{U}$.

Lemma 1.4.1 *For $|x| \leq 1$, define $\{P_n(x)\}$ inductively by $P_0(x) \equiv 0$ and*

$$P_n(x) = P_{n-1}(x) + \frac{1}{2}\big[x^2 - P_{n-1}^2(x)\big]. \tag{1.4.2}$$

Then $P_n(x) \geq 0$ and $P_n(x) \uparrow |x|$ as $n \to \infty$.

Proof. If $0 \le P_{n-1}(x) \le |x| \le 1$, then clearly $P_n(x) \ge P_{n-1}(x)$ and

$$P_n(x) = P_{n-1}(x) + \frac{1}{2}\big[|x| + P_{n-1}(x)\big]\big[|x| - P_{n-1}(x)\big] \le P_{n-1}(x) + \big[|x| - P_{n-1}(x)\big] \le |x|.$$

By induction in $n \ge 0$ we have $0 \le P_n(x) \le |x|$ and $P_n(x) \uparrow$. Let $P(x) = \lim_{n\to\infty} P_n(x)$. From (1.4.2) we obtain $0 = x^2 - P^2(x)$ and hence $P(x) = |x|$. □

Theorem 1.4.3 *Let $\mathscr{U}$ be a family of bounded non-negative functions on Ω which is closed under multiplication. Then $\Lambda(\mathscr{U})$ contains all non-negative $\sigma(\mathscr{U})$-measurable functions.*

Proof. Step 1) For any non-negative function g on Ω, let

$$\mathscr{L}_g = \{f \in \Lambda(\mathscr{U}) : fg \in \Lambda(\mathscr{U})\}.$$

It is easy to show that $\mathscr{L}_g$ is a λ-system for $g \in \Lambda(\mathscr{U})$. (Homework: Prove this fact.)

Step 2) Let $g \in \mathscr{U}$. For any $f \in \mathscr{U}$, we have $fg \in \mathscr{U} \subseteq \Lambda(\mathscr{U})$ and hence $f \in \mathscr{L}_g$. That is, $\mathscr{L}_g \supseteq \mathscr{U}$. Since $\mathscr{L}_g$ is a λ-system, it follows that $\mathscr{L}_g \supseteq \Lambda(\mathscr{U})$.

Step 3) Let $g \in \Lambda(\mathscr{U})$. For any $f \in \mathscr{U}$, we have $\mathscr{L}_f \supseteq \Lambda(\mathscr{U})$ and hence $g \in \mathscr{L}_f$. This implies that $f \in \mathscr{L}_g$. It follows that $\mathscr{L}_g \supseteq \mathscr{U}$. Since $\mathscr{L}_g$ is a λ-system, we have $\mathscr{L}_g \supseteq \Lambda(\mathscr{U})$. Then the definition of $\mathscr{L}_g$ implies that $\Lambda(\mathscr{U})$ is closed under multiplication.

Step 4) For bounded $f, g \in \Lambda(\mathscr{U})$, let us prove $f \wedge g \in \Lambda(\mathscr{U})$. Obviously, we may assume $0 \le f, g \le 1$ and hence $|f - g| \le 1$. From the last step and the definition of the λ-system it follows that

$$(f - g)^2 = f^2 + g^2 - 2fg \in \Lambda(\mathscr{U}).$$

Let $\{P_n(x)\}$ be defined as in Lemma 1.4.1. By induction in $n \ge 0$ we have $P_n(f - g) \in \Lambda(\mathscr{U})$ and hence $|f - g| \in \Lambda(\mathscr{U})$. It then follows that

$$f \wedge g = \frac{1}{2}\big[f + g - |f - g|\big] \in \Lambda(\mathscr{U}).$$

Step 5) Let $\mathscr{F} = \{A \subseteq \Omega : 1_A \in \Lambda(\mathscr{U})\}$. It is easily seen that $\mathscr{F}$ is a λ-class and a π-class, so it is a σ-algebra. Let $f \in \mathscr{U}$ and $\alpha > 0$. Then $(\alpha^{-1}f)\wedge 1 \in \Lambda(\mathscr{U})$ and so $1 - [(\alpha^{-1}f)\wedge 1]^n \in \Lambda(\mathscr{U})$. It follows that

$$1 - \big[(\alpha^{-1}f) \wedge 1\big]^n \uparrow 1_{\{f<\alpha\}} \in \Lambda(\mathscr{U})$$

by the definition of the λ-system. That means that $\{f < \alpha\} \in \mathscr{F}$ and so f is $\mathscr{F}$-measurable. By Proposition 1.1.4 we have $\sigma(\mathscr{U}) \subseteq \mathscr{F}$. That is, $1_A \in \Lambda(\mathscr{U})$ for every $A \in \sigma(\mathscr{U})$.

Step 6) Let $f \ge 0$ be $\sigma(\mathscr{U})$-measurable. For $n \ge 1$ define f_n as in the proof of Theorem 1.4.1. By the last step and the definition of the λ-system it is easy to show that $f_n \in \Lambda(\mathscr{U})$. Since $f_n \uparrow f$, we have $f \in \Lambda(\mathscr{U})$. □

Let Ω be a non-empty set and $(E, \mathscr{E})$ a measurable space. Recall that the σ-algebra generated by a mapping $g : \Omega \to E$ is defined as $\sigma(g) := g^{-1}(\mathscr{E})$.

Theorem 1.4.4 *A function $\phi : (\Omega, \sigma(g)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is measurable if and only if there is a measurable function $f : (E, \mathscr{E}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\phi = f \circ g$.*

Proof. "$\Leftarrow$" Since g is $\sigma(g)/\mathscr{E}$-measurable, this follows from Proposition 1.1.5.

"$\Rightarrow$" The proof of this part is a typical application of the monotone system method. Let $\mathscr{L} = \{f \circ g : f$ is a non-negative and measurable function on $(E, \mathscr{E})\}$. We shall prove that $\mathscr{L}$ is a monotone system containing the indicator of every set in $\sigma(g)$.

(a) Suppose that $A \in \sigma(g)$. Then we have $A = g^{-1}(B)$ for some $B \in \mathscr{E}$. It follows that

$$1_A(\omega) = 1_{g^{-1}(B)}(\omega) = 1_B(g(\omega)) = 1_B \circ g(\omega), \qquad \omega \in \Omega.$$

Thus $\mathscr{L}$ contains the indicators of all sets in $\sigma(g)$. In particular, $1 \in \mathscr{L}$.

(b) Suppose that $\{\phi_1, \phi_2\} \in \mathscr{L}$ and $\{a_1, a_2\} \in \mathbb{R}$ with $a_1\phi_1 + a_2\phi_2 \geq 0$. Then there are non-negative and measurable functions f_1 and f_2 on $(E, \mathscr{E})$ such that $\phi_1 = f_1 \circ g$ and $\phi_2 = f_2 \circ g$. Let $h = a_1 f_1 + a_2 f_2$ and $f = h 1_{\{h \geq 0\}}$. It is easy to see that f is a non-negative and measurable function on $(E, \mathscr{E})$ and

$$a_1\phi_1(\omega) + a_2\phi_2(\omega) = f \circ g(\omega) = h \circ g(\omega), \qquad \omega \in \Omega.$$

Thus we have $a_1\phi_1 + a_2\phi_2 \in \mathscr{L}$.

(c) Suppose that $\{\phi_n\} \subseteq \mathscr{L}$ and $0 \leq \phi_n \uparrow \phi$. For each $n \geq 1$, there is a non-negative and measurable function f_n on $(E, \mathscr{E})$ such that $\phi_n = f_n \circ f$. Let $h = \sup_{n \geq 1} f_n$ and $f = h 1_{\{h < \infty\}}$. Then f is a non-negative and measurable function on $(E, \mathscr{E})$ and $\phi = f \circ g$. It follows that $\phi \in \mathscr{L}$.

We have shown that $\mathscr{L}$ is a monotone system containing the indicator of every set in $\sigma(g)$. By Theorem 1.4.1 we see that $\mathscr{L}$ contains all non-negative $\sigma(g)$ measurable functions on Ω. If ϕ is a measurable function on $(\Omega, \sigma(g))$, both ϕ^+ and ϕ^- are non-negative measurable functions. Then ϕ^+ and $\phi^- \in \mathscr{L}$ by the above arguments. Suppose that $\phi^+ = f_1 \circ g$ and $\phi^- = f_2 \circ g$, where f_1 and f_2 and both non-negative and measurable functions on $(E, \mathscr{E})$. Then $f = f_1 - f_2$ is a measurable function on $(E, \mathscr{E})$ and $\phi = f \circ g$. □

Theorem 1.4.5 *Let (E, ρ) be a metric space and $\mathscr{L}$ a λ-system which contains all bounded, non-negative and continuous functions. Then $\mathscr{L}$ contains all non-negative Borel functions.*

Proof. Clearly, the class $\mathscr{U}$ of bounded, non-negative and continuous functions on E is closed under multiplication. By Proposition 1.2.4 we have $\mathscr{B}(E) = \sigma(\mathscr{U})$. Then Theorem 1.4.3 implies that $\mathscr{L} \supseteq \Lambda(\mathscr{U})$ contains all non-negative $\mathscr{B}(E)$-measurable functions on E. □

Chapter 2

Random Variables and Distributions

2.1 Extension theorem of measures

Let $\mathscr{S}$ be a class of subsets of the non-empty set Ω. A mapping $\mu : \mathscr{S} \to [-\infty, \infty]$ is called a *set function* on $\mathscr{S}$. It is said to be *σ-finite* if there is a sequence $\{A_1, A_2, \cdots\} \subseteq \mathscr{S}$ such that $\bigcup_{n=1}^{\infty} A_n = \Omega$ and $-\infty < \mu(A_n) < \infty$ for each $n \geq 1$. Suppose that μ_1 and μ_2 are set functions on the set classes $\mathscr{S}_1$ and $\mathscr{S}_2$ on Ω, respectively. If $\mathscr{S}_1 \subseteq \mathscr{S}_2$ and $\mu_1(A) = \mu_2(A)$ for every $A \in \mathscr{S}_1$, we say μ_2 is an *extension* of μ_1 on $\mathscr{S}_2$.

Let μ be a set function on the set class $\mathscr{S}$. We say it is *finitely additive* if for any finite sequence of disjoint sets $\{A_1, \cdots, A_n\} \subseteq \mathscr{S}$ such that $A := \bigcup_{k=1}^{n} A_k \in \mathscr{S}$ we have $\mu(A) = \sum_{k=1}^{n} \mu(A_k)$. We say it is *σ-additive* if for any countable sequence of disjoint sets $\{A_1, A_2, \cdots\} \subseteq \mathscr{S}$ such that $A := \bigcup_{k=1}^{\infty} A_k \in \mathscr{S}$ we have $\mu(A) = \sum_{k=1}^{\infty} \mu(A_k)$.

Definition 2.1.1 Let $\mathscr{S}$ be a set class on Ω containing the empty set $\emptyset$. A σ-additive set function μ on $\mathscr{S}$ satisfying $\mu(\emptyset) = 0$ is called a *signed measure*. A non-negative signed measure is called a *measure*. If μ is a measure on the measurable space $(\Omega, \mathscr{F})$, we call $(\Omega, \mathscr{F}, \mu)$ a *measure space*. If $\mu(\Omega) = 1$ in addition, we call $(\Omega, \mathscr{F}, \mu)$ a *probability space*.

Proposition 2.1.1 *Suppose that μ is a signed measure on $(\Omega, \mathscr{F})$ and $\{A_n\} \subseteq \mathscr{F}$. If $A_n \uparrow A$, then $\mu(A_n) \uparrow \mu(A)$. If $\mu(A_1) < \infty$ and $A_n \downarrow A$, then $\mu(A_n) \downarrow \mu(A)$.*

Proof. (Homework.) □

Let (E, ρ) be a metric space. A measure μ on $(E, \mathscr{B}(E))$ is said to be *regular* if

$$\mu(B) = \sup\{\mu(C) : C \subseteq B \text{ is closed}\} = \inf\{\mu(U) : U \supseteq B \text{ is open}\}.$$

Theorem 2.1.1 *A finite measure μ on $(E, \mathscr{B}(E))$ is regular.*

Proof. Let $\mathscr{R}$ be the class of regular sets in $\mathscr{B}(E)$. We need to show that $\mathscr{R} = \mathscr{B}(E)$. We first prove that $\mathscr{R}$ is a σ-algebra. Clearly, $\emptyset$ and $\Omega \in \mathscr{R}$. Suppose that $A \in \mathscr{R}$ and $\varepsilon > 0$. There exists an open set $U_\varepsilon \supseteq A$ and a closed set $C_\varepsilon \subseteq A$ such that

$$\mu(U_\varepsilon) - \varepsilon < \mu(A) < \mu(C_\varepsilon) + \varepsilon.$$

It then follows that

$$\mu(A^c) = \mu(E) - \mu(A) < \mu(E) - \mu(U_\varepsilon) + \varepsilon = \mu(U_\varepsilon^c) + \varepsilon$$

and

$$\mu(A^c) = \mu(E) - \mu(A) > \mu(E) - \mu(C_\varepsilon) - \varepsilon = \mu(C_\varepsilon^c) - \varepsilon.$$

Note that that $U_\varepsilon^c \subseteq A^c \subseteq C_\varepsilon^c$, where U_ε^c is closed and C_ε^c is open. Since $\varepsilon > 0$ is arbitrary, we have $A^c \in \mathscr{R}$. Next we let $\{B_n\} \subseteq \mathscr{R}$ and $B = \bigcup_{n=1}^\infty B_n$. For $\varepsilon > 0$ there are open sets $U_{n,\varepsilon} \supseteq B_n$ and closed sets $C_{n,\varepsilon} \subseteq B_n$ such that

$$\mu(U_{n,\varepsilon} \setminus C_{n,\varepsilon}) < \varepsilon/2^{n+1}, \qquad n \geq 1.$$

Set $U_\varepsilon = \bigcup_{n=1}^\infty U_{n,\varepsilon}$ and $F_\varepsilon = \bigcup_{n=1}^\infty C_{n,\varepsilon}$. We have $\bigcup_{k=1}^n C_{k,\varepsilon} \uparrow F_\varepsilon$. Since μ is finite, there is $n_0 = n_0(\varepsilon) \geq 1$ such that

$$\mu\Big(F_\varepsilon \setminus \bigcup_{n=1}^{n_0} C_{n,\varepsilon}\Big) = \mu(F_\varepsilon) - \mu\Big(\bigcup_{n=1}^{n_0} C_{n,\varepsilon}\Big) < \varepsilon/2.$$

Setting $C_\varepsilon = \bigcup_{n=1}^{n_0} C_{n,\varepsilon}$, we have $C_\varepsilon \subseteq B \subseteq U_\varepsilon$ and

$$\begin{aligned}\mu(U_\varepsilon \setminus C_\varepsilon) &\leq \mu(U_\varepsilon \setminus F_\varepsilon) + \mu(F_\varepsilon \setminus C_\varepsilon) \\ &\leq \mu\Big(\bigcup_{n=1}^\infty (U_{n,\varepsilon} \setminus C_{n,\varepsilon})\Big) + \varepsilon/2 \\ &\leq \sum_{n=1}^\infty \varepsilon/2^{n+1} + \varepsilon/2 = \varepsilon.\end{aligned}$$

This proves that $B \in \mathscr{R}$. If follows that $\mathscr{R}$ is a sub-σ-algebra of $\mathscr{B}(E)$. The theorem will follow if we can show $\mathscr{R}$ contains all closed subsets of E. Let $F \subseteq E$ be closed. Of course, we have $\mu(F) = \sup\{\mu(C) : C \subseteq F \text{ is closed}\}$. For each $n \geq 1$, the set $U_n = \{x \in E : \rho(x, F) < 1/n\}$ is open. On the other hand, we have $F = \bigcap_{n=1}^\infty U_n$ and so $\lim_{n\to\infty} \mu(U_n) = \mu(F)$. It follows that $\mu(F) = \inf\{\mu(U) : U \supseteq F \text{ is open}\}$. □

Corollary 2.1.1 *Suppose that μ and ν are finite measures on $(E, \mathscr{B}(E))$. (i) If $\mu(U) = \nu(U)$ for all open sets $U \subseteq E$, then $\mu(B) = \nu(B)$ for all $B \in \mathscr{B}(E)$; (ii) If $\mu(C) = \nu(C)$ for all closed sets $C \subseteq E$, then $\mu(B) = \nu(B)$ for all $B \in \mathscr{B}(E)$.*

Definition 2.1.2 A class $\mathscr{S}$ of subsets of Ω is called a *semi-algebra* if

(i) $\emptyset \in \mathscr{S}$ and $\Omega \in \mathscr{S}$;

(ii) $A \in \mathscr{S}$ and $B \in \mathscr{S}$ imply $A \cap B \in \mathscr{S}$;

(iii) If $A_1, B \in \mathscr{S}$ and $A_1 \subseteq B$, there is a finite family $\{A_2, \cdots, A_n\} \subseteq \mathscr{S}$ such that $B = \bigcup_{k=1}^n A_k$.

Theorem 2.1.2 (Extension Theorem) *A measure μ on a semi-algebra $\mathscr{S}$ has an extension $\bar{\mu}$ on $\sigma(\mathscr{S})$, that is, $\bar{\mu}$ is a measure on $\sigma(\mathscr{S})$ and $\bar{\mu}(A) = \mu(A)$ for every $A \in \mathscr{S}$. If μ is also σ-finite on $\mathscr{S}$, its extension on $\sigma(\mathscr{S})$ is unique.*

Proof. See Chow and Teicher (1988, pp.159-162). □

2.2 Distributions of random variables

Suppose that $(\Omega, \mathscr{F})$ and $(E, \mathscr{E})$ are measurable spaces and $f : \Omega \to E$ is an $\mathscr{F}/\mathscr{E}$-measurable transformation. Let μ be a measure on $(\Omega, \mathscr{F})$ and let $\mu_f(B) = \mu(f^{-1}(B))$ for $B \in \mathscr{E}$. It is easy to show that μ_f is a measure on $(E, \mathscr{E})$, which is called the *measure induced by* f.

Definition 2.2.1 Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space. A measurable transformation X from $(\Omega, \mathscr{F})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is called a *random variable*. The measure P_X on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ induced by X is called the *distribution* of X.

By Proposition 1.1.7, a real function X defined on $(\Omega, \mathscr{F})$ is a random variable if and only if $X^{-1}((-\infty, b]) \in \mathscr{F}$ for every $b \in \mathbb{R}$.

Definition 2.2.2 A non-decreasing and right continuous function F on $\mathbb{R}$ is called a *distribution function*. Then the limits $F(\infty) := \lim_{x\to\infty} F(x)$ and $F(-\infty) := \lim_{x\to-\infty} F(x)$ exist for a distribution function F. If $F(\infty) = 1$ and $F(-\infty) = 0$ in addition, we call F a *probability distribution function* on $\mathbb{R}$.

In particular, if P_X is the distribution on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ of a one-dimensional random variable X, then $F_X(x) := P_X((-\infty, x])$ defines a probability distribution function F_X, which is called the *distribution function* of X. The following theorem shows that the distribution of a one-dimensional random variable is uniquely determined by its distribution function.

Theorem 2.2.1 *For each distribution function F on $\mathbb{R}$, there is a unique σ-finite measure μ_F on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that*

$$\mu_F((a, b]) = F(b) - F(a), \qquad a \le b \in \mathbb{R}. \tag{2.2.1}$$

To prove the above theorem, we need some lemmas. Set $\mathscr{S} = \{(a, b] : -\infty \le a \le b \le \infty\}$, where $(a, b] = (a, \infty)$ for $b = \infty$ by convention. Clearly, $\mathscr{S}$ is a semi-algebra.

Lemma 2.2.1 *Let $E_0 \in \mathscr{S}$ and let $\{E_k\} \subseteq \mathscr{S}$ be a sequence of disjoint sets such that $E_k \subseteq E_0$. Then we have*

$$\sum_{k=1}^{\infty} \mu_F(E_k) \le \mu_F(E_0).$$

Proof. We first consider the finite subsequence $\{E_1, \cdots, E_n\} \subseteq \mathscr{S}$. Write $E_k = (a_k, b_k]$ for $0 \le k \le n$. By re-enumerating the sequence, we may assume that $a_1 \le a_2 \le \cdots \le a_n$. Since the intervals $\{(a_k, b_k] : k \ge 1\}$ are disjoint and $(a_k, b_k] \subseteq (a_0, b_0]$ for each $k \ge 1$, we have

$$a_0 \le a_1 \le b_1 \le a_2 \le b_2 \le \cdots \le a_n \le b_n \le b_0.$$

It follows that

$$\sum_{k=1}^{n} \mu_F(E_k) = \sum_{k=1}^{n} [F(b_k) - F(a_k)] \le F(b_0) - F(a_0) = \mu_F(E_0).$$

Then we get the desired inequality by letting $n \to \infty$. □

Lemma 2.2.2 *Let $E_0 \in \mathscr{S}$ and $\{E_k\} \subseteq \mathscr{S}$. If $E_0 \subseteq \bigcup_{k=1}^{\infty} E_k$, we have*

$$\mu_F(E_0) \leq \sum_{k=1}^{\infty} \mu_F(E_k).$$

Proof. Write $E_k = (a_k, b_k]$ for $k \geq 0$. We first consider the case where $a_0 < b_0 \in \mathbb{R}$. Choose a constant $0 < \varepsilon < b_0 - a_0$. Set $K_0 = [a_0 + \varepsilon, b_0]$ and $V_k = (a_k, b_k + \varepsilon_k)$, where ε_k is to be specified. Then we have $K_0 \subseteq E_0 \subseteq \bigcup_{k=1}^{\infty} E_k \subseteq \bigcup_{k=1}^{\infty} V_k$. By the Heine-Borel theorem, K_0 has a finite covering $\{V_1, \cdots, V_n\} \subseteq \{V_1, V_2, \cdots\}$. It is easily seen that

$$F(b_0) - F(a_0 + \varepsilon) \leq \sum_{k=1}^{n} [F(b_k + \varepsilon_k) - F(a_k)] \leq \sum_{k=1}^{\infty} [F(b_k + \varepsilon_k) - F(a_k)].$$

Since F is right continuous, we can choose ε_k so that

$$F(b_k + \varepsilon_k) \leq F(b_k) + \varepsilon/2^k.$$

Then we get

$$F(b_0) - F(a_0 + \varepsilon) \leq \sum_{k=1}^{\infty} \{[F(b_k) - F(a_k)] + \varepsilon/2^k\} = \sum_{k=1}^{\infty} [F(b_k) - F(a_k)] + \varepsilon.$$

Letting $\varepsilon \downarrow 0$ we obtain

$$F(b_0) - F(a_0) \leq \sum_{k=1}^{\infty} [F(b_k) - F(a_k)],$$

as desired. In the general case $a_0 < b_0 \in \bar{\mathbb{R}}$, the result follows by a limit procedure. □

Proof of Theorem 2.2.1. By Lemmas 2.2.1 and 2.2.2 we find that μ_F is a measure on the semi-algebra $\mathscr{S}$. Since μ_F is clearly σ-finite, it has a unique extension on $\mathscr{B}(\mathbb{R}) = \sigma(\mathscr{S})$ by Theorem 2.1.2. □

The measure defined by (2.2.1) with $F(x) \equiv x$ is called the *Lebesgue measure* on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Let λ denote the Lebesgue measure. It is easy to see that $\lambda(\{x_0\}) = 0$ for any singleton set $\{x_0\} \subseteq \mathbb{R}$. Consequently, $\lambda(\{x_1, x_2, \cdots\}) = 0$ for any sequence $\{x_1, x_2, \cdots\} \subseteq \mathbb{R}$.

2.3 Examples of distribution functions

Given the finite or countable sets $\{x_1, x_2, \cdots\} \subseteq \mathbb{R}$ and $\{p(x_1), p(x_2), \cdots\} \subseteq (0, \infty)$ such that $\sum_k p(x_k) = 1$, we can define a probability distribution function F by

$$F(x) = \sum_{x_k \leq x} p(x_k) = \sum_k p_i(x_i) 1_{(-\infty, x]}(x_i), \qquad x \in \mathbb{R}, \tag{2.3.1}$$

which is called a *step probability distribution function.* If F is defined by (2.3.1) and if μ_F is the corresponding σ-finite measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given by Theorem 2.2.1, then we have

$$\mu_F(B) = \sum_k p_i(x_i) 1_B(x_i), \qquad B \in \mathscr{B}(\mathbb{R}). \tag{2.3.2}$$

Let λ denote the Lebesgue measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. A probability distribution function F is said to be *singular* if there is $E \in \mathscr{B}(\mathbb{R})$ such that $\lambda(E) = 0$ and $\mu_F(E) = 1$. Clearly, a step distribution function is singular. A probability distribution function F is said to be *absolutely continuous* if

$$F(x) = \int_{-\infty}^{x} p(y)dy, \qquad x \in \mathbb{R}, \tag{2.3.3}$$

where $p(\cdot)$ is a non-negative function on $\mathbb{R}$ such that

$$\int_{-\infty}^{\infty} p(y)dy = 1.$$

(For the moment, we understand the integrals in the Riemannian sense.) In this case, the function $p(\cdot)$ is called the *density* of F.

Example 2.3.1 For fixed $x_0 \in \mathbb{R}$, let

$$F(x) = \begin{cases} 0 & \text{if } x < x_0, \\ 1 & \text{if } x \geq x_0. \end{cases}$$

Then F is a step probability distribution function on $\mathbb{R}$, which gives a *degenerate distribution.* Clearly, we have

$$\mu_F(B) = \begin{cases} 1 & \text{if } x_0 \in B \\ 0 & \text{if } x_0 \notin B. \end{cases}$$

We shall write $P_F = \delta_{x_0}$.

Example 2.3.2 For any parameter $\lambda > 0$, we can define a step probability distribution function F on $\mathbb{R}$ by

$$F(x) = \sum_{i \leq x} e^{-\lambda} \frac{\lambda^i}{i!}, \qquad x \in \mathbb{R},$$

which gives the *Poissonian distribution.*

Example 2.3.3 For any $m \in \mathbb{R}$ and $\sigma > 0$, we can define an absolutely continuous probability distribution function F on $\mathbb{R}$ by

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

which gives the *Gaussian distribution* $N(\mu, \sigma^2)$.

Example 2.3.4 A continuous singular distribution function can be defined as follows. Let

$$\begin{aligned}
G_1 &= \Big(\frac{1}{3}, \frac{2}{3}\Big), \\
G_2 &= \Big(\frac{1}{3^2}, \frac{2}{3^2}\Big) \bigcup \Big(\frac{2}{3} + \frac{1}{3^2}, \frac{2}{3} + \frac{2}{3^2}\Big), \\
&\cdots\cdots \\
G_n &= \bigcup \Big\{ \Big(\sum_{k=1}^{n-1} \frac{a_k}{3^k} + \frac{1}{3^n}, \sum_{k=1}^{n-1} \frac{a_k}{3^k} + \frac{2}{3^n} \Big) : \ a_k = 0 \text{ or } 2 \Big\}.
\end{aligned}$$

Then $G := \bigcup_{n=1}^{\infty} G_n$ is an open subset of $[0,1]$. The closed set $E := [0,1] \setminus G$ is called the *Cantor set*. Let

$$F(u) = \sum_{k=1}^{n-1} \frac{a_k}{2^k} + \frac{1}{2^n}, \qquad \sum_{k=1}^{n-1} \frac{a_k}{3^k} + \frac{1}{3^n} < u < \sum_{k=1}^{n-1} \frac{a_k}{3^k} + \frac{2}{3^n}.$$

We can extend F to a probability distribution function on $\mathbb{R}$ by setting $F(x) = 1$ for $x \geq 1$ and

$$F(x) = \inf\{F(y) : \ y \in G \text{ and } y > x\}$$

for $x < 1$. Clearly, $\mu_F([0,1]) = 1$ and $\mu_F(G) = 0$. It follows that $\mu_F(E) = \mu_F([0,1]) - \mu_F(G) = 1$. On the other hand, we have

$$\lambda(G) = \sum_{n=1}^{\infty} \lambda(G_n) = \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1.$$

Thus $\lambda(E) = \lambda([0,1]) - \lambda(G) = 0$, so F is *singular*. It is not hard to show that F is continuous. (Homework.)

2.4 Multi-dimensional random variables

Definition 2.4.1 Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space. A measurable transformation X from $(\Omega, \mathscr{F})$ to $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ is called a *d-dimensional random variable*. The probability measure P_X on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ induced by X is called the *distribution* of X.

Definition 2.4.2 A real-valued function F on $\mathbb{R}^d$ is called a *probability distribution function* if

(i) $F(x_1, \cdots, x_n) \to 0$ as $x_i \to -\infty$ for some i, and $F(x_1, \cdots, x_n) \to 1$ as $x_i \to \infty$ for all i;

(ii) F is right-continuous in each x_i;

(iii) for all $h_i \geq 0$ and $x_i \in \mathbb{R}$, the following inequality holds:

$$\begin{aligned}
&F(x_1 + h_1, x_2 + h_2, \cdots, x_d + h_d) \\
&\quad -[F(x_1, x_2 + h_2, \cdots, x_d + h_d) + \cdots \\
&\qquad +F(x_1 + h_1, x_2 + h_2, \cdots, x_{d-1} + h_{d-1}, x_d)] \\
&\quad +[F(x_1, x_2, x_3 + h_3, \cdots, x_d + h_d) + \cdots \\
&\qquad +F(x_1 + h_1, x_2 + h_2, \cdots, x_{d-2} + h_{d-2}, x_{d-1}, x_d)] \\
&\quad -\cdots \\
&\quad +(-1)^d F(x_1, x_2, \cdots, x_d) \geq 0.
\end{aligned}$$

The following result can be proved similarly as Theorem 2.2.1.

Theorem 2.4.1 *For each distribution function F on $\mathbb{R}^d$, there is a unique σ-finite measure μ_F on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that*

$$\begin{aligned}
\mu_F((x, x+h]) &= F(x_1+h_1, x_2+h_2, \cdots, x_d+h_d) \\
&\quad -[F(x_1, x_2+h_2, \cdots, x_d+h_d) + \cdots \\
&\qquad +F(x_1+h_1, x_2+h_2, \cdots, x_{d-1}+h_{d-1}, x_d)] \\
&\quad +[F(x_1, x_2, x_3+h_3, \cdots, x_d+h_d) + \cdots \\
&\qquad +F(x_1+h_1, x_2+h_2, \cdots, x_{d-2}+h_{d-2}, x_{d-1}, x_d)] \\
&\quad -\cdots \\
&\quad +(-1)^d F(x_1, x_2, \cdots, x_d) \geq 0.
\end{aligned}$$

for $x \in \mathbb{R}^d$ and $h \in \mathbb{R}^d_+$.

Proposition 2.4.1 *A d-dimensional function $X = (X_1, \cdots, X_d)$ defined on $(\Omega, \mathscr{F}, \mathbf{P})$ is a random variable if and only if each X_j $(1 \leq j \leq d)$ is a random variable.*

Proof. "$\Rightarrow$" Suppose X is a d-dimensional random variable. Then $X^{-1}(B) \in \mathscr{F}$ for each $B \in \mathscr{B}(\mathbb{R}^d)$. Fix $\beta \in \mathbb{R}$ and $1 \leq i \leq d$. Let $B_i = (-\infty, \beta]$ and $B_k = \mathbb{R}$ for all $k \neq i$. Clearly, we have $B = \prod_{k=1}^d B_k \in \mathscr{B}(\mathbb{R}^d)$ and hence $X_i^{-1}((-\infty, \beta]) = X^{-1}(B) \in \mathscr{F}$, proving that X_i is a one-dimensional random variable.

"$\Leftarrow$" Suppose that each X_k is a random variable. For $b \in \mathbb{R}^d$ we have

$$X^{-1}((-\infty, b]) = \bigcap_{k=1}^{d} X_k^{-1}((-\infty, b_k]) \in \mathscr{F}.$$

Thus X is a d-dimensional random variable. □

Definition 2.4.3 Let $X = (X_1, \cdots, X_d)$ be a d-dimensional random variable. The *distribution* P_X of X is a probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ defined by

$$P_X(B) = \mathbf{P}(\{\omega : X(\omega) \in B\}), \qquad B \in \mathscr{B}(\mathbb{R}^d). \tag{2.4.1}$$

The the *distribution function* F_X of X is defined by

$$F_X(x) = P_X\big((-\infty, x]\big), \qquad x \in \mathscr{B}(\mathbb{R}^d). \tag{2.4.2}$$

We sometimes write $P_{X_1, X_2, \cdots, X_d}$ and $F_{X_1, X_2, \cdots, X_d}$ instead of P_X and F_X, respectively. It is easy to show that

$$F_{X_1, X_2, \cdots, X_d}(\infty, x_2, \cdots, x_d) = F_{X_2, \cdots, X_d}(x_2, \cdots, x_d),$$

$$F_{X_1, X_2, \cdots, X_d}(\infty, \infty, \cdots, x_d) = F_{X_3, \cdots, X_d}(x_3, \cdots, x_d),$$

etc. Moreover, for any permutation $\{i_1, i_2, \cdots, i_d\}$ of $\{1, 2, \cdots, d\}$ we have

$$F_{X_{i_1}, X_{i_2}, \cdots, X_{i_d}}(x_{i_1}, x_{i_2}, \cdots, x_{i_d}) = F_{X_1, X_2, \cdots, X_d}(x_1, x_2, \cdots, x_d).$$

2.5 Independence

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a fixed probability space. We assume that all random variables are defined on this space. Let T be a nonempty index set.

Definition 2.5.1 Suppose $F_t \in \mathscr{F}$ for each $t \in T$. We say the events $\{F_t : t \in T\}$ are *independent* if for any finite subset $\{t_1, \cdots, t_n\} \subseteq T$,

$$\mathbf{P}\Big(\bigcap_{j=1}^{n} F_{t_j}\Big) = \prod_{j=1}^{n} \mathbf{P}(F_{t_j}). \tag{2.5.1}$$

Definition 2.5.2 Suppose that $\mathscr{D}_t \subseteq \mathscr{F}$ for each $t \in T$. We say the classes $\{\mathscr{D}_t : t \in T\}$ are *independent* if the events $\{F_t : t \in T\}$ are independent whenever $F_t \in \mathscr{D}_t$ for every $t \in T$.

Clearly, if $\{\mathscr{D}_t : t \in T\}$ are independent event classes and if $\mathscr{C}_t \subseteq \mathscr{D}_t$ for each $t \in T$, then the classes $\{\mathscr{C}_t : t \in T\}$ are also independent.

Definition 2.5.3 Suppose that for each $t \in T$ we have a random variable X_t. Let $\sigma(X_t)$ denote the σ-algebra generated by X_t. We say the random variables $\{X_t : t \in T\}$ are *independent* if $\{\sigma(X_t) : t \in T\}$ are independent classes of events.

Theorem 2.5.1 *Let X and Y be respectively m-dimensional and m-dimensional random variables. Then X and Y are independent if and only if*

$$F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y) \tag{2.5.2}$$

for every $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$.

Proof. If X and Y are independent, by (2.5.1) we have

$$F_{(X,Y)}(x, y) = \mathbf{P}(\{X \leq x\} \cap \{X \leq x\}) = \mathbf{P}\{X \leq x\} \cdot \mathbf{P}\{X \leq x\} = F_X(x) \cdot F_Y(y),$$

proving (2.5.2). Conversely, suppose (2.5.2) holds for every $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$. We shall prove the independence of X and Y in two steps.

Step 1) Let $\mathscr{C} = \{(-\infty, x] : x \in \mathbb{R}^m\}$. Then $\mathscr{C}$ is a π-class on $\mathbb{R}^m$ and $\lambda(\mathscr{C}) = \sigma(\mathscr{C}) = \mathscr{B}(\mathbb{R}^m)$. For fixed $y \in \mathbb{R}^n$ let

$$\mathscr{D} = \{A \in \mathscr{B}(\mathbb{R}^m) : \mathbf{P}\{X \in A, Y \leq y\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \leq y\} \text{ holds}\}.$$

By (2.5.2) we have $\mathscr{C} \subseteq \mathscr{D}$. We claim that $\mathscr{D}$ is a λ-class. Indeed, we clearly have $\mathbb{R}^m \in \mathscr{D}$. Moreover, if $A, B \in \mathscr{D}$ and $A \subseteq B$, we have

$$\begin{aligned}
\mathbf{P}\{X \in B \setminus A, Y \leq y\} &= \mathbf{P}(\{X \in B,\ Y \leq y\} \setminus \{X \in B, Y \leq y\}) \\
&= \mathbf{P}\{X \in B, Y \leq y\} - \mathbf{P}\{X \in A, Y \leq y\} \\
&= \mathbf{P}\{X \in B\} \cdot \mathbf{P}\{Y \leq y\} - \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \leq y\} \\
&= \mathbf{P}\{X \in B \setminus A\} \cdot \mathbf{P}\{Y \leq y\}.
\end{aligned}$$

Consequently, $B \setminus A \in \mathscr{D}$. If $\{A_n\} \subseteq \mathscr{D}$ and $A_n \uparrow A$, we have

$$\begin{aligned} \mathbf{P}\{X \in A, Y \leq y\} &= \lim_{n\to\infty} \mathbf{P}\{X \in A_n, Y \leq y\} \\ &= \lim_{n\to\infty} \mathbf{P}\{X \in A_n\} \cdot \mathbf{P}\{Y \leq y\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \leq y\}, \end{aligned}$$

and hence $A \in \mathscr{D}$. Then $\mathscr{D}$ is a λ-class and the monotone class theorem implies that $\mathscr{D} \supseteq \lambda(\mathscr{C}) = \mathscr{B}(\mathbb{R}^m)$. By the definition of $\mathscr{D}$ we have $\mathscr{D} \subseteq \mathscr{B}(\mathbb{R}^m)$ and so $\mathscr{D} = \mathscr{B}(\mathbb{R}^m)$.

Step 2) Let $\mathscr{E} = \{(-\infty, x] : x \in \mathbb{R}^n\}$. Then $\mathscr{E}$ is a π-class on $\mathbb{R}^n$ and $\lambda(\mathscr{C}) = \sigma(\mathscr{C}) = \mathscr{B}(\mathbb{R}^n)$. Fix $A \in \mathscr{B}(\mathbb{R}^m)$ and let

$$\mathscr{U} = \{B \in \mathscr{B}(\mathbb{R}^n) : \mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\} \text{ holds}\}.$$

The result proved in the last step implies that $\mathscr{E} \subseteq \mathscr{U}$. By repeating the above arguments one shows that $\mathscr{U}$ is a λ-class and so $\mathscr{U} = \mathscr{B}(\mathbb{R}^n)$. In other words,

$$\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}.$$

holds for every $A \in \mathscr{B}(\mathbb{R}^m)$ and $B \in \mathscr{B}(\mathbb{R}^n)$. That proves the theorem. □

Chapter 3

Integration and Mathematical Expectation

3.1 Definition of integrals

Definition 3.1.1 Let $(\Omega, \mathscr{F})$ be a measurable space. An extended real-valued function f on $(\Omega, \mathscr{F})$ is said to be *simple* if there exists a sequence of disjoint sets $\{A_1, \cdots, A_n\} \subseteq \mathscr{F}$ and $\{a_1, \cdots, a_n\} \subseteq \bar{\mathbb{R}}$ such that

$$f(\omega) = \sum_{k=1}^{n} a_k 1_{A_k}(\omega), \qquad \omega \in \Omega, \tag{3.1.1}$$

where $\pm\infty \cdot 0 = 0$ by convention. (We may assume $\bigcup_{k=1}^{n} A_k = \Omega$, so that $\{A_k\}$ is a partition of Ω.)

Proposition 3.1.1 *A simple function is measurable.*

Proof. Suppose f has the representation (3.1.1) with $\bigcup_{k=1}^{n} A_k = \Omega$. For any $B \in \mathscr{B}(\bar{\mathbb{R}})$, let $\{a_{k_1}, \cdots, a_{k_m}\} = B \cap \{a_1, \cdots, a_n\}$. It is easy to see that $f^{-1}(B) = \bigcup_{i=1}^{m} A_{k_i} \in \mathscr{F}$. Then f is measurable. $\square$

Proposition 3.1.2 *(i) A non-negative measurable function is the limit of an increasing sequence of non-negative simple functions; (ii) An extended real-valued measurable function is the limit of a sequence of simple functions.*

Proof. Suppose f is a non-negative measurable function. For $n \geq 1$ and $k \geq 1$, let

$$f_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{\{(k-1)/2^n \leq f < k/2^n\}}(\omega) + n 1_{\{f \geq n\}}(\omega).$$

Clearly, f_n is a simple function and $f_n \uparrow f$. That proves first assertion. The second assertion follows as an immediate consequence. $\square$

Definition 3.1.2 Let $(\Omega, \mathscr{F}, \mu)$ be a σ-finite measure space. The integrals of extended real measurable functions are defined in the following way:

(i) For a simple function $f = \sum_{k=1}^{n} a_k 1_{A_k}$ we define

$$\int_\Omega f d\mu = \sum_{k=1}^{n} a_k \mu(A_k),$$

which is clearly independent of the particular form of the representation of f.

(ii) For a non-negative measurable function f we define

$$\int_\Omega f d\mu = \sup\left\{ \int_\Omega h d\mu : h \text{ is a simple function and } 0 \le h \le f \right\}.$$

(iii) For a measurable function f we set $f^+ = 0 \vee f$ and $f^- = 0 \vee (-f)$. If $\int_\Omega f^+ d\mu < \infty$ or $\int_\Omega f^- d\mu < \infty$, we define

$$\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu.$$

If $\int_\Omega f^+ d\mu + \int_\Omega f^- d\mu < \infty$, we say f is *integrable*. If $\int_\Omega f^+ d\mu = \int_\Omega f^- d\mu = \infty$, we say the integral $\int_\Omega f d\mu$ *does not exist.*

(iv) For $A \in \mathscr{F}$ and a measurable function f we set

$$\int_A f d\mu = \int_\Omega f 1_A d\mu$$

if the integral on the right hand side exists.

The integrals of complex functions can be introduced in the obvious way: If f_1 and f_2 are extended real measurable functions and if $f = f_1 + i f_2$, we define

$$\int_\Omega f d\mu = \int_\Omega f_1 d\mu + i \int_\Omega f_2 d\mu.$$

In this chapter, we shall only discuss integrals of extended real functions. To express explicitly the integration variable, we sometimes write

$$\int_A f d\mu = \int_A f(\omega) d\mu(\omega) = \int_A f(\omega) \mu(d\omega).$$

3.2 Convergence theorems of integrals

In this section, we prove three important convergence theorems of integrals. Let $(\Omega, \mathscr{F}, \mu)$ be a σ-finite measure space.

Theorem 3.2.1 (Monotone Convergence Theorem) *Let f and f_n be non-negative measurable functions on $(\Omega, \mathscr{F}, \mu)$. If $f_n \uparrow f$, then*

$$\int_\Omega f_n d\mu \uparrow \int_\Omega f d\mu.$$

Proof. Since $f_n \leq f_{n+1} \leq f$, it is easy to show that

$$\int_\Omega f_n d\mu \leq \int_\Omega f_{n+1} d\mu \leq \int_\Omega f d\mu$$

and hence

$$a := \lim_{n\to\infty} \int_\Omega f_n d\mu \leq \int_\Omega f d\mu.$$

We will prove the desired result in three steps.

Step 1) Suppose that $\int_\Omega f d\mu < \infty$. For $\epsilon > 0$, let h be a simple function such that $0 \leq h \leq f$ and

$$\int_\Omega f d\mu - \epsilon \leq \int_\Omega h d\mu \leq \int_\Omega f d\mu.$$

Suppose h has the representation $h = \sum_{k=1}^n a_k 1_{A_k}$. Note that if $a_i = \infty$, we must have $\mu(A_i) = 0$. Then we may assume $0 \leq a_i < \infty$ for all $1 \leq i \leq n$. For $0 < c < 1$ let $\Omega_{m,c} = \{\omega \in \Omega : f_m(\omega) \geq ch(\omega)\}$. Then $f_m \geq f_m 1_{\Omega_{m,c}} \geq ch 1_{\Omega_{m,c}}$. It follows that

$$\int_\Omega f_m d\mu \geq \int_\Omega ch 1_{\Omega_{m,c}} d\mu = c \int_\Omega h 1_{\Omega_{m,c}} d\mu. \tag{3.2.1}$$

By Definition 3.1.1,

$$\int_\Omega h 1_{\Omega_{m,c}} d\mu = \int_\Omega \sum_{k=1}^n a_k 1_{A_k \cap \Omega_{m,c}} d\mu = \sum_{k=1}^n a_k \mu(A_k \cap \Omega_{m,c}). \tag{3.2.2}$$

Since $0 \leq h \leq f$ and $f_m \uparrow f$, we have $\Omega_{m,c} \uparrow \Omega$. From (3.2.1) and (3.2.2) it follows that

$$\lim_{m\to\infty} \int_\Omega f_m d\mu \geq \lim_{m\to\infty} c \sum_{k=1}^n a_k \mu(A_k \cap \Omega_{m,c}) = c \sum_{k=1}^n a_k \mu(A_k) = c \int_\Omega h d\mu.$$

Since $0 < c < 1$ is arbitrary, we get

$$\lim_{m\to\infty} \int_\Omega f_m d\mu \geq \int_\Omega h d\mu \geq \int_\Omega f d\mu - \epsilon,$$

and hence

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

Step 2) Suppose that $\int_\Omega f d\mu = \infty$ and $\mu(\{f = \infty\}) = 0$. For $N > 0$, let h be a simple function such that $0 \leq h \leq f$ and $\int_\Omega h d\mu \geq N$. Clearly, $\mu(\{h = \infty\}) = 0$ so we may assume $0 \leq h < \infty$. As in the proof of the last step, we have

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \int_\Omega h d\mu \geq N.$$

It follows that

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \infty = \int_\Omega f d\mu.$$

Step 3) Suppose that $\int_\Omega f d\mu = \infty$ and $b := \mu\{f = \infty\} > 0$. Let $A_{N,n} = \{\omega : f_n(\omega) \geq N\}$. Since $f_n \uparrow f$, we have $A_{N,n} \subseteq A_{N,n+1}$ and $A_N := \bigcup_{n=1}^\infty A_{N,n} \supseteq \{f = \infty\}$. It follows that $\lim_{n\to\infty} \mu(A_{N,n}) = \mu(A_N) \geq b$. In view of the relation

$$\int_\Omega f_n d\mu \geq \int_\Omega f_n 1_{A_{N,n}} d\mu \geq \int_\Omega N 1_{A_{N,n}} d\mu = N\mu(A_{N,n}),$$

we have

$$\lim_{n\to\infty} \int_\Omega f_n d\mu \geq \lim_{n\to\infty} N\mu(A_{N,n}) \geq Nb.$$

Since $N > 0$ is arbitrary, the desired result follows. □

Corollary 3.2.1 *Let f be a measurable function and $\{A_n\} \subseteq \mathscr{F}$ be such that $A_n \uparrow A \in \mathscr{F}$. If $\int_A f d\mu$ exists, then $\int_{A_n} f d\mu$ exists for every $n \geq 1$ and*

$$\int_A f d\mu = \lim_{n\to\infty} \int_{A_n} f d\mu.$$

Proof. (Homework.) □

Corollary 3.2.2 *For any non-negative measurable function f we have*

$$\int_\Omega f d\mu = \int_0^\infty \mu\{f \geq t\} dt$$

where dt denote the integral with respect to the Lebesgue measure.

Proof. Since f is non-negative, we have

$$\begin{aligned}
\int_\Omega f d\mu &= \lim_{n\to\infty} \int_\Omega \sum_{i=1}^\infty \Big(\frac{i-1}{2^n}\Big) 1_{\{(i-1)/2^n \leq f < i/2^n\}} d\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^\infty \Big(\frac{i-1}{2^n}\Big) \mu\Big\{\frac{i-1}{2^n} \leq f < \frac{i}{2^n}\Big\} \\
&= \lim_{n\to\infty} \sum_{i=1}^\infty \sum_{j=1}^{i-1} \frac{1}{2^n} \mu\Big\{\frac{i-1}{2^n} \leq f < \frac{i}{2^n}\Big\} \\
&= \lim_{n\to\infty} \sum_{j=1}^\infty \sum_{i=j+1}^\infty \frac{1}{2^n} \mu\Big\{\frac{i-1}{2^n} \leq f < \frac{i}{2^n}\Big\} \\
&= \lim_{n\to\infty} \sum_{j=1}^\infty \frac{1}{2^n} \mu\Big\{\frac{j}{2^n} \leq f\Big\} \\
&= \lim_{n\to\infty} \int_0^\infty \sum_{j=1}^\infty \mu\Big\{\frac{j}{2^n} \leq f\Big\} 1_{\{(j-1)/2^n \leq t < j/2^n\}} dt \\
&= \int_0^\infty \mu\{t \leq f\} dt,
\end{aligned}$$

where we have used the monotone convergence theorem for two times. □

Proposition 3.2.1 *Suppose that $\alpha \in \mathbb{R}$ and f is a measurable function on $(\Omega, \mathscr{F}, \mu)$. Then we have*

$$\int_\Omega \alpha f d\mu = \alpha \int_\Omega f d\mu.$$

(This means if one of the integrals exists, so does the other and the equality holds.)

Proof. When f is a simple function, the equality follows from Definition 3.1.2. In the case where f is a non-negative measurable function, we have by Proposition 3.1.2 a sequence of simple functions $\{f_n\}$ such that $0 \leq f_n \uparrow f$. If $\alpha \geq 0$, we can use the monotone convergence theorem to the sequence of simple functions $\{\alpha f_n\}$ to see

$$\int_\Omega \alpha f d\mu = \lim_{n\to\infty} \int_\Omega \alpha f_n d\mu = \lim_{n\to\infty} \alpha \int_\Omega f_n d\mu = \alpha \int_\Omega f d\mu.$$

If $\alpha < 0$, we have $(\alpha f)^+ = 0$ and $(\alpha f)^- = -|\alpha| f$. Then we can use Definition 3.1.2 and the result just proved to get

$$\int_\Omega \alpha f d\mu = -\int_\Omega |\alpha| f d\mu = -|\alpha| \int_\Omega f_n d\mu = \alpha \int_\Omega f d\mu.$$

Finally, we obtain the general result by considering the decomposition $\alpha f = (\alpha f)^+ - (\alpha f)^-$ in the two cases $\alpha \geq 0$ and $\alpha < 0$. □

Proposition 3.2.2 *Let g and f be two measurable functions on $(\Omega, \mathscr{F}, \mu)$ such that $f + g$ is well-defined. If*

$$\int_\Omega f d\mu, \quad \int_\Omega g d\mu \quad \text{and} \quad \int_\Omega f d\mu + \int_\Omega g d\mu$$

all exist, then we have

$$\int_\Omega (f+g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu.$$

Proof. Step 1) Suppose that f and g are simple functions given by

$$f = \sum_{i=1}^m a_i 1_{A_i} \quad \text{and} \quad g = \sum_{j=1}^n b_j 1_{B_j}.$$

In this case, we have

$$\begin{aligned}
\int_\Omega (f+g) d\mu &= \int_\Omega \left(\sum_{i=1}^m a_i 1_{A_i} + \sum_{j=1}^n b_j 1_{B_j} \right) d\mu \\
&= \int_\Omega \left(\sum_{i,j} a_i 1_{A_i \cap B_j} + \sum_{i,j} b_j 1_{A_i \cap B_j} \right) d\mu \\
&= \int_\Omega \left[\sum_{i,j} (a_i + b_j) 1_{A_i \cap B_j} \right] d\mu \\
&= \sum_{i,j} (a_i + b_j) \mu(A_i \cap B_j) \\
&= \cdots\cdots \\
&= \int_\Omega f d\mu + \int_\Omega g d\mu.
\end{aligned}$$

Step 2) Suppose that f and g are non-negative measurable functions. By Proposition 3.1.2 there are simple functions f_n and g_n such that $0 \le f_n \uparrow f$ and $0 \le g_n \uparrow g$. By the monotone convergence theorem and the last step,

$$\begin{aligned}\int_\Omega (f+g)d\mu &= \int_\Omega \lim_{n\to\infty}(f_n+g_n)d\mu = \lim_{n\to\infty}\int_\Omega (f_n+g_n)d\mu \\ &= \lim_{n\to\infty}\int_\Omega f_n d\mu + \lim_{n\to\infty}\int_\Omega g_n d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu.\end{aligned}$$

Step 3) In the general case, from the relation

$$(f+g)^+ - (f+g)^- = f+g = f^+ - f^- + g^+ - g^-$$

we obtain

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

Then the result of Step 2) implies that

$$\int_\Omega (f+g)^+ d\mu + \int_\Omega f^- d\mu + \int_\Omega g^- d\mu = \int_\Omega (f+g)^- d\mu + \int_\Omega f^+ d\mu + \int_\Omega g^+ d\mu. \tag{3.2.3}$$

By the assumption, we have $\int f^- d\mu + \int g^- d\mu < \infty$ or $\int f^+ d\mu + \int g^+ d\mu < \infty$. Suppose that

$$\int_\Omega f^- d\mu + \int_\Omega g^- d\mu < \infty. \tag{3.2.4}$$

It is easy to show that $(f+g)^- \le f^- + g^-$. Then (3.2.4) implies that

$$\int_\Omega (f+g)^- d\mu < \infty. \tag{3.2.5}$$

From (3.2.3), (3.2.4) and (3.2.5) we get

$$\int_\Omega (f+g)^+ d\mu - \int_\Omega (f+g)^- d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu + \int_\Omega g^+ d\mu - \int_\Omega g^- d\mu,$$

giving the desired result. □

Corollary 3.2.3 *Let a and b be real numbers and g and f be measurable functions on $(\Omega, \mathscr{F}, \mu)$ such that $af + bg$ is well-defined. If*

$$\int_\Omega f d\mu, \quad \int_\Omega g d\mu \quad \text{and} \quad a\int_\Omega f d\mu + b\int_\Omega g d\mu$$

all exist, then we have

$$\int_\Omega (af+bg)d\mu = a\int_\Omega f d\mu + b\int_\Omega g d\mu.$$

Proof. This follows immediately by Propositions 3.2.2 and 3.2.2. □

Corollary 3.2.4 *Suppose that f is a measurable function on $(\Omega, \mathscr{F}, \mu)$ such that $\int_\Omega f d\mu$ exists. If $\{A_k\}$ is a sequence of disjoint measurable sets and if we let $A = \bigcup_{k=1}^\infty A_k$, then*

$$\int_A f d\mu = \sum_{k=1}^\infty \int_{A_k} f d\mu.$$

Proof. This follows immediately by Corollary 3.2.1 and Proposition 3.2.2. □

Proposition 3.2.3 *Suppose that $g : (\Omega, \mathscr{F}) \to (E, \mathscr{B})$ is a measurable mapping and $f : (E, \mathscr{B}) \to (\bar{\mathbb{R}}, \mathscr{B}(\bar{\mathbb{R}}))$ is a measurable function. Then we have*

$$\int_\Omega f \circ g d\mu = \int_E f d\mu_g,$$

where μ_g is the measure on $(E, \mathscr{B})$ induced by μ and g.

Proof. *Step 1)* Suppose that f is a simple function on $(E, \mathscr{B})$ given by

$$f(x) = \sum_{k=1}^m a_k 1_{A_k}(x), \qquad x \in E.$$

Then we have

$$f \circ g(\omega) = \sum_{k=1}^m a_k 1_{A_k}(g(\omega)) = \sum_{k=1}^m a_k 1_{g^{-1}(A_k)}(\omega),$$

which is a simple function on $(\Omega, \mathscr{F})$. It follows that

$$\int_\Omega f \circ g d\mu = \sum_{k=1}^m a_k \mu(g^{-1}(A_k)) = \sum_{k=1}^m a_k \mu_g(A_k) = \int_E f d\mu_g.$$

Step 2) Suppose that f is a non-negative measurable function. By Proposition 3.1.2 there is a sequence of simple functions $\{f_n\}$ such that $0 \le f_n \uparrow f$. Clearly, each $f_n \circ g$ is a simple function on $(\Omega, \mathscr{F})$ and $0 \le f_n \circ g \uparrow f \circ g$. By the monotone convergence theorem and the result proved in Step 1),

$$\int_\Omega f \circ g d\mu = \lim_{n\to\infty} \int_\Omega f_n \circ g d\mu = \lim_{n\to\infty} \int_E f_n d\mu_g = \int_E f d\mu_g.$$

Step 3) For a general measurable function f, it is easy to show that $(f \circ g)^\pm = f^\pm \circ g$. By the result proved in the last step we have

$$\int_\Omega (f \circ g)^\pm d\mu = \int_\Omega (f^\pm \circ g) d\mu = \int_\Omega f^\pm d\mu_g,$$

yielding the desired result. □

Theorem 3.2.2 (Fatou's Lemma) *Let g and f_n be measurable functions on $(\Omega, \mathscr{F}, \mu)$.*

(i) If g is integrable and $g \le f_n$, then

$$\liminf_{n\to\infty} \int_\Omega f_n d\mu \ge \int_\Omega \liminf_{n\to\infty} f_n d\mu.$$

(ii) If g is integrable and $g \ge f_n$, then

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \le \int_\Omega \limsup_{n\to\infty} f_n d\mu.$$

Proof. We shall only give the proof of (i), which implies (ii) as a consequence. Let $h_n = \inf_{k\ge n} f_k$. Then $0 \le (h_n - g) \uparrow$ and the monotone convergence theorem yields

$$\lim_{n\to\infty} \int_\Omega (h_n - g) d\mu = \int_\Omega \lim_{n\to\infty} (h_n - g) d\mu,$$

that is,

$$\lim_{n\to\infty} \int_\Omega h_n d\mu - \int_\Omega g d\mu = \int_\Omega \lim_{n\to\infty} h_n d\mu - \int_\Omega g d\mu.$$

It follows that

$$\lim_{n\to\infty} \int_\Omega h_n d\mu = \int_\Omega \lim_{n\to\infty} h_n d\mu = \int_\Omega \liminf_{n\to\infty} f_n d\mu. \tag{3.2.6}$$

On the other hand, for any $j \ge n$ we have

$$\int_\Omega f_j d\mu \ge \int_\Omega \inf_{k\ge n} f_k d\mu = \int_\Omega h_n d\mu$$

and hence

$$\inf_{j\ge n} \int_\Omega f_j d\mu \ge \int_\Omega h_n d\mu. \tag{3.2.7}$$

Then we may use (3.2.6) and (3.2.7) to see that

$$\liminf_{n\to\infty} \int_\Omega f_n d\mu = \lim_{n\to\infty} \inf_{j\ge n} \int_\Omega f_j d\mu \ge \lim_{n\to\infty} \int_\Omega h_n d\mu = \int_\Omega \liminf_{n\to\infty} f_n d\mu.$$

That proves the desired result. □

Theorem 3.2.3 (Dominated Convergence Theorem) *Let g and h be integrable functions on $(\Omega, \mathscr{F}, \mu)$. If $\{f_n\}$ is a sequence of measurable functions such that $g \le f_n \le h$ for each n and $f = \lim_{n\to\infty} f_n$, then*

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega \lim_{n\to\infty} f_n d\mu = \int_\Omega f d\mu.$$

Proof. Since $f = \liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n$, we get the result easily by Fatou's lemma. □

Proposition 3.2.4 *If f and g are two measurable functions on $(\Omega, \mathscr{F}, \mu)$ satisfying $f \overset{\text{a.e.}}{=} g$, that is, $\mu(\{f \ne g\}) = 0$, then we have*

$$\int_\Omega f d\mu = \int_\Omega g d\mu.$$

Proof. *Step 1)* Suppose that f and g are simple functions given by

$$f = \sum_{i=1}^{m} a_i 1_{A_i} \quad \text{and} \quad g = \sum_{j=1}^{n} b_j 1_{B_j},$$

where $\{A_i\}$ and $\{B_j\} \subseteq \mathscr{F}$ are two partitions of Ω. If $\mu(A_i \cap B_j) > 0$, then $A_i \cap B_j \neq \emptyset$ and we must have $a_i = b_j$. It follows that

$$\begin{aligned} \int_\Omega f d\mu &= \sum_{i=1}^{m} a_i \mu(A_i) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i \mu(A_i \cap B_j) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{n} b_j \mu(B_j) = \int_\Omega g d\mu. \end{aligned}$$

Step 2) Suppose that f and g are non-negative measurable functions. By Proposition 3.1.2 there are simple functions f_n and g_n such that $0 \le f_n \uparrow f$ and $0 \le g_n \uparrow g$. Let $N = \{\omega : f(\omega) \neq g(\omega)\}$ and $h_n = f_n 1_N + g_n 1_{N^c}$. Then h_n is a simple function and $0 \le h_n \uparrow f 1_N + g 1_{N^c} = f$. Under the assumption, we have $\mu(N) = 0$. It follows that $h_n \overset{\text{a.e.}}{=} g_n$. By the monotone convergence theorem and the result proved in Step 1),

$$\int_\Omega f d\mu = \lim_{n\to\infty} \int_\Omega h_n d\mu = \lim_{n\to\infty} \int_\Omega g_n d\mu = \int_\Omega g d\mu.$$

Step 3) For general measurable functions f and g, from $f \overset{\text{a.e.}}{=} g$ we have $f^\pm \overset{\text{a.e.}}{=} g^\pm$. By the result of the last step,

$$\int_\Omega f^\pm d\mu = \int_\Omega g^\pm d\mu,$$

which implies the desired result. □

Definition 3.2.1 Let h be an extended real-valued function on $(\Omega, \mathscr{F}, \mu)$. If there is an $\mathscr{F}$-measurable function f and a set $N \in \mathscr{F}$ such that $f 1_{N^c} = h 1_{N^c}$ and $\mu(N) = 0$, we say h is μ-a.e. *$\mathscr{F}$-measurable.* If the integral of f exists, we say the *integral of h exists* and define

$$\int_\Omega h d\mu = \int_\Omega f d\mu.$$

If f is μ-integrable, we say h is *μ-integrable.*

Clearly, all the results in this section can be extended to μ-a.e. measurable functions.

Definition 3.2.2 Let F be a distribution function on $\mathbb{R}^d$ and f a measurable function on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Then F defines a unique σ-finite measure μ_F on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ via Theorem 2.4.1. We define the *Lebesgue-Stieltjes integral*

$$\int_{\mathbb{R}^d} f dF = \int_{\mathbb{R}^d} f d\mu_F.$$

Proposition 3.2.5 *Suppose that F is a bounded distribution function on and f is a bounded continuous function on $\mathbb{R}^d$. Then the Lebesgue-Stieltjes integral defined above coincides with the Riemannian-Stieltjes integral.*

Proof. This follows from the dominated convergence theorem. (Homework: Give the details.) □

3.3 Mathematical expectation

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and X a random variable defined on this space. Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel function. Then $g(X)$ is also a random variable. Let F_X denotes the distribution function of X and write

$$\mathbf{E}[g(X)] = \int_{\Omega} g(X(\omega))d\mathbf{P}(\omega) = \int_{\mathbb{R}} g(y)dF_X(y). \tag{3.3.1}$$

Definition 3.3.1 We say that the mathematical expectation of $g(X)$ exists if $\mathbf{E}[|g(X)|] < \infty$. In this case, we call $\mathbf{E}[g(X)]$ the *expectation* of $g(X)$.

Example 3.3.1 Suppose that F_X is a step distribution function with discontinuity points $\{x_1, x_2, \cdots\}$ and jump sizes $\{p(x_1), p(x_2), \cdots\}$. Then $\mathbf{E}[g(X)]$ exists if and only if

$$\sum_{n=1}^{\infty} |g(x_n)|p(x_n) < \infty.$$

In this case, we have

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(y)dF_X(y) = \sum_{n=1}^{\infty} g(x_n)p(x_n).$$

Example 3.3.2 Let F_X be absolutely continuous with density $p(y)$. Then $\mathbf{E}[g(X)]$ exists if and only if $\int_{\mathbb{R}} |g(y)|p(y)dy < \infty$. In this case,

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(y)p(y)dy.$$

Let $\mathscr{L}_1 = \mathscr{L}_1(\Omega, \mathscr{F}, \mathbf{P}) = \{\text{all random variables } X \text{ such that } \mathbf{E}[|X|] < \infty\}$.

Proposition 3.3.1 *(i) For any $X \in \mathscr{L}_1$, we have $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$; (ii) For $X, Y \in \mathscr{L}_1$ and $\alpha, \beta \in \mathbb{R}$, we have $\alpha X + \beta Y \in \mathscr{L}_1$ and $\mathbf{E}[\alpha X + \beta Y] = \alpha\mathbf{E}[X] + \beta\mathbf{E}[Y]$; (iii) If $\mathbf{E}[|X|^p] < \infty$ for some $p > 0$, then $\mathbf{E}[|X|^q] < \infty$ for $0 \leq q \leq p$.*

Proof. (Homework.) □

Proposition 3.3.2 *Let X be a random variable and let $p > 0$. If $\mathbf{E}[|X|^p] < \infty$, then $x^p\mathbf{P}\{|X| > x\} \to 0$ as $x \to \infty$.*

Proof. Let F denote the distribution function of X. Observe that

$$\mathbf{E}[|X|^p] = \int_{\{|X|\geq x\}} |X|^p d\mathbf{P} + \int_{\{|X|<x\}} |X|^p d\mathbf{P}.$$

Under our assumption, the second term converges to $\mathbf{E}[|X|^p]$ as $x \to \infty$. Then we have

$$0 = \lim_{x\to\infty} \int_{\{|X|\geq x\}} |X|^p d\mathbf{P} = \lim_{x\to\infty} x^p \int_{\{|X|\geq x\}} (|X|/x)^p d\mathbf{P} \geq \limsup_{x\to\infty} x^p \mathbf{P}\{|X| \geq x\},$$

giving the desired result. □

Proposition 3.3.3 *Let X be a non-negative random variable with distribution function F. Then we have*

$$\mathbf{E}[X] = \int_0^\infty [1 - F(x)]dx = \int_0^\infty \mathbf{P}\{X > x\}dx = \int_0^\infty \mathbf{P}\{X \geq x\}dx. \tag{3.3.2}$$

Proof. Note that

$$\mathbf{E}[X] = \int_0^\infty x dF(x) = \lim_{n\to\infty} \int_0^n x dF(x).$$

By integration by parts,

$$\begin{aligned} \int_0^n x dF(x) &= nF(n) - \int_0^n F(x)dx \\ &= -n[1 - F(n)] + \int_0^n [1 - F(x)]dx \\ &= -n\mathbf{P}\{X > n\} + \int_0^n [1 - F(x)]dx. \end{aligned} \tag{3.3.3}$$

If $\mathbf{E}[X] < \infty$, Proposition 3.3.2 implies $n\mathbf{P}\{X > n\} \to 0$ as $n \to \infty$. Then we obtain the first equality in (3.3.2) by letting $n \to \infty$ in (3.3.3). On the other hand, if

$$\int_0^\infty [1 - F(x)]dx < \infty,$$

we get from (3.3.3) that

$$\int_0^n x dF(x) \leq \int_0^n [1 - F(x)]dx \leq \int_0^\infty [1 - F(x)]dx < \infty.$$

Letting $n \to \infty$ gives

$$\mathbf{E}[X] = \int_0^\infty x dF(x) \leq \int_0^\infty [1 - F(x)]dx < \infty.$$

Then we obtain the first equality in (3.3.2) again from (3.3.3). The second equality is immediate. The last equality holds since $\mathbf{P}\{X > x\} \neq \mathbf{P}\{X \geq x\}$ for at most countably many points. □

Corollary 3.3.1 *For any random variable X with distribution function F,*

$$\mathbf{E}[|X|] = \int_0^\infty [1 - F(x)]dx + \int_{-\infty}^0 F(x)dx.$$

If $\mathbf{E}[|X|] < \infty$, we have

$$\mathbf{E}[X] = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

Proof. Recall that $X^+ = 0 \vee X$ and $X^- = 0 \vee (-X)$. Let $Y = -X^-$. Let $F^\pm$ and G denote the distribution functions of $X^\pm$ and Y, respectively. We have

$$F^+(x) = \begin{cases} F(x) & x \geq 0 \\ 0 & x < 0 \end{cases}, \qquad G(x) = \begin{cases} 1 & x \geq 0 \\ F(x) & x < 0. \end{cases}$$

Then Proposition 3.3.3 implies that

$$\mathbf{E}[X^+] = \int_0^\infty [1 - F(x)]dx$$

and

$$\begin{aligned} \mathbf{E}[X^-] &= \int_0^\infty \mathbf{P}\{X^- \geq x\}dx = \int_0^\infty \mathbf{P}\{Y \leq -x\}dx \\ &= \int_0^\infty G(-x)dx = -\int_0^{-\infty} G(y)dy = \int_{-\infty}^0 F(y)dy. \end{aligned}$$

Since $|X| = X^+ + X^-$ and $X = X^+ - X^-$, we have the desired result. □

Corollary 3.3.2 *Let X be a random variable and let $0 < p < \infty$. Then*

$$\mathbf{E}[|X|^p] = \int_0^\infty \mathbf{P}\{|X| \geq x^{1/p}\}dx < \infty$$

if and only if

$$\sum_{n=1}^\infty \mathbf{P}\{|X| \geq n^{1/p}\} < \infty.$$

Proof. By Proposition 3.3.3, we have

$$\mathbf{E}[|X|^p] = \int_0^\infty \mathbf{P}\{|X|^p \geq x\}dx = \int_0^\infty \mathbf{P}\{|X| \geq x^{1/p}\}dx.$$

Since $x \mapsto \mathbf{P}\{|X| \geq x^{1/p}\}$ is non-increasing, the desired result follows. □

Corollary 3.3.3 *Let X be a random variable and let $0 < p < \infty$. If $n^p\mathbf{P}\{|X| \geq n\} \to 0$ as $n \to \infty$, we have $\mathbf{E}[|X|^q] < \infty$ for any $0 \leq q < p$.*

Proof. By Proposition 3.3.3 it follows that

$$\begin{aligned}\mathbf{E}[|X|^q] &= \int_0^\infty \mathbf{P}\{|X| \geq x^{1/q}\}dx = \int_0^\infty \mathbf{P}\{|X| \geq y\}qy^{q-1}dy \\ &= q\int_0^\infty y^{-1-(p-q)}y^p\mathbf{P}\{|X| \geq y\}dy. \end{aligned} \tag{3.3.4}$$

Under the condition, we have $y^p\mathbf{P}\{|X| \geq y\} \to 0$ as $y \to \infty$. Then the right hand side of (3.3.4) is finite. □

3.4 Some inequalities

A real-valued function ϕ defined on an open interval $I \subseteq \mathbb{R}$ is said to be *convex* if

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y), \qquad x, y \in I, 0 \leq \lambda \leq 1.$$

Lemma 3.4.1 *If ϕ is a convex function on $I \subseteq \mathbb{R}$, it has left and right derivatives $\phi_l'(x)$ and $\phi_r'(x)$ at every $x \in I$ and*

$$\phi_l'(x) \leq \phi_r'(x) \leq \phi_l'(y) \leq \phi_r'(y), \qquad x \leq y \in I.$$

Proof. Suppose that $x_1 < x_2 < x < y \in I$. It is easy to show that

$$\frac{\phi(x_1) - \phi(x)}{x_1 - x} \leq \frac{\phi(x_2) - \phi(x)}{x_2 - x} \leq \frac{\phi(y) - \phi(x)}{y - x}.$$

Then the left derivative exists

$$\phi_l'(x) = \uparrow \lim_{z \uparrow x} \frac{\phi(z) - \phi(x)}{z - x} \leq \frac{\phi(y) - \phi(x)}{y - x}.$$

The remaining assertions hold by similar arguments. □

Theorem 3.4.1 (Jessen's inequality) *Let X be a random variable and ϕ a convex function on $\mathbb{R}$ such that $\mathbf{E}[X]$ and $\mathbf{E}[\phi(X)]$ exit. Then $\phi(\mathbf{E}[X]) \leq \mathbf{E}[\phi(X)]$.*

Proof. By Lemma 3.4.1, the convex function ϕ is continuous on $\mathbb{R}$. The convexity implies that

$$\phi(z) - \phi(y) \geq \phi_r'(y)(z - y).$$

In particular, we have a.s.

$$\phi(X) - \phi(\mathbf{E}[X]) \geq \phi_r'(\mathbf{E}[X])(X - \mathbf{E}[X]).$$

Taking the expectation in both sides we get the result. □

Corollary 3.4.1 *If $p \geq 1$ and $\mathbf{E}[|X|^p] < \infty$, then $|\mathbf{E}[X]|^p \leq (\mathbf{E}[|X|])^p \leq \mathbf{E}[|X|^p]$.*

Proposition 3.4.1 *Let X be a random variable and $g(\cdot)$ is a non-negative, Borel and even function which is non-decreasing on $[0,\infty)$. Then for any $\epsilon > 0$ we have*

$$\mathbf{P}\{|X| > \epsilon\} \le \frac{\mathbf{E}[g(X)]}{g(\epsilon)}.$$

Proof. Let $E = \{\omega \in \Omega, |X(\omega)| \ge \epsilon\}$. Then

$$\mathbf{E}[g(X)] = \int_E g(X(\omega))d\mathbf{P}(\omega) + \int_{E^c} g(X(\omega))d\mathbf{P}(\omega)$$

For each $\omega \in E$, we have $|X(\omega)| \ge \epsilon$, so $g(X(\omega)) \ge g(\epsilon)$. It follows that

$$\mathbf{E}[g(X)] \ge \int_E g(X(\omega))d\mathbf{P}(\omega) + 0 = g(\epsilon)\mathbf{P}(E),$$

giving the desired inequality. □

Corollary 3.4.2 (Chebyshev/Markov) *For any $p > 0$ and $\epsilon > 0$,*

$$\mathbf{P}\{|X| > \epsilon\} \le \frac{\mathbf{E}[|X|^p]}{\epsilon^p}.$$

Theorem 3.4.2 (Hölder) *Suppose $1 < p, q < \infty$ and $1/p + 1/q = 1$. Let X and Y be random variables such that $\mathbf{E}[|X|^p] + \mathbf{E}[|X|^q] < \infty$. Then we have*

$$\mathbf{E}[|XY|] \le \{\mathbf{E}[|X|^p]\}^{1/p}\{\mathbf{E}[|Y|^q]\}^{1/q}. \tag{3.4.1}$$

Proof. We first show

$$\alpha^{1/p}\beta^{1/q} \le \frac{\alpha}{p} + \frac{\beta}{q}, \qquad \alpha \ge 0, \beta \ge 0. \tag{3.4.2}$$

Indeed, (3.4.2) is equivalent to

$$\frac{1}{p}\ln\alpha + \frac{1}{q}\ln\beta \le \ln\Big(\frac{\alpha}{p} + \frac{\beta}{q}\Big),$$

which is obvious since $\ln x$ is a concave function. To show (3.4.1), we may certainly assume $\mathbf{E}[|X|^p]\mathbf{E}[|X|^q] \ne 0$. For any $\omega \in \Omega$, we set

$$\alpha = \frac{|X(\omega)|^p}{\mathbf{E}[|X|^p]} \quad \text{and} \quad \beta = \frac{|X(\omega)|^q}{\mathbf{E}[|X|^q]}$$

in (3.4.2) to get

$$\frac{|X(\omega)Y(\omega)|}{\{\mathbf{E}[|X|^p]\}^{\frac{1}{p}}\{\mathbf{E}[|X|^q]\}^{\frac{1}{q}}} \le \frac{|X(\omega)|^p}{p\mathbf{E}[|X|^p]} + \frac{|Y(\omega)|^q}{q\mathbf{E}[|Y|^q]}$$

It follows that

$$|X(\omega)Y(\omega)| \le \{\mathbf{E}[|X|^p]\}^{\frac{1}{p}}\{\mathbf{E}[|X|^q]\}^{\frac{1}{q}}\left(\frac{|X(\omega)|^p}{p\mathbf{E}[|X|^p]} + \frac{|Y(\omega)|^q}{q\mathbf{E}[|Y|^q]}\right)$$

Taking the expectations in both sides, we obtain the desired inequality. □

Corollary 3.4.3 (Schwarz) *For any random variables X and Y, we have*

$$\{\mathbf{E}[|XY|]\}^2 \le \mathbf{E}[X^2]\mathbf{E}[Y^2].$$

Theorem 3.4.3 (Minkowsky) *Let $1 \le p < \infty$. Then*

$$\{\mathbf{E}[|X+Y|^p]\}^{1/p} \le \{\mathbf{E}[|X|^p]\}^{1/p} + \{\mathbf{E}[|Y|^p]\}^{1/p}. \tag{3.4.3}$$

Proof. It suffices to assume $1 < p < \infty$ and $\mathbf{E}[|X|^p + |X|^q] < \infty$. Let q be such that $1/p + 1/q = 1$. By Hölder's Inequality,

$$\begin{aligned} \mathbf{E}[|X+Y|^p] &\le \mathbf{E}[|X+Y|^{p-1}|X|] + \mathbf{E}[|X+Y|^{p-1}|Y|] \\ &\le \{\mathbf{E}[|X+Y|^{(p-1)q}]\}^{1/q}(\{\mathbf{E}[|X|^p]\}^{1/p} + \{\mathbf{E}[|Y|^p]\}^{1/p}). \end{aligned}$$

From $1/p + 1/q = 1$ we get $p = (p-1)q$. Then (3.4.3) follows. □

Chapter 4

Product Spaces

4.1 Product measurable spaces

Let E and F be two non-empty sets. We call $E \times F := \{(x, y) : x \in E, y \in F\}$ the *Cartesian* product of E and F. For $A \subseteq E$ and $B \subseteq F$, we call $A \times B$ a *rectangle* in $E \times F$ and refer A and B as its *sides*. A typical product space is the Euclidean plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

The *section* of $C \subseteq E \times F$ determined by $x_0 \in E$ refers to the set $C_{x_0} := \{y \in F : (x_0, y) \in C\} \subseteq F$. The *section* of a function f on $E \times F$ determined by $x_0 \in E$ refers to the the function f_{x_0} on F defined by $f_{x_0}(y) := f(x_0, y)$ for $y \in F$. Similarly, we define the sections C^{y_0} and f^{y_0} determined by $y_0 \in F$. In particular, we have $(A \times B)_{x_0} = B$ or $\emptyset$ according as $x_0 \in A$ or $\notin A$, and $(A \times B)^{y_0} = A$ or $\emptyset$ according as $y_0 \in B$ or $\notin B$.

Let $(E, \mathscr{E})$ and $(F, \mathscr{F})$ be two measurable spaces. If $A \in \mathscr{E}$ and $B \in \mathscr{F}$, we call $A \times B$ a *measurable rectangle*. It is easy to check that the class $\mathscr{R}$ of all measurable rectangles on $E \times F$ is a π-class. We call $\mathscr{E} \times \mathscr{F} := \sigma(\mathscr{R})$ the *product σ-algebra* on $E \times F$.

Theorem 4.1.1 *If $C \in \mathscr{E} \times \mathscr{F}$, we have $C_{x_0} \in \mathscr{F}$ for each $x_0 \in E$, and $C^{y_0} \in \mathscr{E}$ for each $y_0 \in F$.*

Proof. Let $x_0 \in E$ be fixed and let $\mathscr{A} = \{C \subseteq E \times F : C_{x_0} \in \mathscr{F}\}$. If $C = A \times B$ for $A \in \mathscr{E}$ and $B \in \mathscr{F}$, then $C_{x_0} = B$ or ϕ according as $x_0 \in A$ or $\in A^c$. In both cases, we have $C_{x_0} \in \mathscr{F}$. It follows that $\mathscr{A} \supseteq \mathscr{R}$. Furthermore, it is easy to show that $(D^c)_{x_0} = (D_{x_0})^c$ for $D \subseteq E \times F$ and $\left(\bigcup_{n=1}^{\infty} C_n\right)_{x_0} = \bigcup_{n=1}^{\infty} (C_n)_{x_0}$ for $\{C_n\} \subseteq E \times F$. Then $\mathscr{A}$ is closed under the operations of complements and countable union. Thus $\mathscr{A}$ is a σ-algebra, and hence $\mathscr{A} \supseteq \sigma(\mathscr{R}) = \mathscr{E} \times \mathscr{F}$. That shows $C_{x_0} \in \mathscr{F}$ for every $C \in \mathscr{E} \times \mathscr{F}$. The proof of the second assertion is similar. □

Theorem 4.1.2 *Let f be an extended real-valued measurable function on $(E \times F, \mathscr{E} \times \mathscr{F})$. Then f_{x_0} is an $\mathscr{F}$-measurable function for each $x_0 \in E$, and f^{y_0} is and $\mathscr{E}$-measurable function for each $y_0 \in F$.*

Proof. This follows from Theorem 4.1.1 and the identities

$$(f_{x_0})^{-1}(D) = \left(f^{-1}(D)\right)_{x_0} \quad \text{and} \quad (f^{y_0})^{-1}(D) = \left(f^{-1}(D)\right)^{y_0}$$

for $D \in \mathscr{B}(\bar{\mathbb{R}})$. (Homework: Prove one of the above equalities.) □

Proposition 4.1.1 *If E and F are separable metric spaces, we have $\mathscr{B}(E\times F) = \mathscr{B}(E)\times\mathscr{B}(F)$.*

Proof. (Homework; see e.g. Cohn (1980).) □

The definition of product σ-algebras can be generalized to higher dimensions in a obvious way. With such an extension, one may use Proposition 4.1.1 inductively to see that $\mathscr{B}(\mathbb{R}^d) = \mathscr{B}(\mathbb{R})^d$ for any integer $d \geq 1$. (Homework.)

4.2 Product measures and Fubini's theorem

Let us consider two σ-finite measure spaces $(E, \mathscr{E}, \mu)$ and $(F, \mathscr{F}, \nu)$.

Proposition 4.2.1 *If $C \in \mathscr{E} \times \mathscr{F}$, then $x \mapsto \nu(C_x)$ is an $\mathscr{E}$-measurable function on E, and $y \mapsto \nu(C^y)$ is an $\mathscr{F}$-measurable function on F.*

Proof. We only give the proof of the first assertion since the second one follows by symmetry. By Theorem 4.1.1 we have $C_x \in \mathscr{F}$ for each $x \in E$. Suppose that ν is a finite measure and let

$$\mathscr{A} = \{C \in \mathscr{E} \times \mathscr{F} : x \mapsto \nu(C_x) \text{ is } \mathscr{E}\text{-measurable}\}.$$

One sees easily that $\mathscr{A}$ is a λ-class. For $A \in \mathscr{E}$ and $B \in \mathscr{F}$, we have $\nu((A \times B)_x) = 1_A(x)\nu(B)$ and so $A\times B \in \mathscr{A}$. Then $\mathscr{A}$ contains the π-class $\mathscr{R}$ of measurable rectangles. From the monotone class theorem it follows that $\mathscr{A} \supseteq \sigma(\mathscr{R}) = \mathscr{E}\times\mathscr{F}$. That is, $x \mapsto \nu(C_x)$ is $\mathscr{E}$-measurable for every $C \in \mathscr{E} \times \mathscr{F}$. In the case where ν is a general σ-finite measure, there is a partition $\{F_n\} \subseteq \mathscr{F}$ of F such that $\nu(F_n) < \infty$ for every $n \geq 1$. For $n \geq 1$ we define the finite measure ν_n on $(F, \mathscr{F})$ by $\nu_n(B) = \nu(B \cap F_n)$ for $B \in \mathscr{F}$. Then $\nu = \sum_{n=1}^{\infty} \nu_n$. By the result proved above, $x \mapsto \nu_n(C_x)$ is $\mathscr{E}$-measurable for every every $n \geq 1$. It follows that $x \mapsto \nu(C_x) = \sum_{n=1}^{\infty} \nu_n(C_x)$ is $\mathscr{E}$-measurable. □

Proposition 4.2.2 *For each $C \in \mathscr{E} \times \mathscr{F}$, we have*

$$\int_E \nu(C_x)\mu(dx) = \int_F \mu(C^y)\nu(dy). \tag{4.2.1}$$

Proof. We first consider the case where both μ and ν are finite measures. Let $\mathscr{C} = \{C \in \mathscr{E}\times\mathscr{F} :$ the equality (4.2.1) holds$\}$. By the properties of integrals it is easy to show that $\mathscr{C}$ is a λ-class. If $C = A \times B$ for $A \in \mathscr{E}$ and $B \in \mathscr{F}$, then $\nu(C_x) = 1_A(x)\nu(B)$ and $\mu(C^y) = 1_B(y)\mu(A)$ so that both sides of (4.2.1) are equal to $\mu(A)\nu(B)$. Then $\mathscr{C} \supseteq \mathscr{R}$ so that $\mathscr{C} \supseteq \sigma(\mathscr{R}) = \mathscr{E} \times \mathscr{F}$. If μ and ν are general σ-finite measures, there are partitions $\{E_n\} \subseteq \mathscr{E}$ and $\{F_n\} \subseteq \mathscr{F}$ of E and F, respectively, such that $\nu(E_n)+\nu(F_n) < \infty$ for every $n \geq 1$. For $n \geq 1$ define $\mu_n(A) = \mu(A\cap E_n)$ for $A \in \mathscr{E}$ and define $\nu_n(B) = \nu(B \cap F_n)$ for $B \in \mathscr{F}$. Then μ_n and ν_n are finite measures on $(E, \mathscr{E})$ and $(F, \mathscr{F})$, respectively. By the result proved above,

$$\int_{E_m} \nu_n(C_x)\mu_m(dx) = \int_{F_n} \mu_m(C^y)\nu_n(dy).$$

Taking the summations over $m, n \geq 1$ we obtain (4.2.1). □

Theorem 4.2.1 *There is a unique σ-finite measure λ on $(E \times F, \mathscr{E} \times \mathscr{F})$ such that*

$$\lambda(A \times B) = \mu(A)\nu(B), \qquad A \in \mathscr{E}, B \in \mathscr{F}. \tag{4.2.2}$$

Furthermore, for each $C \in \mathscr{E} \times \mathscr{F}$ we have

$$\lambda(C) = \int_E \nu(C_x)\mu(dx) = \int_F \mu(C^y)\nu(dy). \tag{4.2.3}$$

Proof. By Proposition 4.2.2 and the monotone convergence theorem of integrals it is easy to show that (4.2.3) defines a non-negative and σ-additive set function λ satisfying $\lambda(\emptyset) = 0$. That is, λ is a measure on $\mathscr{E} \times \mathscr{F}$. Clearly, (4.2.2) holds. Now suppose μ and ν are finite and γ is a measure on $\mathscr{E} \times \mathscr{F}$ such that

$$\gamma(A \times B) = \mu(A)\nu(B), \qquad A \in \mathscr{E}, B \in \mathscr{F}.$$

Let $\mathscr{C} = \{C \in \mathscr{E} \times \mathscr{F} : \lambda(C) = \gamma(C)\}$. It is easy to show that $\mathscr{C}$ is a λ-class and $\mathscr{C} \supseteq \mathscr{R}$. Then $\mathscr{C} \supseteq \sigma(\mathscr{R}) = \mathscr{E} \times \mathscr{F}$, which gives the uniqueness assertion. The general case can be treated by the decomposition arguments. □

The measure λ defined by (4.2.3) is called the *product* of μ and ν is denoted by $\mu \times \nu$.

Corollary 4.2.1 (Fubini) *The following conditions are equivalent*

(i) $\mu \times \nu(C) = 0$;

(ii) $\nu(C_x) = 0$ *for* μ-a.e. $x \in E$;

(iii) $\mu(C^y) = 0$ *for* ν-a.e. $y \in F$.

Theorem 4.2.2 (Fubini) *Let f be a non-negative extended real-valued measurable function on $(E \times F, \mathscr{E} \times \mathscr{F})$. Then the functions*

$$x \mapsto \int_F f(x,y)\nu(dy) \quad \text{and} \quad y \mapsto \int_E f(x,y)\nu(dx) \tag{4.2.4}$$

are $\mathscr{E}$- and $\mathscr{F}$-measurable, respectively. Moreover, we have

$$\int_{E\times F} f d(\mu \times \nu) = \int_E \left[\int_F f(x,y)\nu(dy)\right]\mu(dx) = \int_F \left[\int_E f(x,y)\mu(dx)\right]\nu(dy). \tag{4.2.5}$$

Proof. Let $C \in \mathscr{E} \times \mathscr{F}$. By Proposition 4.2.1 the function

$$x \mapsto \nu(C_x) = \int_F 1_{C_x}(y)\nu(dy) = \int_F 1_C(x,y)\nu(dy)$$

is $\mathscr{E}$-measurable. That is, the first function in (4.2.4) is $\mathscr{E}$-measurable if $f = 1_C$. By the linearity of the integral, it is $\mathscr{E}$-measurable if f is a $\mathscr{E} \times \mathscr{F}$-measurable simple function. Then we get the $\mathscr{E}$-measurability of the first function in (4.2.4) by approximating the general non-negative $\mathscr{E} \times \mathscr{F}$-measurable function by an increasing sequence of non-negative simple functions. By Theorem 4.2.1 we have

$$\int_{E\times F} 1_C d(\mu \times \nu) = \mu \times \nu(C) = \int_E \nu(C_x)\mu(dx) = \int_E \left[\int_F 1_C(x,y)\nu(dy)\right]\mu(dx).$$

Thus the first equality in (4.2.5) holds if $f = 1_C$. By the linearity of the integrals, it also holds if f is a $\mathscr{E} \times \mathscr{F}$-measurable simple function. For a general non-negative $\mathscr{E} \times \mathscr{F}$-measurable function the equality follows by the monotone convergence theorem. Other assertions hold by symmetry. ☐

Theorem 4.2.3 (Fubini) *Let f be an integrable extended real-valued function on the product measure space $(E \times F, \mathscr{E} \times \mathscr{F}, \mu \times \nu)$. Then for μ-a.e. $x \in E$ the section f_x is a integrable function on $(F, \mathscr{F}, \nu)$, and for ν-a.e $y \in F$ the section f^y is a integrable function on $(E, \mathscr{E}, \mu)$. Moreover, the functions*

$$x \mapsto \int_F f(x,y)\nu(dy) \quad \text{and} \quad y \mapsto \int_E f(x,y)\nu(dx)$$

are a.e. $\mathscr{E}$- and $\mathscr{F}$-measurable, respectively, and

$$\int_{E\times F} f d(\mu \times \nu) = \int_E \left[\int_F f(x,y)\nu(dy)\right]\mu(dx) = \int_F \left[\int_E f(x,y)\mu(dx)\right]\nu(dy). \tag{4.2.6}$$

Proof. By Theorem 4.2.2, we have

$$\begin{aligned} \infty > \int_{E\times F} |f| d(\mu \times \nu) &= \int_E \left[\int_F |f(x,y)|\nu(dy)\right]\mu(dx) \\ &= \int_E \left[\int_F (f^+(x,y) + f^-(x,y))\nu(dy)\right]\mu(dx) \\ &= \int_E \left[\int_F f^+(x,y)\nu(dy)\right]\mu(dx) + \int_E \left[\int_F f^-(x,y)\nu(dy)\right]\mu(dx). \end{aligned}$$

Then for μ-a.e. $x \in E$ we have

$$\int_F |f(x,y)|\nu(dy) = \int_F f^+(x,y)\nu(dy) + \int_F f^-(x,y)\nu(dy) < \infty.$$

In other words, f^x is a integrable function on $(F, \mathscr{F}, \nu)$ for μ-a.e. $x \in E$. It follows that

$$x \mapsto \int_F f(x,y)\nu(dy) = \int_F f^+(x,y)\nu(dy) - \int_F f^-(x,y)\nu(dy)$$

is a μ-a.e. well-defined measurable function. By Theorem 4.2.2 we have

$$\begin{aligned} &\int_E \left[\int_F f(x,y)\nu(dy)\right]\mu(dx) \\ = &\int_E \left[\int_F f^+(x,y)\nu(dy)\right]\mu(dx) - \int_E \left[\int_F f^-(x,y)\nu(dy)\right]\mu(dx) \\ = &\int_{E\times F} f^+ d(\mu \times \nu) - \int_{E\times F} f^- d(\mu \times \nu) \\ = &\int_{E\times F} f d(\mu \times \nu), \end{aligned}$$

which gives the first equality in (4.2.6). Other assertions follow by similar arguments. ☐

The theory of product measures plays a very important role in probability theory. This can be seen from the following situation. Suppose that μ and ν are two probability measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. We can construct a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ by setting $\Omega = \mathbb{R}$, $\mathscr{F} = \mathscr{B}(\mathbb{R}^2)$ and $\mathbf{P} = \mu \times \nu$. For $\omega = (\omega_1, \omega_2) \in \Omega$ let $X_1(\omega) = \omega_1$ and $X_2(\omega) = \omega_2$. Then (X_1, X_2) is a two-dimensional random variable defined on $(\Omega, \mathscr{F}, \mathbf{P})$ and X_1 and X_2 are independent. We remark that all the results given above can be developed for multi-dimensional product spaces and measures; see e.g. Halmos (1974).

4.3 Measures defined by kernels

Definition 4.3.1 Let $(E, \mathscr{E})$ and $(F, \mathscr{F})$ be two measurable spaces. A function $K : E \times \mathscr{F} \to [0, \infty]$ is called a *kernel* from $(E, \mathscr{E})$ to $(F, \mathscr{F})$ if

(i) for each fixed $B \in \mathscr{F}$, the function $x \mapsto K(x, B)$ is $\mathscr{E}$-measurable;

(ii) for each fixed $x \in \mathscr{E}$, the function $B \mapsto K(x, B)$ is a measure on $(F, \mathscr{F})$.

A kernel K is said to be *σ-finite* if there are $\{E_n\} \subseteq \mathscr{E}$ and $\{F_n\} \subseteq \mathscr{F}$ such that $E = \bigcup_{m=1}^{\infty} E_m$, $F = \bigcup_{m=1}^{\infty} F_n$ and

$$\sup_{x \in E_m} K(x, F_n) < \infty, \quad m, n \geq 1.$$

We call K a *probability kernel* if $K(x, F) = 1$ for all $x \in E$ in addition.

Example 4.3.1 For each $t > 0$ we can define a probability kernel $p_t(\cdot, \cdot)$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ by

$$p_t(x, B) = \frac{1}{\sqrt{2\pi t}} \int_B e^{|y-x|^2/2t} dy.$$

Proposition 4.3.1 *Let K be a σ-finite kernel from $(E, \mathscr{E})$ to $(F, \mathscr{F})$. Then for any $C \in \mathscr{E} \times \mathscr{F}$, the function $x \mapsto K(x, C_x)$ is $\mathscr{E}$-measurable.*

Proposition 4.3.2 *Let μ be a σ-finite measure on $(E, \mathscr{E})$ and K a σ-finite kernel from $(E, \mathscr{E})$ to $(F, \mathscr{F})$. Then*

$$\lambda(C) := \int_E K(x, C_x)\mu(dx), \qquad C \in \mathscr{E} \times \mathscr{F}, \tag{4.3.1}$$

defines a σ-finite measure λ on $(E \times F, \mathscr{E} \times \mathscr{F})$.

In the situation of Proposition 4.3.2 we shall write $\lambda(dx, dy) = \mu(dx)K(x, dy)$.

Theorem 4.3.1 *Let f be a non-negative extended real-valued measurable function on $(E \times F, \mathscr{E} \times \mathscr{F})$. Then the function*

$$x \mapsto \int_F f(x, y)K(x, dy)$$

is $\mathscr{E}$-measurable. Moreover, if λ is defined by (4.3.1) then

$$\int_{E \times F} f d\lambda = \int_E \left[\int_F f(x, y)K(x, dy) \right] \mu(dx).$$

Theorem 4.3.2 *Let f be an integrable extended real-valued function on the $(E\times F, \mathscr{E}\times\mathscr{F}, \lambda)$, where λ is defined by (4.3.1). Then for μ-a.e. $x\in E$ the section f_x is a integrable function on $(F, \mathscr{F}, K(x,\cdot))$. Moreover, the function*

$$x\mapsto \int_F f(x,y)K(x,dy)$$

is a.e. $\mathscr{E}$-measurable, and

$$\int_{E\times F} f d\lambda = \int_E\left[\int_F f(x,y)K(x,dy)\right]\mu(dx). \tag{4.3.2}$$

Those results can be proved by obvious modifications of the proofs in the last two sections.

4.4 Kolmogorov's Consistency theorem

Given an non-empty index set T, we note $\mathbb{R}^T$ the set of all functions $w: t\mapsto w_t = w(t)$ from T to $\mathbb{R}$. In the sequel, we may write an element $w\in\mathbb{R}^T$ more explicitly as $w.$ or $w(\cdot)$ or $\{w_t : t\in T\}$. In particular, when $T=\{1,\cdots,n\}$, we identify $\mathbb{R}^T$ with the n-dimensional space $\mathbb{R}^n$.

For any ordered finite set $\{t_1,\cdots,t_n\}\subseteq T$ we define the *projection* $\pi_{t_1,\cdots,t_n}:\mathbb{R}^T\to\mathbb{R}^n$ by setting $\pi_{t_1,\cdots,t_n}(w) = (w_{t_1},\cdots,w_{t_n})$ for $w\in\mathbb{R}^T$.

Definition 4.4.1 A set $A\subseteq\mathbb{R}^T$ is called a *cylinder* if there exist $\{t_1,\cdots,t_n\}\subseteq T$ and $M\subseteq\mathbb{R}^n$ such that

$$A = \pi^{-1}_{t_1,\cdots,t_n}(M) = \{w\in\mathbb{R}^T : (w_{t_1},\cdots,w_{t_n})\in M\}. \tag{4.4.1}$$

We remark that the representation (4.4.1) for the cylinder A is not unique. Indeed, if A has representation (4.4.1) and if $t_{n+1}\notin\{t_1,\cdots,t_n\}$, we also have $A=\pi^{-1}_{t_1,\cdots,t_n,t_{n+1}}(M\times\mathbb{R})$. Of course, if both A and $\{t_1,\cdots,t_n\}$ are given, the set M satisfying (4.4.1) is uniquely determined. The cylinder A defined by (4.4.1) is said to be *Borel* if $M\in\mathscr{B}(\mathbb{R}^n)$. Let $\mathscr{C}_T$ denote the collection of all Borel cylinders on $\mathbb{R}^T$.

Lemma 4.4.1 *The collection $\mathscr{C}_T$ of Borel cylinders is an algebra on $\mathbb{R}^T$.*

Proof. Clearly, we have $\mathbb{R}^T\in\mathscr{C}_T$ and $\emptyset\in\mathscr{C}_T$. If A is given by (4.4.1) for $\{t_1,\cdots,t_n\}\subseteq T$ and $M\in\mathscr{B}(\mathbb{R})^n$, we have $A^c=\pi^{-1}_{t_1,\cdots,t_n}(M^c)\in\mathscr{C}_T$. Now suppose that $A_k\in\mathscr{C}_T$ for $k=1,\cdots,m$. Then for each k there exist $\{t_{k,1},\cdots,t_{k,n_k}\}\subseteq T$ and $M_k\in\mathscr{B}(\mathbb{R}^{n_k})$ such that $A_k=\pi^{-1}_{t_{k,1},\cdots,t_{k,n_k}}(M_k)$. Let $\{t_1,\cdots,t_n\}=\bigcup_{k=1}^m\{t_{k,1},\cdots,t_{k,n_k}\}$. Clearly, for each k there exists $\bar{M}_k\in\mathscr{B}(\mathbb{R}^n)$ such that $A_k=\pi^{-1}_{t_1,\cdots,t_n}(\bar{M}_k)$. Thus

$$\bigcup_{k=1}^m A_k = \bigcup_{k=1}^m \pi^{-1}_{t_1,\cdots,t_n}(\bar{M}_k) = \pi^{-1}_{t_1,\cdots,t_n}\Big(\bigcup_{k=1}^m\bar{M}_k\Big) \tag{4.4.2}$$

is a Borel cylinder. That shows that $\mathscr{C}_T$ is an algebra on $\mathbb{R}^T$. □

Let $\mathscr{B}(\mathbb{R})^T=\sigma(\mathscr{C}_T)$. Clearly, $\pi_{t_1,\cdots,t_n}$ is a measurable mapping from $(\mathbb{R}^T,\mathscr{B}(\mathbb{R})^T)$ to $(\mathbb{R}^n,\mathscr{B}(\mathbb{R})^n)$. Moreover, if we understand $\mathbb{R}^T$ as a topological product space, then $\mathscr{B}(\mathbb{R})^T=\mathscr{B}(\mathbb{R}^T)$. (Homework.)

Example 4.4.1 A one-dimensional *stochastic process* with index set $[0, \infty)$ is a random element X taking values in $(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R})^{[0,\infty)})$. Suppose that X is defined on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Given $\omega \in \Omega$ we have $X(\omega) = X.(\omega) \in \mathbb{R}^{[0,\infty)}$. In other words, for each fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is an element of $\mathbb{R}^{[0,\infty)}$, which is called the *path* of X. We can also regard the stochastic process as a family of random variables $\{X_t : t \in [0, \infty)\}$, where each variable X_t takes values in $\mathbb{R}$. There is another way is to understand the stochastic process X, namely, we consider it as a function $X : (t, \omega) \mapsto X_t(\omega)$ from $[0, \infty) \times \Omega$ to $\mathbb{R}$.

Definition 4.4.2 Suppose for each ordered finite set $\{t_1, \cdots, t_n\} \subseteq T$ there is a probability measure $P_{t_1,\cdots,t_n}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. We say $\{P_{t_1,\cdots,t_n} : n \geq 1, \{t_1, \cdots, t_n\} \subseteq T\}$ are *consistent* if they have the following properties:

(i) if $B_i \in \mathscr{B}(\mathbb{R})$ for $1 \leq i \leq n$ and $\{i_1, \cdots, i_n\}$ is a permutation of $\{1, \cdots, n\}$, then

$$P_{t_{i_1},\cdots,t_{i_n}}(B_{i_1} \times \cdots \times B_{i_n}) = P_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n);$$

(ii) if $m \leq n$ and $B_i \in \mathscr{B}(\mathbb{R})$ for $1 \leq i \leq m$, then

$$P_{t_1,\cdots,t_m}(B_1 \times \cdots \times B_m) = P_{t_1,\cdots,t_m,t_{m+1},\cdots,t_n}(B_1 \times \cdots \times B_m \times \mathbb{R} \times \cdots \times \mathbb{R}).$$

Proposition 4.4.1 *Suppose that P is a probability measure on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$. Then for any $\{t_1, \cdots, t_n\} \subseteq T$ we can define a probability measure $P_{t_1,\cdots,t_n}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ by*

$$P_{t_1,\cdots,t_n}(B) = P(\pi^{-1}_{t_1,\cdots,t_n}(B)), \qquad B \in \mathscr{B}(\mathbb{R}^n). \tag{4.4.3}$$

Moreover, the probability measures $\{P_{t_1,\cdots,t_n} : n \geq 1, \{t_1, \cdots, t_n\} \subseteq T\}$ are consistent.

Proof. (Homework.) □

The following important theorem shows that the converse of Proposition 4.2.1 is also true.

Theorem 4.4.1 (Kolmogorov) *If the probability measures $\{P_{t_1,\cdots,t_n} : n \geq 1, \{t_1, \cdots, t_n\} \subseteq T\}$ are consistent, there is a unique probability measure P on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R}^T))$ such that (4.4.3) holds.*

Proof. Step 1) For any $A \in \mathscr{C}_T$ we can find $\{t_1, \cdots, t_n\} \subseteq T$ and $M \in \mathscr{B}(\mathbb{R}^n)$ such that $A = \pi^{-1}_{t_1,\cdots,t_n}(M)$. By the consistency we see that the value $P_T(A) := P_{t_1,\cdots,t_n}(M)$ is independent of the particular choice of $\{t_1, \cdots, t_n\}$ and M. We claim that P_T is a finitely additive set function on the algebra $\mathscr{C}_T$. To see that, suppose that $\{A_1, \cdots, A_m\}$ is a sequence of disjoint subsets of $\mathscr{C}_T$. As shown in the proof of Lemma 4.4.1, there exist $\{t_1, \cdots, t_n\} \subseteq T$ and $\{\bar{M}_1, \cdots, \bar{M}_m\} \subseteq \mathscr{B}(\mathbb{R}^n)$ such that $A_k = \pi^{-1}_{t_1,\cdots,t_n}(\bar{M}_k)$ for each $1 \leq k \leq m$. Clearly, the sets $\{\bar{M}_1, \cdots, \bar{M}_m\}$ are disjoint. By (4.4.2) we have

$$P_T\Big(\bigcup_{k=1}^m A_k\Big) = P_{t_1,\cdots,t_n}\Big(\bigcup_{k=1}^m \bar{M}_k\Big) = \sum_{k=1}^m P_{t_1,\cdots,t_n}(\bar{M}_k) = \sum_{k=1}^m P_T(A_k),$$

giving the finite additivity of P_T.

Step 2) We shall prove that P_T is σ-additive so it is a measure on $\mathscr{C}_T$. This will follow if we can show $\{A_k\} \subseteq \mathscr{C}_T$ and $A_k \downarrow \emptyset$ imply $P_T(A_k) \downarrow 0$. In other words, we only need to show: if $\{A_k\} \subseteq \mathscr{C}_T$ is a decreasing sequence and $P_T(A_k) \downarrow \varepsilon_0 > 0$, then $\bigcap_{k=1}^\infty A_k \neq \emptyset$.

Let us write $A_k = \pi_{T_k}^{-1}(M_k)$ for some $T_k = \{t_1, \cdots, t_{n_k}\} \subseteq T$ and $M_k \in \mathscr{B}(\mathbb{R}^{n_k})$. In view of the relation $A_k \supseteq A_{k+1}$, we may assume $T_k \subseteq T_{k+1}$, so that $n_k \leq n_{k+1}$. Let $T_\infty = \{t_1, t_2, \cdots\}$. By the definition of P_T, we have $P_T(A_k) = P_{t_1,\cdots,t_{n_k}}(M_k) \geq \varepsilon_0 > 0$. Since $P_{t_1,\cdots,t_{n_k}}$ is a probability on $\mathscr{B}(\mathbb{R}^{n_k})$, there is a compact set $F_k = F_k(\varepsilon_0) \subseteq M_k$ such that $P_{t_1,\cdots,t_{n_k}}(M_k \setminus F_k) < \varepsilon_0/2^{k+1}$. Let $B_k = \pi_{t_1,\cdots,t_{n_k}}^{-1}(F_k) \in \mathscr{C}_T$ and $C_k = B_1 \cap \cdots \cap B_k$. Then $C_k \subseteq B_k \subseteq A_k$. Since P_T is finite additive, we have

$$\begin{aligned} P_T(A_k \setminus C_k) &= P_T\Big(A_k \cap \Big(\bigcup_{i=1}^k B_i^c\Big)\Big) = P_T\Big(\bigcup_{i=1}^n (A_k \setminus B_i)\Big) \\ &\leq P_T\Big(\bigcup_{i=1}^n (A_i \setminus B_i)\Big) \leq \sum_{i=1}^n P_T(A_i \setminus B_i) \\ &\leq \sum_{i=1}^n P_{t_1,\cdots,t_{n_i}}(M_i \setminus F_i) \leq \sum_{i=1}^n \frac{\varepsilon_0}{2^{i+1}} < \frac{\varepsilon_0}{2}. \end{aligned}$$

It follows that $P_T(C_k) = P_T(A_k) - P_T(A_k \setminus C_k) > P_T(A_k) - \varepsilon_0/2 > 0$. In particular, we have $C_k \neq \emptyset$, so there exists some $w_k = w_k(\cdot) \in C_k$. For $m \geq k \geq 1$ we have $w_m \in C_m \subseteq C_k \subseteq B_k$ and hence $\pi_{t_1,\cdots,t_{n_k}}(w_m) = (w_m(t_1), \cdots, w_m(t_{n_k})) \in F_k$. By the method of diagonalization, it is easy to find a subsequence $\{w_{k_i}\} \subseteq \{w_k\}$ such that $w_{k_i}(t_j) \to$ some x_j for every $j \geq 1$. Since F_k is compact, we have $(x_1, \cdots, x_{n_k}) \in F_k$. Now we define $w_0 \in \mathbb{R}^T$ by

$$w_0(t) = \begin{cases} x_j & \text{if } t = t_j, \\ 0 & \text{if } t \in T \setminus T_\infty. \end{cases}$$

Then $\pi_{t_1,\cdots,t_{n_k}}(w_0) = (x_1, \cdots, x_{n_k}) \in F_k$, and so $w_0 \in B_k \subseteq A_k$. That implies $w_0 \in \bigcap_{k=1}^\infty A_k$ and so $\bigcap_{k=1}^\infty A_k \neq \emptyset$.

Step 3) Since $\mathscr{C}_T$ is an algebra, by measure extension theorem P_T has a unique extension P on $\mathscr{B}(\mathbb{R}^T)$. The equality (4.4.3) follows from the definition of P_T. □

Example 4.4.2 Let μ be a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and let $p_t(\cdot,\cdot)$ be the kernel defined as in Example 4.3.1. For any $0 = t_0 < t_1 < \cdots < t_n$ let

$$P_{t_0,t_1,\cdots,t_n}(A) = \int_A \mu(dx_0) p_{t_1-t_0}(x_0, dx_1) \cdots p_{t_n-t_{n-1}}(x_{n-1}, dx_n), \qquad A \in \mathscr{B}(\mathbb{R}^{n+1}),$$

and

$$P_{t_1,\cdots,t_n}(B) = \int_{\mathbb{R}} \mu(dx_0) \int_B p_{t_1-t_0}(x_0, dx_1) \cdots p_{t_n-t_{n-1}}(x_{n-1}, dx_n), \qquad B \in \mathscr{B}(\mathbb{R}^n).$$

If $0 \leq t_0 < t_1 < \cdots < t_n$ and if $\{i_1, \cdots, i_n\}$ is a permutation of $\{1, \cdots, n\}$, we define the probability measure $P_{t_{i_1},\cdots,t_{i_n}}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ by

$$P_{t_{i_1},\cdots,t_{i_n}}(B_{i_1} \times \cdots \times B_{i_n}) = P_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n), \qquad B_i \in \mathscr{B}(\mathbb{R}).$$

Then the family of probability measures $\{P_{t_1,\cdots,t_n} : n \geq 1, \{t_1, \cdots, t_n\} \subseteq [0,\infty)\}$ are consistent. By Theorem 4.4.1 there is a unique probability measure P on $(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)}))$ such that (4.4.3) holds for any $\{t_1, \cdots, t_n\} \subseteq [0,\infty)$. Let $X_t(w) = w_t$ for $t \geq 0$ and $w \in \mathbb{R}^{[0,\infty)}$. Then $w \mapsto X_\cdot(w)$ is a stochastic process defined on the probability space $(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)}), P)$. This process is known as a *one-dimensional standard Brownian motion* with initial distribution μ.

By obvious modifications of the arguments, one can generalize the results in this section to the case where $\mathbb{R}$ is replaced by the d-dimensional space $\mathbb{R}^d$. We leave the consideration of those to the reader.

Chapter 5

Convergence of Random Variables

5.1 Borel-Cantelli lemma

In this section, we prove the very important Borel-Cantelli lemma which will be used in our subsequent investigation. Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a given probability space and let $\{E_n\} \subseteq \mathscr{F}$ be a sequence of events. We often need to evaluate the probability of the the event $E = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k$. Clearly, $\omega \in E$ if and only if $\omega \in E_k$ for infinitely many $k \geq 1$.

Lemma 5.1.1 (Borel-Cantelli) *(i) If $\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty$, then $\mathbf{P}(E) = 0$. (ii) If $\{E_n\}$ are independent, then $\mathbf{P}(E) = 0$ or $= 1$ according as $\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty$ or $= \infty$.*

Proof. Under the assumption of (i), we have

$$\mathbf{P}(E) = \mathbf{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\Big) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcup_{k=n}^{\infty} E_k\Big) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbf{P}(E_k) = 0.$$

If the events $\{E_n\}$ are independent, their complements $\{E_n^c\}$ are also independent. Then for any $N \geq n \geq 1$ we have

$$\begin{aligned} \mathbf{P}\Big(\bigcap_{k=n}^{\infty} E_k^c\Big) &\leq \mathbf{P}\Big(\bigcap_{k=n}^{N} E_k^c\Big) = \prod_{k=n}^{N} [1 - \mathbf{P}(E_k)] \\ &\leq \prod_{k=n}^{N} e^{-\mathbf{P}(E_k)} = \exp\Big\{-\sum_{k=n}^{N} \mathbf{P}(E_k)\Big\}. \end{aligned}$$

Letting $N \to \infty$ we obtain

$$\mathbf{P}\Big(\bigcap_{k=n}^{\infty} E_n^c\Big) \leq \exp\Big\{-\sum_{k=n}^{\infty} \mathbf{P}(E_k)\Big\} = 0.$$

It follows that

$$\mathbf{P}(E^c) = \mathbf{P}\Big(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k^c\Big) \leq \sum_{n=1}^{\infty} \mathbf{P}\Big(\bigcap_{k=n}^{\infty} E_k^c\Big) = 0,$$

and consequently $\mathbf{P}(E) = 1$. □

5.2 Almost sure convergence

Definition 5.2.1 Let X, X_n, $n = 1, 2, \cdots$ be random variables defined on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$. We say $X_n \to X$ almost surely if there is $N \in \mathscr{F}$ such that $\mathbf{P}(N) = 0$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in N^c := \Omega \setminus N$. In this case, we write $X_n \overset{\text{a.s}}{\to} X$ or $\lim_{n\to\infty} X_n \overset{\text{a.s}}{=} X$.

Definition 5.2.2 We say the sequence of random variables $\{X_n\}$ is almost surely Cauchy if there is $N \in \mathscr{F}$ such that $\mathbf{P}(N) = 0$ and $\{X_n(\omega)\}$ is a Cauchy sequence for every $\omega \in N^c$.

Proposition 5.2.1 *(i) A sequence $\{X_n\}$ is a.s. Cauchy if and only if $X_n \overset{\text{a.s.}}{\to}$ some X; (ii) If $X_n \overset{\text{a.s.}}{\to} X$ and $X_n \overset{\text{a.s.}}{\to} Y$, then $Y \overset{\text{a.s.}}{=} X$.*

Proof. (Homework.) □

Proposition 5.2.2 *Let $\{X_n\}$ be a sequence of random variables. Then: (i) $X_n \overset{\text{a.s}}{\to} X$ if and only if for each $\varepsilon > 0$ we have*

$$\lim_{n\to\infty} \mathbf{P}\left(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \varepsilon\} \right) = 0;$$

(ii) $\{X_n\}$ is a.s. Cauchy if and only if for each $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbf{P}\left(\bigcup_{m=n}^{\infty} \{|X_m - X_n| \geq \varepsilon\} \right) = 0.$$

Proof. (i) Let $D = \{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ does not hold}\}$. Then $\omega \in D$ if and if only there is some $k \geq 1$ such that for any $n \geq 1$ there exist some $m \geq n$ such that $|X_m(\omega) - X(\omega)| \geq 1/k$. In other words,

$$D = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\omega \in \Omega : |X_m(\omega) - X(\omega)| \geq 1/k\}.$$

Then $X_n \overset{\text{a.s}}{\to} X$ if and only if $\mathbf{P}(D) = 0$ or, equivalently,

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m(\omega) - X(\omega)| \geq 1/k\} \right) = 0$$

for every $k \geq 1$. By the upper continuity of the probability measure, the above equality holds if and only if

$$\lim_{n\to\infty} \mathbf{P}\left(\bigcup_{m=n}^{\infty} \{|X_m(\omega) - X(\omega)| \geq 1/k\} \right) = 0$$

for every $k \geq 1$. Then we have the assertion (i).

(ii) Recall that a non-random sequence $\{a_n\}$ is Cauchy if and only if for each $k \geq 1$ there is some $n \geq 1$ such that $|a_m - a_n| < 1/k$ for every $m \geq n$. Let $D = \{\omega \in \Omega : \text{the sequence } \{X_n(\omega)\}$ is not Cauchy$\}$. The proof is similar to that of (i). □

Corollary 5.2.1 *(i) If* $\sum_{n=1}^{\infty} \mathbf{P}\{|X_n - X| \geq \varepsilon\} < \infty$ *for every* $\varepsilon > 0$*, then* $X_n \overset{\text{a.s}}{\to} X$*; (ii) If* $\sum_{n=1}^{\infty} \mathbf{E}[|X_n - X|^2] < \infty$*, then* $X_n \overset{\text{a.s}}{\to} X$*.*

Proof. Under the assumption of (i) we have

$$\mathbf{P}\Big(\bigcup_{k=n}^{\infty}\{|X_k - X| \geq \varepsilon\}\Big) \leq \sum_{k=n}^{\infty} \mathbf{P}\{|X_k - X| \geq \varepsilon\} \to 0.$$

Then the result follows from Proposition 5.2.2. Under the condition of (ii),

$$\sum_{n=1}^{\infty} \mathbf{P}\{|X_n - X| \geq \varepsilon\} \leq \sum_{n=1}^{\infty} \varepsilon^{-2}\mathbf{E}[|X_n - X|^2] < \infty.$$

Then the a.s. convergence follows by (i). □

5.3 Convergence in probability

Definition 5.3.1 Let $\{X_n\}$ be a sequence of random variables defined on $(\Omega, \mathscr{F}, \mathbf{P})$. We say X_n converges to X in probability if $\mathbf{P}\{|X_n - X| \geq \varepsilon\} \to 0$ as $n \to \infty$ for every $\varepsilon > 0$. In this case, we write $X_n \overset{\text{P}}{\to} X$.

Proposition 5.3.1 *The sequence* $\{X_n\}$ *converges to zero in probability if and only if*

$$\mathbf{E}\bigg[\frac{|X_n|^r}{1 + |X_n|^r}\bigg] \to 0 \tag{5.3.1}$$

for some $r > 0$*.*

Proof. Observe that $Y_n := |X_n|^r/(1 + |X_n|^r) \leq |X_n|^r$. If $X_n \overset{\text{P}}{\to} 0$, we have $|X_n|^r \overset{\text{P}}{\to} 0$ and hence $Y_n \overset{\text{P}}{\to} 0$. Then (5.3.1) follows by the dominated convergence theorem. (Homework: Prove the dominated convergence theorem for a sequence of random variables that converges in probability.) Conversely, suppose that (5.3.1) holds. For $\delta > 0$ and $\epsilon > 0$, let $N = N(\delta, \epsilon)$ be such that

$$\mathbf{E}\bigg[\frac{|X_n|^r}{1 + |X_n|^r}\bigg] < \frac{\epsilon\delta^r}{1 + \delta^r}$$

for $n \geq N$. By Chebyshev's inequality we have

$$\mathbf{P}\{|X_n| \geq \delta\} = \mathbf{P}\bigg\{\frac{|X_n|^r}{1 + |X_n|^r} \geq \frac{\delta^r}{1 + \delta^r}\bigg\} \leq \frac{1 + \delta^r}{\delta^r}\mathbf{E}\bigg[\frac{|X_n|^r}{1 + |X_n|^r}\bigg] < \epsilon$$

for $n \geq N$. That proves $X_n \overset{\text{P}}{\to} 0$. □

Proposition 5.3.2 *If* $X_n \overset{\text{a.s.}}{\to} X$*, then* $X_n \overset{\text{P}}{\to} X$*.*

Proof. By Proposition 5.2.2, if $X_n \stackrel{\text{a.s.}}{\to} X$, for each $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbf{P}\bigg(\bigcup_{m=n}^{\infty} \{|X_m - X| \geq \varepsilon\} \bigg) = 0.$$

Then $\lim_{n\to\infty} \mathbf{P}\{|X_n - X| \geq \varepsilon\} = 0.$ □

Definition 5.3.2 We say $\{X_n\}$ is Cauchy in probability if $\mathbf{P}\{|X_n - X_m| \geq \varepsilon\} \to 0$ as $m, n \to \infty$ for each $\varepsilon > 0$.

Proposition 5.3.3 *(i) If $X_n \stackrel{\text{P}}{\to} X$, there is a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \stackrel{\text{a.s.}}{\to} X$; (ii) If $\{X_n\}$ is Cauchy in probability, there is a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \stackrel{\text{a.s.}}{\to}$ some X.*

Proof. (i) Since $X_n \stackrel{\text{P}}{\to} X$, for each $k \geq 1$ there is an integer n_k such that

$$\mathbf{P}\{|X_{n_k} - X| \geq 1/k\} \leq 1/2^k.$$

Let $\varepsilon > 0$ and choose an integer $m > 1/\varepsilon$. We have

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon\} &\leq \sum_{k=m}^{\infty} \mathbf{P}\{|X_{n_k} - X| \geq 1/m\} + \sum_{k=1}^{m-1} \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon\} \\
&\leq \sum_{k=m}^{\infty} \mathbf{P}\{|X_{n_k} - X| \geq 1/k\} + \sum_{k=1}^{m-1} \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon\} \\
&\leq \sum_{k=m}^{\infty} \frac{1}{2^k} + \sum_{k=1}^{m-1} \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon\} \\
&= \frac{1}{2^{m-1}} + \sum_{k=1}^{m-1} \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon\} < \infty.
\end{aligned}$$

Then $X_{n_k} \stackrel{\text{a.s.}}{\to} X$ by Corollary 5.2.1.

(ii) Since $\{X_n\}$ is Cauchy in probability, for each $k \geq 1$ we can choose m_k so that

$$\mathbf{P}\{|X_m - X_n| \geq 1/2^k\} < 1/2^k$$

for all $m, n \geq m_k$. Define $\{n_k\}$ inductively by setting $n_1 = m_1$ and $n_{k+1} = \max\{n_k + 1, m_{k+1}\}$ for $k \geq 1$. Clearly, $n_k \to \infty$. Let

$$F_m = \bigcup_{k=m}^{\infty} \{|X_{n_k} - X_{n_{k+1}}| \geq 1/2^k\}.$$

Then $\{F_m\}$ is a decreasing sequence and

$$\mathbf{P}(F_m) \leq \sum_{k=m}^{\infty} \mathbf{P}\{|X_{n_k} - X_{n_{k+1}}| \geq 1/2^k\} \leq \sum_{k=m}^{\infty} \frac{1}{2^k} = \frac{1}{2^{m-1}}.$$

Let $N = \bigcap_{m=1}^{\infty} F_m$. We have $\mathbf{P}(N) = \lim_{m\to\infty} \mathbf{P}(F_m) = 0$. For any

$$\omega \in F_m^c = \bigcap_{k=m}^{\infty} \{|X_{n_k} - X_{n_{k+1}}| < 1/2^k\}$$

we have $|X_{n_k} - X_{n_{k+1}}| < 1/2^k$ for $k \geq m$, so $\{X_{n_k}(\omega)\}$ is a Cauchy sequence. Consequently, for each $\omega \in N^c = \bigcup_{m=1}^{\infty} F_m^c$ the sequence $\{X_{n_k}(\omega)\}$ is Cauchy. It follows that $X_{n_k}(\omega) \to$ some $X(\omega)$ for every $\omega \in N^c$. Setting $X(\omega) = 0$ for $\omega \in N$, we have $X_{n_k} \xrightarrow{\text{a.s.}} X$. □

Proposition 5.3.4 *The sequence of random variables $\{X_n\}$ is Cauchy in probability if and only if $X_n \xrightarrow{\text{P}}$ some X.*

Proof. Suppose that $X_n \xrightarrow{\text{P}} X$. Then for any $\varepsilon > 0$ and $\eta > 0$, there is $N = N(\varepsilon, \eta)$ such that

$$\mathbf{P}\{|X_n - X| \geq \varepsilon/2\} < \eta/2$$

for $n \geq N$. When $m, n \geq N$, we have

$$\mathbf{P}\{|X_m - X_n| \geq \varepsilon\} \leq \mathbf{P}(\{|X_m - X| \geq \varepsilon/2\} \cup \{|X_n - X| \geq \varepsilon/2\}) \leq \eta/2 + \eta/2 = \eta.$$

Thus $\{X_n\}$ is Cauchy in probability. Conversely, suppose that $\{X_n\}$ is Cauchy in probability. By Proposition 5.3.3, there is a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \xrightarrow{\text{a.s.}}$ some X. Consequently, we have $X_{n_k} \xrightarrow{\text{P}} X$. For $\delta > 0$ and $\varepsilon > 0$, there exists $N \geq 1$ such that $\mathbf{P}\{|X_n - X_m| \geq \varepsilon/2\} < \delta$ for $m \geq n \geq N$. For any $n \geq N$ we choose $n_k \geq n$ so that

$$\begin{aligned} \mathbf{P}\{|X_n - X| \geq \varepsilon\} &\leq \mathbf{P}\{|X_n - X_{n_k}| \geq \varepsilon/2\} + \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon/2\} \\ &< \delta + \mathbf{P}\{|X_{n_k} - X| \geq \varepsilon/2\}. \end{aligned}$$

Then we may let $k \to \infty$ to see that $\mathbf{P}\{|X_n - X| \geq \varepsilon\} < \delta$. □

Remark 5.3.1 (i) It is not hard to construct a sequence $\{X_n\}$ such that $X_n \xrightarrow{\text{P}} X$, but $X_n \xrightarrow{\text{a.s.}} X$ does not hold. (ii) If $X_n \xrightarrow{\text{P}} X$ and $X_n \xrightarrow{\text{P}} Y$, then $X \overset{\text{a.s.}}{=} Y$. (Homework.)

5.4 Convergence in mean

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and $\mathscr{L}_1 = \mathscr{L}_1(\Omega, \mathscr{F}, \mathbf{P})$ be the set of all random variables X on $(\Omega, \mathscr{F}, \mathbf{P})$ such that $\mathbf{E}[|X|] < \infty$.

Definition 5.4.1 Let $X, X_n \in \mathscr{L}_1$. We say $\{X_n\}$ converges to X in mean if $\mathbf{E}[|X_n - X|] \to 0$ as $n \to \infty$. In this case, we write $X_n \xrightarrow{\mathscr{L}_1} X$. We say $\{X_n\}$ is Cauchy in mean if $\mathbf{E}[|X_m - X_n|] \to 0$ as $m, n \to \infty$.

Proposition 5.4.1 *(i) If $X_n \xrightarrow{\mathscr{L}_1} X$, then $X_n \xrightarrow{\text{P}} X$; (ii) If $\{X_n\}$ is Cauchy in mean, then it is Cauchy in probability.*

Proof. Both results follow immediately from Markov's inequality. □

Proposition 5.4.2 *The sequence $\{X_n\} \subseteq \mathscr{L}_1$ is Cauchy in mean if and only if $X_n \stackrel{\mathscr{L}_1}{\to}$ some X.*

Proof. Suppose $X_n \stackrel{\mathscr{L}_1}{\to} X$. For $\epsilon > 0$, there is $N = N(\epsilon)$ such that $\mathbf{E}[|X_n - X|] < \epsilon/2$ for $n \geq N$. It follows that

$$\mathbf{E}[|X_m - X_n|] \leq \mathbf{E}[|X_m - X|] + \mathbf{E}[|X - X_m|] \leq \epsilon/2 + \epsilon/2 = \epsilon$$

for $m \geq n \geq N$. That shows that $\{X_n\}$ is Cauchy in mean. Conversely, suppose $\{X_n\}$ is Cauchy in mean. By Proposition 5.4.1, $\{X_n\}$ is Cauchy in probability. By Proposition 5.3.3, there exists a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \stackrel{\text{a.s.}}{\to} X$. For $\epsilon > 0$, let $N = N(\epsilon)$ be such that $\mathbf{E}[|X_m - X_n|] < \epsilon$ for $m \geq n \geq N$. In particular, we have $\mathbf{E}[|X_{n_k} - X_n|] < \epsilon$ for $n_k \geq n \geq N$. By Fatou's lemma,

$$\mathbf{E}[|X - X_n|] = \mathbf{E}\Big[\liminf_{k\to\infty} |X_{n_k} - X_n|\Big] \leq \liminf_{k\to\infty} \mathbf{E}[|X_{n_k} - X_n|] \leq \epsilon.$$

It follows that $X_n \stackrel{\mathscr{L}_1}{\to} X$. □

Lemma 5.4.1 *Let $X \in \mathscr{L}_1$. Then to each $\epsilon > 0$ there corresponds $\delta = \delta(\epsilon) > 0$ such that*

$$\int_E |X| d\mathbf{P} < \epsilon$$

for every $E \in \mathscr{F}$ with $\mathbf{P}(E) < \delta$.

Proof. Since $\mathbf{E}[|X|] < \infty$, we may apply the monotone convergence theorem to see

$$\lim_{n\to\infty} \int_{\{|X|>n_0\}} |X| d\mathbf{P} = \lim_{n\to\infty} \int_\Omega 1_{\{|X|>n_0\}} |X| d\mathbf{P} = 0.$$

Then for any $\epsilon > 0$, there exists $n_0 = n_0(\epsilon) \geq 1$ such that $\int_{\{|X|>n_0\}} |X| d\mathbf{P} < \epsilon/2$. Let $\delta = \epsilon/2n_0$. For $E \in \mathscr{F}$ with $\mathbf{P}(E) < \delta$ we have

$$\int_E |X| d\mathbf{P} \leq \int_{E\cap\{|X|\leq n_0\}} |X| d\mathbf{P} + \int_{\{|X|>n_0\}} |X| d\mathbf{P} < n_0 \cdot \epsilon/2n_0 + \epsilon/2 = \epsilon,$$

as desired. □

Proposition 5.4.3 *Suppose that $\{X_n\} \subseteq \mathscr{L}_1$. Then $\{X_n\}$ is Cauchy in mean if and only if it is Cauchy in probability and for each $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that*

$$\sup_{n\geq 1} \int_E |X_n| d\mathbf{P} < \epsilon \tag{5.4.1}$$

for each $E \in \mathscr{F}$ satisfying $\mathbf{P}(E) < \delta$.

Proof. Suppose that $\{X_n\}$ is Cauchy in mean. By Proposition 5.4.1, the sequence is Cauchy in probability. Moreover, for $\epsilon > 0$ there exists $n_0 = n_0(\epsilon) \geq 1$ such that $\mathbf{E}|X_n - X_{n_0}| < \epsilon/2$ for $n \geq n_0$. By Lemma 5.4.1, there exists $\delta = \delta(\epsilon) > 0$ such that

$$\int_E |X_n| d\mathbf{P} < \epsilon/2$$

when $\mathbf{P}(E) < \delta$ and $n \leq n_0$. If $\mathbf{P}(E) < \delta$ and $n > n_0$, we have

$$\begin{aligned}
\int_E |X_n| d\mathbf{P} &\leq \int_E |X_n - X_{n_0}| d\mathbf{P} + \int_E |X_{n_0}| d\mathbf{P} \\
&\leq \mathbf{E}[|X_n - X_{n_0}|] + \int_E |X_{n_0}| d\mathbf{P} \leq \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

That proves (5.4.1). Conversely, suppose that $\{X_n\}$ is Cauchy in probability and (5.4.1) holds for $\epsilon > 0$ and $\delta > 0$. Let $N = N(\delta) \geq 1$ be such that $\mathbf{P}\{|X_n - X_m| \geq \epsilon\} < \delta$ for $m, n \geq N$. Then we have

$$\sup_{k \geq 1} \int_{\{|X_n - X_m| \geq \epsilon\}} |X_k| d\mathbf{P} < \epsilon.$$

It follows that, for $m, n \geq N$,

$$\begin{aligned}
\mathbf{E}[|X_n - X_m|] &= \int_{\{|X_n - X_m| < \epsilon\}} |X_n - X_m| d\mathbf{P} + \int_{\{|X_n - X_m| \geq \epsilon\}} |X_n - X_m| d\mathbf{P} \\
&\leq \epsilon + \int_{\{|X_n - X_m| \geq \epsilon\}} |X_n| d\mathbf{P} + \int_{\{|X_n - X_m| \geq \epsilon\}} |X_m| d\mathbf{P} \\
&\leq \epsilon + \epsilon + \epsilon = 3\epsilon.
\end{aligned}$$

Thus $\{X_n\}$ is Cauchy in mean. □

Chapter 6

Laws of Large Numbers

6.1 Tail events and tail functions

Let $\{X_n\}$ be a sequence of independent random variables on $(\Omega, \mathscr{F}, \mathbf{P})$. An interesting problem in probability is the following: what is the probability that the series $\sum_{i=1}^{\infty} X_i$ converges? A remarkable result that we shall prove is that this probability must be either 0 or 1. The property is shared by many other events associated with independent variables. Results of these type are called zero-one laws.

Definition 6.1.1 For any infinite sequence $\{X_n\}$ of random variables,

$$\mathscr{T} := \bigcap_{n=1}^{\infty} \sigma(\{X_n, X_{n+1}, \cdots\}),$$

is a σ-algebra, which is called the *tail σ-algebra* of $\{X_n\}$. Any event $A \in \mathscr{T}$ is called a *tail event*. A function $f : \Omega \to \mathbb{R}$ is called a *tail function* if it is $\mathscr{T}$-measurable, that is, $f^{-1}(B) \in \mathscr{T}$ for every $B \in \mathscr{B}(\mathbb{R})$.

Proposition 6.1.1 *Let $\{X_n\}$ be a sequence of independent random variables. Then for any $n \geq 1$ the classes $\sigma(\{X_1, \cdots, X_n\})$ and $\sigma(\{X_{n+1}, X_{n+2}, \cdots\})$ are independent.*

Proof. Let $n \geq 1$ and $A \in \sigma(\{X_1, \cdots, X_n\})$ be fixed and let $\mathscr{D}$ be the class of sets $C \in \sigma(\{X_{n+1}, X_{n+2}, \cdots\})$ such that $\mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C)$. It is easy to show that $\mathscr{D}$ is a λ-class. According to the definition of the independence, for any $m \geq 1$ the random variables $\{X_1, \cdots, X_n, X_{n+1}, \cdots, X_{n+m}\}$ are independent. Then the classes $\{X_1, \cdots, X_n\}$ and $\{X_{n+1}, \cdots, X_{n+m}\}$ are independent. It follows that $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ for every $B \in \sigma(\{X_{n+1}, \cdots, X_{n+m}\})$. In other words, we have $\mathscr{D} \supseteq \mathscr{C} := \bigcup_{m \geq 1} \sigma(\{X_{n+1}, \cdots, X_{n+m}\})$. Since $\mathscr{D}$ is clearly a π-class, we get $\mathscr{D} \supseteq \sigma(\mathscr{C}) = \sigma(\{X_{n+1}, X_{n+2}, \cdots\})$ by the monotone class theorem. That yields $\mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C)$ for all $C \in \sigma(\{X_{n+1}, X_{n+2}, \cdots\})$. Then $\sigma(\{X_1, \cdots, X_n\})$ and $\sigma(\{X_{n+1}, X_{n+2}, \cdots\})$ are independent. □

Theorem 6.1.1 (Kolmogorov) *Let $\{X_n\}$ be a sequence of independent random variables. Then $\mathbf{P}(A) = 0$ or 1 for any $A \in \mathscr{T}$.*

Proof. By Proposition 6.1.1, for $E \in \sigma(\{X_1, \cdots, X_n\})$ and $A \in \mathscr{T} \subseteq \sigma(\{X_{n+1}, X_{n+2}, \cdots\})$ we have $\mathbf{P}(E \cap A) = \mathbf{P}(E)\mathbf{P}(A)$. By a monotone class argument, one shows that $\mathbf{P}(E \cap A) = \mathbf{P}(E)\mathbf{P}(A)$ for $E \in \sigma(\{X_1, X_2, \cdots\})$ and $A \in \mathscr{T}$. In particular, taking $E = A$ we get $\mathbf{P}(A) = \mathbf{P}(A)^2$ and hence $\mathbf{P}(A) = 0$ or 1. □

Corollary 6.1.1 *Let $\{X_n\}$ be a sequence of independent random variables. Then any tail function of $\{X_n\}$ is a.s. constant.*

Proof. Let ξ be a tail function of $\{X_n\}$. By Theorem 6.1.1, for any $a \leq b \in \mathbb{R}$ we have $\mathbf{P}\{a \leq \xi \leq b]\} = 0$ or $= 1$. Then to each $k \geq 1$ there corresponds an unique n_k such that $\mathbf{P}\{n_k/2^k \leq \xi \leq (n_k + 1)/2^k\} = 1$. Clearly, $[n_{k+1}/2^{k+1}, (n_{k+1} + 1)/2^{k+1}] \subset [n_k/2^k, (n_k + 1)/2^k]$ for every $k \geq 1$. Let $\alpha \in \mathbb{R}$ be the unique point contained in all the intervals $[n_k/2^k, (n_k+1)/2^k]$ for $k \geq 1$. Then we must have a.s. $\xi = a$. □

Corollary 6.1.2 *Let $\{X_n\}$ be a sequence of independent random variables. Then we have*

(i) $\{X_n\}$ either a.s. converges to a finite limit or a.s. diverges.

(ii) $\sum_{i=1}^n X_i$ either a.s. converges to a finite limit or a.s. diverges.

(iii) if $b_n \to \infty$, then $b_n^{-1} \sum_{i=1}^n X_i$ either a.s. converges to a finite limit or a.s. diverges.

Moreover, if the sequence as (i) or (iii) a.s. converges, the limit is a.s. constant.

Proof. For any $m \geq 1$ the event

$$\bigcap_{i,j \geq n} \{\omega : |X_i(\omega) - X_j(\omega)| < 1/m\}$$

decreases in $n \geq 1$. Then we have

$$\bigcup_{n\geq 1} \bigcap_{i,j \geq n} \{\omega : |X_i(\omega) - X_j(\omega)| < 1/m\} = \bigcup_{n\geq k} \bigcap_{i,j \geq n} \{\omega : |X_i(\omega) - X_j(\omega)| < 1/m\}$$

for every $k \geq 1$. Consequently, the above event is belong to the tail σ-algebra $\mathscr{T}$, then so is

$$\{\omega : X_n(\omega) \text{ converges}\} = \bigcap_{m\geq 1} \bigcup_{n\geq 1} \bigcap_{i,j \geq n} \{\omega : |X_i(\omega) - X_j(\omega)| < 1/m\}.$$

By Theorem 6.1.1 we have $\mathbf{P}\{X_n \text{ converges}\} = 0$ or $= 1$. If this probability is one, let $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ when the limit exists and let $X(\omega) = 0$ when $X_n(\omega)$ diverges. Clearly, X is a tail function. Then Corollary 6.1.1 implies that X is a.s. constant. That proves (i). The proofs of (ii) and (iii) are similar. (Homework.) □

Definition 6.1.2 We say two sequences of random variables $\{X_n\}$ and $\{Y_n\}$ are *tail equivalent* if they differ a.s. only by a finite number of term, that is, for a.e. $\omega \in \Omega$ there corresponds some $N(\omega)$ such that $X_n(\omega) = Y_n(\omega)$ for all $n \geq N(\omega)$. The two sequences $\{X_n\}$ and $\{Y_n\}$ are said to be *convergence equivalent* if $\mathbf{P}(\{X_n \text{ converges and } Y_n \text{ diverges}\}) = \mathbf{P}(\{X_n \text{ diverges and } Y_n \text{ converges}\}) = 0$.

Proposition 6.1.2 *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables such that*

$$\sum_{n=1}^{\infty} \mathbf{P}\{X_n \neq Y_n\} < \infty. \tag{6.1.1}$$

Then we have

(i) $\{X_n\}$ *and* $\{Y_n\}$ *are tail equivalent;*

(ii) $\sum_{n=1}^{\infty} X_n$ *and* $\sum_{n=1}^{\infty} Y_n$ *are convergence equivalent;*

(iii) for any $b_n \uparrow \infty$*, the sequences* $b_n^{-1}\sum_{i=1}^{n} X_i$ *and* $b_n^{-1}\sum_{i=1}^{n} Y_i$ *are convergence equivalent and their limits a.s. coincide.*

Proof. By (6.1.1) and Borel-Cantalli lemma,

$$\mathbf{P}\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=1}^{\infty}\{X_k \neq Y_k\}\Big) = 0,$$

or, equivalently,

$$\mathbf{P}\Big(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{X_k = Y_k\}\Big) = 1.$$

Thus $\{X_n\}$ and $\{Y_n\}$ are tail equivalent, giving (i). The assertions (ii) and (iii) are immediate. □

6.2 Weak law of large numbers

Let $\{X_n\}$ be a sequence of random variables defined on $(\Omega, \mathscr{F}, P)$. In this section, we study weak laws of large numbers, which deal with condition of convergence in probability of the partial sums

$$S_n := \sum_{k=1}^{n} X_k, \qquad n = 1, 2, \cdots .$$

Definition 6.2.1 If there are sequences $\{A_n\}$ and $\{B_n\}$ with $0 < B_n \to \infty$ such that $(S_n - A_n)/B_n \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, we say $\{X_n\}$ satisfies a weak law of large numbers.

Theorem 6.2.1 (Chebyshev) *Let $\{X_n\}$ be a sequence of independent random variables. Suppose that there is a constant $\gamma > 0$ such that $\mathbf{Var}(X_n) \leq \gamma$ for all $n \geq 1$. Then we have $(S_n - \mathbf{E}[S_n])/n \xrightarrow{\mathrm{P}} 0$.*

Proof. By Chebyshev's inequality, for any $\varepsilon > 0$ we have

$$\mathbf{P}\Big\{\frac{1}{n}|S_n - \mathbf{E}[S_n]| \geq \varepsilon\Big\} \leq \frac{1}{n^2}\sum_{j=1}^{n}\mathbf{Var}(X_j) \leq \frac{1}{n}\gamma.$$

The right hand side goes to zero as $n \to \infty$. It follows that $(S_n - \mathbf{E}[S_n])/n \xrightarrow{\mathrm{P}} 0$. □

Corollary 6.2.1 *If* $\{X_n\}$ *are i.i.d. random variables with* $\mathbf{E}[X_n^2] < \infty$*, then*

$$\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\text{P}} \mu := \mathbf{E}[X_1].$$

Theorem 6.2.2 *Let* $\{X_n\}$ *be i.i.d random variables with common mean* $\mu = \mathbf{E}[X_n]$*. Then* $S_n/n \xrightarrow{\text{P}} \mu$ *as* $n \to \infty$*.*

Proof. The following argument is a typical application of the truncation method. Let $\delta > 0$ and $n \geq 1$ be fixed. For $0 \leq k \leq n$, set

$$X_k^{(n)} = \begin{cases} X_k & \text{if } |X_k| < n\delta, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\{X_k^{(n)} : 1 \leq k \leq n\}$ are i.i.d. random variables. Write $\beta = \mathbf{E}[|X_1|]$ and define $E_n = \{\omega : |X_1(\omega)| < n\delta\}$. By the dominated convergence theorem,

$$\mu_n^* := \mathbf{E}[X_1^{(n)}] = \mathbf{E}[X_1 1_{E_n}] \to \mu.$$

Moreover, we have

$$\mathbf{Var}(X_1^{(n)}) \leq \mathbf{E}[(X_1^{(n)})^2] = \int_{E_n} X_1^2 d\mathbf{P} \leq n\delta \int_{E_n} |X_1| d\mathbf{P} \leq n\delta\beta.$$

Writing $S_n^* = \sum_{k=1}^n X_k^{(n)}$, we have $\mathbf{E}[S_n^*/n] = \mu_n^*$ and $\mathbf{Var}(S_n^*/n) = \mathbf{Var}(X_1^{(n)})/n \leq \delta\beta$. Choose $N = N(\varepsilon) \geq 1$ such that $|\mu_n^* - \mu| < \varepsilon$ for all $n \geq N$. Then we can see by Chebyshev's inequality that

$$\mathbf{P}\Big\{\Big|\frac{S_n^*}{n} - \mu\Big| \geq 2\varepsilon\Big\} \leq \mathbf{P}\Big\{\Big|\frac{S_n^*}{n} - \mu_n^*\Big| + |\mu_n^* - \mu| \geq \varepsilon\Big\} \leq \mathbf{P}\Big\{\Big|\frac{S_n^*}{n} - \mu_n^*\Big| \geq \varepsilon\Big\} \leq \frac{\beta\delta}{\varepsilon^2}.$$

For $1 \leq k \leq n$, we have

$$\mathbf{P}\{X_k \neq X_k^{(n)}\} = \mathbf{P}\{|X_k| \geq \delta n\} = \mathbf{P}\{|X_1| \geq \delta n\} \leq \frac{1}{n\delta}\int_{E_n^c} |X_1| d\mathbf{P}$$

and consequently

$$\mathbf{P}\{S_n \neq S_n^*\} \leq \mathbf{P}\Big(\bigcup_{k=1}^n \{X_k \neq X_k^{(n)}\}\Big) \leq \sum_{k=1}^n \mathbf{P}\{X_k \neq X_k^{(n)}\} \leq \frac{1}{\delta}\int_{E_n^c} |X_1| d\mathbf{P}.$$

It then follows that

$$\mathbf{P}\Big\{\Big|\frac{S_n}{n} - \mu\Big| \geq 2\varepsilon\Big\} \leq \mathbf{P}\Big(\Big\{\Big|\frac{S_n^*}{n} - \mu\Big| \geq 2\varepsilon\Big\} \cup \{S_n \neq S_n^*\}\Big) \leq \frac{\beta\delta}{\varepsilon^2} + \frac{1}{\delta}\int_{E_n^c} |X_1| d\mathbf{P}. \qquad (6.2.1)$$

By the dominated convergence theorem,

$$\lim_{n\to\infty} \int_{E_n^c} |X_1| d\mathbf{P} = \lim_{n\to\infty} \int_{\Omega} |X_1| 1_{E_n^c} d\mathbf{P} = 0.$$

Then (6.2.1) implies that

$$\limsup_{n\to\infty} \mathbf{P}\Big\{\Big|\frac{S_n}{n} - \mu\Big| \geq 2\varepsilon\Big\} \leq \frac{\beta\delta}{\varepsilon^2}.$$

Since $\delta > 0$ is arbitrary, we must have

$$\lim_{n\to\infty} \mathbf{P}\Big\{\Big|\frac{S_n}{n} - \mu\Big| \geq 2\varepsilon\Big\} = 0,$$

which yields the desired convergence. □

6.3 Kolmogorov's inequalities

Proposition 6.3.1 *Suppose that* $\{X_1, \cdots, X_n\}$ *is a finite sequence of independent random variables such that* $\mathbf{E}[X_k] = 0$ *and* $\sigma_k^2 = \mathbf{Var}(X_k) < \infty$ *for* $1 \le k \le n$. *Let* $S_k = \sum_{i=1}^k X_i$. *Then, for every* $\epsilon > 0$,

$$\mathbf{P}\Big\{\max_{1\le k\le n} |S_k| \ge \epsilon\Big\} = \mathbf{P}\Big(\bigcup_{k=1}^n \{|S_k| \ge \epsilon\}\Big) \le \frac{1}{\epsilon^2}\sum_{k=1}^n \sigma_k^2.$$

Proof. Let $E = \{\max_{1\le k\le n} |S_k| \ge \epsilon\}$ and $E_1 = \{|S_1| \ge \epsilon\}$. For $2 \le k \le n$ let

$$E_k = \{|S_k| \ge \epsilon\} \cap \bigcap_{j=1}^{k-1} \{|S_j| < \epsilon\}.$$

Then $\{E_1, \cdots, E_n\}$ are disjoint and $E = \bigcup_{k=1}^n E_k$. By the independence of the sequence,

$$\sum_{k=1}^n \sigma_k^2 = \mathbf{Var}(S_n) = \int_\Omega S_n^2 d\mathbf{P} \ge \int_E S_n^2 d\mathbf{P} = \sum_{k=1}^n \int_{E_k} S_n^2 d\mathbf{P}. \tag{6.3.1}$$

For any $1 \le k \le n$ we have

$$\begin{aligned}\int_{E_k} S_n^2 d\mathbf{P} &= \int_{E_k} \Big(S_k + \sum_{j=k+1}^n X_j\Big)^2 d\mathbf{P} \\ &= \int_{E_k} S_k^2 d\mathbf{P} + \sum_{j=k+1}^n \int_{E_k} X_j^2 d\mathbf{P} + 2\sum_{j=k+1}^n \int_{E_k} S_k X_j d\mathbf{P} \\ &\qquad + 2\sum_{k+1\le i<j\le n} \int_{E_k} X_i X_j d\mathbf{P}, \end{aligned} \tag{6.3.2}$$

where

$$\int_{E_k} S_k X_j d\mathbf{P} = \mathbf{E}[1_{E_k} S_k X_j] = \mathbf{E}[1_{E_k} S_k]\mathbf{E}[X_j] = 0$$

and

$$\int_{E_k} X_i X_j d\mathbf{P} = \mathbf{E}[1_{E_k} X_i X_j] = \mathbf{E}[1_{E_k} X_i]\mathbf{E}[X_j] = 0.$$

From (6.3.1) and (6.3.2) it follows that

$$\sum_{k=1}^n \sigma_k^2 \ge \sum_{k=1}^n \int_{E_k} S_k^2 d\mathbf{P} \ge \epsilon^2 \sum_{k=1}^n \mathbf{P}(E_k) = \epsilon^2 \mathbf{P}(E),$$

giving the desired inequality. □

Proposition 6.3.2 *Let* $\{X_1, \cdots, X_n\}$ *be a set of centered and independent random variables. Suppose there is a constant* $\gamma > 0$ *such that* $\mathbf{P}\{|X_k| \le \gamma\} = 1$ *for all* $1 \le k \le n$. *Let* $S_k = \sum_{i=1}^k X_i$. *Then for every* $\epsilon > 0$,

$$\mathbf{P}\Big\{\max_{1\le k\le n} |S_k| \ge \epsilon\Big\} = \mathbf{P}\Big(\bigcup_{k=1}^n \{|S_k| \ge \epsilon\}\Big) \ge 1 - \frac{(\epsilon+\gamma)^2}{\gamma^2 + \mathbf{Var}(S_n)}.$$

Proof. Let E and E_k be defined as in the proof of Proposition 6.3.1. Let $F_0 = \Omega$ and $F_k = \bigcap_{j=1}^k \{|S_j| < \epsilon\}$ for $k \geq 1$. Clearly, $F_{k-1} = E_k \cup F_k$ and $E_k \cap F_k = \emptyset$ for $1 \leq k \leq n$. Note also that

$$\begin{aligned} I_k &:= \int_{F_k} S_k^2 d\mathbf{P} - \int_{F_{k-1}} S_{k-1}^2 d\mathbf{P} \\ &= \int_{F_{k-1}} S_k^2 d\mathbf{P} - \int_{E_k} S_k^2 d\mathbf{P} - \int_{F_{k-1}} S_{k-1}^2 d\mathbf{P} \\ &= \int_{F_{k-1}} [(S_{k-1} + X_k)^2 - S_{k-1}^2] d\mathbf{P} - \int_{E_k} S_k^2 d\mathbf{P} \\ &= \int_{F_{k-1}} (X_k^2 + 2X_k S_{k-1}) d\mathbf{P} - \int_{E_k} S_k^2 d\mathbf{P}, \end{aligned} \tag{6.3.3}$$

where

$$\int_{F_{k-1}} X_k^2 d\mathbf{P} = \mathbf{E}[1_{F_{k-1}} X_k^2] = \mathbf{P}(F_{k-1})\mathbf{E}[X_k^2] \geq \mathbf{P}(F_n)\mathbf{E}[X_k^2]$$

and

$$\int_{F_{k-1}} X_k S_{k-1} d\mathbf{P} = \mathbf{E}[1_{F_{k-1}} S_{k-1} X_k] = \mathbf{E}[1_{F_{k-1}} S_{k-1}]\mathbf{E}[X_k] = 0.$$

For any $\omega \in E_k$ we have

$$|S_k(\omega)| \leq |S_{k-1}(\omega)| + |X_k(\omega)| \leq \epsilon + \gamma,$$

so that

$$\int_{E_k} S_k^2 d\mathbf{P} \leq (\epsilon + \gamma)^2 \mathbf{P}(E_k).$$

From (6.3.3) it follows that

$$I_k \geq \mathbf{P}(F_n)\mathbf{E}[X_k^2] - (\epsilon + \gamma)^2 \mathbf{P}(E_k)$$

for $1 \leq k \leq n$. Since $E = \bigcup_{k=1}^n E_k = F_n^c$, we can take the summation in both sides to get

$$\begin{aligned} \int_{F_n} S_n^2 d\mathbf{P} &\geq \mathbf{P}(F_n) \sum_{k=1}^n \mathbf{E}[X_k^2] - (\epsilon + \gamma)^2 \mathbf{P}(E) \\ &= \mathbf{P}(F_n)\mathbf{Var}(S_n) - (\epsilon + \gamma)^2 (1 - \mathbf{P}(F_n)). \end{aligned}$$

On the other hand, we have

$$\int_{F_n} S_n^2 d\mathbf{P} \leq \epsilon^2 \mathbf{P}(F_n).$$

It then follows that

$$\epsilon^2 \mathbf{P}(F_n) \geq \mathbf{P}(F_n)\mathbf{Var}(S_n) - (\epsilon + \gamma)^2 (1 - \mathbf{P}(F_n))$$

and so

$$(\epsilon + \gamma)^2 \geq [\mathbf{Var}(S_n) + (\epsilon + \gamma)^2 - \epsilon^2]\mathbf{P}(F_n) \geq [\mathbf{Var}(S_n) + \gamma^2]\mathbf{P}(F_n).$$

We may rewrite this into

$$(\epsilon + \gamma)^2 \geq [\mathbf{Var}(S_n) + \gamma^2](1 - \mathbf{P}(E)),$$

from which the desired inequality follows. □

6.4 Random series

Proposition 6.4.1 *Let $\{X_n\}$ be a sequence of independent random variables. If*

$$\sum_{n=1}^{\infty} \mathbf{Var}(X_n) < \infty,$$

then

$$\sum_{n=1}^{\infty} (X_n - \mathbf{E}[X_n])$$

a.s. converges.

Proof. Set $\sigma_n^2 = \mathbf{Var}(X_n)$ and $S_n = \sum_{j=1}^{n} X_j$. By Proposition 6.3.1 applied to the sequence

$$X_{n+1} - \mathbf{E}[X_{n+1}], X_{n+2} - \mathbf{E}[X_{n+2}], \cdots$$

we have

$$\mathbf{P}\Big(\bigcup_{k=1}^{m} \Big\{ \Big| \sum_{j=n+1}^{n+k} (X_j - \mathbf{E}[X_j]) \Big| \ge \epsilon \Big\} \Big) \le \frac{1}{\epsilon^2} \sum_{j=n+1}^{n+m} \sigma_j^2 \le \frac{1}{\epsilon^2} \sum_{j=n+1}^{\infty} \sigma_j^2.$$

By the lower continuity of the probability measure, we may let $m \to \infty$ to get

$$\mathbf{P}\Big(\bigcup_{k=1}^{\infty} \Big\{ \Big| \sum_{j=n+1}^{n+k} (X_j - \mathbf{E}[X_j]) \Big| \ge \epsilon \Big\} \Big) \le \frac{1}{\epsilon^2} \sum_{j=n+1}^{\infty} \sigma_j^2.$$

It follows that

$$\lim_{n\to\infty} \mathbf{P}\Big(\bigcup_{k=1}^{\infty} \Big\{ \Big| \sum_{j=n+1}^{n+k} (X_j - \mathbf{E}[X_j]) \Big| \ge \epsilon \Big\} \Big) = 0.$$

By Proposition 5.2.2 we see that the sequence $\{S_n - ES_n\}$ is a.s. Cauchy and hence a.s. converges.
□

Proposition 6.4.2 *Let $\{X_n\}$ be a sequence of independent random variables such that $|X_n| \le \gamma$ a.s. for some constant $\gamma \ge 0$. If*

$$\sum_{n=1}^{\infty} \mathbf{Var}(X_n) = \infty, \tag{6.4.1}$$

then the series

$$\sum_{n=1}^{\infty} (X_n - \mathbf{E}[X_n]) \tag{6.4.2}$$

a.s. converges.

Proof. Note that $\big|X_n - \mathbf{E}[X_n]\big| \le 2\gamma$. By Proposition 6.3.2,

$$\mathbf{P}\Big\{ \sup_{k\ge1} \Big| \sum_{j=n+1}^{n+k} (X_j - \mathbf{E}[X_j]) \Big| \ge \epsilon \Big\} \ge 1 - \frac{(\epsilon + 2\gamma)^2}{4\gamma^2 + \sum_{j=n+1}^{n+k} \sigma_j^2}.$$

If (6.4.1) holds, we have

$$\mathbf{P}\Big\{\sup_{k\geq 1}\Big|\sum_{j=n+1}^{n+k}(X_j-\mathbf{E}[X_j])\Big|\geq\epsilon\Big\}=1,$$

which clearly implies the a.s. divergence of (6.4.2). (Homework: Give a detailed proof of the last step.) □

Corollary 6.4.1 *Let $\{X_n\}$ be a sequence of independent random variables such that $|X_n|\leq\gamma$ a.s. for some constant $\gamma\geq 0$. Then the series*

$$\sum_{n=1}^{\infty}(X_n-\mathbf{E}[X_n])$$

converges a.s. or diverges a.s. according as

$$\sum_{n=1}^{\infty}\mathbf{Var}(X_n)<\infty \quad \text{or} \quad =\infty.$$

Proposition 6.4.3 *Let $\{X_n\}$ be a sequence of independent random variables such that $|X_n|\leq\gamma$ a.s. for some constant $\gamma\geq 0$. Then $\sum_{n=1}^{\infty}X_n$ converges a.s. if and only if the following condition hold:*

(i) $\sum_{n=1}^{\infty}\mathbf{E}[X_n]$ *converges;*

(ii) $\sum_{n=1}^{\infty}\mathbf{Var}(X_n)<\infty$.

Proof. By Proposition 6.4.2, conditions (i) and (ii) imply the a.s. converges of the series $\sum_{n=1}^{\infty}X_n$. Then it suffices to show that the converse assertion. Suppose that $\sum_{n=1}^{\infty}X_n$ a.s. converges. On an extension of the original probability space, we may construct a sequence of random variables $\{Y_n\}$ which is i.i.d. with $\{X_n\}$. Let $Z_n=X_n-Y_n$. Then $\{Z_n\}$ is a sequence of independent random variables. Moreover, we have $|Z_n|\leq 2\gamma$ a.s. with $\mathbf{E}[Z_n]=0$ and $\mathbf{Var}(Z_n)=2\mathbf{Var}(X_n)$. Since $\sum_{n=1}^{\infty}X_n$ a.s. converges, so does $\sum_{n=1}^{\infty}Y_n$. It follows that

$$\sum_{n=1}^{\infty}Z_n=\sum_{n=1}^{\infty}(X_n-Y_n)$$

also a.s. converges. From Proposition 6.4.2 it follows that $\sum_{n=1}^{\infty}\mathbf{Var}(Z_n)<\infty$, and hence $\sum_{n=1}^{\infty}\mathbf{Var}(X_n)<\infty$. By Proposition 6.4.1, $\sum_{n=1}^{\infty}(X_n-\mathbf{E}[X_n])$ converges a.s. and hence

$$\sum_{n=1}^{\infty}\mathbf{E}[X_n]=\sum_{n=1}^{\infty}X_n-\sum_{n=1}^{\infty}(X_n-\mathbf{E}[X_n])$$

converges. □

Theorem 6.4.1 (Three-Series Criterion) *Let $\{X_n\}$ be a sequence of independent random variables. Then $\sum_{n=1}^{\infty}X_n$ converges a.s. if and only if for some constant $c>0$ the following three series converge:*

(i) $\sum_{n=1}^{\infty} \mathbf{P}\{|X_n| \geq c\}$;

(ii) $\sum_{n=1}^{\infty} \mathbf{E}[X_n^c]$;

(iii) $\sum_{n=1}^{\infty} \mathbf{Var}(X_n^c)$,

where $X_n^c = 1_{\{|X_n|<c\}} X_n$. *Moreover, if (i), (ii) and (iii) converge for some* $c > 0$, *they converge for all* $c > 0$.

Proof. By Borel-Cantelli Lemma, the series (i) converges if and only if

$$\mathbf{P}\Big(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k| \geq c\} \Big) = 0 \tag{6.4.3}$$

or, equivalently,

$$\mathbf{P}\Big(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{X_k = X_k^c\} \Big) = 1. \tag{6.4.4}$$

Suppose $\sum_{n=1}^{\infty} X_n$ a.s. converges. Then $X_n \stackrel{\text{a.s.}}{\to} 0$, and hence we have (6.4.3) for each $c > 0$. It follows that (i) converges. By (6.4.4), the a.s. convergence of $\sum_{n=1}^{\infty} X_n$ implies that of $\sum_{n=1}^{\infty} X_n^c$. By Proposition 6.4.3, the series (ii) and (iii) are both convergent. Conversely, suppose (i), (ii) and (iii) converge for some $c > 0$. By Proposition 6.4.3, we know that $\sum_{n=1}^{\infty} X_n^c$ a.s. converges. In view of (6.4.4), the series $\sum_{n=1}^{\infty} X_n$ a.s. converges. □

6.5 Strong law of large numbers

Lemma 6.5.1 (Toeplitz) *If* $\{a_n\}$ *be a sequence such that* $a_n \to$ *some* a *as* $n \to \infty$, *we have*

$$\frac{1}{n} \sum_{k=1}^{n} a_k \to a \qquad (n \to \infty).$$

Proof. Well-known. □

Lemma 6.5.2 (Kronecker) *If* $\sum_{n=1}^{\infty} a_n$ *converges, then*

$$\frac{1}{n} \sum_{k=1}^{n} k a_k \to 0 \quad (n \to \infty).$$

Proof. Let $s_0 = 0$ and $s_n = \sum_{k=1}^{n} a_k$. Suppose that $s_n \to s$. Clearly,

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n} k a_k &= \frac{1}{n} \sum_{k=1}^{n} k(s_k - s_{k-1}) \\ &= \frac{1}{n} \sum_{k=1}^{n} [k s_k - (k-1) s_{k-1}] - \frac{1}{n} \sum_{k=1}^{n} s_{k-1} \\ &= s_n - \frac{1}{n} \sum_{k=1}^{n-1} s_{k-1} \to s - s = 0. \end{aligned}$$

That gives the result. □

Proposition 6.5.1 *Let $\{X_n\}$ be a sequence of independent random variables with $\sigma_n^2 := \mathbf{Var}(X_n) < \infty$ for each $n \geq 1$. If $\sum_{n=1}^{\infty} \sigma_n^2/n^2 < \infty$, then*

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbf{E}[X_k]) \stackrel{\text{a.s.}}{\to} 0 \qquad (n \to \infty).$$

Proof. Let $Y_n = (X_n - \mathbf{E}[X_n])/n$. Then $\mathbf{E}[Y_n] = 0$ and $\mathbf{Var}(Y_n) = \sigma_n^2/n^2$. It follows that $\sum_{n=1}^{\infty} \mathbf{Var}(Y_n) < \infty$. From Proposition 6.4.1 it follow that $\sum_{n=1}^{\infty} Y_n$ converges a.s. Then

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbf{E}[X_k]) = \frac{1}{n}\sum_{k=1}^{n} kY_k \stackrel{\text{a.s.}}{\to} 0 \qquad (n \to \infty)$$

by Kronecker's Lemma. □

Definition 6.5.1 Let $\{X_n\}$ be a sequence of independent random variables and let $S_n = \sum_{k=1}^{n} X_k$. If there are sequences $\{A_n\}$ and $\{B_n\}$ with $0 < B_n \to \infty$ such that

$$\frac{1}{B_n}(S_n - A_n) \stackrel{\text{a.s.}}{\to} 0 \quad (n \to \infty),$$

we say that $\{X_n\}$ satisfies a strong law of large numbers.

Theorem 6.5.1 (Kolmogorov) *Let $\{X_n\}$ be a sequence of i.i.d. random variables and let $S_n = \sum_{k=1}^{n} X_k$. Then the sequence $\{S_n/n\}$ converges a.s. to a finite limit α if and only if $\mathbf{E}[|X_n|] < \infty$. In this case, we have $\alpha = \mathbf{E}[X_n]$.*

Lemma 6.5.3 *For any random variable X, we have*

$$\sum_{n=1}^{\infty} \mathbf{P}\{|X| \geq n\} \leq \mathbf{E}[|X|] \leq \sum_{n=0}^{\infty} \mathbf{P}\{|X| \geq n\}.$$

Proof. For the first inequality we have

$$\begin{aligned}
\mathbf{E}[|X|] &\geq \sum_{k=1}^{\infty} k\mathbf{P}\{k \leq |X| < k+1\} = \sum_{k=1}^{\infty}\sum_{n=1}^{k} \mathbf{P}\{k \leq |X| < k+1\} \\
&= \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} \mathbf{P}\{k \leq |X| < k+1\} = \sum_{n=1}^{\infty} \mathbf{P}\{|X| \geq n\}.
\end{aligned}$$

The second inequality follows similarly. □

Proof of Theorem 6.5.1. Suppose that $\mathbf{E}[|X_1|] < \infty$. Let $E_n = \{|X_1| \geq n\}$ for $n \geq 0$. Then $E_n \downarrow \emptyset$ and $E_n = \bigcup_{k=n}^{\infty} E_k \setminus E_{k+1}$, which is a union of disjoint events. By Lemma 6.5.3 we have $\sum_{n=1}^{\infty} \mathbf{P}(E_n) < \infty$. For $n \geq 1$, let $X_n^* = 1_{\{|X_n|<n\}} X_n$. It follows that

$$\mathbf{Var}(X_n^*) \leq \int_{\{|X_n|<n\}} X_n^2 d\mathbf{P} \leq \sum_{k=1}^{n} k^2 \mathbf{P}(E_{k-1} \setminus E_k),$$

and hence

$$\sum_{n=1}^{\infty} \frac{\mathbf{Var}(X_n^*)}{n^2} \le \sum_{n=1}^{\infty}\sum_{k=1}^{n} \frac{k^2}{n^2}\mathbf{P}(E_{k-1}\setminus E_k) = \sum_{k=1}^{\infty} k^2\mathbf{P}(E_{k-1}\setminus E_k)\sum_{n=k}^{\infty}\frac{1}{n^2},$$

where

$$\sum_{n=k}^{\infty}\frac{1}{n^2} \le \frac{1}{k^2} + \int_k^{\infty}\frac{1}{x^2}dx = \frac{1}{k^2}+\frac{1}{k} \le \frac{2}{k}.$$

Then we have

$$\sum_{n=1}^{\infty} \frac{\mathbf{Var}(X_n^*)}{n^2} \le 2\sum_{k=1}^{\infty} k\mathbf{P}(E_{k-1}\setminus E_k) \le 2(1+\mathbf{E}[|X_1|]) < \infty.$$

Set $S_n^* = \sum_{k=1}^n X_k^*$. By Proposition 6.5.1 we conclude that $(S_n^* - \mathbf{E}[S_n^*])/n \overset{\text{a.s.}}{\to} 0$. For each $n \ge 1$, we have

$$\mathbf{E}[X_n^*] = \mathbf{E}[1_{\{|X_n|<n\}}X_n] = \mathbf{E}[1_{\{|X_1|<n\}}X_1].$$

Then the dominated convergence theorem implies that $\mathbf{E}[X_n^*] \to \mathbf{E}[X_1]$ as $n\to\infty$. By Toeplitz' Lemma,

$$\frac{1}{n}\mathbf{E}[S_n^*] = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}[X_k^*] \to \mathbf{E}[X_1] \quad (n\to\infty).$$

On the other hand, since

$$\sum_{n=1}^{\infty}\mathbf{P}\{X_n \ne X_n^*\} = \sum_{n=1}^{\infty}\mathbf{P}(E_n) \le \mathbf{E}[|X_1|] < \infty,$$

by Proposition 6.1.2 we have $(S_n - S_n^*)/n \overset{\text{a.s.}}{\to} 0$ and hence $S_n/n \overset{\text{a.s.}}{\to} \mathbf{E}[X_1]$. Conversely, suppose that $\{S_n/n\}$ converges to a finite limit α as $n\to\infty$. We have

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1},$$

so that $X_n/n \overset{\text{a.s.}}{\to} 0$ $(n\to\infty)$. It follows that, for $\epsilon > 0$,

$$\mathbf{P}\Big(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_k/k| \ge \epsilon\}\Big) = 0,$$

By Borel-Cantelli Lemma,

$$\sum_{n=1}^{\infty}\mathbf{P}\{|X_n/n|\ge\epsilon\} = \sum_{n=1}^{\infty}\mathbf{P}\{|X_n|\ge n\epsilon\} < \infty.$$

In particular, we have

$$\sum_{n=1}^{\infty}\mathbf{P}\{|X_n|\ge n\} = \sum_{n=1}^{\infty}\mathbf{P}(E_n) \le \infty$$

and hence

$$\mathbf{E}[X_1] \le \sum_{n=0}^{\infty}\mathbf{P}(E_n) \le \infty.$$

From the first part of the proof it follows that $\alpha = \mathbf{E}[X_1]$. □

Chapter 7

Convergence of Distributions

7.1 Convergence of distribution functions

Definition 7.1.1 Let F and F_n, $n = 1, 2, \cdots$ be distribution functions on $\mathbb{R}$ and let C_F denote the set of continuity points of F. We say F_n *converges weakly* to F if $F_n(x) \to F(x)$ as $n \to \infty$ for every $x \in C_F$. In this case, we write $F_n \xrightarrow{\text{w}} F$.

Clearly, if $F_n \xrightarrow{\text{w}} F$ and $F_n \xrightarrow{\text{w}} G$, then $F \equiv G$.

Proposition 7.1.1 *Suppose that $F_n \xrightarrow{\text{w}} F$. Then $F_n(\pm\infty) \to F(\pm\infty)$ if and only if $F_n(+\infty) - F_n(-\infty) \to F(+\infty) - F(-\infty)$.*

Proof. We only need to prove the "if" part of the proposition. Since $F_n \xrightarrow{\text{w}} F$, for any $x \in C_F$ we have

$$\limsup_{n\to\infty} F_n(\infty) \geq \liminf_{n\to\infty} F_n(\infty) \geq \lim_{n\to\infty} F_n(x) = F(x).$$

By letting $x \to \infty$ in C_F we obtain

$$\limsup_{n\to\infty} F_n(\infty) \geq \liminf_{n\to\infty} F_n(\infty) \geq F(\infty). \tag{7.1.1}$$

Similarly, we can see that

$$\liminf_{n\to\infty} F_n(-\infty) \leq \limsup_{n\to\infty} F_n(-\infty) \leq F(-\infty). \tag{7.1.2}$$

If there is a strict inequality in (7.1.1), we can choose a subsequence $\{n_k\} \subseteq \{n\}$ such that $\lim_{k\to\infty} F_{n_k}(\infty) > F(\infty)$. In view of (7.1.2), we can find $\{m_k\} \subseteq \{n_k\}$ such that $\lim_{k\to\infty} F_{n_k}(-\infty) \leq F(-\infty)$. It then follows that

$$\lim_{k\to\infty} [F_{m_k}(\infty) - F_{m_k}(-\infty)] > F(\infty) - F(-\infty),$$

which is in contradiction to the assumption. There is a similar contradiction if there is a strict inequality in (7.1.2). Then we have the desired result. □

Definition 7.1.2 Let F and F_n, $n = 1, 2, \cdots$ be distribution functions on $\mathbb{R}$ and suppose that $F(\pm\infty)$ are both finite. We say F_n *converges completely* to F and write $F_n \xrightarrow{c} F$ provided $F_n \xrightarrow{w} F$ and $F_n(\pm\infty) \to F(\pm\infty)$ as $n \to \infty$.

Definition 7.1.3 Suppose that X_n and X are random variables with distribution functions F_n and F, respectively. If $F_n \xrightarrow{c} F$, we say X_n *converges in law* to X and write $X_n \xrightarrow{L} X$.

Theorem 7.1.1 (Kolmogorov) *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{L} X$.*

Proof. For any $x < y \in \mathbb{R}$ we have

$$\begin{aligned} \mathbf{P}\{X \le x\} &\le \mathbf{P}\{X_n \le y\} + \mathbf{P}\{X \le x, X_n > y\} \\ &\le \mathbf{P}\{X_n \le y\} + \mathbf{P}\{|X - X_n| > y - x\}. \end{aligned}$$

Let F and F_n denote respectively the distribution of X and X_n. Since $X_n \xrightarrow{P} X$, we obtain $F(x) \le \liminf_{n\to\infty} F_n(y)$ and hence

$$F(y-) \le \liminf_{n\to\infty} F_n(y). \tag{7.1.3}$$

By considering $y < z$ and interchanging X_n and X in the above arguments, we obtain

$$\mathbf{P}\{X_n \le y\} \le \mathbf{P}\{X \le z\} + \mathbf{P}\{|X_n - X| > z - y\},$$

implying $\limsup_{n\to\infty} F_n(y) \le F(z)$ and so

$$\limsup_{n\to\infty} F_n(y) \le F(y+) = F(y). \tag{7.1.4}$$

From (7.1.3) and (7.1.4) we have $F(y) = \lim_{n\to\infty} F_n(y)$ for all $y \in C_F$. Since F and F_n are both probability distribution functions, we get $F_n \xrightarrow{c} F$. □

Corollary 7.1.1 *Let $c \in \mathbb{R}$ be a constant. Then $X_n \xrightarrow{L} c$ if and only if $X_n \xrightarrow{P} c$.*

Proof. (Homework.) □

Theorem 7.1.2 (Helly) *Suppose that $\{F_n\}$ is a uniformly bounded sequence of distribution functions. Then there is a subsequence $\{F_{n_k}\} \subseteq \{F_n\}$ which converges weakly to a bounded distribution function F.*

Proof. Let $D = \{x_1, x_2, \cdots\} \subseteq \mathbb{R}$ be a countable set which is everywhere dense in $\mathbb{R}$. For each $x_i \in D$, the sequence $\{F_n(x_i) : n \ge 1\}$ is bounded. Then we may choose subsequences

$$\{F_n\} \supseteq \{F_n^{(1)}\} \supseteq \{F_n^{(2)}\} \supseteq \{F_n^{(3)}\} \supseteq \cdots$$

such that $F_n^{(i)}(x_i) \to$ some $G(x_i)$ as $n \to \infty$. Let $G_n = F_n^{(n)}$. We shall prove $G_n \xrightarrow{w}$ some F so the theorem will follow. It is easily seen that $G_n(x_i) = F_n^{(n)}(x_i) \to G(x_i)$ for every $i \ge 1$ as $n \to \infty$. For $x \in \mathbb{R} \setminus D$, set

$$G(x) = \inf_{x<z\in D} G(z).$$

Then G is a bounded non-decreasing function on $\mathbb{R}$. For any $x \in \mathbb{R}$, there are $\{y_k, z_k : k \geq 1\} \subseteq D$ such that $y_k \uparrow x$ and $z_k \downarrow x$. We have

$$G(y_k) = \lim_{n\to\infty} G_n(y_k) \leq \liminf_{n\to\infty} G_n(x) \leq \limsup_{n\to\infty} G_n(x) \leq \lim_{n\to\infty} G_n(z_k) = G(z_k).$$

Letting $k \to \infty$ we see that $\lim_{n\to\infty} G_n(x) = G(x)$ for all $x \in C_G$. Now we define $F(x) = G(x+)$ for $x \in \mathbb{R}$. Clearly, F is a bounded distribution function and $F_n \xrightarrow{w} F$. □

7.2 Convergence of integrals

Theorem 7.2.1 *Let F and F_n be distribution functions on $\mathbb{R}$ such that $F_n \xrightarrow{w} F$. If g is a continuous functions on $[a, b]$ with $a, b \in C_F$, then*

$$\lim_{n\to\infty} \int_{(a,b]} g dF_n = \int_{(a,b]} g dF.$$

Proof. For each $k \geq 1$, let $\pi_k = \{a = x_{k,0} < x_{k,1} < \cdots < x_{k,m_k} = b\}$ be a sub-division of $[a, b]$. We choose π_k in the way that $x_{k,i} \in C_F$ for all $0 \leq i \leq m_k$ and $k \geq 1$ and

$$\Delta_k := \max_{1\leq i\leq m_k} (x_{k,i} - x_{k,i-1}) \to 0 \qquad (k \to \infty).$$

For each $k \geq 1$, define the simple function

$$g_k(x) = \sum_{i=1}^{m_k} g(x_{k,i}) 1_{(x_{k,i-1}, x_{k,i}]}(x).$$

Since g is uniformly continuous on $[a, b]$, we have

$$M_k := \sup_{a\leq x\leq b} |g_k(x) - g(x)| \to 0 \qquad (k \to \infty). \tag{7.2.1}$$

By the dominated convergence theorem,

$$\int_{(a,b]} g_k dF_n \to \int_{(a,b]} g dF_n \quad \text{and} \quad \int_{(a,b]} g_k dF \to \int_{(a,b]} g dF \qquad (k \to \infty).$$

Observe that

$$\begin{aligned} \left| \int_{(a,b]} g dF_n - \int_{(a,b]} g dF \right| &\leq \int_{(a,b]} |g - g_k| dF_n + \int_{(a,b]} |g - g_k| dF \\ &\quad + \left| \int_{(a,b]} g_k dF_n - \int_{(a,b]} g_k dF \right| \\ &\leq M_k[F_n(b) - F_n(a)] + M_k[F(b) - F(a)] \\ &\quad + \left| \int_{(a,b]} g_k dF_n - \int_{(a,b]} g_k dF \right|. \end{aligned} \tag{7.2.2}$$

Since $\{x_{k,i}\} \subseteq C_F$, we have

$$\begin{aligned} \lim_{n\to\infty} \int_{(a,b]} g_k dF_n &= \lim_{n\to\infty} \sum_{i=1}^{m_k} g(x_{k,i})[F_n(x_{k,i}) - F_n(x_{k,i-1})] \\ &= \sum_{i=1}^{m_k} g(x_{k,i})[F(x_{k,i}) - F(x_{k,i-1})] = \int_{(a,b]} g_k dF. \end{aligned}$$

Then we get from (7.2.2) that

$$\limsup_{n\to\infty}\left|\int_{(a,b]} gdF_n - \int_{(a,b]} gdF\right| \le 2M_k[F(b)-F(a)].$$

In view of (7.2.1) we may let $k \to \infty$ in the above to see that

$$\limsup_{n\to\infty}\left|\int_a^b gdF_n - \int_a^b gdF\right| = 0.$$

That gives the desired result. □

Theorem 7.2.2 *Let F and F_n be bounded distribution functions on $\mathbb{R}$, and let g be a bounded continuous functions on $\mathbb{R}$. If $F_n \xrightarrow{c} F$, then*

$$\lim_{n\to\infty}\int_{\mathbb{R}} gdF_n = \int_{\mathbb{R}} gdF.$$

Proof. Let $M = \sup_x |g(x)|$. For any $a, b \in C_F$ we have

$$\begin{aligned}\left|\int_{\mathbb{R}} gdF_n - \int_{\mathbb{R}} gdF\right| &\le \int_{(b,\infty)} |g|dF_n + \int_{(-\infty,a]} |g|dF_n + \int_{(b,\infty)} |g|dF \\ &\quad + \int_{(-\infty,a]} |g|dF + \left|\int_{(a,b]} gdF_n - \int_{(a,b]} gdF\right| \\ &\le M[F_n(\infty)-F_n(b)] + M[F_n(a)-F_n(-\infty)] \\ &\quad + M[F(\infty)-F(b)] + M[F(a)-F(-\infty)] \\ &\quad + \left|\int_{(a,b]} gdF_n - \int_{(a,b]} gdF\right|.\end{aligned}$$

It follows that

$$\limsup_{n\to\infty}\left|\int_{-\infty}^{\infty} gdF_n - \int_{-\infty}^{\infty} gdF\right| \le 2M[F(\infty)-F(b)] + 2M[F(a)-F(-\infty)].$$

Letting $a \to -\infty$ and $b \to \infty$ we obtain the desired result. □

7.3 Weak convergence on metric spaces

Let (E, ρ) be a metric space with the Borel σ-algebra denoted by $\mathscr{B}(E)$.

Definition 7.3.1 Let $C(E)$ denotes the space of bounded continuous functions on E. Suppose that μ_n and μ are finite Borel measures on $(E, \mathscr{B}(E))$. We say μ_n *converges weakly* to μ and write $\mu_n \Rightarrow \mu$ provided

$$\int_E f d\mu_n \to \int_E f d\mu \tag{7.3.1}$$

for all $f \in C(E)$.

Proposition 7.3.1 *Let μ_n and μ be finite Borel measures on E such that $\mu(E) > 0$. Then $\mu_n \Rightarrow \mu$ if and only if $\mu_n(E) \to \mu(E)$ and $\mu_n(E)^{-1}\mu_n \Rightarrow \mu(E)^{-1}\mu$.*

Proof. (Homework.) □

Thus the discussion of weak convergence of finite measures can often be reduced to that of probability measures. The following theorem presents a number of equivalent conditions for the weak convergence of probability measures.

Theorem 7.3.1 *Let μ_n and μ be probability measures on $(E, \mathscr{B}(E))$. Then the following statements are equivalent:*

(i) $\mu_n \Rightarrow \mu$;

(ii) (7.3.1) holds for all uniformly continuous $f \in C(E)$;

(iii) $\mu(U) \le \liminf_{n\to\infty} \mu_n(U)$ for all open sets $U \subseteq E$;

(iv) $\mu(C) \ge \limsup_{n\to\infty} \mu_n(C)$ for all closed sets $C \subseteq E$;

(v) $\mu_n(B) \to \mu(B)$ for every continuity set B of μ, that is, $\mu(\partial B) = 0$.

Proof. The implications "(i) ⇒ (ii)" and "(iii) ⇔ (iv)" are obvious.

"(ii) ⇒ (iii)" Suppose that (7.3.1) holds for all uniformly continuous $f \in C(E)$. Let $U \subseteq E$ be an open set and let $h(x) = 1 \wedge \rho(x, U^c)$ and $h_k(x) = h(x)^{1/k}$. Then h_k is uniformly continuous on E and $h_k \uparrow 1_U$ as $k \to \infty$. Note that

$$\liminf_{n\to\infty} \mu_n(U) \ge \lim_{n\to\infty} \int_E h_k d\mu_n = \int_E h_k d\mu.$$

Then we get (iii) by letting $k \to \infty$ in the above inequality and applying the monotone convergence theorem.

"(iv) ⇒ (v)" Suppose that (iv) holds and $B \in \mathscr{B}(E)$ is a continuity set of μ. By the equivalence of (iii) and (iv) we have

$$\mu(\bar{B}) \ge \limsup_{n\to\infty} \mu_n(\bar{B}) \ge \limsup_{n\to\infty} \mu_n(B)$$

and

$$\mu(B^\circ) \le \liminf_{n\to\infty} \mu_n(B^\circ) \le \liminf_{n\to\infty} \mu_n(B).$$

Since $\mu(\partial B) = \mu(\bar{B} \setminus B^\circ) = 0$, we have $\mu(\bar{B}) = \mu(B^\circ) = \mu(B)$. It follows that $\mu(B) = \lim_{n\to\infty} \mu_n(B)$.

"(v) ⇒ (i)" Suppose that (v) holds. It suffices to show (7.3.1) for a non-negative function $f \in C(E)$. Let $a = \sup_{x\in E} f(x)$. By Corollary 3.2.2, we have

$$\int_E f d\mu_n = \int_{[0,a]} \mu_n\{f \ge t\} dt \quad \text{and} \quad \int_E f d\mu = \int_{[0,a]} \mu\{f \ge t\} dt. \tag{7.3.2}$$

It is easy to show that $\partial\{f \ge t\} \subseteq \{f = t\}$. Then $\{f \ge t\}$ is a continuity set of μ if $s \mapsto \mu\{f \ge s\}$ is continuous at $s = t$. Since the non-increasing function $s \mapsto \mu\{f \ge s\}$ has at most countably many discontinuity points, we conclude that $\mu_n\{f \ge t\} \to \mu\{f \ge t\}$ as $n \to \infty$ for a.e. $t \ge 0$. Thus (7.3.1) follows from (7.3.2) and dominated convergence. □

Corollary 7.3.1 *Let F and F_n, $n = 1, 2, \cdots$ be probability distribution functions on $\mathbb{R}$. Then $F_n \xrightarrow{w} F$ if and only if*

$$\int_{\mathbb{R}} f dF_n \to \int_{\mathbb{R}} f dF$$

for all $f \in C(\mathbb{R})$.

Proof. (Homework.)

Chapter 8

Characteristic Functions

8.1 Definition of and basic properties

Definition 8.1.1 Given a probability distribution function F on $\mathbb{R}$ we define its *characteristic function* by

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} dF(x) = \int_{\mathbb{R}} \cos(tx) dF(x) + i \int_{\mathbb{R}} \sin(tx) dF(x), \qquad t \in \mathbb{R}. \tag{8.1.1}$$

We also call φ the *characteristic function* of the probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ determined by F.

Proposition 8.1.1 *The characteristic function φ is uniformly continuous on $\mathbb{R}$ and has the following properties: (i) $\varphi(0) = 1$; (ii) $|\varphi(t)| \leq 1$; (iii) $\varphi(-t) = \bar{\varphi}(t)$, where $\bar{\varphi}$ is the complex conjugate of φ.*

Proof. For $t \in \mathbb{R}$ and $h \in \mathbb{R}$ we have

$$\begin{aligned} |\varphi(t+h) - \varphi(t)| &= \Big| \int_{-\infty}^{\infty} (e^{i(t+h)x} - e^{itx}) dF(x) \Big| \leq \int_{-\infty}^{\infty} |e^{ihx} - 1| dF(x) \\ &= \int_{-\infty}^{\infty} |\cos(hx) - 1 + i \sin(hx)| dF(x) \\ &= \int_{-\infty}^{\infty} \sqrt{2 - 2\cos(hx)}\, dF(x) \\ &= 2 \int_{-\infty}^{\infty} \Big| \sin\Big(\frac{hx}{2}\Big) \Big| dF(x), \end{aligned}$$

where the right side is independent of $t \in \mathbb{R}$. Using dominated convergence we see that $\sup_{t \in \mathbb{R}} |\varphi(t+h) - \varphi(t)| \to 0$ as $h \to 0$, giving the uniform continuity of φ on $\mathbb{R}$. The properties (i), (ii) and (iii) follow immediately from the definition. $\square$

Theorem 8.1.1 *Let F be a probability distribution with finite moments up to order $n \geq 1$. Then the corresponding characteristic function φ has continuous derivatives up the order n and*

$\varphi^{(k)}(0) = i^k\alpha_k$, where

$$\alpha_k = \int_{-\infty}^{\infty} x^k dF(x), \qquad k = 1, \cdots, n. \tag{8.1.2}$$

Moreover, φ admits the expansion

$$\varphi(t) = 1 + \sum_{k=1}^{n} \frac{(it)^k}{k!}\alpha_k + o(t^n) \qquad (t \to 0). \tag{8.1.3}$$

Conversely, suppose that the characteristic function φ has an expression of the form (8.1.3) for an integer $n \geq 1$. Then F has finite moments up to order n if n is even, and up to order $n-1$ if n is odd.

Proof. Suppose that F has finite moments

$$\int_{-\infty}^{\infty} |x|^k dF(x) < \infty, \qquad k = 1, \cdots, n.$$

Observe that

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \int_{-\infty}^{\infty} e^{itx} \frac{e^{ihx} - 1}{h} dF(x)$$

and $|e^{ihx} - 1| \leq |hx|$. By dominated convergence we obtain

$$\frac{d\varphi(t)}{dt} = \lim_{h\to 0} \int_{-\infty}^{\infty} e^{itx} \frac{e^{ihx} - 1}{h} dF(x) = i \int_{-\infty}^{\infty} e^{itx} x dF(x).$$

Thus the first derivative of φ exists and $\varphi^{(1)}(0) = i\alpha_1$. Proceeding inductively we conclude that φ has derivatives up to order n and

$$\varphi^{(k)}(t) = i^k \int_{-\infty}^{\infty} e^{itx} x^k dF(x), \qquad k = 1, \cdots, n.$$

Then $\varphi^{(k)}(0) = i^k\alpha_k$. By dominated convergence we see that $\varphi^{(k)}(t)$ is continuous in $t \in \mathbb{R}$. By Taylor's expansion, for some $0 < \theta < 1$ we have

$$\begin{aligned} \varphi^{(k)}(t) &= 1 + \sum_{k=1}^{n-1} \varphi^{(k)}(0)\frac{t^k}{k!} + \varphi^{(n)}(\theta t)\frac{t^n}{n!} \\ &= 1 + \sum_{k=1}^{n} \varphi^{(k)}(0)\frac{t^k}{k!} + R_n(t) \end{aligned}$$

with

$$R_n(t) = [\varphi^{(n)}(\theta t) - \varphi^{(n)}(0)]\frac{t^n}{n!} = o(t^n).$$

That proves the first part of the theorem. Conversely, suppose that φ has an expansion of the form (8.1.3) where $n = 2m$ is even. Then φ has finite derivative of order $2m$ at $t = 0$. If we define the difference operator Δ_h by $\Delta_h f(t) = f(t+h) - f(t-h)$, it is easy to check that

$$\Delta_h^n \varphi(t) = \int_{-\infty}^{\infty} \left(e^{ihx} - e^{-ihx}\right)^n e^{-itx} dF(x).$$

It then follows that

$$\begin{aligned}\varphi^{(2m)}(0) &= \lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{e^{ihx}-e^{-ihx}}{2h}\right)^{2m} dF(x)\\ &= (-1)^m \lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{\sin(hx)}{h}\right)^{2m} dF(x).\end{aligned}$$

By Fatou's lemma,

$$\int_{-\infty}^{\infty} x^{2m} dF(x) \le \lim_{h\to 0}\int_{-\infty}^{\infty}\left(\frac{\sin(hx)}{h}\right)^{2m} dF(x) = (-1)^m\varphi^{(2m)}(0) < \infty.$$

That proves the existence of the moment of order $2m$. If φ has an expansion of the form (8.1.3) where $n = 2m+1$ is odd, we conclude by the same procedure that F has finite moment of order $2m = n-1$. □

Corollary 8.1.1 *The characteristic function φ has continuous derivatives of all orders if and only if F has finite moments of all orders.*

Corollary 8.1.2 *If φ is a characteristic function and $\varphi(t) = 1 + o(t^{2+\delta})$ for $\delta > 0$ as $t \to 0$. Then φ corresponds to the degenerate distribution at zero.*

Proof. Since $\varphi(t) = 1 + o(t^2)$, by Theorem 8.1.1 the first and the second moments of the distribution of φ are both zero. Thus we have the result. □

We remark that if $\{\varphi_k\}$ is a sequence of characteristic functions and $\{\alpha_k\}$ is a sequence of non-negative numbers such that $\sum_{k=1}^{\infty}\alpha_k = 1$, then $\sum_{k=1}^{\infty}\alpha_k\varphi_k$ is also a characteristic function.

8.2 Inversion Theorems

In this section, we prove two results which show the importance of characteristic functions in the study of distributions. The first of these enables us to compute the distribution function from its characteristic function. The second result gives a sufficient condition for the distribution function to have a density.

Lemma 8.2.1 *For any $\alpha \in \mathbb{R}$ we have*

$$\lim_{T\to\infty}\int_0^T \frac{\sin(\alpha x)}{x}dx = \frac{\pi}{2}\text{sign}(\alpha),$$

where $\text{sign}(\alpha) = 1, -1$ *or* 0 *according as* $\alpha > 0, < 0$ *or* $= 0$.

Proof. By Fubini's Theorem we have

$$\int_0^T \frac{\sin x}{x}dx = \int_0^T dx\int_0^{\infty} e^{-ux}\sin x du = \int_0^{\infty} du\int_0^T e^{-ux}\sin x dx.$$

By integration by parts it is not hard to show that

$$\int_0^T e^{-ux} \sin x dx = \frac{1}{1+u^2} - \frac{e^{-uT}}{1+u^2}(u \sin T + \cos T).$$

It follows that

$$\begin{aligned}\int_0^T \frac{\sin x}{x} dx &= \arctan u \Big|_{u=0}^{u=\infty} - \int_0^\infty \frac{e^{-uT}}{1+u^2}(u \sin T + \cos T) du \\ &= \frac{\pi}{2} - \int_0^\infty \frac{e^{-t}}{T^2+t^2}(u \sin T + T \cos T) dt.\end{aligned}$$

By dominated convergence one sees that the send term on the right hand side goes to zero as $T \to \infty$. Then the desired result follows by a simple change of the integration variable. □

Theorem 8.2.1 (Inversion Theorem) *Let F be a probability distribution function and φ its characteristic function. Then the relation*

$$F(a+h) - F(a-h) = \lim_{T\to\infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin(ht)}{t} e^{-ita} \varphi(t) dt \tag{8.2.1}$$

holds for $a \in \mathbb{R}$ and $h > 0$ whenever the points $a \pm h \in C_F$.

Proof. By Lemma 8.2.1 we have

$$\theta(h,T) := \frac{2}{\pi} \int_0^T \frac{\sin(ht)}{t} dt \to \operatorname{sign}(h)$$

as $T \to \infty$. By Fubini's Theorem we have

$$\begin{aligned}I_T(a,h) &:= \frac{1}{\pi} \int_{-T}^T \frac{\sin(ht)}{t} e^{-ita} \varphi(t) dt \\ &= \frac{1}{\pi} \int_{-T}^T dt \int_{-\infty}^\infty \frac{\sin(ht)}{t} e^{it(x-a)} dF(x) \\ &= \frac{1}{\pi} \int_{-\infty}^\infty dF(x) \int_{-T}^T \frac{\sin(ht)}{t} e^{it(x-a)} dt \\ &= \frac{2}{\pi} \int_{-\infty}^\infty dF(x) \int_0^T \frac{\sin(ht)}{t} \cos[(x-a)t] dt \\ &= \int_{-\infty}^\infty g(x,T) dF(x),\end{aligned}$$

where

$$\begin{aligned}g(x,T) &= \frac{2}{\pi} \int_0^T \frac{\sin(ht)}{t} \cos[(x-a)t] dt \\ &= \frac{1}{\pi} \int_0^T \frac{\sin[(x-a+h)t]}{t} dt - \frac{1}{\pi} \int_0^T \frac{\sin[(x-a-h)t]}{t} dt \\ &= \frac{1}{2}\theta(x-a+h,T) - \frac{1}{2}\theta(x-a-h,T).\end{aligned}$$

Note that both $\theta(\cdot,\cdot)$ and $g(\cdot,\cdot)$ are bounded functions on $\mathbb{R}\times\mathbb{R}_+$ and

$$\lim_{T\to\infty} g(x,T) = \begin{cases} 0 & \text{if } x < a-h, \\ 1/2 & \text{if } x = a-h, \\ 1 & \text{if } a-h < x < a+h, \\ 1/2 & \text{if } x = a+h, \\ 0 & \text{if } x > a+h. \end{cases}$$

Since $a \pm h \in C_F$, we can use dominated convergence to obtain

$$\lim_{T\to\infty} I_T(a,h) = \int_{a-h}^{a+h} dF(x) = F(a+h) - F(a-h).$$

That completes the proof. □

Corollary 8.2.1 *For any $\alpha, \beta \in C_F$ such that $\alpha < \beta$ we have*

$$F(\alpha) - F(\beta) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-it\alpha} - e^{-it\beta}}{it} \varphi(t)dt. \tag{8.2.2}$$

Proof. Letting $a = (\beta+\alpha)/2$ and $h = (\beta-\alpha)/2$ in (8.2.1) we obtain

$$F(\beta) - F(\alpha) = F(a+h) - F(a-h) = \lim_{T\to\infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\sin[(\beta-\alpha)t/2]}{t} e^{-it(\beta+\alpha)/2} \varphi(t)dt.$$

By Euler Formula we have

$$\begin{aligned} & \sin[(\beta-\alpha)t/2]e^{-it(\beta+\alpha)/2} \\ = \; & \sin[(\beta-\alpha)t/2]\cos[t(\beta+\alpha)/2] - i\sin[(\beta-\alpha)t/2]\sin[t(\beta+\alpha)/2] \\ = \; & \frac{1}{2}[\sin(\beta t) - \sin(\alpha t)] + \frac{i}{2}[\cos(\beta t) - i\cos(\alpha t)] \\ = \; & \frac{e^{-it\alpha} - e^{-it\beta}}{2i}. \end{aligned}$$

The proof is completed. □

Corollary 8.2.2 *Let F_1 and F_2 be two probability distribution functions with characteristic functions φ_1 and φ_2, respectively. If $\varphi_1(t) = \varphi_2(t)$ for every $t \in \mathbb{R}$, then $F_1 \equiv F_2$.*

Proof. Let $a, b \in C_{F_1} \cap C_{F_2}$ be such that $a < b$. Then (2) yields

$$F_1(b) - F_1(a) = F_2(b) - F_2(a).$$

Letting $a \to -\infty$ we obtain $F_1(b) = F_2(b)$, so that $F_1 = F_2$ on $C_{F_1} \cap C_{F_2}$ and hence $F_1 \equiv F_2$. □

Corollary 8.2.3 *A characteristic function φ is real and even if and only if the corresponding probability distribution function F is symmetric, that is, $F(x-) = 1 - F(-x)$ for all $x \in \mathbb{R}$.*

Proof. If φ is real, we have $\varphi(t) = \varphi(-t) = \bar{\varphi}(t)$. By the uniqueness we see that φ and $\bar{\varphi}$ has the same distribution function F. Then F is symmetric. Conversely, if F is symmetric, we have $\varphi(t) = [\varphi(t) + \bar{\varphi}(t)]/2$. Thus

$$\varphi(t) = \int_{-\infty}^{\infty} \cos(tx)dF(x),$$

which is real and even. □

Theorem 8.2.2 (Fourier Inversion Theorem) *Suppose that the characteristic function φ is absolutely integrable on $\mathbb{R}$. Then the corresponding distribution function F is absolutely continuous. Moreover, the density function $f = F'$ is bounded and uniformly continuous on $\mathbb{R}$ and it is given by*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\varphi(t)dt, \qquad x \in \mathbb{R}. \tag{8.2.3}$$

Proof. Under the assumption, the integrand on the right-hand side of (8.2.1) is dominated by an absolutely integrable function. Then we can write

$$F(x+h) - F(x-h) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin(ht)}{t} e^{-itx}\varphi(t)dt$$

whenever $x \pm h \in C_F$. Taking the limits of both sides as $h \to 0+$, we obtain $F(x) - F(x-0) = 0$. It follows that F is continuous on $\mathbb{R}$. Using a similar argument on

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin(ht)}{ht} e^{-itx}\varphi(t)dt,$$

we conclude that F is differentiable and

$$f(x) := F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h\to 0+} \frac{\sin(ht)}{ht} e^{-itx}\varphi(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\varphi(t)dt.$$

Finally, we have

$$\begin{aligned} |f(x+h) - f(x)| &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|e^{-it(x+h)} - e^{-itx}\right| |\varphi(t)|dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|e^{ith/2} - e^{-ith/2}\right| |\varphi(t)|dt \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left|\sin(th/2)\right| |\varphi(t)|dt. \end{aligned}$$

Then the uniform continuity of f follows by dominated convergence. □

8.3 Convolution of distributions

Let μ_1 and μ_1 be two probability measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. We can define a probability measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ by

$$\int_{\mathbb{R}} f(y)\mu(dy) = \int_{\mathbb{R}} \mu_1(dx_1) \int_{\mathbb{R}} f(x_1 + x_2)\mu_2(dx_2), \qquad f \in C(\mathbb{R}). \tag{8.3.1}$$

It is not hard to show that μ is the image of the product probability measure $\mu_1 \times \mu_2$ under the mapping $(x_1, x_2) \mapsto x_1 + x_2$. (Homework.)

Definition 8.3.1 The probability measure μ defined by (8.3.1) is called the *convolution* of μ_1 and μ_2 and denoted by $\mu_1 * \mu_2$. We also write $F = F_1 * F_2$, where F, F_1 and F_2 denote the distribution functions of μ, μ_1 and μ_2, respectively.

Theorem 8.3.1 *Let μ, μ_1 and μ_2 be three probability measures with characteristic functions φ, φ_1 and φ_2, respectively. Then $\mu = \mu_1 * \mu_2$ if and only if $\varphi = \varphi_1\varphi_2$.*

Proof. Suppose first that $\mu = \mu_1 * \mu_2$. Then we have (8.3.1). Clearly, this equality also holds for a bounded continuous complex function f. In particular, we have

$$\varphi(t) = \int_{\mathbb{R}} e^{itx}\mu(dx) = \int_{\mathbb{R}} \mu_1(dx_1)\int_{\mathbb{R}} e^{it(x_1+x_2)}\mu_2(dx_2) = \varphi_1(t)\varphi_2(t)$$

for all $t \in \mathbb{R}$. Conversely, suppose that $\varphi = \varphi_1\varphi_2$. Let $\nu = \mu_1 * \mu_2$, and let θ be the characteristic function of ν. Then from what we have shown above it follows that $\theta = \varphi_1\varphi_2 = \varphi$. By the uniqueness of the characteristic function we have $\nu = \mu$. □

By applying Theorem 8.3.1 successively we see that, if φ is a characteristic function, so is φ^k for each integer $k \geq 0$. Then for any $\lambda \geq 0$, the function

$$e^{\lambda(\varphi-1)} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\varphi^k \tag{8.3.2}$$

is also a characteristic function.

Proposition 8.3.1 *If $F = F_1 * F_2$, then*

$$F(x) = \int_{-\infty}^{\infty} F_1(x-y)dF_2(y), \qquad x \in \mathbb{R}. \tag{8.3.3}$$

Proof. Suppose that F is defined by (8.3.3). Let $[a, b]$ be an arbitrary closed bounded interval in $\mathbb{R}$, and let

$$a = x_{n,0} < x_{n,1} < \cdots < x_{n,k_n} = b$$

be a sequence of subdivisions of $[a, b]$ such that

$$\Delta_n := \max_{1\leq k\leq k_n}(x_{n,k} - x_{n,k-1}) \to 0$$

as $n \to \infty$. From Proposition 3.2.5 it follows that

$$\begin{aligned}
\int_a^b e^{itx}dF(x) &= \lim_{n\to\infty}\sum_{k=1}^{k_n} e^{itx_{n,k}}[F(x_{n,k}) - F(x_{n,k-1})] \\
&= \lim_{n\to\infty}\int_{-\infty}^{\infty}\sum_{k=1}^{k_n} e^{it(x_{n,k}-y)}[F_1(x_{n,k}-y) - F_1(x_{n,k-1}-y)]e^{ity}dF_2(y) \\
&= \int_{-\infty}^{\infty}\left[\int_{a-y}^{b-y} e^{itx}dF_1(x)\right]e^{ity}dF_2(y).
\end{aligned}$$

Let φ, φ_1 and φ_2 denote the characteristic functions of F, F_1 and F_2, respectively. Then we can take the limits of both sides of the above equality as $a \to -\infty$ and $b \to \infty$ to obtain $\varphi = \varphi_1\varphi_2$. Then the result follows from Theorem 8.3.1 and the uniqueness of the characteristic function. □

Proposition 8.3.2 *Let $p > 0$ and let μ and ν be probability measures on $\mathbb{R}$. Then*

$$\int_{\mathbb{R}} |x|^p \mu * \nu(dx) < \infty \tag{8.3.4}$$

holds if and only if

$$\int_{\mathbb{R}} |x|^p \mu(dx) + \int_{\mathbb{R}} |x|^p \nu(dx) < \infty. \tag{8.3.5}$$

Proof. Suppose that (8.3.5) holds. Let $c_p = 1$ for $0 < p \leq 1$ and $c_p = 2^{p-1}$ for $p > 1$. From Fubini's theorem it follows that

$$\begin{aligned} \int_{\mathbb{R}} |x|^p \mu * \nu(dx) &= \int_{\mathbb{R}} \mu(dx) \int_{\mathbb{R}} |x+y|^p \nu(dy) \\ &\leq c_p \int_{\mathbb{R}} \mu(dx) \int_{\mathbb{R}} (|x|^p + |y|^p) \nu(dy) \\ &= c_p \Big[\int_{\mathbb{R}} |x|^p \mu(dx) + \int_{\mathbb{R}} |y|^p \nu(dx) \Big] < \infty. \end{aligned}$$

Conversely, if (8.3.4) holds, we have

$$\int_{\mathbb{R}} \mu(dx) \int_{\mathbb{R}} |x+y|^p \nu(dy) = \int_{\mathbb{R}} |x|^p \mu * \nu(dx) < \infty.$$

It follows that

$$\int_{\mathbb{R}} |x+y|^p \nu(dy) < \infty$$

for μ-a.e. $x \in \mathbb{R}$. Taking any $x \in \mathbb{R}$ for which the above is true we find

$$\begin{aligned} \int_{\mathbb{R}} |y|^p \nu(dy) &\leq c_p \int_{\mathbb{R}} (|x+y|^p + |x|^p) \nu(dy) \\ &= c_p \int_{\mathbb{R}} |x+y|^p \nu(dy) + c_p |x|^p < \infty. \end{aligned}$$

By symmetry we see that

$$\int_{\mathbb{R}} |x|^p \mu(dx) < \infty.$$

That proves the desired result. □

8.4 Continuity Theorem

In this section, we prove Lévy's continuity theorem, which gives a necessary and sufficient condition for the complete convergence of a sequence of probability distribution functions.

Lemma 8.4.1 *For any $\alpha \in \mathbb{R}$ we have*

$$\int_0^\infty \frac{1-\cos x}{x^2} dx = \frac{\pi}{2}.$$

Proof. By integration by parts,

$$\int_0^T \frac{1-\cos x}{x^2}dx = -\frac{1-\cos x}{x}\bigg|_0^T + \int_0^T \frac{\sin x}{x}dx.$$

Then the desired result is reduced to Lemma 8.2.1. □

Lemma 8.4.2 *Let F be a probability distribution function with characteristic function φ. Then we have*

$$\int_0^h F(y)\,dy - \int_{-h}^0 F(y)dy = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos(ht)}{t^2}\varphi(t)dt \tag{8.4.1}$$

for every $h > 0$.

Proof. Let $a > 0$ and let G be the uniform distribution on $[-a, a]$ with characteristic function $\theta(t) = \sin(at)/at$. The convolution $H = F * G$ is given by

$$H(x) = \int_{-\infty}^{\infty} F(x-y)G(dy) = \frac{1}{2a}\int_{-a}^{a} F(x-y)dy = \frac{1}{2a}\int_{x-a}^{x+a} F(z)dz, \tag{8.4.2}$$

which is continuous on $\mathbb{R}$. Let ψ be the characteristic function of H. From Theorem 8.3.1 it follows that

$$\psi(t) = \varphi(t)\theta(t) = \varphi(t)\frac{\sin(at)}{at}.$$

Applying the inversion theorem to H and ψ we obtain

$$\begin{aligned} H(x+a) - H(x-a) &= \lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\sin^2(at)}{at^2}e^{-itx}\varphi(t)dt \\ &= \frac{1}{2\pi a}\int_{-\infty}^{\infty}\frac{1-\cos(2at)}{t^2}e^{-itx}\varphi(t)dt. \end{aligned}$$

In particular, we have

$$H(h/2) - H(-h/2) = \frac{1}{\pi h}\int_{-\infty}^{\infty}\frac{1-\cos(ht)}{t^2}\varphi(t)dt.$$

On the other hand, from (8.4.2) we obtain

$$H(h/2) - H(-h/2) = \frac{1}{h}\int_0^h F(z)dz - \frac{1}{h}\int_{-h}^0 F(z)dz,$$

Then the desired equality follows. □

Theorem 8.4.1 (Lévy) *Let $\{F_n\}$ be a sequence of probability distribution functions with characteristic functions $\{\varphi_n\}$. Then $\{F_n\}$ converges completely to a probability distribution function F if and only if $\varphi_n \to$ some φ on $\mathbb{R}$ as $n \to \infty$ and φ is continuous at $t = 0$. In this case, the limit function φ is the characteristic function of F.*

Proof. Suppose that $F_n \xrightarrow{c} F$. From Theorem 7.2.2 it follows that $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$ as $n \to \infty$, where φ is the characteristic function of F. Conversely, suppose that $\varphi_n \to \varphi$ on $\mathbb{R}$ and φ is continuous at $t = 0$. Since $\{F_n\}$ is a sequence of probability distribution functions, Helly's theorem implies the existence of a subsequence $\{F_{n_k}\}$ which converges weakly to a bounded distribution function F. We show that F is a probability distribution function. From Lemma 8.4.2 we obtain

$$\int_0^h F_{n_k}(y)dy - \int_{-h}^0 F_{n_k}(y)dy = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos(ht)}{t^2}\varphi_{n_k}(t)dt$$

for every $h > 0$. By dominated convergence we can let $k \to \infty$ in the above equality to obtain

$$\int_0^h F(y)dy - \int_{-h}^0 F(y)dy = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos(ht)}{t^2}\varphi(t)dt.$$

Dividing both sides by h, we have

$$\frac{1}{h}\left[\int_0^h F(y)dy - \int_{-h}^0 F(y)dy\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos t}{t^2}\varphi\Big(\frac{t}{h}\Big)dt. \tag{8.4.3}$$

Since φ is continuous at $t = 0$, we have

$$\lim_{h\to\infty} \varphi(t/h) = \varphi(0) = \lim_{n\to\infty} \varphi_n(0) = 1.$$

Letting $h \to \infty$ in (8.4.3), we obtain

$$F(+\infty) - F(-\infty) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos t}{t^2}dt = 1.$$

Then F is a probability distribution function. From the first part of the proof we conclude that φ is the characteristic function corresponding to F. Now suppose that $\{F_n\}$ contains another subsequence which converges to a limit, say F^*. Proceeding as above, we see that F^* is a probability distribution function and φ is the corresponding characteristic function. By the uniqueness theorem it follows that $F = F^*$, which implies that every weakly convergent subsequence of $\{F_n\}$ has the same limit F. This shows that $F_n \xrightarrow{c} F$ and φ is the characteristic function of F. □

Theorem 8.4.2 *Let $\{F_n\}$ be a sequence of probability distribution functions with characteristic functions $\{\varphi_n\}$. Then $\{F_n\}$ converges completely to a probability distribution function F if and only if $\varphi_n \to$ some φ uniformly on each bounded interval. In this case, the limit function φ is the characteristic function of F.*

Proof. See Laha and Rohatgi (1979, p.156-158). □

8.5 Criteria for characteristic functions

In this section we derive some important necessary and sufficient conditions for a complex-valued function on $\mathbb{R}$ to be a characteristic function. For this purpose we first introduce the concept of a *positive definite function* on $\mathbb{R}$, which is due to Bochner.

Definition 8.5.1 Let φ be a complex-valued function defined on $\mathbb{R}$. Then φ is said to be *positive definite* on $\mathbb{R}$ if the inequality

$$\sum_{k=1}^{n}\sum_{l=1}^{n}\omega_k\bar{\omega}_l\varphi(t_k-t_l)\geq 0 \tag{8.5.1}$$

holds for all finite sets $\{t_1,\cdots,t_n\}\subset\mathbb{R}$ and $\{\omega_1,\cdots,\omega_n\}\subset\mathbb{C}$.

We note the following elementary property of a positive definite function.

Proposition 8.5.1 *If φ is a positive definite function on $\mathbb{R}$, we have: (a) $\varphi(0)\geq 0$; (b) $\varphi(-t)=\bar{\varphi}(t)$; (c) $|\varphi(t)|\leq\varphi(0)$.*

Proof. (a) This follows by setting $n=1$, $t_1=0$ and $\omega_1=1$ in (8.5.1).

(b) Setting $n=2$, $t_1=0$, $t_2=t$, $\omega_1=\omega$ and $\omega_2=1$ in (8.5.1) we see

$$(1+|\omega|^2)\varphi(0)+\omega\varphi(-t)+\bar{\omega}\varphi(t)\geq 0.$$

Then we use (a) to see that $\omega\varphi(-t)+\bar{\omega}\varphi(t)$ is real. Now set $\omega=1$ and $\omega=i=\sqrt{-1}$ successively to see that $\varphi(-t)+\varphi(t)=a$ and $\varphi(t)-\varphi(-t)=ib$ for real numbers a and b. It follows that $2\varphi(t)=a+ib$ and $2\varphi(-t)=a-ib$, yielding (b).

(c) We first consider the case where $\varphi(0)=0$. Setting $n=2$, $t_1=0$, $t_2=t$, $\omega_1=\varphi(t)$ and $\omega_2=-1$ in (8.5.1) and using (b) we get

$$0\leq(|\varphi(t)|+1)\varphi(0)-\varphi(t)\varphi(-t)-\bar{\varphi}(t)\varphi(t)=-2|\varphi(t)|^2.$$

Then $\varphi(t)=0$ for all $t\in\mathbb{R}$. When $\varphi(0)>0$, we set $n=2$, $t_1=0$, $t_2=t$, $\omega_1=\varphi(t)/\varphi(0)$ and $\omega_2=-1$ to get

$$0\leq\left(\frac{|\varphi(t)|^2}{|\varphi(0)|^2}+1\right)\varphi(0)-\frac{\varphi(t)}{\varphi(0)}\varphi(-t)-\frac{\bar{\varphi}(t)}{\varphi(0)}\varphi(t)=\varphi(0)-\frac{|\varphi(t)|^2}{\varphi(0)},$$

which yields (c). □

By Proposition 8.5.1, if a positive definite function φ satisfies $\varphi(0)=0$, then $\varphi(t)=0$ for all $t\in\mathbb{R}$. We say a positive definite function φ is is *normalized* if $\varphi(0)=1$. Note that a positive definite function need not be continuous. As an example, we can consider the function defined by $\varphi(0)=1$ and $\varphi(t)=0$ for $t\neq 0$. We shall see, however, that if a positive definite function is continuous at the origin, it is a characteristic function and hence uniformly continuous.

Lemma 8.5.1 *Let $\{\theta(s):s=0,\pm1,\pm2,...\}$ be a sequence of complex numbers such that $\theta(0)=1$ and*

$$\sum_{k=0}^{n}\sum_{l=0}^{n}\omega_k\bar{\omega}_l\theta(k-l)\geq 0 \tag{8.5.2}$$

for every finite set $\{\omega_0,\omega_1,...,\omega_n\}\subset\mathbb{C}$. Then there exists a probability distribution function G concentrated on $[-\pi,\pi]$ such that

$$\theta(s)=\int_{[-\pi,\pi]}e^{isx}dG(x),\qquad s=0,\pm1,2,\cdots. \tag{8.5.3}$$

Proof. From (8.5.2) we obtain

$$g_n(x) := \frac{1}{n}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} e^{-i(k-l)x}\theta(k-l) \geq 0, \qquad n \geq 1, x \in \mathbb{R}. \tag{8.5.4}$$

Observe that the integer $e^{-irx}\theta(r)$ occurs in $N - |r|$ terms of the sum in (8.5.4) for $-N+1 \leq r \leq N-1$. Hence we can rewrite (8.5.4) as

$$g_n(x) = \sum_{r=-n}^{n} \Big(1 - \frac{|r|}{n}\Big) e^{-irx}\theta(r) \geq 0. \tag{8.5.5}$$

Multiplying both sides of (8.5.5) by e^{isx} for an integer $s \in [-n, n]$ and integrating with respect to the Lebesgue measure on $[-\pi, \pi]$ we obtain

$$\int_{-\pi}^{\pi} e^{isx} g_n(x)dx = \sum_{r=-n}^{n} \Big(1 - \frac{|r|}{n}\Big)\theta(r) \int_{-\pi}^{\pi} e^{i(s-r)x}dx = 2\pi\Big(1 - \frac{|s|}{n}\Big)\theta(s).$$

Since $\theta(0) = 1$, the formula

$$G_n(x) = 1_{[\pi,\infty)}(x) + 1_{[-\pi,\pi)}(x)\frac{1}{2\pi}\int_{-\pi}^{x} g_n(y)dy$$

defines a distribution function G_n which determines a probability measure concentrated on $[-\pi, \pi]$. It follows that

$$\Big(1 - \frac{|s|}{n}\Big)\theta(s) = \int_{[-\pi,\pi]} e^{isx}dG_n(x), \qquad -n \leq s \leq n. \tag{8.5.6}$$

By Helly's theorem, it is easy to see that $\{G_n\}$contains a subsequence $\{G_{n_k}\}$ that converges weakly to a probability distribution G concentrated on $[-\pi, \pi]$. From (8.5.6) and the Helly-Bray theorem we get

$$\theta(s) = \int_{[-\pi,\pi]} e^{isx}dG(x), \qquad s = 0, \pm 1, \cdots.$$

This completes the proof. □

Theorem 8.5.1 (Bochner) *Let φ be a complex-valued function defined on $\mathbb{R}$. Then φ is a continuous, normalized and positive definite function on $\mathbb{R}$ if and only if φ is the characteristic function of a probability distribution function.*

Proof. Suppose that φ is the characteristic function of a probability distribution function F. Clearly, φ is continuous on $\mathbb{R}$ and $\varphi(0) = 1$. For $\{t_1, \cdots, t_n\} \subset \mathbb{R}$ and $\{\omega_1, \cdots, \omega_n\} \subset \mathbb{C}$ we have

$$\begin{aligned}\sum_{k=1}^{n}\sum_{l=1}^{n} \omega_k\bar{\omega}_l\varphi(t_k - t_l) &= \sum_{k=1}^{n}\sum_{l=1}^{n} \omega_k\bar{\omega}_l \int_{\mathbb{R}} e^{i(t_k-t_l)x}dF(x) \\ &= \int_{\mathbb{R}} \Big|\sum_{k=1}^{n} \omega_k e^{it_kx}\Big|^2 dF(x) \geq 0,\end{aligned}$$

which proves (8.5.1). Conversely, let φ be a continuous, normalized and positive definite function on $\mathbb{R}$. We note that for each integer $n \geq 1$ the sequence $\{\varphi(s/n) : s = 0, \pm 1, \pm 2, \cdots\}$ satisfies the condition of Lemma 8.5.1. It follows that there exists a probability distribution function G_n concentrated on $[-\pi, \pi]$ such that

$$\varphi(s/n) = \int_{[-\pi,\pi]} e^{isx} dG_n(x), \qquad s = 0, \pm 1, \pm 2, \cdots \tag{8.5.7}$$

Set $F_n(x) = G_n(x/n)$. Then F_n is a probability distribution function concentrated on $[-n\pi, n\pi]$. Let φ_n be the characteristic function of F_n. We have

$$\varphi_n(t) = \int_{[-n\pi,n\pi]} e^{itx} dF_n(x) = \int_{[-\pi,\pi]} e^{itny} dG_n(y). \tag{8.5.8}$$

It follows from (8.5.7) and (8.5.8) that

$$\varphi(s/n) = \varphi_n(s/n), \qquad s = 0, \pm 1, \pm 2, \cdots \tag{8.5.9}$$

Let $t \in \mathbb{R}$ and $n \geq 1$ be fixed. Then there exists an integer $k = k(t, n)$ such that $0 \leq \theta_n(t) := t - k/n \leq 1/n$. We also have

$$\begin{aligned}
\Big|\varphi_n(t) - \varphi_n\Big(\frac{k}{n}\Big)\Big| &= \Big|\varphi_n\Big(\theta_n(t) + \frac{k}{n}\Big) - \varphi_n\Big(\frac{k}{n}\Big)\Big| \\
&\leq \int_{[-n\pi,n\pi]} |e^{i\theta_n(t)x} - 1| dF_n(x) \\
&\leq \Big[\int_{[-n\pi,n\pi]} |e^{i\theta_n(t)x} - 1|^2 dF_n(x)\Big]^{1/2} \\
&= \Big[2\int_{[-n\pi,n\pi]} [1 - \cos(\theta_n(t)x)] dF_n(x)\Big]^{1/2}.
\end{aligned}$$

Note that $1 - \cos(\theta x) \leq 1 - \cos(x/n)$ for $0 \leq \theta < 1/n$ and $-n\pi \leq x \leq n\pi$. It follows that

$$\begin{aligned}
\Big|\varphi_n(t) - \varphi_n\Big(\frac{k}{n}\Big)\Big| &\leq \Big[2\int_{[-n\pi,n\pi]} \Big[1 - \cos\Big(\frac{x}{n}\Big)\Big] dF_n(x)\Big]^{1/2} \\
&= \Big\{2\Big[1 - \mathrm{Re}\ \varphi_n\Big(\frac{1}{n}\Big)\Big]\Big\}^{1/2} \\
&= \Big\{2\Big[1 - \mathrm{Re}\ \varphi\Big(\frac{1}{n}\Big)\Big]\Big\}^{1/2},
\end{aligned}$$

where we used (8.5.9) for the last equality. Since φ is continuous on $\mathbb{R}$ and $\varphi(0) = 1$, we conclude that

$$\lim_{n\to\infty} \Big|\varphi_n(t) - \varphi_n\Big(\frac{k}{n}\Big)\Big| = 0.$$

Now we note that

$$\begin{aligned}
\varphi_n(t) &= \Big[\varphi_n(t) - \varphi_n\Big(\frac{k}{n}\Big)\Big] + \varphi_n\Big(\frac{k}{n}\Big) \\
&= \Big[\varphi_n(t) - \varphi_n\Big(\frac{k}{n}\Big)\Big] + \varphi\Big(\frac{k}{n}\Big) \to \varphi(t)
\end{aligned}$$

as $n \to \infty$. Then the continuity theorem implies that φ is a probability characteristic function. □

By Bochner's theorem, if φ is a characteristic function, $\bar{\varphi}$, $|\varphi|^2$ and Re φ are also characteristic functions.

Theorem 8.5.2 (Cramer) *Let φ be a bounded, continuous and complex-valued function on $\mathbb{R}$. Then φ is a probability characteristic function if and only if $\varphi(0) = 1$ and*

$$\psi(x,T) = \int_0^T \int_0^T \varphi(s-t)e^{ix(s-t)}dsdt \geq 0 \tag{8.5.10}$$

for every $T > 0$ and $x \in \mathbb{R}$.

Proof. Let φ be the characteristic function of a probability distribution function F. Then $\varphi(0) = 1$ and by Fubini's theorem,

$$\begin{aligned} \psi(x,T) &= \int_0^T \int_0^T \left[\int_{\mathbb{R}} e^{i(s-t)y}dF(y)\right] e^{i(s-t)x}dsdt \\ &= \int_{\mathbb{R}} \left[\int_0^T e^{is(x+y)}ds\right]\left[\int_0^T e^{-it(x+y)}dt\right] dF(y). \end{aligned}$$

Writing $z = x + y$ we have

$$\begin{aligned} \int_0^T e^{isz}ds &= \int_0^T [\cos(sz) + i\sin(sz)]ds = \frac{1}{z}[\sin(sz) - i\cos(sz)]\Big|_{s=0}^{s=T} \\ &= \frac{1}{z}\{\sin(Tz) + i[1-\cos(Tz)]\}. \end{aligned}$$

It follows that

$$\int_0^T e^{isz}ds \int_0^T e^{-itz}dt = \frac{2}{z^2}\{\sin^2(Tz) + [1-\cos(Tz)]^2\} = \frac{2}{z^2}[1-\cos(Tz)].$$

Thus we have

$$\psi(x,T) = 2\int_{\mathbb{R}} \frac{1-\cos[T(x+y)]}{(x+y)^2}dF(y) \geq 0.$$

This proves (8.5.10). Conversely, suppose that φ is a bounded, continuous and complex-valued function such that $\varphi(0) = 1$ and (8.5.10) holds. We have

$$f(x,T) := \frac{1}{2\pi T}\psi(x,T) = \frac{1}{2\pi T}\int_0^T \int_0^T \varphi(s-t)e^{i(s-t)x}dsdt \geq 0.$$

Writing $u = s - t$ and $v = t$ we get by some simple computation that

$$\begin{aligned} f(x,T) &= \frac{1}{2\pi T}\int_{-T}^0 du \int_{-u}^T \varphi(u)e^{iux}dv + \frac{1}{2\pi T}\int_0^T du \int_0^{T-u} \varphi(u)e^{iux}dv \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_T(u)e^{iux}du, \end{aligned}$$

where

$$\varphi_T(u) = 1_{\{|u|\leq T\}}(1 - |u|/T)\varphi(u).$$

For every $W > 0$ set

$$J(t;T,W) = \int_{-W}^W \left(1 - \frac{|x|}{W}\right) f(x,T)e^{itx}dx. \tag{8.5.11}$$

By interchanging the order of the integration and using integration by parts it is not hard to show that

$$\begin{aligned} J(t;T,W) &= \frac{1}{2\pi}\int_{-W}^{W}\Big(1-\frac{|x|}{W}\Big)\Big[\int_{-\infty}^{\infty}\varphi_T(u)e^{i(u+t)x}du\Big]dx \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos[(t+u)W]}{(t+u)^2W}\varphi_T(u)du. \end{aligned}$$

Now writing $v=(t+u)W$ we obtain

$$J(t;T,W)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos v}{v^2}\varphi_T\Big(\frac{v}{W}-t\Big)dv.$$

Then the dominated convergence theorem implies

$$\begin{aligned} \lim_{W\to\infty} J(t;T,W) &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos v}{v^2}\lim_{W\to\infty}\varphi_T(\frac{v}{W}-t)dv \\ &= \frac{2}{\pi}\int_{0}^{\infty}\frac{1-\cos v}{v^2}\varphi_T(-t)dv=\varphi_T(-t), \end{aligned}$$

where we have used Lemma 8.4.1 for the last equality. Since $f(x,T)\geq 0$, we see from (8.5.11) that $J(t;T,W)/J(0;T,W)$ is a probability characteristic function for every $W>0$. Since $\varphi_T(0)=1$, by the continuity theorem we conclude that $\varphi_T(-t)$ and hence $\varphi_T(t)$ is a probability characteristic function. Clearly, $\varphi(t)=\lim_{T\to\infty}\varphi_T(t)$, so that φ is also a probability characteristic function. □

8.6 Criteria for absolute continuity

In this section, we give an additional set of useful sufficient conditions for a function to be the characteristic function of an absolutely continuous distribution; see also Theorem 8.2.2.

Theorem 8.6.1 (Pólya) *Suppose that φ is a real-valued continuous function on $\mathbb{R}$ satisfying the following condition: (i) $\varphi(0)=1$; (ii) $\varphi(-t)=\varphi(t)$; (iii) φ is convex on $(0,\infty)$; (iv) $\lim_{t\to\infty}\varphi(t)=0$. Then φ is the characteristic function of an absolutely continuous probability distribution function.*

Lemma 8.6.1 *Under the assumption of Theorem 8.6.1, the function φ is a.e. differentiable and the derivative has a version φ' that is non-positive and non-decreasing on $(0,\infty)$ and $\lim_{t\to\infty}\varphi'(t)=0$.*

Proof. By Lemma 3.4.1, the function φ has a right-hand derivative ψ on $(0,\infty)$. Suppose that $\psi(t_0)>0$ for some $t_0>0$. Then $\psi(t)>0$ for all $t\geq t_0$. Consequently φ is strictly increasing for $t\geq t_0$. For $t_2\geq t_2\geq t_0$ we have by the convexity of φ that

$$\varphi\Big(\frac{t_1+t_2}{2}\Big)\leq\frac{1}{2}[\varphi(t_1)+\varphi(t_2)].$$

Letting $t_2\to\infty$ and using property (iv) we see that $\varphi(t_1)\geq 0$ for all $t_1\geq t_0$, which contradicts (iv) since φ is strictly increasing for $t\geq t_0$. It follows that $\psi(t)\leq 0$ for every $t>0$. Then φ is non-increasing on $(0,\infty)$, so it differentiable a.e. on $(0,\infty)$. Clearly, the derivative has a version φ' that is non-positive and non-decreasing. Let $\alpha = \lim_{t\to\infty}\varphi'(t)$. Suppose that $\alpha < 0$. From $\varphi'(t) \le \alpha$ for all $t > 0$ we obtain

$$\varphi(t) = \varphi(0) + \int_0^t \varphi'(s)ds \le 1 + \alpha t.$$

Letting $t \to \infty$, we see this inequality contradicts (iv). That proves $\lim_{t\to\infty}\varphi'(t) = 0$. ☐

Lemma 8.6.2 (Pringsheim's Lemma) *Let φ be a non-increasing function on $(0,\infty)$ which is integrable over every finite interval $(0,a)$. Suppose further that $\lim_{t\to\infty}\varphi(t) = 0$. Then for every $t > 0$ we have the inversion formula*

$$\frac{1}{2}[\varphi(t+0) + \varphi(t-0)] = \frac{2}{\pi}\int_0^\infty \cos(tu)\Big[\int_0^\infty \varphi(y)\cos(uy)dy\Big]du.$$

Proof. See e.g. Titchmarsh (1937, p.16). ☐

Proof of Theorem 8.6.1. Let φ' be given by Lemma 8.6.1. For $T > 0$ we can use integrating by parts to see that

$$\int_0^T \cos(tx)\varphi(t)dt = \varphi(T)\frac{\sin(Tx)}{x} - \frac{1}{x}\int_0^T \sin(tx)\varphi'(t)dt.$$

Since $\varphi'(t) \le 0$, the properties (i) and (iv) imply that $-\varphi'(t)dt$ determines a probability measure on $[0,\infty)$. Then the function

$$f(x) := \frac{1}{2\pi}\int_{-\infty}^\infty e^{-itx}\varphi(t)dt = \frac{1}{\pi}\int_0^\infty \cos(tx)\varphi(t)dt$$

is well-defined. By Pringsheim's Lemma, we obtain

$$\varphi(t) = 2\int_0^\infty \cos(tx)f(x)dx = \int_{-\infty}^\infty e^{itx}f(x)dx.$$

We shall prove that f is a probability density function. From (i) we get

$$\int_{-\infty}^\infty f(x)dx = \varphi(0) = 1.$$

For $x > 0$ we can integrate by parts to obtain

$$\begin{aligned} f(x) &= \frac{1}{\pi x}\int_0^\infty \sin(tx)[-\varphi'(t)]dt = \frac{1}{\pi x^2}\int_0^\infty \sin u\Big[-\varphi'\Big(\frac{u}{x}\Big)\Big]du \\ &= \frac{1}{\pi x^2}\sum_{k=0}^\infty \int_{k\pi}^{(k+1)\pi} \sin u\Big[-\varphi'\Big(\frac{u}{x}\Big)\Big]du \\ &= \frac{1}{\pi x^2}\int_0^\pi \sin u\Big\{\sum_{k=0}^\infty (-1)^k\Big[-\varphi'\Big(\frac{u+k\pi}{x}\Big)\Big]\Big\}du. \end{aligned}$$

By Lemma 8.6.1, the function $-\varphi'$ is non-negative and non-increasing on $(0,\infty)$. It follows that $f(x) \ge 0$ for $x > 0$. Since f is even, we also have $f(x) \ge 0$ for $x < 0$. We have proved that φ is the characteristic function of the probability density function f. ☐

Example 8.6.1 It is not hard to show that the conditions of Theorem 8.6.1 are satisfied by the following functions:

(i) $$\varphi(t) = \begin{cases} 1 - |t| & \text{for } 0 < |t| \leq 1/2, \\ 1/4|t| & \text{for } |t| > 1/2. \end{cases}$$

(ii) $$\varphi(t) = \begin{cases} 1 - |t| & \text{for } |t| \leq 1, \\ 0 & \text{for } |t| > 1. \end{cases}$$

Note that the two distinct characteristic functions coincide over a finite interval.

Chapter 9

Signed-Measures and Decompositions

9.1 Hahn and Jordan decompositions

Definition 9.1.1 Let $(\Omega, \mathscr{F})$ be a measurable space. A mapping $\psi : \mathscr{F} \to \bar{\mathbb{R}}$ is called a *signed measure* if $\psi(\emptyset) = 0$ and for any countable sequence of disjoint sets $A_n \subseteq \mathscr{F}$ we have

$$\psi\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \psi(A_n). \tag{9.1.1}$$

We say $D \in \mathscr{F}$ is a *positive set* of ψ if $\psi(A) \geq 0$ for all $A \in \mathscr{F}$ satisfying $A \subseteq D$. Similarly, we call $G \in \mathscr{F}$ a *negative set* if $\psi(A) \leq 0$ for all $A \in \mathscr{F}$ satisfying $A \subseteq G$.

Proposition 9.1.1 *Suppose that ψ is a signed measure on $(\Omega, \mathscr{F})$ and $\psi(A) < 0$ for $A \in \mathscr{F}$. Then there is a negative set $G \in \mathscr{F}$ such that $G \subseteq A$ and $\psi(G) \leq \psi(A)$.*

Proof. Let $\alpha_0 = \sup\{\psi(B) : B \in \mathscr{F} \text{ and } B \subseteq A\}$. Then we have $\alpha_0 \geq \psi(\emptyset) = 0$. If $\alpha_0 = 0$, we can take $G = A$ so that $\psi(G) = \psi(A)$. Otherwise, we have $\alpha_0 > 0$. For each integer $k > 1/\alpha_0$ we have $1/k < \alpha_0$ so there exists $B \in \mathscr{F}$ such that $B \subseteq A$ and $\psi(B) \geq 1/k$. Let k_1 be the smallest positive integer such that there is $A_1 \in \mathscr{F}$ with $A_1 \subseteq A$ and $\psi(A_1) \geq 1/k_1$. It follows that

$$\psi(A \setminus A_1) = \psi(A) - \psi(A_1) \leq \psi(A) - 1/k_1 < \psi(A) < 0.$$

Let $\alpha_1 = \sup\{\psi(B) : B \in \mathscr{F} \text{ and } B \subseteq A \setminus A_1\}$. If $\alpha_1 = 0$, we can take $G = A \setminus A_1$ and the procedure finishes here. In this case, we have $\psi(G) < \psi(A)$. Otherwise, we have $\alpha_1 > 0$. Then the argument just applied to A is applicable to $A \setminus A_1$. We may continue the procedure. If it finishes at a finite time, we can take $G = A \setminus \bigcup_{i=1}^{n} A_i$ for some $n \geq 1$. Otherwise, we get an infinite sequence of positive integers $\{k_n\}$ and an infinite sequence of disjoint sets $\{A_n\} \subseteq \mathscr{F}$ with $A_n \subseteq A_0$. Setting $G = A \setminus \bigcup_{n=1}^{\infty} A_n$ we get

$$0 > \psi(A) = \psi(G) + \sum_{n=1}^{\infty} \psi(A_n) \geq \psi(G) + \sum_{n=1}^{\infty} \frac{1}{k_n},$$

so that $1/k_n \to 0$ as $n \to \infty$. Clearly, for each $B \in \mathscr{F}$ satisfying $B \subseteq G$ we have $\psi(B) \leq 0$. Then $G \in \mathscr{F}$ is a negative set of ψ and $\psi(G) < \psi(A)$. □

Theorem 9.1.1 (Hahn) *Let ψ be a signed measure on $(\Omega, \mathscr{F})$. Then there exists $D \in \mathscr{F}$ such that $\psi(A \bigcap D) \geq 0$ and $\psi(A \bigcap D^c) \leq 0$ for all $A \in \mathscr{F}$.*

Proof. From the additivity of ψ it follows that ether $-\infty \leq \psi(A) < \infty$ for all $A \in \mathscr{F}$, or $-\infty < \psi(A) \leq \infty$ for all $A \in \mathscr{F}$. Without loss of generality, we assume $\psi(A) > -\infty$ for all $A \in \mathscr{F}$. Let $\mathscr{G}$ denote the class of negative sets of ψ. Clearly, $\mathscr{G}$ is closed under countable unions. Let $\beta = \inf\{\psi(B) : B \in \mathscr{G}\}$. We choose a sequence $\{G_n\} \subseteq \mathscr{G}$ such that $\lim_{n\to\infty} \psi(G_n) = \beta$ and let $G = \bigcup_{n=1}^{\infty} G_n$ and $D = G^c$. Clearly, $G \in \mathscr{G}$ and

$$\beta \leq \psi(G) = \psi(G_n) + \psi(G \setminus G_n) \leq \psi(G_n)$$

for each $n \geq 1$. It follows that $0 \geq \beta = \psi(G) > -\infty$. To complete the proof it suffices to show $\psi(A) \geq 0$ for all $A \in \mathscr{F}$ satisfying $A \subseteq D$. If this is not true, there exists some $A_0 \in \mathscr{F}$ such that $A_0 \subseteq D$ and $\psi(A_0) < 0$. By Proposition 9.1.1, there exists $G_0 \in \mathscr{G}$ satisfying $G_0 \subseteq A_0$ and $\psi(G_0) \leq \psi(A_0) < 0$. Then $G_0 \cup G \in \mathscr{G}$ and

$$\psi(G_0 \cup G) = \psi(G_0) + \psi(G) < \beta,$$

which contradicts the definition of β. □

The decomposition $\Omega = D \cup D^c$ given in Theorem 9.1.1 is called a *Hahn decomposition* of Ω with respect to ψ.

Example 9.1.1 Let f be a measurable function on the measure space $(\Omega, \mathscr{F}, \mu)$ such that $\int_\Omega f d\mu$ exists. Then

$$\psi(A) := \int_A f d\mu, \qquad A \in \mathscr{F}, \tag{9.1.2}$$

defines a signed measure ψ. In this case, $\{f \geq 0\} \cup \{f < 0\}$ and $\{f > 0\} \cup \{f \leq 0\}$ are both Hahn decompositions of Ω.

The last example shows that the Hahn decomposition is usually not unique. However, we have the following

Proposition 9.1.2 *Suppose $\Omega = D_1 \cup D_1^c = D_2 \cup D_2^c$ are two Hahn decompositions of Ω with respect to ψ. Then we have*

$$\psi(A \cap D_1) = \psi(A \cap D_2) \quad \textit{and} \quad \psi(A \cap D_1^c) = \psi(A \cap D_2^c) \tag{9.1.3}$$

for all $A \in \mathscr{F}$.

Proof. Observe that $A \cap (D_1 \setminus D_2) = A \cap D_1 \cap D_2^c \subseteq D_1$, so that $\psi(A \cap (D_1 \setminus D_2)) \geq 0$. From $A \cap (D_1 \setminus D_2) \subseteq D_2^c$ we have $\psi(A \cap (D_1 \setminus D_2)) \leq 0$. It follows that $\psi(A \cap (D_1 \setminus D_2)) = 0$. By symmetry we get $\psi(A \cap (D_2 \setminus D_1)) = 0$. Consequently,

$$\psi(A \cap D_1) = \psi((A \cap D_1) \cup (A \cap (D_2 \setminus D_1)) = \psi(A \cap (D_1 \cup D_2)).$$

Similarly we get $\psi(A \cap D_2) = \psi(A \cap (D_1 \cup D_2))$ and so $\psi(A \cap D_1) = \psi(A \cap D_2)$. The second equality in (9.1.3) follows by symmetry. □

Theorem 9.1.2 (Jordan) *Let $\Omega = D \cup D^c$ be a Hahn decomposition with respect to the signed measure ψ. Then*

$$\psi^+(A) = \psi(A \cap D) \quad \text{and} \quad \psi^-(A) = -\psi(A \cap D^c), \quad A \in \mathscr{F},$$

define two measures ψ^+ and ψ^- on $(\Omega, \mathscr{F})$. Moreover, ψ^+ and ψ^- are independent of the particular choice of the Hahn decomposition, at least one of them is finite and $\psi = \psi^+ - \psi^-$.

The representation $\psi = \psi^+ - \psi^-$ is called the *Jordan decomposition* of ψ. The measure $|\psi| := \psi^+ + \psi^-$ is called the *total variation* of ψ. Let f be a measurable function on $(\Omega, \mathscr{F})$. If both

$$\int_\Omega f d\psi^+ \quad \text{and} \quad \int_\Omega f d\psi^-$$

exist and

$$\int_\Omega f d\psi^+ - \int_\Omega f d\psi^-$$

is well-defined, we set

$$\int_\Omega f d\psi = \int_\Omega f d\psi^+ - \int_\Omega f d\psi^-$$

and call it the *integral* of f with respect to the signed measure ψ.

9.2 Radon-Nikodym derivatives

Definition 9.2.1 Let ψ and γ be signed measures on $(\Omega, \mathscr{F})$. We say ψ is *absolutely continuous* with respect to γ and write $\psi \ll \gamma$ if $\psi(A) = 0$ for all $A \in \mathscr{F}$ with $|\gamma|(A) = 0$. If $\psi \ll \gamma$ and $\gamma \ll \psi$, we say ψ and γ are *equivalent* and write $\psi \sim \gamma$.

Example 9.2.1 Let g and h be two measurable functions on $(\Omega, \mathscr{F}, \mu)$ such that $\int_\Omega g d\mu$ and $\int_\Omega h d\mu$ exist. Let ψ_g and ψ_h be two signed measures defined by

$$\psi_g(A) = \int_A g d\mu \quad \text{and} \quad \psi_h(A) = \int_A h d\mu, \quad A \in \mathscr{F}.$$

If $\mu(\{g \neq 0\} \setminus \{h \neq 0\}) = 0$, then $\psi_g \ll \psi_h$. (Homework: Prove this result.)

Theorem 9.2.1 *Let ψ and γ be signed measures on $(\Omega, \mathscr{F})$. Then the following properties are equivalent: (i) $\psi \ll \gamma$; (ii) $\psi^\pm \ll \gamma$; (iii) $\psi^\pm \ll |\gamma|$; (iv) $|\psi| \ll |\gamma|$.*

Proof. "(i) $\Rightarrow$ (ii)" Let $\Omega = D \cup D^c$ be the Hahn decomposition with respect to ψ. If $A \in \mathscr{F}$ and $|\gamma|(A) = 0$, we have $|\gamma|(A \cap D) = |\gamma|(A \cap D^c) = 0$. Then (i) implies that $\psi(A \cap D) = \psi(A \cap D^c) = 0$, that is, $\psi^+(A) = \psi^-(A) = 0$. This shows that $\psi^+ \ll \gamma$ and $\psi^- \ll \gamma$, proving (ii). From the definition of the absolute continuity we know "(ii) $\Rightarrow$ (iii)". The implications "(iii) $\Rightarrow$ (iv)" and "(iv) $\Rightarrow$ (i)" follow from the relations $|\psi|(A) = \psi^+(A) + \psi^-(A)$ and

$$0 \le |\psi(A)| = |\psi^+(A) - \psi^-(A)| \le \psi^+(A) + \psi^-(A) = |\psi|(A)$$

for every $A \in \mathscr{F}$. □

Theorem 9.2.2 *Let ψ and γ be finite measures on $(\Omega, \mathscr{F})$ such that $\psi \ll \gamma$. Then for each $\varepsilon > 0$, there is $\delta > 0$ such that $\gamma(A) < \delta$ implies $\psi(A) < \varepsilon$ for all $A \in \mathscr{F}$.*

Proof. Suppose that for some $\varepsilon_0 > 0$ it is possible to find $\{B_n\} \subseteq \mathscr{F}$ such that $\gamma(B_n) < 1/2^n$ and $\psi(B_n) \geq \varepsilon_0$. Let $B = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k$. Then we have

$$\gamma(B) \leq \sum_{k=n}^{\infty} \gamma(B_k) \to 0 \qquad (n \to \infty)$$

and hence $\gamma(B) = 0$. On the other hand, we have

$$\psi(B) = \lim_{n\to\infty} \psi\Big(\bigcup_{k=n}^{\infty} B_k\Big) \geq \limsup_{n\to\infty} \psi(B_n) \geq \varepsilon_0.$$

Since those contradict the relation $\psi \ll \gamma$, we have proved the desired result. □

Theorem 9.2.3 *Suppose ψ and γ are finite measures on $(\Omega, \mathscr{F})$ with $\psi \ll \gamma$ and $\psi \neq 0$. Then there exist a constant $\varepsilon > 0$ and a positive set $A \in \mathscr{F}$ for $\psi - \varepsilon\gamma$ such that $\gamma(A) > 0$.*

Proof. Let $\Omega = D_n \cup D_n^c$ be a Hahn decomposition with respect to the signed measure $\psi - \gamma/n$. Then D_n is a positive set for $\psi - \gamma/n$. Set $D_0 = \bigcup_{n=1}^{\infty} D_n$ and $D_0^c = \bigcap_{n=1}^{\infty} D_n^c$. Since $D_0^c \subseteq D_n^c$, we have $\psi(D_0^c) - \gamma(D_0^c)/n \leq 0$ and consequently $0 \leq \psi(D_0^c) \leq \gamma(D_0^c)/n$. Then $\psi(D_0^c) = 0$. Since $\psi \neq 0$, we must have $\psi(D_0) > 0$ and so $\gamma(D_0) > 0$ by the absolute continuity $\psi \ll \gamma$. Therefore, there is some $n \geq 1$ such that $\gamma(D_n) > 0$. Now the result follows with $\varepsilon = 1/n$ and $A = D_n$. □

Theorem 9.2.4 (Radon-Nikodym) *Let γ and ψ be σ-finite signed measures on $(\Omega, \mathscr{F})$ such that $\psi \ll \gamma$. Then there is a real-valued measurable function f on $(\Omega, \mathscr{F})$ such that*

$$\psi(A) = \int_A f d\gamma, \qquad A \in \mathscr{F}. \tag{9.2.1}$$

Moreover, the function f is γ-a.e. unique.

Proof. *Step 1)* We first assume that both γ and ψ are finite measures. Let $\mathscr{U}$ be the class of non-negative measurable functions f on $(\Omega, \mathscr{F})$ satisfying

$$\int_A f d\gamma \leq \psi(A), \quad A \in \mathscr{F}.$$

Then we have

$$\alpha := \sup\Big\{\int_\Omega f d\gamma : f \in \mathscr{U}\Big\} \leq \psi(\Omega) < \infty.$$

Choose $\{f_n\} \subseteq \mathscr{U}$ such that

$$\lim_{n\to\infty} \int_\Omega f_n d\gamma = \alpha.$$

Let $g_n = \max_{1\leq k\leq n} f_k$ and

$$E_{n,k} = \{f_1 < g_n\} \cap \cdots \cap \{f_{k-1} < g_n\} \cap \{f_k = g_n\}.$$

Since $\Omega = \bigcup_{k=1}^n E_{n,k}$, for any $A \in \mathscr{F}$ we have

$$\int_A g_n d\gamma = \sum_{k=1}^n \int_{A\cap E_{n,k}} g_n d\gamma = \sum_{k=1}^n \int_{A\cap E_{n,k}} f_k d\gamma \le \sum_{k=1}^n \psi(A\cap E_{n,k}) = \psi(A).$$

Let $f_0 = \sup_{n\ge 1} f_n = \lim_{n\to\infty} g_n$. By the monotone convergence theorem,

$$\int_A f_0 d\gamma = \lim_{n\to\infty} \int_A g_n d\gamma \le \psi(A), \qquad A \in \mathscr{F}$$

and consequently $f_0 \in \mathscr{U}$. Moreover, we have $\int_\Omega f_0 d\gamma = \alpha < \infty$, so γ-a.e. $f_0 < \infty$. Setting $f = f_0 I_{\{f_0<\infty\}}$ we have γ-a.e. $f = f_0$ and so $f \in \mathscr{U}$. Now define the measure ψ_0 on $(\Omega, \mathscr{F})$ by

$$\psi_0(A) = \psi(A) - \int_A f d\gamma = \psi(A) - \int_A f_0 d\gamma, \qquad A \in \mathscr{F}.$$

Clearly, $\psi_0 \ll \psi \ll \gamma$. We shall prove $\psi_0 = 0$, which yields the representation (9.2.1). If this is not true, by Theorem 9.2.3 there exist $\varepsilon > 0$ and a positive set $A \in \mathscr{F}$ for $\psi_0 - \varepsilon\gamma$ such that $\gamma(A) > 0$. Thus

$$\varepsilon\gamma(A\cap B) \le \psi_0(A\cap B) = \psi(A\cap B) - \int_{A\cap B} f d\gamma$$

for each $B \in \mathscr{F}$. It follows that

$$\begin{aligned}\int_B (f+\varepsilon 1_A) d\gamma &= \int_B f d\gamma + \varepsilon\gamma(A\cap B) = \int_{B\setminus A} f d\gamma + \int_{A\cap B} f d\gamma + \varepsilon\gamma(A\cap B) \\ &\le \int_{B\setminus A} f d\gamma + \psi(A\cap B) \le \psi(B\setminus A) + \psi(A\cap B) = \psi(B)\end{aligned}$$

and consequently $g := f + \varepsilon 1_A \in \mathscr{U}$. Since

$$\int_\Omega g d\gamma = \int_\Omega f d\gamma + \varepsilon\gamma(A) = \alpha + \varepsilon\gamma(A) > \alpha,$$

that contradicts the definition of f.

Step 2) Suppose that γ and ψ are both σ-finite measures. Then there is a sequence of disjoint sets $\{\Omega_n\} \subset \mathscr{F}$ such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$ and $\gamma(\Omega_n) + \psi(\Omega_n) < \infty$ for each $n \ge 1$. For any $n \ge 1$ we can define finite measures γ_n and ψ_n by $\gamma_n(A) = \gamma(A\cap\Omega_n)$ and $\psi_n(A) = \psi(A\cap\Omega_n)$ for $A \in \mathscr{F}$. Clearly, we have $\psi_n \ll \gamma_n$. By the last step, there is a sequence of non-negative measurable functions $\{f_n\}$ such that $\psi_n(A) = \int_A f d\gamma_n$ for all $A \in \mathscr{F}$. It follows that

$$\begin{aligned}\psi(A) &= \sum_{n=1}^\infty \psi(A\cap\Omega_n) = \sum_{n=1}^\infty \psi_n(A\cap\Omega_n) = \sum_{n=1}^\infty \int_{A\cap\Omega_n} f_n d\gamma_n \\ &= \sum_{n=1}^\infty \int_{A\cap\Omega_n} f_n d\gamma = \sum_{n=1}^\infty \int_A 1_{\Omega_n} f_n d\gamma = \int_A \sum_{n=1}^\infty 1_{\Omega_n} f_n d\gamma.\end{aligned}$$

Then we get (9.2.1) by setting $f = \sum_{n=1}^\infty 1_{\Omega_n} f_n$.

Step 3) Suppose that γ and ψ are both σ-finite signed measures. Let $\gamma = \gamma^+ - \gamma^-$ and $\psi = \psi^+ - \psi^-$ be the Jordan decompositions for γ and ψ, respectively. Then we have $\psi^\pm \ll |\gamma|$ by Theorem 9.2.1. By the last step, there are non-negative measurable functions $f_\pm$ such that

$\psi^{\pm}(A) = \int_A f_{\pm} d|\gamma|$ for all $A \in \mathscr{F}$. Let $\Omega = D \cup D^c$ be a Hahn decomposition for γ. For any $A \in \mathscr{F}$ we have

$$\begin{aligned}
\psi(A) &= \int_A f_+ d|\gamma| - \int_A f_- d|\gamma| \\
&= \int_{A\cap D} f_+ d\gamma^+ + \int_{A\cap D^c} f_+ d\gamma^- - \int_{A\cap D} f_- d\gamma^+ - \int_{A\cap D^c} f_- d\gamma^- \\
&= \int_{A\cap D} f_+ d\gamma - \int_{A\cap D^c} f_+ d\gamma - \int_{A\cap D} f_- d\gamma + \int_{A\cap D^c} f_- d\gamma \\
&= \int_A \big[(f_+ - f_-)1_D + (f_- - f_+)1_{D^c}\big] d\gamma.
\end{aligned}$$

Then (9.2.1) follows with $f = (f_+ - f_-)1_D + (f_- - f_+)1_{D^c}$.

Step 4) Suppose that f_2 and f_2 are two measurable functions satisfying (9.2.1). Let $\Omega = D \cup D^c$ be a Hahn decomposition for γ. It follows that

$$\int_{\{f_2 > f_1\}} (f_2 - f_1) d\gamma^+ = \int_{\{f_2 > f_1\}\cap D} (f_2 - f_1) d\gamma = 0,$$

which implies $\gamma^+(\{f_2 > f_1\}) = 0$. By symmetry we have $\gamma^+(\{f_1 > f_2\}) = 0$ and hence $\gamma^+(\{f_2 \neq f_1\}) = 0$. A similar arguments shows that $\gamma^-(\{f_2 \neq f_1\}) = 0$. Consequently, we have $|\gamma|(\{f_2 \neq f_1\}) = 0$. That proves that $|\gamma|$-a.e. uniqueness of the function f satisfying (9.2.1). ☐

Definition 9.2.2 If the signed measures ψ and γ are related by (9.2.1), we call f a *Radon-Nikodym derivative* of ψ with respect to γ and write

$$f = \frac{d\psi}{d\gamma} \quad \text{or} \quad d\psi = f d\gamma.$$

Of course, the Radon-Nikodym derivative is only $|\gamma|$-a.e. unique.

Theorem 9.2.5 *If ψ, γ and μ are σ-finite signed measures on $(\Omega, \mathscr{F})$ such that $\psi \ll \gamma$ and $\gamma \ll \mu$, then $\psi \ll \mu$ and μ-a.e.*

$$\frac{d\psi}{d\mu} = \frac{d\psi}{d\gamma}\frac{d\gamma}{d\mu}. \tag{9.2.2}$$

Proof. The assertion $\psi \ll \mu$ is obvious. Thanks to the Hahn decompositions, in showing the relation (9.2.2) we may and do assume μ, γ and ψ are all measures. Accordingly, we have

$$f := \frac{d\psi}{d\gamma} \geq 0 \quad \text{and} \quad g := \frac{d\gamma}{d\mu} \geq 0.$$

Let $\{f_n\}$ be a sequence of simple functions such that $0 \leq f_n \uparrow f$. For each $A \in \mathscr{F}$, we have

$$\lim_{n\to\infty} \int_A f_n d\gamma = \int_A f d\gamma \quad \text{and} \quad \lim_{n\to\infty} \int_A f_n g d\mu = \int_A f g d\mu. \tag{9.2.3}$$

On the other hand, for $B \in \mathscr{F}$ we have

$$\int_A 1_B d\gamma = \gamma(A \cap B) = \int_{A\cap B} g d\mu = \int_A 1_B g d\mu,$$

which implies that

$$\int_A f_n d\gamma = \int_A f_n g d\mu.$$

From (9.2.3) we get

$$\psi(A) = \int_A f d\gamma = \int_A f g d\mu,$$

proving (9.2.2). □

Corollary 9.2.1 *Let ψ and γ be σ-finite signed measures on $(\Omega, \mathscr{F})$ such that $\psi \ll \gamma$. If f is a measurable function on $(\Omega, \mathscr{F})$ for which $\int_\Omega f d\psi$ exists, then*

$$\int_\Omega f d\psi = \int_\Omega f \frac{d\psi}{d\gamma} d\gamma.$$

Proof. (Homework.) □

9.3 Lebesgue decomposition

We say two signed measures ψ and γ on $(\Omega, \mathscr{F})$ are *singular* to each other and write $\gamma \perp \psi$, if there exists $E \in \mathscr{F}$ such that $|\gamma|(E) = |\psi|(E^c) = 0$. Consequently, $\gamma \perp \psi$ if and only if $|\gamma| \perp |\psi|$.

Theorem 9.3.1 (Lebesgue) *For any σ-finite signed measures γ and ψ on $(\Omega, \mathscr{F})$, there are two uniquely determined σ-finite signed measures ψ_0 and ψ_1 such that $\psi_0 \perp \gamma$, $\psi_1 \ll \gamma$ and $\psi_0 + \psi_1 = \psi$.*

Proof. As usual, we may assume both γ and ψ are finite measures. Since $\psi \ll \gamma + \psi$, by Radon-Nikodym theorem, there is a measurable function f_0 such that

$$\psi(A) = \int_A f_0 d(\gamma + \psi) = \int_A f_0 d\gamma + \int_A f_0 d\psi, \qquad A \in \mathscr{F}.$$

It is easy to see that $(\gamma + \psi)$-a.e. $0 \leq f_0 \leq 1$ and so ψ-a.e. $0 \leq f_0 \leq 1$. Let $f = f_0 1_{\{0 \leq f \leq 1\}}$ and write $E = \{f = 1\}$ and $F = E^c = \{0 \leq f < 1\}$. We define the finite measures ψ_0 and ψ_1 by $\psi_0(A) = \psi(A \cap E)$ and $\psi_1(A) = \psi(A \cap F)$ for $A \in \mathscr{F}$. Then $\psi = \psi_0 + \psi_1$. Note that

$$\psi(E) = \int_E f d\gamma + \int_E f d\psi = \gamma(E) + \psi(E).$$

and so $\gamma(E) = 0$. It is then clear that $\psi_0 \perp \gamma$. If $\gamma(A) = 0$, then

$$\int_{A \cap F} d\psi = \psi(A \cap F) = \int_{A \cap F} f d\psi + \int_{A \cap F} f d\gamma = \int_{A \cap F} f d\psi$$

It follows that

$$\int_{A \cap F} (1 - f) d\psi = 0.$$

But $1 - f > 0$ on F, so we must have $\psi(A \cap F) = 0$, that is, $\psi_1(A) = 0$. That proves $\psi_1 \ll \gamma$. Suppose that we also have $\psi = \bar{\psi}_0 + \bar{\psi}_1$ with $\bar{\psi}_0 \perp \gamma$ and $\bar{\psi}_1 \ll \gamma$. Then $\psi_0 - \bar{\psi}_0 = \bar{\psi}_1 - \psi_1$. If we denote this measure by η, then $\eta \perp \gamma$ and $\eta \ll \gamma$. It follows obviously that $\eta = 0$, implying $\psi_0 = \bar{\psi}_0$ and $\psi_1 = \bar{\psi}_1$. □

Chapter 10

Conditional Expectations

10.1 Definition and examples

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. If X is a random variable such that $\mathbf{E}[|X|] < \infty$, then

$$\psi(G) = \int_G X d\mathbf{P}, \qquad G \in \mathscr{G},$$

defines a finite signed measure ψ which is absolutely continuous with respect to the restriction of $\mathbf{P}$ on $\mathscr{G}$. By Radon-Nikodym theorem, there is an a.s. unique $\mathscr{G}$-measurable random variable $\xi : \Omega \to \mathbb{R}$ such that

$$\int_G \xi d\mathbf{P} = \int_G X d\mathbf{P}, \qquad G \in \mathscr{G}. \tag{10.1.1}$$

We call ξ the *conditional expectation* of X given $\mathscr{G}$, and denote it by $\mathbf{E}[X|\mathscr{G}]$. For any $A \in \mathscr{F}$, we call $\mathbf{E}[1_A|\mathscr{G}]$ the *conditional probability* of A given $\mathscr{G}$, and denote it by $\mathbf{P}[A|\mathscr{G}]$. If Y is another random variable, we simply write $\mathbf{E}[X|Y]$ instead of $\mathbf{E}[X|\sigma(Y)]$.

Proposition 10.1.1 *Let X be a random variable such that $\mathbf{E}[|X|] < \infty$. Then we have*

(i) $\mathbf{E}[X|\mathscr{G}] \geq 0$ *a.s. if* $X \geq 0$ *a.s.;*

(ii) $\mathbf{E}[aX|\mathscr{G}] = a\mathbf{E}[X|\mathscr{G}]$ *for any* $a \in \mathbb{R}$*;*

(iii) $\mathbf{E}\{\mathbf{E}[X|\mathscr{G}]\} = \mathbf{E}[X]$*;*

(iv) $\mathbf{E}[X|\mathscr{G}] = X$ *a.s. if* X *is* $\mathscr{G}$*-measurable;*

(v) $\mathbf{E}[X|\mathscr{G}] = \mathbf{E}[X]$ *a.s. if* X *is independent of* $\mathscr{G}$*;*

(vi) $\mathbf{E}[X|\mathscr{G}] = \mathbf{E}[X]$ *a.s. if* $\mathscr{G} = \{\emptyset, \Omega\}$*.*

Proof. These are immediate consequences of the definition. □

Proposition 10.1.2 *Let a and b be real constants and X and Y be random variables. Suppose that $\mathbf{E}[|X|+|Y|]<\infty$. Then we have a.s.*

$$\mathbf{E}[aX+bY|\mathscr{G}] = a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}].$$

Proof. Clearly, $a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}]$ is $\mathscr{G}$-measurable. By Corollary 3.2.3 we have

$$\begin{aligned}\int_G (a\mathbf{E}[X|\mathscr{G}] + b\mathbf{E}[Y|\mathscr{G}])d\mathbf{P} &= a\int_G \mathbf{E}[X|\mathscr{G}]d\mathbf{P} + b\int_G \mathbf{E}[Y|\mathscr{G}]d\mathbf{P} \\ &= a\int_G X d\mathbf{P} + b\int_G Y d\mathbf{P} = \int_G (aX+bY)d\mathbf{P}.\end{aligned}$$

That proves the desired result. □

Theorem 10.1.1 *Let X and Y be random variables such that $\mathbf{E}[|Y|+|XY|]<\infty$ exist. If X is $\mathscr{G}$-measurable, we have*

$$\mathbf{E}[XY|\mathscr{G}] = X\mathbf{E}[Y|\mathscr{G}].$$

Proof. Since X is $\mathscr{G}$-measurable, so is $X\mathbf{E}[Y|\mathscr{G}]$. Then it suffices to prove

$$\int_G X\mathbf{E}[Y|\mathscr{G}]d\mathbf{P} = \int_G XY d\mathbf{P}, \qquad G\in\mathscr{G}. \tag{10.1.2}$$

If $X = 1_H$ for some $H\in\mathscr{G}$, we have $G\cap H\in\mathscr{G}$ and hence

$$\int_G X\mathbf{E}[Y|\mathscr{G}]d\mathbf{P} = \int_{G\cap H} \mathbf{E}[Y|\mathscr{G}]d\mathbf{P} = \int_{G\cap H} Y d\mathbf{P} = \int_G XY d\mathbf{P}.$$

By taking the linear combinations we see that (10.1.2) holds if X is a $\mathscr{G}$-measurable simple function. If X and Y are both non-negative random variables, we may take a sequence of simple functions $\{X_n\}$ such that $0\le X_n\uparrow X$. By applying the monotone convergence theorem, we still get (10.1.2). Finally, for a general X and Y we have

$$\begin{aligned}\int_G XY d\mathbf{P} &= \int_G X^+Y^+d\mathbf{P} - \int_G X^-Y^+d\mathbf{P} - \int_G X^+Y^-d\mathbf{P} + \int_G X^-Y^-d\mathbf{P} \\ &= \int_G X^+\mathbf{E}[Y^+|\mathscr{G}]d\mathbf{P} - \int_G X^-\mathbf{E}[Y^+|\mathscr{G}]d\mathbf{P} - \int_G X^+\mathbf{E}[Y^-|\mathscr{G}]d\mathbf{P} \\ &\quad + \int_G X^-\mathbf{E}[Y^-|\mathscr{G}]d\mathbf{P} \\ &= \int_G (X^+-X^-)\mathbf{E}\big[(Y^+-Y^-)|\mathscr{G}\big]d\mathbf{P} \\ &= \int_G X\mathbf{E}[Y|\mathscr{G}]d\mathbf{P}\end{aligned}$$

by Proposition 10.1.2, which proves the result. □

Theorem 10.1.2 *Let X be a random variable such that $\mathbf{E}[|X|]<\infty$, and let $\mathscr{H}$ and $\mathscr{G}$ be σ-algebras such that $\mathscr{H}\subseteq\mathscr{G}\subseteq\mathscr{F}$. Then we have*

$$\mathbf{E}[X|\mathscr{H}] = \mathbf{E}\{\mathbf{E}[X|\mathscr{G}]|\mathscr{H}\}. \tag{10.1.3}$$

Proof. Let $H \in \mathscr{H}$. By the definition of the conditional expectation,

$$\int_H X d\mathbf{P} = \int_H \mathbf{E}[X|\mathscr{H}]d\mathbf{P}.$$

Since $\mathscr{H} \subseteq \mathscr{G}$, we have $H \in \mathscr{G}$ and hence

$$\int_H X d\mathbf{P} = \int_H \mathbf{E}[X|\mathscr{G}]d\mathbf{P}.$$

It then follows that

$$\int_H \mathbf{E}[X|\mathscr{H}]d\mathbf{P} = \int_H \mathbf{E}[X|\mathscr{G}]d\mathbf{P}.$$

Certainly, $\mathbf{E}[X|\mathscr{H}]$ is $\mathscr{H}$-measurable, so (10.1.3) follows. □

Example 10.1.1 Suppose that $(\Omega, \mathscr{F}, \mathbf{P})$ is a probability space. Let $\mathscr{G} = \sigma(\{U_i : i \in I\})$ for a finite or countable partition $\{U_i : i \in I\} \subseteq \mathscr{F}$ of Ω with $\mathbf{P}(U_i) > 0$ for each $i \in I$. Then for any $A \in \mathscr{F}$ we have

$$\mathbf{P}(A|\mathscr{G})(\omega) = \mathbf{P}(A|U_i) := \frac{\mathbf{P}(A \cap U_i)}{\mathbf{P}(U_i)}, \qquad \omega \in U_i. \tag{10.1.4}$$

This gives an interpretation for the random variable $\mathbf{P}(A|\mathscr{G})(\omega)$. The σ-algebra $\mathscr{G}$ can be interpreted as the information obtained by observing a random system with different states $\{U_i : i \in I\}$ which has some influence on the event A. In this situation, (10.1.4) simply means that the probability of A varies according to the different status of the system. To show (10.1.4), define

$$\eta(\omega) = \mathbf{P}(A|U_i), \qquad \omega \in U_i.$$

From Example 1.1.2 we know that η is a $\mathscr{G}$-measurable random variable. By Example 1.1.1, each $G \in \mathscr{G}$ can be represented as $G = \bigcup_{j \in J} U_j$ for a (finite or countable) set $J \subseteq I$. It follows that

$$\begin{aligned} \int_G 1_A d\mathbf{P} &= \sum_{j \in J} \mathbf{P}(A \cap U_j) = \sum_{j \in J} \mathbf{P}(U_j)\mathbf{P}(A|U_j) \\ &= \int_{\bigcup_{j \in J} U_j} \eta d\mathbf{P} = \int_G \eta d\mathbf{P}. \end{aligned}$$

Then $\eta = \mathbf{E}[1_A|\mathscr{G}] = \mathbf{P}(A|\mathscr{G})$ by the definition of conditional probability.

Example 10.1.2 Consider a probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Let $\mathscr{G} = \sigma(\{U_i : i \in I\})$ for a countable partition $\{U_i : i \in I\} \subseteq \mathscr{F}$ of Ω with $\mathbf{P}(U_i) > 0$ for each $i \in I$. Clearly, for each $i \in I$,

$$\mathbf{P}_i(A) := \mathbf{P}(A|U_i), \qquad A \in \mathscr{F} \tag{10.1.5}$$

defines a probability measure on $(\Omega, \mathscr{F})$. Suppose that X is a random variable such that $\mathbf{E}[|X|] < \infty$. Then we have

$$\mathbf{E}[X|\mathscr{G}](\omega) = \int_\Omega X d\mathbf{P}_i, \qquad \omega \in U_i, \tag{10.1.6}$$

which gives a representation for the conditional expectation. Indeed, by Example 1.1.2,

$$\xi(\omega) = \int_\Omega X d\mathbf{P}_i, \qquad \omega \in U_i$$

defines a $\mathscr{G}$-measurable random variable ξ. By (10.1.5), we have

$$\mathbf{P}_i(A) = \frac{\mathbf{P}(A \cap U_i)}{\mathbf{P}(U_i)}, \qquad A \in \mathscr{F}$$

and hence

$$\int_\Omega X(\omega) d\mathbf{P}_i(\omega) = \mathbf{P}(U_i)^{-1} \int_{U_i} X(\omega) \mathbf{P}(d\omega).$$

If $G \in \mathscr{G}$ has the representation $G = \bigcup_{j\in J} U_j$ for $J \subseteq I$, then

$$\begin{aligned} \int_G X d\mathbf{P} &= \sum_{j\in J} \int_{U_j} X d\mathbf{P} = \sum_{j \in J} \mathbf{P}(U_j) \int_\Omega X d\mathbf{P}_j \\ &= \int_{\bigcup_{j\in J} U_j} \xi d\mathbf{P} = \int_G \xi d\mathbf{P}. \end{aligned}$$

By the definition of conditional expectation we get $\xi = \mathbf{E}[X|\mathscr{G}]$.

10.2 Some properties

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and let $\mathscr{G} \subseteq \mathscr{F}$ be a σ-algebra.

Theorem 10.2.1 (Conditional Monotone Convergence Theorem) *If X_n and X are non-negative random variables such that $\mathbf{E}[|X|] < \infty$ and $X_n \uparrow X$ a.s., then $\mathbf{E}[X_n|\mathscr{G}] \uparrow \mathbf{E}[X|\mathscr{G}]$ a.s.*

Proof. Clearly, $\mathbf{E}[X_n|\mathscr{G}] \uparrow$ some $\mathscr{G}$-measurable random variable $Y \le \mathbf{E}[X|\mathscr{G}]$ a.s. For any $G \in \mathscr{G}$, the monotone convergence theorem implies

$$\int_G Y d\mathbf{P} =\uparrow \lim_{n\to\infty} \int_G \mathbf{E}[X_n|\mathscr{G}] d\mathbf{P} =\uparrow \lim_{n\to\infty} \int_G X_n d\mathbf{P} = \int_G X d\mathbf{P}.$$

Then we have a.s. $Y = \mathbf{E}[X|\mathscr{G}]$. □

Corollary 10.2.1 *For any sequence of disjoint events $\{A_n\} \subseteq \mathscr{F}$, we have a.s.*

$$\mathbf{P}\Big(\bigcup_{n=1}^\infty A_n \Big| \mathscr{G}\Big) = \sum_{n=1}^\infty \mathbf{P}(A_n|\mathscr{G}).$$

Proof. (Homework.) □

Theorem 10.2.2 (Conditional Fatou's Lemma) *Let $\{X_n\}$ be a sequence of non-negative random variables such that $X := \liminf_{n\to\infty} X_n < \infty$ a.s. and $\mathbf{E}[|X_n| + |X|] < \infty$ for every $n \geq 1$. Then we have a.s.*

$$\mathbf{E}[X|\mathscr{G}] \leq \liminf_{n\to\infty} \mathbf{E}[X_n|\mathscr{G}].$$

Proof. (Homework.) □

Theorem 10.2.3 (Conditional Dominated Convergence) *Let X_n and X be random variables such that $X_n \to X$ a.s. Suppose there is a non-negative random variable Y such that $\mathbf{E}[|Y|] < \infty$ and $|X_n| \leq Y$ a.s. for each $n \geq 1$. Then $\mathbf{E}[X_n|\mathscr{G}] \to \mathbf{E}[X|\mathscr{G}]$ a.s.*

Proof. (Homework.) □

Theorem 10.2.4 *Let X_1 and X_2 be two random variables such that X_1 is $\mathscr{G}$-measurable and X_2 is independent of $\mathscr{G}$. For a given bounded Borel function $H(\cdot,\cdot)$ on $\mathbb{R}^2$ define*

$$h(x_1) = \mathbf{E}[H(x_1, X_2)], \qquad x_1 \in \mathbb{R}.$$

Then $h(\cdot)$ is a Borel function on $\mathbb{R}$ and a.s.

$$\mathbf{E}[H(X_1, X_2)|\mathscr{G}] = h(X_1).$$

Proof. Clearly, X_1 and X_2 are independent of each other. Let Q_2 denote the probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ induced by X_2. Then

$$h(x_1) = \int_{E_2} H(x_1, x_2) Q_2(dx_2)$$

By Theorem 4.2.2 we see that $h(\cdot)$ is a Borel function on $\mathbb{R}$. Let Y be a bounded $\mathscr{G}$-measurable random variable and let Q denote the probability measure on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$ induced by (X_1, Y). By the independence of (X_1, Y) and X_2, we have

$$\begin{aligned} \mathbf{E}[H(X_1, X_2)Y] &= \int_{\mathbb{R}^2} Q(dx_1, dy) \int_{\mathbb{R}} H(x_1, x_2) y Q_2(dx_2) \\ &= \int_{\mathbb{R}^2} h(x_1) y Q(dx_1, dy) = \mathbf{E}[h(X_1)Y], \end{aligned}$$

which implies the desired result. □

Theorem 10.2.5 (Jessen's inequality) *Let ϕ be a convex function on $\mathbb{R}$ and X a random variable such that $\mathbf{E}[|X| + |\phi(X)|] < \infty$ exist. Then for any σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ we have a.s.*

$$\phi(\mathbf{E}[X|\mathscr{G}]) \leq \mathbf{E}[\phi(X)|\mathscr{G}].$$

Proof. (Homework.) □

Corollary 10.2.2 *If $p \geq 1$ and $\mathbf{E}[X]$ exits, then a.s.*

$$|\mathbf{E}[X|\mathscr{G}]|^p \leq \mathbf{E}[|X||\mathscr{G}]^p \leq \mathbf{E}[|X|^p|\mathscr{G}].$$

Proof. (Homework.) □

Proposition 10.2.1 *Suppose that X and Y are random variables such that $\mathbf{E}[X^2 + Y^2] < \infty$. If Y is $\mathscr{G}$-measurable, then*

$$\mathbf{E}[(X - Y)^2] = \mathbf{E}[(X - \mathbf{E}[X|\mathscr{G}])^2] + \mathbf{E}[(Y - \mathbf{E}[X|\mathscr{G}])^2].$$

Proof. Observe that

$$(X - Y)^2 = (X - \mathbf{E}[X|\mathscr{G}])^2 + (Y - \mathbf{E}[X|\mathscr{G}])^2 - 2(X - \mathbf{E}[X|\mathscr{G}])(Y - \mathbf{E}[X|\mathscr{G}]). \quad (10.2.1)$$

Under the assumption, each term in the above equation is integrable, so we may take the conditional expectations given $\mathscr{G}$. Since $Y - \mathbf{E}[X|\mathscr{G}]$ is $\mathscr{G}$-measurable, we have

$$\mathbf{E}\big\{(X - \mathbf{E}[X|\mathscr{G}])(Y - \mathbf{E}[X|\mathscr{G}])|\mathscr{G}\big\} = (Y - \mathbf{E}[X|\mathscr{G}])\mathbf{E}[(X - \mathbf{E}[X|\mathscr{G}])|\mathscr{G}] = 0.$$

Then the last term on the right hand side of (10.2.1) has expectation zero. That shows the desired result. □

Corollary 10.2.3 *If $\mathbf{E}[X^2] < \infty$, then*

$$\mathbf{Var}(X) \geq \mathbf{Var}(\mathbf{E}[X|\mathscr{G}]) \quad (10.2.2)$$

with equality if and only if $X \overset{\text{a.s.}}{=} Z$ for a $\mathscr{G}$-measurable random variable Z.

Proof. Setting $Y = \mathbf{E}[X]$ in Proposition 10.2.1 we obtain

$$\mathbf{Var}(X) = \mathbf{E}[(X - \mathbf{E}[X|\mathscr{G}])^2] + \mathbf{Var}(\mathbf{E}[X|\mathscr{G}]).$$

Then we have (10.2.2) with equality if and only if $X \overset{\text{a.s.}}{=} \mathbf{E}[X|\mathscr{G}]$, which holds if and only if $X = Z$ for a $\mathscr{G}$-measurable random variable Z. □

10.3 Regular conditional probabilities

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. Recall that the conditional probability $\mathbf{P}(A|\mathscr{G})$ of an event $A \in \mathscr{F}$ given $\mathscr{G}$ is defined by

$$\mathbf{P}(A|\mathscr{G}) = \mathbf{E}[1_A|\mathscr{G}].$$

We emphasis that $\mathbf{P}(A|\mathscr{G})$ is only uniquely determined almost surely. It is easy to show that a.s.

$$0 \leq \mathbf{P}(A|\mathscr{G}) \leq 1 \quad \text{and} \quad \mathbf{P}(\Omega|\mathscr{G}) = 1.$$

By Corollary 10.2.1, for any sequence of disjoint events $\{A_n\} \subseteq \mathscr{F}$ we have a.s.

$$\mathbf{P}\Big(\bigcup_{n=1}^{\infty} A_n \Big| \mathscr{G}\Big) = \sum_{n=1}^{\infty} \mathbf{P}(A_n|\mathscr{G}). \tag{10.3.1}$$

However, these properties do not always imply that for a.e. $\omega \in \Omega$, the set function $A \mapsto \mathbf{P}(A|\mathscr{G})(\omega)$ is a probability measure on $(\Omega, \mathscr{F})$. The problem is that the exceptional null set for (10.3.1) depends on the sequence $\{A_n\}$, and there are usually uncountably many these sequences in the σ-algebra $\mathscr{F}$. We shall give some sufficient conditions that allow us to choose the random variables $\mathbf{P}(A|\mathscr{G})$ suitably so that $A \mapsto \mathbf{P}(A|\mathscr{G})$ is a.s. a probability measure on $\mathscr{F}$.

Definition 10.3.1 A function $Q : \Omega \times \mathscr{F} \to [0,1]$ is called a *regular conditional probability* given $\mathscr{G}$ if

(i) for each $\omega \in \Omega$, the set function $A \mapsto Q(\omega, A)$ is a probability measure on $(\Omega, \mathscr{F})$;

(ii) for each fixed $A \in \mathscr{F}$, the function $\omega \mapsto Q(\omega, A)$ is $\mathscr{G}$-measurable;

(iii) for each $A \in \mathscr{F}$, we have a.s.

$$\mathbf{P}(A|\mathscr{G}) = Q(\cdot, A).$$

Proposition 10.3.1 *Suppose there is a regular conditional probability $Q(\cdot,\cdot)$ on $(\Omega, \mathscr{F})$ given $\mathscr{G}$. Let X be a random variable such that $\mathbf{E}[|X|] < \infty$. Then we have a.s.*

$$\mathbf{E}[X|\mathscr{G}](\omega) = \int_{\Omega} X(\omega')Q(\omega, d\omega').$$

Proof. If $X = 1_A$ for some $A \in \mathscr{F}$, for a.e. $\omega \in \Omega$ we have

$$\mathbf{E}[1_A|\mathscr{G}](\omega) = \mathbf{P}(A|\mathscr{G})(\omega) = Q(\omega, A) = \int_{\Omega} 1_A(\omega')Q(\omega, d\omega').$$

The proof for the general case can be carried out by approximating arguments. □

Definition 10.3.2 A function $Q_X : \Omega \times \mathscr{B}(\mathbb{R}) \to [0,1]$ is called a *regular conditional distribution* of the random variable X given $\mathscr{G}$ if

(i) for each fixed $\omega \in \Omega$, the set function $B \mapsto Q_X(\omega, B)$ is a probability measure on $\mathscr{B}(\mathbb{R})$;

(ii) for each fixed $B \in \mathscr{B}(\mathbb{R})$, the function $\omega \mapsto Q_X(\omega, B)$ is $\mathscr{G}$-measurable;

(iii) for every $B \in \mathscr{B}(\mathbb{R})$, we have a.s.

$$\mathbf{P}(X^{-1}(B)|\mathscr{G}) = Q_X(\omega, B).$$

In particular, if a regular conditional probability function $Q(\cdot,\cdot)$ on $(\Omega, \mathscr{F})$ given $\mathscr{G}$ exists, we can define a regular conditional distribution of X by setting

$$Q_X(\omega, B) = Q(\omega, X^{-1}(B)), \qquad \omega \in \Omega, B \in \mathscr{B}(\mathbb{R}).$$

Definition 10.3.3 A function $F_X : \Omega \times \mathbb{R} \to [0,1]$ is called a *regular conditional distribution function* of X given $\mathscr{G}$ if

(i) for each fixed $\omega \in \Omega$, the function $x \mapsto F(\omega, x)$ is a probability distribution function on $\mathbb{R}$;

(ii) for each fixed $x \in \mathbb{R}$, the function $\omega \mapsto F(\omega, x)$ is $\mathscr{G}$-measurable;

(iii) for each $x \in \mathbb{R}$, we have a.s.

$$F_X(\cdot, x) = \mathbf{P}(\{X \le x\}|\mathscr{G}).$$

Proposition 10.3.2 *A regular conditional distribution of X given $\mathscr{G}$ exits if and only if there is a regular conditional distribution function of X given $\mathscr{G}$.*

Proof. If $Q_X(\cdot,\cdot)$ is a regular conditional distribution of X given $\mathscr{G}$, we may define a regular conditional distribution function F_X by setting $F_X(\omega, x) = Q_X(\omega, (-\infty, x])$. Conversely, suppose there is a regular conditional distribution function F_X of X given $\mathscr{G}$. Then for each $\omega \in \Omega$, the probability distribution function $F_X(\omega, \cdot)$ determines uniquely a probability measure $Q_X(\omega, \cdot)$ on $\mathscr{B}(\mathbb{R})$. Let $\mathscr{C}$ denote the class of sets $B \in \mathscr{B}(\mathbb{R})$ such that $\omega \mapsto Q_X(\omega, B)$ is $\mathscr{G}$-measurable, and $\mathbf{P}(X^{-1}(B)|\mathscr{G}) = Q_X(\omega, B)$ a.s. Then $\mathscr{C} \supseteq \mathscr{A} :=$ {finite unions of left open and right closed intervals}. Clearly, $\mathscr{C}$ is a monotone class and $\mathscr{A}$ is an algebra. By the monotone class theorem we have $\mathscr{C} \supseteq \sigma(\mathscr{A}) = \mathscr{B}(\mathbb{R})$. Consequently, Q_X is a regular conditional distribution of X given $\mathscr{G}$. □

Theorem 10.3.1 *A regular conditional distribution of X given $\mathscr{G}$ always exists.*

Proof. By Proposition 10.3.2, it suffices to prove the existence of a regular conditional distribution function F_X of X given $\mathscr{G}$. Let $\mathbb{Q} = \{r_1, r_2, \cdots\}$ be an enumeration of all rational numbers. For each $r \in \mathbb{Q}$, we fix a random variable $\eta(r)$ such that $\eta(r) \overset{\text{a.s.}}{=} \mathbf{P}(X \le r|\mathscr{G})$. For $m, n \ge 1$, let

$$A_{m,n} = \{\omega \in \Omega : \eta(r_m)(\omega) > \eta(r_n)(\omega)\}$$

and let $A = \bigcup_{r_m < r_n} A_{m,n}$. Since $r_m < r_n$ implies a.s. $\eta(r_m) \le \eta(r_n)$, we have $\mathbf{P}(A) = 0$. Next we set

$$B_n = \{\omega \in \Omega : \limsup_{k\to\infty} \eta(r_n + 1/k)(\omega) \ne \eta(r_n)(\omega)\}$$

and $B = \bigcup_{n=1}^{\infty} B_n$. From the conditional monotone convergence theorem we have a.s.

$$\lim_{k\to\infty} \eta(r_n + 1/k) = \lim_{k\to\infty} \mathbf{P}(\{X \le r_n + 1/k\}|\mathscr{G}) = \mathbf{P}(\{X \le r_n\}|\mathscr{G}) = \eta(r_n).$$

It follows that $\mathbf{P}(B_n) = 0$ for each $n \ge 1$, and so $\mathbf{P}(B) = 0$. Similarly, letting

$$E = \{\omega \in \Omega : \limsup_{n\to\infty} \eta(n)(\omega) \ne 1 \text{ or } \limsup_{n\to\infty} \eta(-n)(\omega) \ne 0\},$$

we have $\mathbf{P}(E) = 0$. Thus for each $\omega \in A^c \cap B^c \cap E^c$, the function $\eta(r)(\omega)$ of $r \in \mathbb{Q}$ is non-decreasing, right continuous, $\lim_{r\to\infty} \eta(r)(\omega) = 1$ and $\lim_{r\to-\infty} \eta(r)(\omega) = 0$. Let G be an arbitrary probability distribution function on $\mathbb{R}$ and let

$$F_X(\omega, x) = \begin{cases} \lim_{\mathbb{Q}\ni r\downarrow x} \eta(r)(\omega) & \text{for } \omega \in A^c \cap B^c \cap E^c \\ G(x) & \text{for } \omega \in A \cup B \cup E. \end{cases}$$

Then for any $x \in \mathbb{R}$ we have a.s.

$$F_X(\omega, x) = \lim_{\mathbb{Q} \ni r \downarrow x} \eta(r)(\omega) = \lim_{\mathbb{Q} \ni r \downarrow x} \mathbf{P}(\{X \le r\}|\mathscr{G}) = \mathbf{P}(\{X \le x\}|\mathscr{G})$$

by the conditional monotone convergence theorem. That is, F_X is a regular conditional distribution function of X given $\mathscr{G}$. □

Corollary 10.3.1 *Let Q_X be a regular conditional probability distribution of X given $\mathscr{G}$. If ϕ is a Borel function on $\mathbb{R}$ such that $\mathbf{E}[\phi(X)]$ exists, then a.s*

$$\mathbf{E}[\phi(X)|\mathscr{G}](\omega) = \int_{\mathscr{R}} \phi(x) Q_X(\omega, dx).$$

Proof. If $\phi = 1_B$ for some $B \in \mathscr{B}(\mathbb{R})$, this is just (iii) of Definition 10.3.2. The general case follows by an approximating argument. □

Theorem 10.3.2 *Suppose that there is a Borel set $F \subseteq \mathbb{R}$ such that $(\Omega, \mathscr{F})$ is isomorphic to $(F, \mathscr{B}(F))$. Then for any sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$, a regular conditional probability function on $(\Omega, \mathscr{F})$ given $\mathscr{G}$ exists.*

Proof. Let $X : \Omega \to F$ be the isomorphism. By Theorem 10.3.1, a regular conditional distribution Q_X of the random variable X given $\mathscr{G}$ exists. For $\omega \in \Omega$ and $A \in \mathscr{F}$, let $Q(\omega, A) = Q_X(\omega, X(A))$. Since X is an isomorphism, $Q(\omega, \cdot)$ is a probability measure on $\mathscr{F}$. Clearly, $\omega \mapsto Q(\omega, A)$ is $\mathscr{F}$-measurable. Moreover, we have a.s.

$$\mathbf{P}(A|\mathscr{G})(\omega) = \mathbf{P}(X^{-1}(X(A))|\mathscr{G})(\omega) = Q_X(\omega, X(A)) = Q(\omega, A).$$

Then $Q(\cdot, \cdot)$ is a regular conditional probability on $(\Omega, \mathscr{F})$ given $\mathscr{G}$. □

Corollary 10.3.2 *Let Ω be a Borel subset of some complete separable metric space with $\mathscr{F} = \mathscr{B}(\Omega)$. Then for any sub-$\sigma$-algebra $\mathscr{G}$ of $\mathscr{F}$, a regular conditional probability function Q on $(\Omega, \mathscr{F})$ given $\mathscr{G}$ exists.*

Proof. Under the assumption, Ω is isomorphic to a closed subset of the unit interval $[0, 1]$ furnished with the Borel σ-algebra; see e.g. Parthasarathy (1967, p.14). Then the result follows by Theorem 10.3.2. □

Chapter 11

Infinitely Divisible Distributions

11.1 Definition and properties

Recall that the characteristic function ϕ_μ of a finite measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is defined by

$$\phi_\mu(t) = \int_{\mathbb{R}} e^{itx}\mu(dx), \qquad u \in \mathbb{R}, \tag{11.1.1}$$

which determined μ uniquely. In particular, if μ is supported by $\mathbb{R}_+$, it is also uniquely determined by its *Laplace transformation* L_μ defined by

$$L_\mu(t) = \int_{\mathbb{R}_+} e^{-tx}\mu(dx), \qquad t \in \mathbb{R}_+, \tag{11.1.2}$$

Recall also that given two probability measures μ and ν on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, we can define a probability measure γ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ by

$$\int_{\mathbb{R}} f(z)\gamma(dz) = \int_{\mathbb{R}\times\mathbb{R}} f(x+y)\mu \times \nu(dx, dy), \qquad f \in C(\mathbb{R}), \tag{11.1.3}$$

which is called the convolution of μ and ν and denoted by $\mu * \nu$. Moreover, (11.1.3) holds if and only if

$$\phi_\gamma(t) = \phi_\mu(t)\phi_\nu(t), \qquad t \in \mathbb{R}. \tag{11.1.4}$$

Similarly, we can define the n-fold convolution $\mu_1 \times \mu_2 \times \cdots \times \mu_n$. A probability measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is called *infinitely divisible* if for each integer $n \geq 2$ there is a probability measure μ_n on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mu = \mu_n^{*n} := \mu_n * \mu_n * \cdots * \mu_n$.

The characteristic function ϕ_μ of a probability measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is said to be *infinitely divisible* if so is μ. Clearly, the characteristic function ϕ is infinitely divisible if and only if for each integer $n \geq 1$ there is a probability μ_n with characteristic function ϕ_n such that $\phi(u) = \phi_n(u)^n$ for all $u \in \mathbb{R}$.

Example 11.1.1 Let μ be the normal distribution $N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$. We have

$$\phi_\mu(t) = \exp\{i\mu t - \sigma^2 t^2/2\}.$$

Let μ_n be the normal distribution $N(\mu/n, \sigma^2/n)$. Then $\phi_\mu(u) = \phi_{\mu_n}(u)^n$ and so $\mu = \mu_n^{*n}$. Thus μ is infinitely divisible.

Example 11.1.2 Let μ be Poisson distribution with parameter $\lambda > 0$. Then

$$\phi_\mu(t) = \exp\{\lambda(e^{it} - 1)\}.$$

It follows that $\mu = \mu_n^{*n}$, where μ_n is the Poisson distribution with parameter λ/n. Thus μ is infinitely divisible.

Example 11.1.3 Let μ be the Gamma distribution $\Gamma(\alpha, \beta)$ for $\alpha > 0$ and $\beta > 0$. We have

$$\begin{aligned}
\phi_\mu(t) &= \int_0^\infty e^{itx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{ity/\beta} y^{\alpha-1} e^{-y} dy \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y(1-it/\beta)} dy \\
&= \frac{1}{\Gamma(\alpha)(1-it/\beta)^\alpha} \int_0^\infty z^{\alpha-1} e^{-z} dz \\
&= \Big(1 - \frac{it}{\beta}\Big)^{-\alpha}.
\end{aligned}$$

Then $\mu = \mu_n^{*n}$ with $\mu_n = \Gamma(\alpha/n, \beta)$ and so μ is infinitely divisible.

Proposition 11.1.1 *If both μ and ν are infinitely divisible, so is $\mu * \nu$.*

Proof. If $\mu = \mu_n^{*n}$ and $\nu = \nu_n^{*n}$, then $\mu * \nu = (\mu_n * \nu_n)^{*n}$. □

Corollary 11.1.1 *If ϕ is an infinitely divisible characteristic function, so is $|\phi|$.*

Proof. If ϕ is an infinitely divisible characteristic function, so is $\bar{\phi}(t) = \phi(-t)$. Then the non-negative real-valued function $\phi\bar{\phi}$ is a infinitely divisible characteristic function. Now the conclusion follows as we notice $|\phi| = [\phi\bar{\phi}]^{1/2}$. □

Proposition 11.1.2 *If ϕ is an infinitely divisible characteristic function, then $\phi(t) \neq 0$ for all $t \in \mathbb{R}$.*

Proof. Suppose $\phi = (\phi_n)^n$ for $n \geq 1$, where ϕ_n is a characteristic function. Then both $\psi := |\phi|$ and $\psi_n := |\phi_n|$ are non-negative real-valued characteristic functions. Since $\psi = (\psi_n)^n$, we must have $\psi_n = \psi^{1/n}$. Thus $0 \leq \psi \leq 1$ implies $\theta(t) := \lim_{n\to\infty} \psi_n(t) = 0$ or 1 according as $\psi(t) = 0$ or 1. Recall that $\psi(0) = 1$. Then there is a neighborhood U of the origin so that $\psi(t) > 0$ for all $t \in U$. It follows that $\theta(t) = 1$ for all $t \in U$. Now the continuity theorem implies that θ is a characteristic function. The continuity of θ thus implies that $\theta(t) = 1$ for all $t \in \mathbb{R}$. Then we must have $\psi(t) > 0$ and hence $\phi(t) \neq 0$ for all $t \in \mathbb{R}$. □

Proposition 11.1.3 *If f is a continuous and non-vanishing complex function on $\mathbb{R}$ with $f(0) = 1$, there is a unique (single-valued) continuous function λ on $\mathbb{R}$ with $\lambda(0) = 0$ and $f(t) = e^{\lambda(t)}$ for all $t \in \mathbb{R}$.*

Proof. Let $T > 0$ be fixed and let $\rho_T = \inf |f(t)|$. Since f is a continuous and non-vanishing on $[-T, T]$, there is $\delta_T \in (0, \rho_T)$ such that for any $-T \le r \le t \le T$ satisfying $|t - r| \le \delta_T$ we have $|f(t) - f(r)| \le \rho_T/2 \le 1/2$. For any integer $m \ge 1$, define the sequence $\{t_j : j = 0, \pm 1, \pm 2, \cdots, \pm m\}$ by $t_j = jT/m$. Note that

$$L(z) := \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j}(z-1)^j$$

is the unique determination (principle value) of $\log z$ on $D := \{z : |z-1| < 1\}$ vanishing at $z = 1$. For any $t \in [t_{-1}, t_1]$ we have $|f(t) - 1| = |f(t) - f(t_0)| \le 1/2$ and so $\lambda(t) := L(f(t))$ is well-defined and $\exp\{\lambda(t)\} = f(t)$. Clearly, $\lambda(t)$ is continuous in $t \in [t_{-1}, t_1]$ and $\lambda(0) = 0$. If $1 \le k \le m-1$ and $t \in [t_k, t_{k+1}]$, then

$$\left|\frac{f(t)}{f(t_k)} - 1\right| = \frac{|f(t) - f(t_k)|}{|f(t_k)|} \le \frac{\rho_T}{2\rho_T} = \frac{1}{2}.$$

Thus the definition of λ can be extended from $[t_{-k}, t_k]$ to $[t_k, t_{k+1}]$ by $\lambda(t) = \lambda(t_k) + L(f(t)/f(t_k))$. Analogously, we may extend the definition to $[t_{-k-1}, t_{-k}]$. Now the function λ is defined and continuous on $[-T, T]$, and for each $1 \le k \le m-1$ and $t \in [t_k, t_{k+1}]$,

$$\begin{aligned} e^{\lambda(t)} &= \exp\left\{L\left(\frac{f(t)}{f(t_k)}\right) + \lambda(t_k)\right\} \\ &= \exp\left\{L\left(\frac{f(t)}{f(t_k)}\right) + \sum_{j=0}^{k-1} L\left(\frac{f(t_{j+1})}{f(t_j)}\right)\right\} \\ &= f(t). \end{aligned}$$

A similar statement holds in $[t_{-k-1}, t_{-k}]$. Next, given λ on $[-T, T]$ it can be extended by the prior method to $[-T-1, T+1]$ and hence by induction to $(-\infty, \infty)$. Finally, if two such functions λ_1 and λ_2 exist, we have $e^{\lambda_1(t)} = e^{\lambda_2(t)}$ and hence $\lambda_1(t) - \lambda_1(t) = 2\pi i k(t)$ for an integer $k(t)$. Since $k(t)$ is continuous with $k(0) = 0$, it is necessary that $k(t) \equiv 0$. That is, λ is unique. □

Definition 11.1.1 The function λ defined in Proposition 11.1.3 is called the *distinguished logarithm* of f and is denoted by $\mathrm{Log}\, f$. The function $\exp\{\alpha\lambda\}$ is called the *distinguished αth power* of f and is denoted by f^α.

Note that $\mathrm{Log}\,(fg) = \mathrm{Log}\, f + \mathrm{Log}\, g$ and $\mathrm{Log}\,(f^\alpha) = \alpha \mathrm{Log}\, f$.

Corollary 11.1.2 *A characteristic function ϕ is infinitely divisible if and only if it does not vanish on $\mathbb{R}$ and $\phi^{1/n}$ is a characteristic function for each integer $n \ge 1$.*

Proposition 11.1.4 *Let ϕ_k and ϕ be characteristic functions. If each ϕ_k is infinitely divisible and $\phi_k \to \phi$, then ϕ is infinitely divisible.*

Proof. Let $\psi_k = |\phi_k|^2$ and $\psi = |\phi|^2$. Then both ψ_k and ψ are characteristic functions and each ψ_k is infinitely divisible. By Corollary 11.1.2, $\psi_k^{1/n}$ is a characteristic function for each $n \ge 1$. Since $\lim_{k\to\infty} \psi_k^{1/n} = \psi^{1/n}$ and the limit is a continuous function, it is a characteristic function. Consequently, the characteristic function ψ is infinitely divisible. By Proposition 11.1.2, ψ does not vanish, so neither does ϕ. Then $\phi^{1/n}$ is a well-defined continuous function. Since $\phi^{1/n} = \lim_{k\to\infty} \phi_k^{1/n}$, it is a characteristic function. That yields the infinite divisibility of ϕ. □

11.2 Poisson type distribution

Let $\alpha, \beta \in \mathbb{R}$ and $\lambda \geq 0$ be fixed constants. A probability measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is said to be of *Poisson type* if

$$\mu(\{\beta + n\alpha\}) = \frac{\lambda^n}{n!} e^{-\lambda}, \qquad n = 0, 1, 2, \cdots \tag{11.2.1}$$

(In the case $\alpha = 0$, we take $\mu(\{\beta\}) = 1$.) If μ is given by (11.2.1), the corresponding characteristic function is given by

$$\begin{aligned} \phi_\mu(t) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} e^{i(\beta+n\alpha)t} \\ &= e^{i\beta t - \lambda} \sum_{n=0}^{\infty} \frac{\lambda^n e^{in\alpha t}}{n!} \\ &= \exp\{i\beta t + \lambda(e^{i\alpha t} - 1)\} \end{aligned} \tag{11.2.2}$$

In view of (11.2.2), a Poisson type distribution is infinitely divisible.

Proposition 11.2.1 *Let $\beta \in \mathbb{R}_+$ and let L be a σ-finite Borel measure on $(0, \infty)$ such that*

$$\int_{(0,\infty)} \frac{\xi}{1+\xi} L(d\xi) < \infty.$$

Then there is an infinitely divisible probability measure μ on $\mathbb{R}_+$ with characteristic function $\phi = e^\psi$, where

$$\psi(t) = i\beta t + \int_{(0,\infty)} (e^{it\xi} - 1) L(d\xi).$$

Proof. The assertion is immediate if L is trivial. Then we assume

$$\gamma := \int_{(0,\infty)} \frac{\xi}{1+\xi} L(d\xi) > 0.$$

in the proof. Let $h(t, 0) = it$ and

$$h(t, \xi) = \big(e^{it\xi} - 1\big) \frac{1+\xi}{\xi}, \qquad t \in \mathbb{R}, \xi > 0.$$

Then h is bounded and uniformly continuous on $[-T, T] \times \mathbb{R}_+$ for each $T > 0$. We fix $T > 0$ and let $C = C_T > 0$ be a constant such that $|h(t, \xi)| \leq C$ when $|t| \leq T$. For each integer $n \geq 1$, choose a sequence $\{0 = \eta_{n,0} < \eta_{n,1} < \cdots < \eta_{n,k_n} = M_n\}$ so that

$$\int_{(M_n,\infty)} \frac{\xi}{1+\xi} L(d\xi) < \frac{1}{2nC}$$

and

$$\sup\{|h(t, \xi) - h(t, \eta_{n,j})| : \ \eta_{n,j-1} \leq \xi \leq \eta_{n,j}, 1 \leq j \leq k_n\} \leq \frac{1}{2n\gamma}$$

when $|t| \leq T$. Let

$$G(d\xi) = \frac{\xi}{1+\xi} L(d\xi) \quad \text{and} \quad \lambda_{n,j} = \frac{1+\eta_{n,j}}{\eta_{n,j}} G(\eta_{n,j-1}, \eta_{n,j}].$$

When $|t| \leq T$, we have

$$\begin{aligned} & \Big| \int_{(0,\infty)} (e^{it\xi} - 1) L(d\xi) - \sum_{j=1}^{k_n} \lambda_{n,j} \big(e^{it\eta_{n,j}} - 1 \big) \Big| \\ \leq & \Big| \int_{(0,\infty)} h(t,\xi) G(d\xi) - \sum_{j=1}^{k_n} h(t, \eta_{n,j}) G(\eta_{n,j-1}, \eta_{n,j}] \Big| \\ \leq & C \cdot G(M_n, \infty) + \sum_{j=1}^{k_n} \frac{1}{2n\gamma} G(\eta_{n,j-1}, \eta_{n,j}] \ < \ \frac{1}{n}. \end{aligned} \tag{11.2.3}$$

It is easy to see that

$$\phi_n(t) = \exp \left\{ i\beta t + \sum_{j=1}^{k_n} \lambda_{n,j} \big(e^{it\eta_{n,j}} - 1 \big) \right\}$$

is the characteristic function of an infinitely divisible probability measure μ_n. On the other hand, from (11.2.3) we have $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ with uniform convergence on $[-T, T]$. It follows that ϕ is the characteristic function of an infinitely divisible probability measure μ on $\mathbb{R}$. Clearly, we have $\mu_n(\mathbb{R}_+) = 1$ for each $n \geq 1$ and so $\mu(\mathbb{R}_+) = 1$. □

Proposition 11.2.2 *Let μ and ψ be given as in Proposition 11.2.1 and let*

$$L_\mu(t) = \int_{\mathbb{R}_+} e^{-t\xi} \mu(d\xi), \qquad t \geq 0. \tag{11.2.4}$$

Then we have $L_\mu(t) = \exp\{-\theta(t)\}$ with

$$\theta(t) = \beta t + \int_{(0,\infty)} (1 - e^{-t\xi}) L(d\xi). \tag{11.2.5}$$

Proof. The result follows from the calculations in the proof of Proposition 11.2.1 with it replaced by $-t$. □

Proposition 11.2.3 *Let μ and ψ be given as in Proposition 11.2.1. Then*

$$\int_{\mathbb{R}_+} \xi \mu(d\xi) < \infty \tag{11.2.6}$$

holds if and only if

$$\int_{(0,\infty)} \xi L(d\xi) < \infty. \tag{11.2.7}$$

Proof. Suppose (11.2.6) holds. From (11.2.4) we have

$$\frac{L_\mu(t) - L_\mu(t+s)}{s} = \int_{\mathbb{R}_+} \frac{1-e^{-s\xi}}{s} e^{-t\xi} \mu(d\xi).$$

By dominated convergence we see from the above equality that $L_\mu(u)$ is continuously differentiable in $u \geq 0$ and

$$-\frac{d}{dt} L_\mu(t) = \int_{\mathbb{R}_+} \xi e^{-t\xi} \mu(d\xi).$$

Then $\theta(t)$ is continuously differentiable in $t \geq 0$. From (11.2.5) it follows that

$$\frac{\theta(t+s) - \theta(t)}{s} = \beta + \int_{\mathbb{R}_+} \frac{1-e^{-s\xi}}{s} e^{-t\xi} L(d\xi).$$

By Fatou's lemma we find that

$$\frac{d}{dt}\theta(t) \geq \beta + \int_{\mathbb{R}_+} \xi e^{-t\xi} L(d\xi).$$

In particular, we have

$$\beta + \int_{\mathbb{R}_+} \xi L(d\xi) \leq \frac{d}{dt}\theta(0) < \infty.$$

The converse assertion follows by similar arguments. □

Proposition 11.2.4 *Let $\beta \in \mathbb{R}$ and let L be a σ-finite measure on $\mathbb{R} \setminus \{0\}$ such that*

$$\int_{\mathbb{R}\setminus\{0\}} \frac{|\xi|}{1+|\xi|} L(d\xi) < \infty.$$

Then there is an infinitely divisible probability measure μ on $\mathbb{R}$ with characteristic function $\phi = e^\psi$, where

$$\psi(t) = i\beta t + \int_{\mathbb{R}\setminus\{0\}} (e^{it\xi} - 1) L(d\xi).$$

Moreover, both β and L are uniquely determined by μ.

Proof. For $t \in \mathbb{R}$ let

$$\psi_1(t) = \int_{(0,\infty)} (e^{it\xi} - 1) L(d\xi)$$

and

$$\psi_2(t) = \int_{(-\infty,0)} (e^{-it\xi} - 1) L(d\xi).$$

By Proposition 11.2.1, there are infinitely divisible probability measures ν_1 and ν_2 on $\mathbb{R}_+$ with characteristic functions $\phi_1(t) = e^{\psi_1(t)}$ and $\phi_2(t) = e^{\psi_2(t)}$. Let γ_2 be the probability measure ν_2 induced from ν_2 by the mapping $x \mapsto -x$. Then γ_2 is an infinitely divisible probability measure supported by $\mathbb{R}_-$ with characteristic function $\phi_2(-t) = e^{\psi_2(-t)}$. We see easily that $\phi = e^\psi$ is the characteristic function of the infinitely divisible probability measure $\mu := \nu_1 * \gamma_2 * \delta_\beta$. To prove the uniqueness of α and L, we introduce the function

$$V(t) := 2\psi(t) - \int_{t-1}^{t+1} \psi(s) ds.$$

By elementary calculations,

$$V(t) = 2\int_{\mathbb{R}\setminus\{0\}} e^{it\xi} G(d\xi),$$

where

$$G(d\xi) = 2\Big(1 - \frac{\sin\xi}{\xi}\Big) L(d\xi).$$

That is, $V(t)$ is the characteristic function of the finite measure G. Then both G and L are uniquely determined by ψ and hence by ϕ. The uniqueness of β follows immediately. □

Corollary 11.2.1 *Let μ and ψ be given as in Proposition 11.2.4. Then we have*

(i) $\mu(-\infty, 0) = 0$ if and only if $\beta \geq 0$ and $L(-\infty, 0) = 0$;

(ii) $\mu(0, \infty) = 0$ if and only if $\beta \leq 0$ and $L(0, \infty) = 0$.

Proof. (Homework.) □

Proposition 11.2.5 *Let μ and L be related as in Proposition 11.2.4. Then*

$$\int_{\mathbb{R}} |\xi| \mu(d\xi) < \infty \tag{11.2.8}$$

holds if and only if

$$\int_{\mathbb{R}\setminus\{0\}} |\xi| L(d\xi) < \infty. \tag{11.2.9}$$

Proof. We use the notation of the proof of Proposition 11.2.4. Suppose (11.2.8) holds. By Proposition 8.3.2, we have

$$\int_{\mathbb{R}} |\xi| \nu_1(d\xi) + \int_{\mathbb{R}} |\xi| \gamma_2(d\xi) < \infty,$$

and so

$$\int_{\mathbb{R}_+} \xi \nu_1(d\xi) + \int_{\mathbb{R}_+} \xi \nu_2(d\xi) < \infty.$$

From Proposition 11.2.3 it follows that

$$\int_{(0,\infty)} \xi L(d\xi) + \int_{(-\infty,0)} (-\xi) L(d\xi) < \infty.$$

Then we have (11.2.9). The converse assertion follows similarly. □

11.3 Lévy-Khintchine representation

Let us define a complex continuous function $K(\cdot,\cdot)$ on $\mathbb{R}\times\mathbb{R}$ by $K(t,0) = -u^2/2$ and

$$K(t,\xi) = \Big(e^{it\xi} - 1 - \frac{it\xi}{1+\xi^2}\Big)\frac{1+\xi^2}{\xi^2}, \qquad t \in \mathbb{R},\ \xi \in \mathbb{R}\setminus\{0\}.$$

For any $\beta \in \mathbb{R}$ and any finite measure G on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, set

$$\psi(t, \beta, G) = i\beta t + \int_{\mathbb{R}} K(t, \xi) G(d\xi), \qquad t \in \mathbb{R}. \tag{11.3.1}$$

Then $\psi(t, \beta, G)$ is a continuous function of $t \in \mathbb{R}$ with $\psi(0, \beta, G) = 0$. It follows that $\psi = \operatorname{Log} e^{\psi}$. Clearly, ψ has representation (11.3.1) if and only if it has representation

$$\psi(t) = i\beta t - \alpha t^2 + \int_{\mathbb{R}\setminus\{0\}} \Big(e^{it\xi} - 1 - \frac{it\xi}{1+\xi^2}\Big) L(d\xi), \tag{11.3.2}$$

where $\beta \in \mathbb{R}$, $\alpha \geq 0$ and L is a σ-finite measure on $\mathbb{R} \setminus \{0\}$ such that

$$\int_{\mathbb{R}\setminus\{0\}} \frac{\xi^2}{1+\xi^2} L(d\xi) < \infty.$$

Theorem 11.3.1 *If ψ is given by (11.3.1) or (11.3.2), then $\phi = e^{\psi}$ is an infinitely divisible characteristic function. Moreover, ϕ uniquely determines the two sets of parameters (β, G) and (β, α, L).*

Proof. For each integer $n \geq 1$, let

$$\psi_n(t) = i\beta t + \int_{\{|\xi|>1/n\}} \Big(e^{it\xi} - 1 - \frac{it\xi}{1+\xi^2}\Big) L(d\xi).$$

We see by Proposition 11.2.4 that $e^{\psi_n(t)}$ is an infinitely divisible characteristic function. Then $\phi_n(t) := \exp\{\psi_n(t) - \alpha t^2\}$ is an infinitely divisible characteristic function. By dominated convergence we have $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ and the limit function is continuous in $t \in \mathbb{R}$. Then ϕ is an infinitely divisible characteristic function. The uniqueness of (ρ, G) follows by arguments similar to those in the proof of Proposition 11.2.4. The uniqueness of (ρ, α, L) is then immediate. □

Theorem 11.3.2 *Let β and $\beta_n \in \mathbb{R}$ and let G and G_n be finite measure on $\mathbb{R}$. If $\beta_n \to \beta$ and $G_n \xrightarrow{\text{w}} G$, then $\psi(t, \beta_n, G_n) \to \psi(t, \beta, G)$ for every $t \in \mathbb{R}$.*

Proof. Since for each fixed $t \in \mathbb{R}$, the function $\xi \mapsto K(t, \xi)$ is bounded and continuous, the result is immediate. □

Theorem 11.3.3 *Suppose $\{\beta_n\} \subseteq \mathbb{R}$ and $\{G_n\}$ are finite measure on $\mathbb{R}$. If $\psi(\cdot, \beta_n, G_n)$ converges to a continuous function $g(\cdot)$ on $\mathbb{R}$, then $\beta_n \to$ some $\beta \in \mathbb{R}$ and $G_n \xrightarrow{\text{w}}$ some finite measure G. Moreover, we have $g(t) = \psi(t, \beta, G)$ for all $t \in \mathbb{R}$.*

Proof. It is easy to check that

$$V_n(t) := 2\psi(t, \beta_n, G_n) - \int_{t-1}^{t+1} \psi(s, \beta_n, G_n) ds = \int_{\mathbb{R}} e^{it\xi} H_n(d\xi),$$

where

$$H_n(d\xi) = 2\Big(1 - \frac{\sin\xi}{\xi}\Big) \frac{1+\xi^2}{\xi^2} G_n(d\xi).$$

Under the assumptions, we have

$$V_n(t) \to V(t) := 2g(t) - \int_{t-1}^{t+1} g(s)ds,$$

where the limit function V is continuous on $\mathbb{R}$. By the continuity theorem, V is the characteristic function of some finite measure H on $\mathbb{R}$ and $H_n \xrightarrow{w} H$. It is easy to find constants $c_1 > c_0 > 0$ such that

$$c_0 < 2\Big(1 - \frac{\sin\xi}{\xi}\Big)\frac{1+\xi^2}{\xi^2} < c_1$$

for all $\xi \in \mathbb{R}$. Then we have $G_n \xrightarrow{w} G$, where

$$G(d\xi) = \frac{1}{2}\Big(1 - \frac{\sin\xi}{\xi}\Big)^{-1}\frac{\xi^2}{1+\xi^2}H(d\xi).$$

It follows that

$$\lim_{n\to\infty} \psi(t, 0, G_n) = \lim_{n\to\infty} \int_{\mathbb{R}} K(t,\xi)G_n(d\xi) = \int_{\mathbb{R}} K(t,\xi)G(d\xi).$$

By the assumption, the limit $\lim_{n\to\infty} \psi(1, \beta_n, G_n)$ exists, then so does the limit

$$\beta := \lim_{n\to\infty} \beta_n = \lim_{n\to\infty} i[\psi(1, 0, G_n) - \psi(1, \beta_n, G_n)].$$

Consequently,

$$\lim_{n\to\infty} \psi(t, \beta_n, G_n) = \lim_{n\to\infty} [i\beta_n t + \psi(t, 0, G_n)] = i\beta t + \int_{\mathbb{R}} K(t,\xi)G(d\xi).$$

That proves the desired result. □

Theorem 11.3.4 (Lévy-Khintchine representation) *A characteristic function ϕ is infinitely divisible if and only if $\psi = \operatorname{Log}\phi$ has representation (11.3.1).*

Proof. By Theorem 11.3.1, if ϕ is given by (11.3.1), it is an infinitely divisible characteristic function. Conversely, suppose ϕ is an infinitely divisible characteristic function. It is easily seen that

$$\phi^{1/n}(t) = \exp\Big\{\frac{1}{n}\operatorname{Log}\phi(t)\Big\} = 1 + \frac{1}{n}\operatorname{Log}\phi(t) + o\Big(\frac{1}{n}\Big).$$

Consequently,

$$\operatorname{Log}\phi(t) = \lim_{n\to\infty} n[\phi^{1/n}(t) - 1] = \lim_{n\to\infty} n \int_{\mathbb{R}} (e^{it\xi} - 1)\mu_n(d\xi),$$

where μ_n denote the probability measure corresponding to $\phi^{1/n}$. Setting

$$\beta_n = \int_{\mathbb{R}} \frac{n\xi}{1+\xi^2}\mu_n(d\xi) \quad \text{and} \quad G_n(d\xi) = \frac{n\xi^2}{1+\xi^2}\mu_n(d\xi),$$

we obtain

$$\operatorname{Log}\phi(t) = \lim_{n\to\infty} \Big[i\beta_n t + \int_{\mathbb{R}} K(t,\xi)G_n(d\xi)\Big].$$

Then Theorem 11.3.3 implies that ϕ has representation (11.3.1). □

11.4 Kolmogorov representation

Proposition 11.4.1 *Let μ be a probability measure on $\mathbb{R}$ with characteristic function ϕ. Then*

$$\int_{\mathbb{R}} \xi^{2n} \mu(d\xi) < \infty \tag{11.4.1}$$

holds for an integer $n \geq 1$ if and only if ϕ is continuously differentiable to the $(2n)$th degree. In this case, we have

$$\phi^{(k)}(t) = i^k \int_{\mathbb{R}} \xi^k e^{it\xi} \mu(d\xi), \qquad u \in \mathbb{R}, \tag{11.4.2}$$

for $0 \leq k \leq 2n$.

Proof. Suppose that (11.4.1) holds for some $n \geq 1$. Observe that

$$\frac{\phi(t+s) - \phi(t)}{s} = \int_{\mathbb{R}} \frac{e^{is\xi} - 1}{s} e^{it\xi} \mu(d\xi),$$

and the integrand on the right hand side is bounded above by $|\xi|$. By dominated convergence we see that

$$\phi'(t) = \lim_{t \to 0} \frac{\phi(t+s) - \phi(t)}{s} = i \int_{\mathbb{R}_+} \xi e^{it\xi} \mu(d\xi).$$

exists and continuous in $t \in \mathbb{R}$. Proceeding inductively we find that ϕ is continuously differentiable to the $(2n)$th degree with derivatives given by (11.4.2).

Conversely, suppose ϕ is continuously differentiable to the $(2n)$th degree for some $n \geq 1$. It is easy to show that

$$\theta(t) := 2\phi(0) - \phi(2t) - \phi(-2t) = 4 \int_{\mathbb{R}} \sin^2(t\xi) \mu(d\xi).$$

Then the left hand side is real and non-negative. Now the monotone convergence theorem implies that

$$\int_{\mathbb{R}} \xi^2 \mu(d\xi) = \lim_{t \to 0} \int_{\mathbb{R}} \frac{\sin^2(t\xi)}{t^2} \mu(d\xi) = \lim_{t \to 0} \frac{\theta(t)}{4t^2} = -\phi''(0) < \infty.$$

Then the first part of the proof shows that ϕ is twice continuously differentiable with ϕ' and ϕ'' given by (11.4.2) with $n = 1$ and 2, respectively. Proceeding inductively we obtain (11.4.1). □

Theorem 11.4.1 *A function ϕ is the characteristic function of infinitely divisible probability measure μ with finite variance if and only if $\psi = \operatorname{Log} \phi$ has the representation*

$$\psi(t) = i\gamma t - \alpha t^2 + \int_{\mathbb{R} \setminus \{0\}} \left(e^{it\xi} - 1 - it\xi\right) L(d\xi), \tag{11.4.3}$$

where $\gamma \in \mathbb{R}$, $\alpha \geq 0$ and L is a σ-finite measure on $\mathbb{R} \setminus \{0\}$ such that

$$\int_{\mathbb{R} \setminus \{0\}} \xi^2 L(d\xi) < \infty. \tag{11.4.4}$$

Proof. Suppose that ψ has representation (11.4.3) with L satisfying (11.4.4). Clearly, we have (11.3.2) with

$$\beta = \gamma + \int_{\mathbb{R}\setminus\{0\}} \Big(\frac{\xi}{1+\xi^2} - \xi\Big) L(d\xi) = \gamma - \int_{\mathbb{R}\setminus\{0\}} \frac{\xi^3}{1+\xi^2} L(d\xi). \tag{11.4.5}$$

From Theorem 11.3.4 we know that $\phi = e^{\psi}$ is the characteristic function of an infinitely divisible probability measure μ. By dominated convergence it is not hard to show that ψ ϕ is twice continuously differentiable. Then ϕ is also twice continuously differentiable. It follows that μ has finite variance. Conversely, suppose ϕ is the characteristic function of an infinitely divisible probability measure μ with finite variance. By Theorem 11.3.4 we know that $\psi = \operatorname{Log}\phi$ has representation (11.3.2). It follows that

$$\theta(t) := 2\psi(0) - \psi(2t) - \psi(-2t) = 8\alpha t^2 + 4\int_{\mathbb{R}\setminus\{0\}} \sin^2(t\xi) L(d\xi).$$

Then the value on the left hand side is real and positive. Since ϕ and hence ψ is twice continuously differentiable, by monotone convergence we have

$$\int_{\mathbb{R}} \xi^2 L(d\xi) = \lim_{t\to 0}\int_{\mathbb{R}} \frac{\sin^2(t\xi)}{t^2} L(d\xi) = \lim_{t\to 0}\frac{\theta(t)}{4t^2} - 8\alpha \le \psi''(0) - 8\alpha < \infty.$$

Now (11.4.3) follows from (11.3.2) with

$$\gamma = \beta + \int_{\mathbb{R}\setminus\{0\}} \Big(\xi - \frac{\xi}{1+\xi^2}\Big) L(d\xi) = \beta + \int_{\mathbb{R}\setminus\{0\}} \frac{\xi^3}{1+\xi^2} L(d\xi). \tag{11.4.6}$$

That proves the desired result. □

Theorem 11.4.2 *A function ϕ is the characteristic function of an infinitely divisible probability measure μ satisfying*

$$\int_{\mathbb{R}} |\xi|\mu(d\xi) < \infty \tag{11.4.7}$$

if and only if $\psi = \operatorname{Log}\phi$ has the representation (11.4.3), where $\gamma \in \mathbb{R}$, $\alpha \ge 0$ and L is a σ-finite measure on $\mathbb{R}\setminus\{0\}$ such that

$$\int_{\mathbb{R}\setminus\{0\}} |\xi| \wedge |\xi|^2 L(d\xi) < \infty. \tag{11.4.8}$$

Proof. Suppose that ϕ is the characteristic function of an infinitely divisible probability measure μ. From Theorem 11.3.4 we know that $\psi = \operatorname{Log}\phi$ has representation (11.3.2). Under condition (11.4.8), we have (11.4.3) with γ given by (11.4.6). Then it remains to prove (11.4.7) is equivalent to (11.4.8). Let μ_0 and μ_1 denote respectively the infinitely divisible probability measures corresponding to

$$\begin{aligned}\psi_0(u) \;:=\;& i\beta u + iu\int_{\{0<|\xi|\le 1\}} \frac{\xi^3}{1+\xi^2} L(d\xi) + iu\int_{\{|\xi|>1\}} \frac{\xi}{1+\xi^2} L(d\xi) \\ & - \alpha u^2 + \int_{\{0<|\xi|\le 1\}} \big(e^{iu\xi} - 1 - iu\xi\big) L(d\xi)\end{aligned}$$

and

$$\psi_1(u) = \int_{\{|\xi|>1\}} (e^{iu\xi} - 1)L(d\xi).$$

It is easily seen that $\psi = \psi_0 + \psi_1$ and so $\mu = \mu_0 * \mu_1$. By Theorem 11.4.1 we have

$$\int_{\mathbb{R}} |\xi|^2 \mu_0(d\xi) < \infty.$$

Then Proposition 11.1.3 implies that (11.4.7) is equivalent to

$$\int_{\mathbb{R}} |\xi| \mu_1(d\xi) < \infty.$$

By Proposition 11.2.5 the above holds if and only if

$$\int_{\{|\xi|>1\}} |\xi| L(d\xi) < \infty.$$

which holds if and only if (11.4.8) is true. □

11.5 Infinitesimal random variables

In applications, we often need to consider classes of random variables such as

$$\{X_{n,j} : 1 \le j \le k_n; n \ge 1\}, \tag{11.5.1}$$

where $k_n \to \infty$ as $n \to \infty$. We say the random variables are *rowwise independent* if the random variables $X_{n,1}, X_{n,2}, \cdots, X_{n,k_n}$ are independent for each $n \ge 1$. The random variables in (11.5.1) are said to be *infinitesimal* if

$$\lim_{n\to\infty} \max_{1\le j\le k_n} \mathbf{P}\{|X_{n,j}| \ge \varepsilon\} = 0 \tag{11.5.2}$$

for every $\varepsilon > 0$.

Theorem 11.5.1 *If the random variables in (11.5.1) are infinitesimal and rowwise independent and*

$$\sum_{j=1}^{k_n} X_{n,j} - A_n \tag{11.5.3}$$

converges in distribution for some sequence $\{A_n\} \subseteq \mathbb{R}$, then the limiting distribution of (11.5.3) is infinitely divisible. Conversely, for each infinitely divisible distribution μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, there is a family of infinitesimal and rowwise independent random variables (11.5.1) and a sequence $\{A_n\} \subseteq \mathbb{R}$ such that the distribution of (11.5.3) converges to μ.

Proof. Chow and Teicher (1988, pp.434-440). □

Theorem 11.5.2 *Suppose that (11.5.1) are infinitesimal and rowwise independent random variables and that (11.5.3) converges in distribution for some sequence $\{A_n\} \subseteq \mathbb{R}$. Then the limiting distribution is Gaussian if and only if*

$$\max_{1 \le j \le k_n} |X_{n,j}| \xrightarrow{\mathrm{P}} 0$$

or equivalently

$$\sum_{j=1}^{k_n} \mathbf{P}\{|X_{n,j}| \ge \varepsilon\} \to 0$$

for every $\varepsilon > 0$.

Proof. Chow and Teicher (1988, pp.444-446). □

Bibliography

[1] Chow, Y.S. and Teicher, H.: *Probability Theory: Independence, Interchangeability, Martingales* (2nd edition). Springer-Verlag, Berlin (1988).

[2] Cohn, D.L.: *Measure Theory.* Birkhäuser Boston, Inc., Boston, MA (1980).

[3] Halmos, P.R.: *Measure Theory.* Springer-Verlag, Berlin (1974).

[4] Hewitt, E. and Stromberg, K. (1965): *Real and Abstract Analysis.* Springer-Verlag, Berlin.

[5] Laha, R.G. and Rohatgi, V.K.: *Probability Theory.* John Wiley and Sons, New York (1979).

[6] Parthasarathy, K.R.: *Probability Measures on Metric Spaces.* Academic Press, New York (1967).

[7] Titchmarsh, E.C.: *Introduction to the Theory of Fourier Integrals.* Clarendon Press, Oxford (1937).