Chapter 15
Characteristic Functions and the Central Limit Theorem

The main goal of this chapter is the central limit theorem (CLT) for sums of independent random variables (Theorem 15.37) and for independent arrays of random variables (Lindeberg–Feller theorem, Theorem 15.43). For the latter, we prove only that one of the two implications (Lindeberg's theorem) that is of interest in the applications.

The ideal tools for the treatment of central limit theorems are so-called characteristic functions; that is, Fourier transforms of probability measures. We start with a more general treatment of classes of test functions that are suitable to characterize weak convergence and then study Fourier transforms in greater detail. The subsequent section proves the CLT for real-valued random variables by means of characteristic functions. In the fifth section, we prove a multidimensional version of the CLT.

15.1 Separating Classes of Functions

Let (E, d) be a metric space with Borel σ-algebra $\mathcal{E} = \mathcal{B}(E)$.

Denote by $\mathbb{C} = \{u + iv : u, v \in \mathbb{R}\}$ the field of complex numbers. Let

$$\operatorname{Re}(u + iv) = u \quad \text{and} \quad \operatorname{Im}(u + iv) = v$$

denote the real part and the imaginary part, respectively, of $z = u + iv \in \mathbb{C}$. Let $\overline{z} = u - iv$ be the complex conjugate of z and $|z| = \sqrt{u^2 + v^2}$ its modulus. A prominent role will be played by the complex exponential function $\exp : \mathbb{C} \to \mathbb{C}$, which can be defined either by Euler's formula $\exp(z) = \exp(u)(\cos(v) + i \sin(v))$ or by the power series $\exp(z) = \sum_{n=0}^{\infty} z^n/n!$. It is well-known that $\exp(z_1 + z_2) = \exp(z_1) \cdot \exp(z_2)$. Note that $\operatorname{Re}(z) = (z + \overline{z})/2$ and $\operatorname{Im}(z) = (z - \overline{z})/2i$ imply

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2} \quad \text{and} \quad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \quad \text{for all } x \in \mathbb{R}.$$

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_15,

A map $f : E \to \mathbb{C}$ is measurable if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are measurable (see Theorem 1.90 with $\mathbb{C} \cong \mathbb{R}^2$). In particular, any continuous function $E \to \mathbb{C}$ is measurable. If $\mu \in \mathcal{M}(E)$, then we define

$$\int f \, d\mu := \int \mathrm{Re}(f) \, d\mu + i \int \mathrm{Im}(f) \, d\mu$$

if both integrals exist and are finite. Let $C_b(E; \mathbb{C})$ denote the Banach space of continuous, bounded, complex-valued functions on E equipped with the supremum norm $\|f\|_\infty = \sup\{|f(x)| : x \in E\}$. We call $\mathcal{C} \subset C_b(E; \mathbb{C})$ a separating class for $\mathcal{M}_f(E)$ if for any two measures $\mu, \nu \in \mathcal{M}_f(E)$ with $\mu \neq \nu$, there is an $f \in \mathcal{C}$ such that $\int f \, d\mu \neq \int f \, d\nu$. The analogue of Theorem 13.34 holds for $\mathcal{C} \subset C_b(E; \mathbb{C})$.

Definition 15.1 Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. A subset $\mathcal{C} \subset C_b(E; \mathbb{K})$ is called an *algebra* if

(i) $1 \in \mathcal{C}$,
(ii) if $f, g \in \mathcal{C}$, then $f \cdot g$ and $f + g$ are in $\mathcal{C}$, and
(iii) if $f \in \mathcal{C}$ and $\alpha \in \mathbb{K}$, then (αf) is in $\mathcal{C}$.

We say that $\mathcal{C}$ *separates points* if for any two points $x, y \in E$ with $x \neq y$, there is an $f \in \mathcal{C}$ with $f(x) \neq f(y)$.

Theorem 15.2 (Stone–Weierstraß) *Let E be a compact Hausdorff space. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Let $\mathcal{C} \subset C_b(E; \mathbb{K})$ be an algebra that separates points. If $\mathbb{K} = \mathbb{C}$, then in addition assume that $\mathcal{C}$ is closed under complex conjugation (that is, if $f \in \mathcal{C}$, then the complex conjugate function $\overline{f}$ is also in $\mathcal{C}$).*
Then $\mathcal{C}$ is dense in $C_b(E; \mathbb{K})$ with respect to the supremum norm.

Proof We follow the exposition in Dieudonné [34, Chapter VII.3]. First consider the case $\mathbb{K} = \mathbb{R}$. We proceed in several steps.

Step 1. By Weierstraß's approximation theorem (Example 5.15), there is a sequence $(p_n)_{n \in \mathbb{N}}$ of polynomials that approach the map $[0, 1] \to [0, 1]$, $t \mapsto \sqrt{t}$ uniformly. If $f \in \mathcal{C}$, then also

$$|f| = \|f\|_\infty \lim_{n \to \infty} p_n\big(f^2 / \|f\|_\infty^2\big)$$

is in the closure $\overline{\mathcal{C}}$ of $\mathcal{C}$ in $C_b(E; \mathbb{R})$.

Step 2. Applying Step 1 to the algebra $\overline{\mathcal{C}}$ yields that, for all $f, g \in \overline{\mathcal{C}}$,

$$f \vee g = \frac{1}{2}\big(f + g + |f - g|\big) \quad \text{and} \quad f \wedge g = \frac{1}{2}\big(f + g - |f - g|\big)$$

are also in $\overline{\mathcal{C}}$.

Step 3. For any $f \in C_b(E; \mathbb{R})$, any $x \in E$ and any $\varepsilon > 0$, there exists a $g_x \in \overline{\mathcal{C}}$ with $g_x(x) = f(x)$ and $g_x(y) \leq f(y) + \varepsilon$ for all $y \in E$. As $\mathcal{C}$ separates points, for any $z \in E \setminus \{x\}$, there exists an $H_z \in \mathcal{C}$ with $H_z(z) \neq H(x) = 0$. For such z, define $h_z \in \mathcal{C}$ by

$$h_z(y) = f(x) + \frac{f(z) - f(x)}{H_z(z)} H_z(y) \quad \text{for all } y \in E.$$

In addition, define $h_x := f$. Then $h_z(x) = f(x)$ and $h_z(z) = f(z)$ for all $z \in E$. Since f and h_z are continuous, for any $z \in E$, there exists an open neighborhood $U_z \ni z$ with $h(y) \le f(y) + \varepsilon$ for all $y \in U_z$. We construct a finite covering $U_{z_1}, \dots, U_{z_n}$ of E consisting of such neighborhoods and define $g_x = \min(h_{z_1}, \dots, h_{z_n})$. By Step 2, we have $g_x \in \overline{\mathcal{C}}$.

Step 4. Let $f \in C_b(E; \mathbb{R})$, $\varepsilon > 0$ and, for any $x \in E$, let g_x be as in Step 3. As f and g_x are continuous, for any $x \in E$, there exists an open neighborhood $V_x \ni x$ with $g_x(y) \ge f(y) - \varepsilon$ for any $y \in V_x$. We construct a finite covering $V_{x_1}, \dots, V_{x_n}$ of E and define $g := \max(g_{x_1}, \dots, g_{x_n})$. Then $g \in \overline{\mathcal{C}}$ by Step 2 and $\|g - f\|_\infty < \varepsilon$ by construction. Letting $\varepsilon \downarrow 0$, we get $\overline{\mathcal{C}} = C_b(E; \mathbb{R})$.

Step 5. Now consider $\mathbb{K} = \mathbb{C}$. If $f \in \mathcal{C}$, then by assumption $\mathrm{Re}(f) = (f + \bar{f})/2$ and $\mathrm{Im}(f) = (f - \bar{f})/2i$ are in $\mathcal{C}$. In particular, $\mathcal{C}_0 := \{\mathrm{Re}(f) : f \in \mathcal{C}\} \subset \mathcal{C}$ is a real algebra that, by assumption, separates points and contains the constant functions. Hence $\mathcal{C}_0$ is dense in $C_b(E; \mathbb{R})$. Since $\mathcal{C} = \mathcal{C}_0 + i\mathcal{C}_0$, $\mathcal{C}$ is dense in $C_b(E; \mathbb{C})$. □

Corollary 15.3 *Let E be a compact metric space. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Let $\mathcal{C} \subset C_b(E; \mathbb{K})$ be a family that separates points; that is, stable under multiplication and that contains 1. If $\mathbb{K} = \mathbb{C}$, then in addition assume that $\mathcal{C}$ is closed under complex conjugation.*

Then $\mathcal{C}$ is a separating family for $\mathcal{M}_f(E)$.

Proof Let $\mu_1, \mu_2 \in \mathcal{M}_f(E)$ with $\int g\, d\mu_1 = \int g\, d\mu_2$ for all $g \in \mathcal{C}$. Let $\mathcal{C}'$ be the algebra of finite linear combinations of elements of $\mathcal{C}$. By linearity of the integral, $\int g\, d\mu_1 = \int g\, d\mu_2$ for all $g \in \mathcal{C}'$.

For any $f \in C_b(E, \mathbb{R})$ and any $\varepsilon > 0$, by the Stone–Weierstraß theorem, there exists a $g \in \mathcal{C}'$ with $\|f - g\|_\infty < \varepsilon$. By the triangle inequality,

$$\begin{aligned}\left|\int f\, d\mu_1 - \int f\, d\mu_2\right| &\le \left|\int f\, d\mu_1 - \int g\, d\mu_1\right| + \left|\int g\, d\mu_1 - \int g\, d\mu_2\right| \\ &\quad + \left|\int g\, d\mu_2 - \int f\, d\mu_2\right| \\ &\le \varepsilon\big(\mu_1(E) + \mu_2(E)\big).\end{aligned}$$

Letting $\varepsilon \downarrow 0$, we get equality of the integrals and hence $\mu_1 = \mu_2$ (by Theorem 13.11). □

The following theorems are simple consequences of Corollary 15.3.

Theorem 15.4 *The distribution of a bounded real random variable X is characterized by its moments.*

Proof Without loss of generality, we can assume that X takes values in $E := [0, 1]$. For $n \in \mathbb{N}$, define the map $f_n : [0, 1] \to [0, 1]$ by $f_n : x \mapsto x^n$. Further, let $f_0 \equiv 1$. The family $\mathcal{C} = \{f_n, n \in \mathbb{N}_0\}$ separates points and is closed under multiplication; hence it is a separating class for $\mathcal{M}_f(E)$. Thus $\mathbf{P}_X$ is uniquely determined by its moments $\mathbf{E}[X^n] = \int x^n \mathbf{P}_X(dx)$, $n \in \mathbb{N}$. □

Example 15.5 (due to [73]) In the preceding theorem, we cannot simply drop the assumption that X is bounded without making other assumptions (see Corollary 15.32). Even if all moments exist, the distribution of X is, in general, not uniquely determined by its moments. As an example consider $X := \exp(Y)$, where $Y \sim \mathcal{N}_{0,1}$. The distribution of X is called the *log-normal distribution*. For every $n \in \mathbb{N}$, nY is distributed as the sum of n^2 independent, standard normally distributed random variables $nY \stackrel{\mathcal{D}}{=} Y_1 + \ldots + Y_{n^2}$. Hence, for $n \in \mathbb{N}$,

$$\begin{aligned}\mathbf{E}\big[X^n\big] = \mathbf{E}\big[e^{nY}\big] = \mathbf{E}\big[e^{Y_1+\ldots+Y_{n^2}}\big] = \prod_{i=1}^{n^2} \mathbf{E}\big[e^{Y_i}\big] = \mathbf{E}\big[e^{Y}\big]^{n^2} \\ = \left(\int_{-\infty}^{\infty} (2\pi)^{-1/2} e^y e^{-y^2/2}\, dy\right)^{n^2} = e^{n^2/2}.\end{aligned} \tag{15.1}$$

We construct a whole family of distributions with the same moments as X. By the transformation formula for densities (Theorem 1.101), the distribution of X has the density

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-1} \exp\left(-\frac{1}{2}\log(x)^2\right) \quad \text{for } x > 0.$$

For $\alpha \in [-1, 1]$, define probability densities f_α on $(0, \infty)$ by

$$f_\alpha(x) = f(x)\big(1 + \alpha \sin\big(2\pi \log(x)\big)\big).$$

In order to show that f_α is a density and has the same moments as f, it is enough to show that, for all $n \in \mathbb{N}_0$,

$$m(n) := \int_0^\infty x^n f(x) \sin\big(2\pi \log(x)\big)\, dx = 0.$$

With the substitution $y = \log(x) - n$, we get (note that $\sin(2\pi(y + n)) = \sin(2\pi y)$)

$$\begin{aligned} m(n) &= \int_{-\infty}^{\infty} e^{yn+n^2} (2\pi)^{-1/2} e^{-(y+n)^2/2} \sin\big(2\pi(y + n)\big)\, dy \\ &= (2\pi)^{-1/2} e^{n^2/2} \int_{-\infty}^{\infty} e^{-y^2/2} \sin(2\pi y)\, dy = 0,\end{aligned}$$

where the last equality holds since the integrand is an odd function. ◇

Theorem 15.6 (Laplace transform) *A finite measure* μ *on* $[0,\infty)$ *is characterized by its Laplace transform*

$$\mathcal{L}_\mu(\lambda) := \int e^{-\lambda x}\,\mu(dx) \quad \textit{for } \lambda \geq 0.$$

Proof We face the problem that the space $[0,\infty)$ is not compact by passing to the one-point compactification $E = [0,\infty]$. For $\lambda \geq 0$, define the continuous function $f_\lambda : [0,\infty] \to [0,1]$ by $f_\lambda(x) = e^{-\lambda x}$ if $x < \infty$ and $f_\lambda(\infty) = \lim_{x\to\infty} e^{-\lambda x}$. Then $\mathcal{C} = \{f_\lambda, \lambda \geq 0\}$ separates points, $f_0 = 1 \in \mathcal{C}$ and $f_\mu \cdot f_\lambda = f_{\mu+\lambda} \in \mathcal{C}$. By Corollary 15.3, $\mathcal{C}$ is a separating class for $\mathcal{M}_f([0,\infty])$ and thus also for $\mathcal{M}_f([0,\infty))$. □

Definition 15.7 For $\mu \in \mathcal{M}_f(\mathbb{R}^d)$, define the map $\varphi_\mu : \mathbb{R}^d \to \mathbb{C}$ by

$$\varphi_\mu(t) := \int e^{i\langle t,x\rangle}\,\mu(dx).$$

φ_μ is called the *characteristic function* of μ.

Theorem 15.8 (Characteristic function) *A finite measure* $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ *is characterized by its characteristic function.*

Proof Let $\mu_1, \mu_2 \in \mathcal{M}_f(\mathbb{R}^d)$ with $\varphi_{\mu_1}(t) = \varphi_{\mu_2}(t)$ for all $t \in \mathbb{R}^d$. By Theorem 13.11(ii), $C_c(\mathbb{R}^d)$ is a separating class for $\mathcal{M}_f(\mathbb{R}^d)$. Hence, it is enough to show that $\int f\,d\mu_1 = \int f\,d\mu_2$ for all $f \in C_c(\mathbb{R}^d)$.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be continuous with compact support and let $\varepsilon > 0$. Assume that $K > 0$ is large enough such that $f(x) = 0$ for $x \notin (-K/2, K/2)^d$ and such that $\mu_i(\mathbb{R}^d \setminus (-K,K)^d) < \varepsilon$, $i = 1,2$. Consider the *torus* $E := \mathbb{R}^d/(2K\mathbb{Z}^d)$ and define $\tilde{f} : E \to \mathbb{R}$ by

$$\tilde{f}\big(x + 2K\mathbb{Z}^d\big) = f(x) \quad \text{for } x \in [-K,K)^d.$$

Since the support of f is contained in $(-K,K)^d$, $\tilde{f}$ is continuous.

For $m \in \mathbb{Z}^d$ define

$$g_m : \mathbb{R}^d \to \mathbb{C}, \quad x \mapsto \exp\big(i\langle \pi m/K, x\rangle\big).$$

Let $\mathcal{C}$ be the algebra of finite linear combinations of the g_m. For $g \in \mathcal{C}$, we have $g(x) = g(x + 2Kn)$ for all $x \in \mathbb{R}^d$ and $n \in \mathbb{Z}^d$. Hence, the map

$$\tilde{g} : E \to \mathbb{C}, \quad \tilde{g}\big(x + 2K\mathbb{Z}^d\big) = g(x)$$

is well-defined, continuous and bounded. Furthermore, $\tilde{\mathcal{C}} := \{\tilde{g} : g \in \mathcal{C}\} \subset C_b(E;\mathbb{C})$ is an algebra that separates points and is closed under complex conjugation. As E is compact, by the Stone–Weierstraß theorem, there is a $g \in \mathcal{C}$ such that $\|\tilde{g} - \tilde{f}\|_\infty < \varepsilon$. We infer

$$\left\|(f - g)\mathbb{1}_{[-K,K]^d}\right\|_\infty < \varepsilon$$

and

$$\left\|(f - g)\mathbb{1}_{\mathbb{R}^d \setminus [-K,K]^d}\right\|_\infty \le \|g\|_\infty = \|\tilde{g}\|_\infty \le \|\tilde{f}\|_\infty + \varepsilon = \|f\|_\infty + \varepsilon.$$

By assumption of the theorem, $\int g \, d\mu_1 = \int g \, d\mu_2$. Hence, using the triangle inequality, we conclude

$$\begin{aligned}\left|\int f \, d\mu_1 - \int f \, d\mu_2\right| &\le \int |f - g| \, d\mu_1 + \int |f - g| \, d\mu_2 \\ &\le \varepsilon\big(2\|f\|_\infty + 2\varepsilon + \mu_1\big(\mathbb{R}^d\big) + \mu_2\big(\mathbb{R}^d\big)\big).\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, the integrals coincide. □

Corollary 15.9 *A finite measure μ on $\mathbb{Z}^d$ is uniquely determined by the values*

$$\varphi_\mu(t) = \int e^{i\langle t,x\rangle} \mu(dx), \quad t \in [-\pi, \pi)^d.$$

Proof This is obvious since $\varphi_\mu(t + 2\pi k) = \varphi_\mu(t)$ for all $k \in \mathbb{Z}^d$. □

While the preceding corollary only yields an abstract uniqueness statement, we will profit also from an explicit inversion formula for Fourier transforms.

Theorem 15.10 (Discrete Fourier inversion formula) *Let $\mu \in \mathcal{M}_f(\mathbb{Z}^d)$ with characteristic function φ_μ. Then, for every $x \in \mathbb{Z}^d$,*

$$\mu\big(\{x\}\big) = (2\pi)^{-d} \int_{[-\pi,\pi)^d} e^{-i\langle t,x\rangle} \varphi_\mu(t) \, dt.$$

Proof By the dominated convergence theorem,

$$\begin{aligned}\int_{[-\pi,\pi)^d} e^{-i\langle t,x\rangle} \varphi_\mu(t) \, dt &= \int_{[-\pi,\pi)^d} e^{-i\langle t,x\rangle} \left(\lim_{n\to\infty} \sum_{|y|\le n} e^{i\langle t,y\rangle} \mu\big(\{y\}\big)\right) dt \\ &= \lim_{n\to\infty} \int_{[-\pi,\pi)^d} e^{-i\langle t,x\rangle} \sum_{|y|\le n} e^{i\langle t,y\rangle} \mu\big(\{y\}\big) \, dt \\ &= \sum_{y\in\mathbb{Z}^d} \mu\big(\{y\}\big) \int_{[-\pi,\pi)^d} e^{i\langle t,y-x\rangle} \, dt.\end{aligned}$$

The claim follows since, for $y \in \mathbb{Z}^d$,

$$\int_{[-\pi,\pi)^d} e^{i\langle t, y-x\rangle}\,dt = \begin{cases} (2\pi)^d, & \text{if } x = y, \\ 0, & \text{else.} \end{cases} \qquad \Box$$

Similar inversion formulas hold for measures μ on $\mathbb{R}^d$. Particularly simple is the case where μ possesses an integrable density $f := \frac{d\mu}{d\lambda}$ with respect to d-dimensional Lebesgue measure λ. In this case, we have the Fourier inversion formula,

$$f(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} \varphi_\mu(t)\lambda(dt). \tag{15.2}$$

Furthermore, by Plancherel's theorem, $f \in \mathcal{L}^2(\lambda)$ if and only if $\varphi_\mu \in \mathcal{L}^2(\lambda)$. In this case, $\|f\|_2 = \|\varphi\|_2$.

Since we will not need these statements in the following, we only refer to the standard literature (e.g., [173, Chapter VI.2] or [54, Theorem XV.3.3 and Eq. (XV.3.8)]).

Exercise 15.1.1 Show that, in the Stone–Weierstraß theorem, compactness of E is essential. *Hint:* Let $E = \mathbb{R}$ and use the fact that $C_b(\mathbb{R}) = C_b(\mathbb{R};\mathbb{R})$ is not separable. Construct a countable algebra $\mathcal{C} \subset C_b(\mathbb{R})$ that separates points.

Exercise 15.1.2 Let $d \in \mathbb{N}$ and let μ be a finite measure on $[0,\infty)^d$. Show that μ is characterized by its Laplace transform $\mathcal{L}_\mu(\lambda) = \int e^{-\langle\lambda,x\rangle}\mu(dx)$, $\lambda \in [0,\infty)^d$.

Exercise 15.1.3 Show that, under the assumptions of Theorem 15.10, *Plancherel's equation* holds:

$$\sum_{x\in\mathbb{Z}^d} \mu\big(\{x\}\big)^2 = (2\pi)^{-d} \int_{[-\pi,\pi)^d} \big|\varphi_\mu(t)\big|^2\,dt.$$

Exercise 15.1.4 (Mellin transform) Let X be a nonnegative real random variable. For $s \ge 0$, define the Mellin transform of $\mathbf{P}_X$ by

$$m_X(s) = \mathbf{E}\big[X^s\big]$$

(with values in $[0,\infty]$).

Assume there is an $\varepsilon_0 > 0$ with $m_X(\varepsilon_0) < \infty$ (respectively $m_X(-\varepsilon_0) < \infty$). Show that, for any $\varepsilon > 0$, the distribution $\mathbf{P}_X$ is characterized by the values $m_X(s)$ (respectively $m_X(-s)$), $s \in [0,\varepsilon]$.

Hint: For continuous $f : [0,\infty) \to [0,\infty)$, let

$$\phi_f(z) = \int_0^\infty t^{z-1} f(t)\,dt$$

for those $z \in \mathbb{C}$ for which the integral is well-defined. By a standard result of complex analysis if $\phi_f(s) < \infty$ for an $s > 1$, then ϕ_f is holomorphic in

$\{z \in \mathbb{C} : \mathrm{Re}(z) \in (1, s)\}$ (and is thus uniquely determined by the values $\phi_f(r)$, $r \in (1, 1+\varepsilon)$ for any $\varepsilon > 0$). Furthermore, for all $r \in (1, s)$,

$$f(t) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} t^{-(r+i\rho)} \phi_f(r + i\rho)\, d\rho.$$

(i) Conclude the statement for X with a continuous density.
(ii) For $\delta > 0$, let $Y_\delta \sim \mathcal{U}_{[1-\delta,1]}$ be independent of X. Show that XY_δ has a continuous density.
(iii) Compute m_{XY_δ}, and show that $m_{XY_\delta} \to m_X$ for $\delta \downarrow 0$.
(iv) Show that $XY_\delta \Longrightarrow X$ for $\delta \downarrow 0$.

Exercise 15.1.5 Let X, Y, Z be independent nonnegative random variables such that $\mathbf{P}[Z > 0] > 0$ and such that the Mellin transform $m_{XZ}(s)$ is finite for some $s > 0$.

Show that if $XZ \stackrel{\mathcal{D}}{=} YZ$ holds, then $X \stackrel{\mathcal{D}}{=} Y$.

Exercise 15.1.6 Let μ be a probability measure on $\mathbb{R}$ with integrable characteristic function φ_μ and hence $\varphi_\mu \in \mathcal{L}^1(\lambda)$, where λ is the Lebesgue measure on $\mathbb{R}$. Show that μ is absolutely continuous with bounded continuous density $f = \frac{d\mu}{d\lambda}$ given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_\mu(t)\, dt \quad \text{for all } x \in \mathbb{R}.$$

Hint: Show this first for the normal distribution $\mathcal{N}_{0,\varepsilon}$, $\varepsilon > 0$. Then show that $\mu * \mathcal{N}_{0,\varepsilon}$ is absolutely continuous with density f_ε, which converges pointwise to f (as $\varepsilon \to 0$).

Exercise 15.1.7 Let (Ω, τ) be a separable topological space that satisfies the $T_{3\frac{1}{2}}$ separation axiom: For any closed set $A \subset \Omega$ and any point $x \in \Omega \setminus A$, there exists a continuous function $f : \Omega \to [0, 1]$ with $f(x) = 0$ and $f(y) = 1$ for all $y \in A$. (Note in particular that every metric space is a $T_{3\frac{1}{2}}$-space.)

Show that $\sigma(C_b(\Omega)) = \mathcal{B}(\Omega)$; that is, the Borel σ-algebra is generated by the bounded continuous functions $\Omega \to \mathbb{R}$.

15.2 Characteristic Functions: Examples

Recall that $\mathrm{Re}(z)$ is the real part of $z \in \mathbb{C}$. We collect some simple properties of characteristic functions.

Lemma 15.11 *Let X be a random variable with values in $\mathbb{R}^d$ and characteristic function $\varphi_X(t) = \mathbf{E}[e^{i\langle t,X\rangle}]$. Then:*

(i) $|\varphi_X(t)| \le 1$ *for all* $t \in \mathbb{R}^d$ *and* $\varphi_X(0) = 1$.
(ii) $\varphi_{aX+b}(t) = \varphi_X(at)e^{i\langle b,t\rangle}$ *for all* $a \in \mathbb{R}$ *and* $b \in \mathbb{R}^d$.
(iii) $\mathbf{P}_X = \mathbf{P}_{-X}$ *if and only if* φ *is real-valued.*
(iv) *If* X *and* Y *are independent, then* $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$.
(v) $0 \le 1 - \mathrm{Re}(\varphi_X(2t)) \le 4(1 - \mathrm{Re}(\varphi_X(t)))$ *for all* $t \in \mathbb{R}^d$.

Proof (i) and (ii) are trivial.

(iii) $\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$.

(iv) As $e^{i\langle t,X\rangle}$ and $e^{i\langle t,Y\rangle}$ are independent random variables, we have

$$\varphi_{X+Y}(t) = \mathbf{E}\big[e^{i\langle t,X\rangle} \cdot e^{i\langle t,Y\rangle}\big] = \mathbf{E}\big[e^{i\langle t,X\rangle}\big]\mathbf{E}\big[e^{i\langle t,Y\rangle}\big] = \varphi_X(t)\varphi_Y(t).$$

(v) By the addition theorem for trigonometric functions,

$$1 - \cos\big(\langle 2t, X\rangle\big) = 2\big(1 - \big(\cos\big(\langle t, X\rangle\big)\big)^2\big) \le 4\big(1 - \cos\big(\langle t, X\rangle\big)\big).$$

Now take the expectations of both sides. □

In the next theorem, we collect the characteristic functions for some of the most important distributions.

Theorem 15.12 (Characteristic functions of some distributions) *For some distributions P with density $x \mapsto f(x)$ on $\mathbb{R}$ or weights $P(\{k\})$, $k \in \mathbb{N}_0$, the characteristic function $\varphi(t)$ is given explicitly in Table 15.1.*

Proof (i) *(Normal distribution)* By Lemma 15.11, it is enough to consider the case $\mu = 0$ and $\sigma^2 = 1$. By virtue of the differentiation lemma (Theorem 6.28) and using partial integration, we get

$$\frac{d}{dt}\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, i x e^{-x^2/2}\, dx = -t\varphi(t).$$

This linear differential equation with initial value $\varphi(0) = 1$ has the unique solution $\varphi(t) = e^{-t^2/2}$.

(ii) *(Uniform distribution)* This is immediate.

(iii) *(Triangle distribution)* Note that $\mathrm{Tri}_a = \mathcal{U}_{[-a/2,a/2]} * \mathcal{U}_{[-a/2,a/2]}$; hence

$$\varphi_{\mathrm{Tri}_a}(t) = \varphi_{\mathcal{U}_{[-a/2,a/2]}}(t)^2 = 4\frac{\sin(at/2)^2}{a^2t^2} = 2\frac{1-\cos(at)}{a^2t^2}.$$

Here we used the fact that by the addition theorem for trigonometric functions

$$1 - \cos(x) = \sin(x/2)^2 + \cos(x/2)^2 - \cos(x) = 2\sin(x/2)^2.$$

Table 15.1 Characteristic functions of some distributions

Distribution				Char. fct.
Name Symbol	Parameter	on	Density / Weights	$\varphi(t)$
Normal $\mathcal{N}_{\mu,\sigma^2}$	$\mu \in \mathbb{R}$, $\sigma^2 > 0$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$e^{i\mu t} \cdot e^{-\sigma^2 t^2/2}$
Uniform $\mathcal{U}_{[0,a]}$	$a > 0$	$[0, a]$	$1/a$	$\frac{e^{iat}-1}{iat}$
Uniform $\mathcal{U}_{[-a,a]}$	$a > 0$	$[-a, a]$	$1/2a$	$\frac{\sin(at)}{at}$
Triangle Tri_a	$a > 0$	$[-a, a]$	$\frac{1}{a}(1 - \lvert x\rvert/a)^+$	$2\frac{1-\cos(at)}{a^2t^2}$
N.N.	$a > 0$	$\mathbb{R}$	$\frac{1}{\pi}\frac{1-\cos(ax)}{ax^2}$	$(1 - \lvert t\rvert/a)^+$
Gamma $\Gamma_{\theta,r}$	$\theta > 0$, $r > 0$	$[0, \infty)$	$\frac{\theta^r}{\Gamma(r)} x^{r-1} e^{-\theta x}$	$(1 - it/\theta)^{-r}$
Exponential $\exp_\theta$	$\theta > 0$	$[0, \infty)$	$\theta e^{-\theta x}$	$\frac{\theta}{\theta - it}$
Two-sided exponential $\exp^2_\theta$	$\theta > 0$	$\mathbb{R}$	$\frac{\theta}{2} e^{-\theta\lvert x\rvert}$	$\frac{1}{1+(t/a)^2}$
Cauchy Cau_a	$a > 0$	$\mathbb{R}$	$\frac{1}{a\pi}\frac{1}{1+(x/a)^2}$	$e^{-a\lvert t\rvert}$
Binomial $b_{n,p}$	$n \in \mathbb{N}$, $p \in [0, 1]$	$\{0, \ldots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$((1-p) + pe^{it})^n$
Negative binomial $b^-_{r,p}$	$r > 0$, $p \in (0, 1]$	$\mathbb{N}_0$	$\binom{-r}{k}(-1)^k p^r (1-p)^k$	$\left(\frac{p}{1-(1-p)e^{it}}\right)^r$
Poisson Poi_λ	$\lambda > 0$	$\mathbb{N}_0$	$e^{-\lambda}\frac{\lambda^k}{k!}$	$\exp(\lambda(e^{it} - 1))$

(iv) *(N.N.)* This can either be computed directly or can be deduced from (iii) by using the Fourier inversion formula (Eq. (15.2)).

(v) *(Gamma distribution)* Again it suffices to consider the case $\theta = 1$. For $0 \le b < c \le \infty$ and $t \in \mathbb{R}$, let $\gamma_{b,c,t}$ be the linear path in $\mathbb{C}$ from $b - ibt$ to $c - ict$, let $\delta_{b,t}$ be the linear path from b to $b - ibt$ and let $\epsilon_{c,t}$ be the linear path from $c - ict$ to c. Substituting $z = (1 - it)x$, we get

$$\varphi(t) = \frac{1}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-x} e^{itx}\, dx = \frac{(1-it)^{-r}}{\Gamma(r)} \int_{\gamma_{0,\infty,t}} z^{r-1} e^{-z}\, dz.$$

Hence, it suffices to show that $\int_{\gamma_{0,\infty,t}} z^{r-1} \exp(-z)\, dz = \Gamma(r)$.

The function $z \mapsto z^{r-1}\exp(-z)$ is holomorphic in the right complex plane. Hence, by the residue theorem for $0 < b < c < \infty$,

$$\int_b^c x^{r-1}\exp(-x)\,dx = \int_{\gamma_{b,c,t}} z^{r-1}\exp(-z)\,dz + \int_{\delta_{b,t}} z^{r-1}\exp(-z)\,dz + \int_{\epsilon_{c,t}} z^{r-1}\exp(-z)\,dz.$$

Recall that $\int_0^\infty x^{r-1}\exp(-x)\,dx =: \Gamma(r)$. Hence, it is enough to show that the integrals along $\delta_{b,t}$ and $\epsilon_{c,t}$ vanish if $b \to 0$ and $c \to \infty$.

However, $|z^{r-1}\exp(-z)| \le (1+t^2)^{(r-1)/2} b^{r-1}\exp(-b)$ for $z \in \delta_{b,t}$. As the path $\delta_{b,t}$ has length $b|t|$, we get the estimate

$$\left|\int_{\delta_{b,t}} z^{r-1}e^{-z}\,dz\right| \le b^r e^{-b}\big(1+t^2\big)^{r/2} \longrightarrow 0 \quad \text{for } b \to 0.$$

Similarly,

$$\left|\int_{\epsilon_{c,t}} z^{r-1}e^{-z}\,dz\right| \le c^r e^{-c}\big(1+t^2\big)^{r/2} \longrightarrow 0 \quad \text{for } c \to \infty.$$

(vi) *(Exponential distribution)* This follows from (v) since $\exp_\theta = \Gamma_{\theta,1}$.

(vii) *(Two-sided exponential distribution)* If X and Y are independent $\exp_\theta$-distributed random variables, then it is easy to check that $X - Y \sim \exp_\theta^2$. Hence

$$\varphi_{\exp_\theta^2}(t) = \varphi_{\exp_\theta}(t)\varphi_{\exp_\theta}(-t) = \frac{1}{1-it/\theta}\,\frac{1}{1+it/\theta} = \frac{1}{1+(t/\theta)^2}.$$

(viii) *(Cauchy distribution)* This can either be computed directly using residue calculus or can be inferred from the statement for the two-sided exponential distribution by the Fourier inversion formula (Eq. (15.2)).

(ix) *(Binomial distribution)* By the binomial theorem,

$$\varphi(t) = \sum_{k=0}^{n}\binom{n}{k}(1-p)^{n-k}\big(pe^{it}\big)^k = \big(1-p+pe^{it}\big)^n.$$

(x) *(Negative binomial distribution)* By the generalized binomial theorem (Lemma 3.5), for all $x \in \mathbb{C}$ with $|x| < 1$,

$$(1-x)^{-r} = \sum_{k=0}^{\infty}\binom{-r}{k}(-x)^k.$$

Using this formula with $x = (1-p)e^{it}$ gives the claim.

(xi) *(Poisson distribution)* Clearly,

$$\varphi_{\mathrm{Poi}_\lambda}(t) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{(\lambda e^{it})^n}{n!} = e^{\lambda(e^{it}-1)}.$$

□

Corollary 15.13 *The following convolution formulas hold.*

(i) $\mathcal{N}_{\mu_1,\sigma_1^2} * \mathcal{N}_{\mu_2,\sigma_2^2} = \mathcal{N}_{\mu_1+\mu_2,\sigma_1^2+\sigma_2^2}$ *for* $\mu_1, \mu_2 \in \mathbb{R}$ *and* $\sigma_1^2, \sigma_2^2 > 0$.
(ii) $\Gamma_{\theta,r} * \Gamma_{\theta,s} = \Gamma_{\theta,r+s}$ *for* $\theta, r, s > 0$.
(iii) $\mathrm{Cau}_a * \mathrm{Cau}_b = \mathrm{Cau}_{a+b}$ *for* $a, b > 0$.
(iv) $b_{m,p} * b_{n,p} = b_{m+n,p}$ *for* $m, n \in \mathbb{N}$ *and* $p \in [0, 1]$.
(v) $b^-_{r,p} * b^-_{s,p} = b^-_{r+s,p}$ *for* $r, s > 0$ *and* $p \in (0, 1]$.
(vi) $\mathrm{Poi}_\lambda * \mathrm{Poi}_\mu = \mathrm{Poi}_{\lambda+\mu}$ *for* $\lambda, \mu \geq 0$.

Proof This follows by Theorem 15.12 and by $\varphi_{\mu*\nu} = \varphi_\mu \varphi_\nu$ (Lemma 15.11). □

The following theorem gives two simple procedures for calculating the characteristic functions of compound distributions.

Theorem 15.14

(i) *Let* $\mu_1, \mu_2, \ldots \in \mathcal{M}_f(\mathbb{R}^d)$ *and let* $p_1, p_2, \ldots$ *be nonnegative numbers with* $\sum_{n=1}^\infty p_n \mu_n(\mathbb{R}^d) < \infty$. *Then the measure* $\mu := \sum_{n=1}^\infty p_n \mu_n \in \mathcal{M}_f(\mathbb{R}^d)$ *has characteristic function*

$$\varphi_\mu = \sum_{n=1}^{\infty} p_n \varphi_{\mu_n}. \tag{15.3}$$

(ii) *Let* $N, X_1, X_2, \ldots$ *be independent random variables. Assume* $X_1, X_2, \ldots$ *are identically distributed on* $\mathbb{R}^d$ *with characteristic function* φ_X. *Assume* N *takes values in* $\mathbb{N}_0$ *and has the probability generating function* f_N. *Then* $Y := \sum_{n=1}^N X_n$ *has the characteristic function* $\varphi_Y(t) = f_N(\varphi_X(t))$.
(iii) *In particular, if we let* $N \sim \mathrm{Poi}_\lambda$ *in* (ii), *then* $\varphi_Y(t) = \exp(\lambda(\varphi_X(t) - 1))$.

Proof (i) Define $\nu_n = \sum_{k=1}^n p_k \mu_k$. By the linearity of the integral, $\varphi_{\nu_n} = \sum_{k=1}^n p_k \varphi_{\mu_k}$. By assumption, $\mu = \text{w-lim}_{n\to\infty} \nu_n$; hence also $\varphi_\mu(t) = \lim_{n\to\infty} \varphi_{\nu_n}(t)$.

(ii) Clearly,

$$\begin{aligned}\varphi_Y(t) &= \sum_{n=0}^{\infty} \mathbf{P}[N = n]\mathbf{E}\big[e^{i\langle t, X_1+\ldots+X_n\rangle}\big] \\ &= \sum_{n=0}^{\infty} \mathbf{P}[N = n]\varphi_X(t)^n = f_N\big(\varphi(t)\big).\end{aligned}$$

(iii) In this special case, $f_N(z) = e^{\lambda(z-1)}$ for $z \in \mathbb{C}$ with $|z| \leq 1$. □

Example 15.15 Let $n \in \mathbb{N}$, and assume that the points $0 = a_0 < a_1 < \ldots < a_n$ and $1 = y_0 > y_1 > \ldots > y_n = 0$ are given. Let $\varphi : \mathbb{R} \to [0, \infty)$ have the properties that

- $\varphi(a_k) = y_k$ for all $k = 0, \ldots, n$ and φ is linearly interpolated between the points a_k,

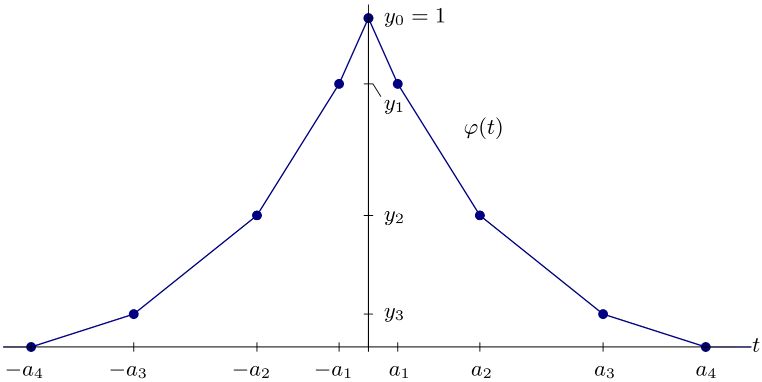

Fig. 15.1 The characteristic function φ from Example 15.15 with $n = 4$

- $\varphi(x) = 0$ for $|x| > a_n$, and
- φ is even (that is, $\varphi(x) = \varphi(-x)$).

Assume in addition that the y_k are chosen such that φ is convex on $[0, \infty)$. This is equivalent to the condition that $m_1 \le m_2 \le \ldots \le m_n \le 0$, where $m_k := \frac{y_k - y_{k-1}}{a_k - a_{k-1}}$ is the slope on the kth interval. We want to show that φ is the characteristic function of a probability measure $\mu \in \mathcal{M}_1(\mathbb{R})$.

Define $p_k = a_k(m_{k+1} - m_k)$ for $k = 1, \ldots, n$.

Let $\mu_k \in \mathcal{M}_1(\mathbb{R})$ be the distribution on $\mathbb{R}$ with density $\frac{1}{\pi} \frac{1-\cos(a_k \pi)}{a_k x^2}$. By Theorem 15.12, μ_k has the characteristic function $\varphi_{\mu_k}(t) = \big(1 - \frac{|t|}{a_k}\big)^+$. The characteristic function φ_μ of $\mu := \sum_{k=1}^n p_k \mu_k$ is then

$$\varphi_\mu(t) = \sum_{k=1}^n p_k \big(1 - |t|/a_k\big)^+.$$

This is a continuous, symmetric, real function with $\varphi_\mu(0) = 1$. It is linear on each of the intervals $[a_{k-1}, a_k]$. See Fig. 15.1 for an example with $n = 4$. By partial summation, for all $k = 1, \ldots, n$ (since $m_{n+1} = 0$),

$$\begin{aligned}
\varphi_\mu(a_l) &= \sum_{k=1}^n a_k(m_{k+1} - m_k)\left(1 - \frac{a_l}{a_k}\right)^+ = \sum_{k=l}^n (a_k - a_l)(m_{k+1} - m_k) \\
&= \big[(a_n - a_l)m_{n+1} - (a_l - a_l)m_l\big] - \sum_{k=l+1}^n (a_k - a_{k-1})m_k \\
&= -\sum_{k=l+1}^n (y_k - y_{k-1}) = y_l = \varphi(a_l).
\end{aligned}$$

Hence $\varphi_\mu = \varphi$. $\Diamond$

Example 15.16 Define the function $\varphi : \mathbb{R} \to [0, 1]$ for $t \in [-\pi, \pi)$ by $\varphi(t) = 1 - 2|t|/\pi$, and assume φ is periodic (with period 2π). By the discrete Fourier inversion formula (Theorem 15.10), φ is the characteristic function of the probability measure $\mu \in \mathcal{M}_1(\mathbb{Z})$ with $\mu(\{x\}) = (2\pi)^{-1} \int_{-\pi}^{\pi} \cos(tx)\varphi(t)\,dt$. In fact, in order that μ be a measure (not only a signed measure), we still have to show that all of the masses $\mu(\{x\})$ are nonnegative. Clearly, $\mu(\{0\}) = 0$. For $x \in \mathbb{Z} \setminus \{0\}$, use partial integration to compute the integral,

$$\begin{aligned}\int_{-\pi}^{\pi} \cos(tx)\varphi(t)\,dt &= 2\int_0^{\pi} \cos(tx)(1 - 2t/\pi)\,dt \\ &= \frac{4}{x}\left(1 - \frac{2}{\pi}\right)\sin(\pi x) - \frac{4}{x}\sin(0) + \frac{4}{\pi x}\int_0^{\pi} \sin(tx)\,dt \\ &= \frac{4}{\pi x^2}\big(1 - \cos(\pi x)\big).\end{aligned}$$

Summing up, we have

$$\mu\big(\{x\}\big) = \begin{cases} \frac{4}{\pi^2 x^2}, & \text{if } x \text{ is odd,} \\ 0, & \text{else.} \end{cases}$$

Since $\mu(\mathbb{Z}) = \varphi(0) = 1$, μ is indeed a probability measure. ◊

Example 15.17 Define the function $\psi : \mathbb{R} \to [0, 1]$ for $t \in [-\pi/2, \pi/2)$ by $\psi(t) = 1 - 2|t|/\pi$. Assume ψ is periodic with period π. If φ is the characteristic function of the measure μ from the previous example, then clearly $\psi(t) = |\varphi(t)|$. On the other hand, $\psi(t) = \frac{1}{2} + \frac{1}{2}\varphi(2t)$. By Theorem 15.14 and Lemma 15.11(ii), we infer that ψ is the characteristic function of the measure ν with $\nu(A) = \frac{1}{2}\delta_0(A) + \frac{1}{2}\mu(A/2)$ for $A \subset \mathbb{R}$. Hence,

$$\nu\big(\{x\}\big) = \begin{cases} \frac{1}{2}, & \text{if } x = 0, \\ \frac{8}{\pi^2 x^2}, & \text{if } \frac{x}{2} \in \mathbb{Z} \text{ is odd,} \\ 0, & \text{else.} \end{cases}$$ ◊

Example 15.18 Let $\varphi(t) = (1 - 2|t|/\pi)^+$ be the characteristic function of the distribution "N.N." from Theorem 15.12 (with $a = \pi/2$) and let ψ be the characteristic function from the preceding example. Note that $\varphi(t) = \psi(t)$ for $|t| \le \pi/2$ and $\varphi(t) = 0$ for $|t| > \pi/2$; hence $\varphi^2 = \varphi \cdot \psi$. Now let X, Y, Z be independent real random variables with characteristic functions $\varphi_X = \varphi_Y = \varphi$ and $\varphi_Z = \psi$. Then $\varphi_X\varphi_Y = \varphi_X\varphi_Z$; hence $X + Y \stackrel{\mathcal{D}}{=} X + Z$. However, the distributions of Y and Z do not coincide. ◊

Exercise 15.2.1 Let φ be the characteristic function of the d-dimensional random variable X. Assume that $\varphi(t) = 1$ for some $t \neq 0$. Show that $\mathbf{P}[X \in H_t] = 1$, where

$$H_t = \{x \in \mathbb{R}^d : \langle x, t\rangle \in 2\pi\mathbb{Z}\}$$
$$= \{y + z \cdot (2\pi t/\|t\|_2^2) : z \in \mathbb{Z},\, y \in \mathbb{R}^d \text{ with } \langle y, t\rangle = 0\}.$$

Infer that $\varphi(t+s) = \varphi(s)$ for all $s \in \mathbb{R}^d$.

Exercise 15.2.2 Show that there are real random variables X, X' and Y, Y' with the properties (i) $X \stackrel{\mathcal{D}}{=} X'$ and $Y \stackrel{\mathcal{D}}{=} Y'$, (ii) X' and Y' are independent, (iii) $X + Y \stackrel{\mathcal{D}}{=} X' + Y'$, and (iv) X and Y are not independent.

Exercise 15.2.3 Let X be a real random variable with characteristic function φ. X is called *lattice distributed* if there are $a, d \in \mathbb{R}$ such that $\mathbf{P}[X \in a + d\mathbb{Z}] = 1$. Show that X is lattice distributed if and only if there exists a $u \neq 0$ such that $|\varphi(u)| = 1$.

Exercise 15.2.4 Let X be a real random variable with characteristic function φ. Assume that there is a sequence $(t_n)_{n\in\mathbb{N}}$ of real numbers such that $|t_n| \downarrow 0$ and $|\varphi(t_n)| = 1$ for any n. Show that there exists a $b \in \mathbb{R}$ such that $X = b$ almost surely. If in addition, $\varphi(t_n) = 1$ for all n, then $X = 0$ almost surely.

15.3 Lévy's Continuity Theorem

The main statement of this section is Lévy's continuity theorem (Theorem 15.23). Roughly speaking, it says that a sequence of characteristic functions converges pointwise to a continuous function if and only if the limiting function is a characteristic function and the corresponding probability measures converge weakly. We prepare for the proof of this theorem by assembling some analytic tools.

Lemma 15.19 *Let $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ with characteristic function φ. Then*

$$\big|\varphi(t) - \varphi(s)\big|^2 \le 2\big(1 - \mathrm{Re}\big(\varphi(t-s)\big)\big) \quad \textit{for all } s, t \in \mathbb{R}^d.$$

Proof By the Cauchy–Schwarz inequality,

$$\begin{aligned}
\big|\varphi(t) - \varphi(s)\big|^2 &= \left|\int_{\mathbb{R}^d} e^{i\langle t,x\rangle} - e^{i\langle s,x\rangle}\,\mu(dx)\right|^2 \\
&= \left|\int_{\mathbb{R}^d} \big(e^{i\langle t-s,x\rangle} - 1\big) e^{i\langle s,x\rangle}\,\mu(dx)\right|^2 \\
&\le \int_{\mathbb{R}^d} \big|e^{i\langle t-s,x\rangle} - 1\big|^2 \mu(dx) \cdot \int_{\mathbb{R}^d} \big|e^{i\langle s,x\rangle}\big|^2 \mu(dx) \\
&= \int_{\mathbb{R}^d} \big(e^{i\langle t-s,x\rangle} - 1\big)\big(e^{-i\langle t-s,x\rangle} - 1\big)\mu(dx) \\
&= 2\big(1 - \mathrm{Re}\big(\varphi(t-s)\big)\big).
\end{aligned}$$

□

Definition 15.20 Let (E, d) be a metric space. A family $(f_i, i \in I)$ of maps $E \to \mathbb{R}$ is called *uniformly equicontinuous* if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $|f_i(t) - f_i(s)| < \varepsilon$ for all $i \in I$ and all $s, t \in E$ with $d(s, t) < \delta$.

Theorem 15.21 *If $\mathcal{F} \subset \mathcal{M}_1(\mathbb{R}^d)$ is a tight family, then $\{\varphi_\mu : \mu \in \mathcal{F}\}$ is uniformly equicontinuous. In particular, every characteristic function is uniformly continuous.*

Proof We have to show that, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $t \in \mathbb{R}^d$, all $s \in \mathbb{R}^d$ with $|t - s| < \delta$ and all $\mu \in \mathcal{F}$, we have

$$\left|\varphi_\mu(t) - \varphi_\mu(s)\right| < \varepsilon.$$

As $\mathcal{F}$ is tight, there exists an $N \in \mathbb{N}$ with $\mu([-N, N]^d) > 1 - \varepsilon^2/6$ for all $\mu \in \mathcal{F}$. Furthermore, there exists a $\delta > 0$ such that, for $x \in [-N, N]^d$ and $u \in \mathbb{R}^d$ with $|u| < \delta$, we have $|1 - e^{i\langle u,x\rangle}| < \varepsilon^2/6$. Hence we get for all $\mu \in \mathcal{F}$

$$\begin{aligned} 1 - \mathrm{Re}\big(\varphi_\mu(u)\big) &\le \int_{\mathbb{R}^d} \left|1 - e^{i\langle u,x\rangle}\right| \mu(dx) \\ &\le \frac{\varepsilon^2}{3} + \int_{[-N,N]^d} \left|1 - e^{i\langle u,x\rangle}\right| \mu(dx) \\ &\le \frac{\varepsilon^2}{3} + \frac{\varepsilon^2}{6} = \frac{\varepsilon^2}{2}. \end{aligned}$$

Thus, for $|t - s| < \delta$ by Lemma 15.19, $|\varphi_\mu(t) - \varphi_\mu(s)| \le \varepsilon$. □

Lemma 15.22 *Let (E, d) be a metric space and let $f, f_1, f_2, \ldots$ be maps $E \to \mathbb{R}$ with $f_n \overset{n\to\infty}{\longrightarrow} f$ pointwise. If $(f_n)_{n\in\mathbb{N}}$ is uniformly equicontinuous, then f is uniformly continuous and $(f_n)_{n\in\mathbb{N}}$ converges to f uniformly on compact sets; that is, for every compact set $K \subset E$, we have*

$$\sup_{s\in K} \left|f_n(s) - f(s)\right| \overset{n\to\infty}{\longrightarrow} 0.$$

Proof Fix $\varepsilon > 0$, and choose $\delta > 0$ such that $|f_n(t) - f_n(s)| < \varepsilon$ for all $n \in \mathbb{N}$ and all $s, t \in E$ with $d(s, t) < \delta$. For these s, t, we thus have

$$\left|f(s) - f(t)\right| = \lim_{n\to\infty} \left|f_n(s) - f_n(t)\right| \le \varepsilon.$$

Hence, f is uniformly continuous.

Now let $K \subset E$ be compact. As compact sets are totally bounded, there exists an $N \in \mathbb{N}$ and points $t_1, \ldots, t_N \in K$ with $K \subset \bigcup_{i=1}^N B_\delta(t_i)$. Choose $n_0 \in \mathbb{N}$ large enough that $|f_n(t_i) - f(t_i)| \le \varepsilon$ for all $i = 1, \ldots, N$ and $n \ge n_0$.

Now let $s \in K$ and $n \ge n_0$. Choose a t_i with $d(s, t_i) < \delta$. Then

$$\left|f_n(s) - f(s)\right| \le \left|f_n(s) - f_n(t_i)\right| + \left|f_n(t_i) - f(t_i)\right| + \left|f(t_i) - f(s)\right| \le 3\varepsilon.$$

As $\varepsilon > 0$ was arbitrary, we infer that $f_n \overset{n\to\infty}{\longrightarrow} f$ uniformly on K. □

A map $f : \mathbb{R}^d \to \mathbb{R}$ is called *partially continuous* at $x = (x_1, \ldots, x_d)$ if, for any $i = 1, \ldots, d$, the map $y_i \mapsto f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d)$ is continuous at $y_i = x_i$.

Theorem 15.23 (Lévy's continuity theorem) *Let $P, P_1, P_2, \ldots \in \mathcal{M}_1(\mathbb{R}^d)$ with characteristic functions $\varphi, \varphi_1, \varphi_2, \ldots$.*

(i) *If $P = \text{w-lim}_{n\to\infty} P_n$, then $\varphi_n \overset{n\to\infty}{\longrightarrow} \varphi$ uniformly on compact sets.*
(ii) *If $\varphi_n \overset{n\to\infty}{\longrightarrow} f$ pointwise for some $f : \mathbb{R}^d \to \mathbb{C}$ that is partially continuous at 0, then there exists a probability measure Q such that $\varphi_Q = f$ and $Q = \text{w-lim}_{n\to\infty} P_n$.*

Proof (i) By the definition of weak convergence, we have $\varphi_n \overset{n\to\infty}{\longrightarrow} \varphi$ pointwise. As the family $(P_n)_{n\in\mathbb{N}}$ is tight, by Theorem 15.21, $(\varphi_n)_{n\in\mathbb{N}}$ is uniformly equicontinuous. By Lemma 15.22, this implies uniform convergence on compact sets.

(ii) By Theorem 13.34, it is enough to show that the sequence $(P_n)_{n\in\mathbb{N}}$ is tight. For this purpose, it suffices to show that, for every $k = 1, \ldots, n$, the sequence $(P_n^k)_{n\in\mathbb{N}}$ of kth marginal distributions is tight. (Here $P_n^k := P_n \circ \pi_k^{-1}$, where $\pi_k : \mathbb{R}^d \to \mathbb{R}$ is the projection on the kth coordinate.) Let e_k be the kth unit vector in $\mathbb{R}^d$. Then $\varphi_{P_n^k}(t) = \varphi_n(te_k)$ is the characteristic function of P_n^k. By assumption, $\varphi_{P_n^k} \overset{n\to\infty}{\longrightarrow} f_k$ pointwise for some function f_k that is continuous at 0. We have thus reduced the problem to the one-dimensional situation and will henceforth assume $d = 1$.

As $\varphi_n(0) = 1$ for all $n \in \mathbb{N}$, we have $f(0) = 1$. Define the map $h : \mathbb{R} \to [0, \infty)$ by $h(x) = 1 - \sin(x)/x$ for $x \neq 0$ and $h(0) = 0$. Clearly, h is continuously differentiable on $\mathbb{R}$. It is easy to see that $\alpha := \inf\{h(x) : |x| \ge 1\} = 1 - \sin(1) > 0$. Now, for $K > 0$, compute (using Markov's inequality and Fubini's theorem)

$$\begin{aligned}
P_n\big([-K, K]^c\big) &\le \alpha^{-1} \int_{[-K,K]^c} h(x/K)\, P_n(dx) \\
&\le \alpha^{-1} \int_{\mathbb{R}} h(x/K)\, P_n(dx) \\
&= \alpha^{-1} \int_{\mathbb{R}} \left(\int_0^1 \big(1 - \cos(tx/K)\big)\, dt \right) P_n(dx) \\
&= \alpha^{-1} \int_0^1 \left(\int_{\mathbb{R}} \big(1 - \cos(tx/K)\big)\, P_n(dx) \right) dt \\
&= \alpha^{-1} \int_0^1 \big(1 - \text{Re}\big(\varphi_n(t/K)\big)\big)\, dt.
\end{aligned}$$

Using dominated convergence, we conclude that

$$\begin{aligned}\limsup_{n\to\infty} P_n\big([-K,K]^c\big) &\le \alpha^{-1}\limsup_{n\to\infty}\int_0^1 \big(1-\mathrm{Re}\big(\varphi_n(t/K)\big)\big)\,dt\\ &= \alpha^{-1}\int_0^1 \Big(\lim_{n\to\infty}\big(1-\mathrm{Re}\big(\varphi_n(t/K)\big)\big)\Big)\,dt\\ &= \alpha^{-1}\int_0^1 \big(1-\mathrm{Re}\big(f(t/K)\big)\big)\,dt.\end{aligned}$$

As f is continuous and $f(0)=1$, the last integral converges to 0 for $K\to\infty$. Hence $(P_n)_{n\in\mathbb{N}}$ is tight. □

Applying Lévy's continuity theorem to Example 15.15, we get a theorem of Pólya.

Theorem 15.24 (Pólya) *Let $f:\mathbb{R}\to[0,1]$ be continuous and even with $f(0)=1$. Assume that f is convex on $[0,\infty)$. Then f is the characteristic function of a probability measure.*

Proof Define f_n by $f_n(k/n):=f(k/n)$ for $k=0,\ldots,n^2$, and assume f_n is linearly interpolated between these points. Furthermore, let f_n be constant to the right of n and for $x<0$, define $f_n(x)=f_n(-x)$. This is an approximation of f on $[0,\infty)$ by convex and piecewise linear functions. By Example 15.15, every f_n is a characteristic function of a probability measure μ_n. Clearly, $f_n \overset{n\to\infty}{\longrightarrow} f$ pointwise; hence f is the characteristic function of a probability measure $\mu=\text{w-lim}_{n\to\infty}\mu_n$ on $\mathbb{R}$. □

Corollary 15.25 *For every $\alpha\in(0,1]$ and $r>0$, $\varphi_{\alpha,r}(t)=e^{-|rt|^\alpha}$ is the characteristic function of a symmetric probability measure $\mu_{\alpha,r}$ on $\mathbb{R}$.*

Remark 15.26 In fact, $\varphi_{\alpha,r}$ is a characteristic function for every $\alpha\in(0,2]$ ($\alpha=2$ corresponds to the normal distribution), see Section 16.2. The distributions $\mu_{\alpha,r}$ are the so-called α-*stable* distributions (see Definition 16.20): If $X_1,X_2,\ldots,X_n$ are independent and $\mu_{\alpha,a}$-distributed, then $\varphi_{X_1+\ldots+X_n}(t)=\varphi_X(t)^n=\varphi_X(n^{1/\alpha}t)$; hence $X_1+\ldots+X_n \overset{\mathcal{D}}{=} n^{1/\alpha}X_1$. ◊

The Stone–Weierstraß theorem implies that a characteristic function determines a probability distribution uniquely. Pólya's theorem gives a sufficient condition for a symmetric real function to be a characteristic function. Clearly, that condition is not necessary, as, for example, the normal distribution does not fulfill it. For general education we present Bochner's theorem that formulates a necessary and sufficient condition for a function $\varphi:\mathbb{R}^d\to\mathbb{C}$ to be the characteristic function of a probability measure.

Definition 15.27 A function $f : \mathbb{R}^d \to \mathbb{C}$ is called *positive semidefinite* if, for all $n \in \mathbb{N}$, all $t_1, \dots, t_n \in \mathbb{R}^d$ and all $y_1, \dots, y_n \in \mathbb{C}$, we have

$$\sum_{k,l=1}^{n} y_k \bar{y}_l f(t_k - t_l) \geq 0,$$

in other words, if the matrix $(f(t_k - t_l))_{k,l=1,\dots,n}$ is positive semidefinite.

Lemma 15.28 *If $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ has characteristic function φ, then φ is positive semidefinite.*

Proof We have

$$\begin{aligned}\sum_{k,l=1}^{n} y_k \bar{y}_l \varphi(t_k - t_l) &= \sum_{k,l=1}^{n} y_k \bar{y}_l \int e^{ix(t_k - t_l)} \mu(dx) \\ &= \int \sum_{k,l=1}^{n} y_k e^{ixt_k} \overline{y_l e^{ixt_l}} \mu(dx) \\ &= \int \left| \sum_{k=1}^{n} y_k e^{ixt_k} \right|^2 \mu(dx) \geq 0.\end{aligned}$$

□

In the case $d = 1$, the following theorem goes back to Bochner (1932) [19].

Theorem 15.29 (Bochner) *A continuous function $\varphi : \mathbb{R}^d \to \mathbb{C}$ is the characteristic function of a probability distribution on $\mathbb{R}^d$ if and only if φ is positive semidefinite and $\varphi(0) = 1$.*
The statement still holds if $\mathbb{R}^d$ is replaced by a locally compact Abelian group.

Proof For the case $d = 1$ see [19, Section 20, Theorem 23] or [54, Chapter XIX.2, p. 622]. For the general case, see, e.g., [71, p. 293, Theorem 33.3]. □

Exercise 15.3.1 (Compare [50] and [4]) Show that there exist two exchangeable sequences $X = (X_n)_{n \in \mathbb{N}}$ and $Y = (Y_n)_{n \in \mathbb{N}}$ of real random variables with $\mathbf{P}_X \neq \mathbf{P}_Y$ but such that

$$\sum_{k=1}^{n} X_k \stackrel{\mathcal{D}}{=} \sum_{k=1}^{n} Y_k \quad \text{for all } n \in \mathbb{N}. \tag{15.4}$$

Hint:

(i) Define the characteristic functions (see Theorem 15.12) $\varphi_1(t) = \frac{1}{1+t^2}$ and $\varphi_2(t) = (1 - t/2)^+$. Use Pólya's theorem to show that

$$\psi_1(t) := \begin{cases} \varphi_1(t), & \text{if } |t| \le 1, \\ \varphi_2(t), & \text{if } |t| > 1, \end{cases}$$

and

$$\psi_2(t) := \begin{cases} \varphi_2(t), & \text{if } |t| \le 1, \\ \varphi_1(t), & \text{if } |t| > 1, \end{cases}$$

are characteristic functions of probability distributions on $\mathbb{R}$.

(ii) Define independent random variables $X_{n,i}, Y_{n,i}, n \in \mathbb{N}, i = 1, 2$, and $\Theta_n, n \in \mathbb{N}$ such that $X_{n,i}$ has characteristic function φ_i, $Y_{n,i}$ has characteristic function ψ_i and $\mathbf{P}[\Theta_n = 1] = \mathbf{P}[\Theta_n = -1] = \frac{1}{2}$. Define $X_n = X_{n,\Theta_n}$ and $Y_n = Y_{n,\Theta_n}$. Show that (15.4) holds.

(iii) Determine $\mathbf{E}[e^{it_1 X_1 + it_2 X_2}]$ and $\mathbf{E}[e^{it_1 Y_1 + it_2 Y_2}]$ for $t_1 = \frac{1}{2}$ and $t_2 = 2$. Conclude that $(X_1, X_2) \stackrel{\mathcal{D}}{\neq} (Y_1, Y_2)$ and thus $\mathbf{P}_X \neq \mathbf{P}_Y$.

Exercise 15.3.2 Show that for any $\delta > 0$ and $\varepsilon > 0$, there is a $C < \infty$ such that for any $\mu \in \mathcal{M}_1(\mathbb{R})$ with characteristic function φ, we have

$$\mu\big([-\delta, \delta]^c\big) \le C \int_0^\varepsilon \big(1 - \mathrm{Re}\big(\varphi(t)\big)\big)\, dt.$$

For $\varepsilon\delta \le 3$ one can choose $C = 12/\delta^2\varepsilon^3$.

Hint: Proceed as in the proof of Lévy's continuity theorem.

Exercise 15.3.3 Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}$ and denote by $(\varphi_n)_{n\in\mathbb{N}}$ the corresponding characteristic functions. Assume that $\varphi_n(t) \stackrel{n\to\infty}{\longrightarrow} 1$ for t in a neighborhood of 0. Use Exercise 15.3.2 to show that $\mu_n \stackrel{n\to\infty}{\longrightarrow} \delta_0$.

15.4 Characteristic Functions and Moments

We want to study the connection between the moments of a real random variable X and the derivatives of its characteristic function φ_X. We start with a simple lemma.

Lemma 15.30 *For $t \in \mathbb{R}$ and $n \in \mathbb{N}$, we have*

$$\left| e^{it} - 1 - \frac{it}{1!} - \ldots - \frac{(it)^{n-1}}{(n-1)!} \right| \le \frac{|t|^n}{n!}.$$

Proof As the nth derivative of e^{it} has modulus 1, this follows by Taylor's formula. □

Theorem 15.31 (Moments and differentiability) *Let X be a real random variable with characteristic function φ.*

(i) *If $\mathbf{E}[|X|^n] < \infty$, then φ is n-times continuously differentiable with derivatives*

$$\varphi^{(k)}(t) = \mathbf{E}\big[(iX)^k e^{itX}\big] \quad \text{for } k = 0, \ldots, n.$$

(ii) *In particular, if $\mathbf{E}[X^2] < \infty$, then*

$$\varphi(t) = 1 + it\mathbf{E}[X] - \frac{1}{2}t^2\mathbf{E}\big[X^2\big] + \varepsilon(t)t^2$$

with $\varepsilon(t) \to 0$ for $t \to 0$.

(iii) *Let $h \in \mathbb{R}$. If $\lim_{n\to\infty} \frac{|h|^n \mathbf{E}[|X|^n]}{n!} = 0$, then, for every $t \in \mathbb{R}$,*

$$\varphi(t+h) = \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} \mathbf{E}\big[e^{itX} X^k\big].$$

In particular, this holds if $\mathbf{E}[e^{|hX|}] < \infty$.

Proof (i) For $t \in \mathbb{R}$, $h \in \mathbb{R} \setminus \{0\}$ and $k \in \{1, \ldots, n\}$, define

$$Y_k(t, h, x) = k! h^{-k} e^{itx} \left(e^{ihx} - \sum_{l=0}^{k-1} \frac{(ihx)^l}{l!} \right).$$

Then

$$\mathbf{E}\big[Y_k(t, h, X)\big] = k! h^{-k} \left(\varphi(t+h) - \varphi(t) - \sum_{l=1}^{k-1} \mathbf{E}\big[e^{itX}(iX)^l\big] \frac{h^l}{l!} \right).$$

If the limit $\varphi_k(t) := \lim_{h\to 0} \mathbf{E}[Y_k(t, h, X)]$ exists, then φ is k-times differentiable at t with $\varphi^{(k)}(t) = \varphi_k(t)$.

However (by Lemma 15.30 with $n = k + 1$), $Y_k(t, h, x) \overset{h\to 0}{\longrightarrow} (ix)^k e^{itx}$ for all $x \in \mathbb{R}$ and (by Lemma 15.30 with $n = k$) $|Y_k(t, h, x)| \le |x|^k$. As $\mathbf{E}[|X|^k] < \infty$ by assumption, the dominated convergence theorem implies

$$\mathbf{E}\big[Y_k(t, h, X)\big] \overset{h\to 0}{\longrightarrow} \mathbf{E}\big[(iX)^k e^{itX}\big] = \varphi^{(k)}(t).$$

Applying the continuity lemma (Theorem 6.27) yields that $\varphi^{(k)}$ is continuous.

(ii) This is a direct consequence of (i).
(iii) By assumption,

$$\begin{aligned}\left|\varphi(t+h) - \sum_{k=0}^{n-1} \frac{(ih)^k}{k!} \mathbf{E}\big[e^{itX} X^k\big]\right| &= \frac{h^n}{n!} \big|\mathbf{E}\big[Y_n(t,h,X)\big]\big| \\ &\le \frac{h^n \mathbf{E}[|X|^n]}{n!} \overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$ □

Corollary 15.32 (Method of moments) *Let X be a real random variable with*

$$\alpha := \limsup_{n\to\infty} \frac{1}{n} \mathbf{E}\big[|X|^n\big]^{1/n} < \infty.$$

Then the characteristic function φ of X is analytic and the distribution of X is uniquely determined by the moments $\mathbf{E}[X^n]$, $n \in \mathbb{N}$. In particular, this holds if $\mathbf{E}[e^{t|X|}] < \infty$ for some $t > 0$.

Proof By Stirling's formula,

$$\lim_{n\to\infty} \frac{1}{n!} n^n e^{-n} \sqrt{2\pi n} = 1.$$

Thus, for $|h| < 1/(3\alpha)$,

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{E}\big[|X|^n\big] \cdot |h|^n / n! &= \limsup_{n\to\infty} \sqrt{2\pi n} \big(\mathbf{E}\big[|X|^n\big]^{1/n} \cdot |h| \cdot e/n\big)^n \\ &\le \limsup_{n\to\infty} \sqrt{2\pi n} (e/3)^n = 0.\end{aligned}$$

Hence the characteristic function can be expanded about any point $t \in \mathbb{R}$ in a power series with radius of convergence at least $1/(3\alpha)$. In particular, it is analytic and is hence determined by the coefficients of its power series about $t = 0$; that is, by the moments of X. □

Example 15.33

(i) Let $X \sim \mathcal{N}_{\mu,\sigma^2}$. Then, for every $t \in \mathbb{R}$,

$$\begin{aligned}\mathbf{E}\big[e^{tX}\big] &= \big(2\pi\sigma^2\big)^{-1/2} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/2\sigma^2}\, dx \\ &= e^{\mu t + t^2\sigma^2/2} \big(2\pi\sigma^2\big)^{-1/2} \int_{-\infty}^{\infty} e^{-(x-\mu-t\sigma^2)^2/2\sigma^2}\, dx \\ &= e^{\mu t + t^2\sigma^2/2} < \infty.\end{aligned}$$

Hence the distribution of X is characterized by its moments. The characteristic function $\varphi(t) = e^{i\mu t} e^{-\sigma^2 t^2/2}$ that we get by the above calculation with t replaced by it is indeed analytic.

(ii) Let X be exponentially distributed with parameter $\theta > 0$. Then, for $t \in (0, \theta)$,

$$\mathbf{E}\big[e^{tX}\big] = \theta \int_0^\infty e^{tx} e^{-\theta x}\, dx = \frac{\theta}{\theta - t} < \infty.$$

Hence the distribution of X is characterized by its moments. The above calculation with t replaced by it yields $\varphi(t) = \theta/(\theta - it)$, and this function is indeed analytic. The fact that in the complex plane φ has a singularity at $t = -i\theta$ implies that the power series of φ about 0 has radius of convergence θ. In particular, this implies that not all exponential moments are finite. This is reflected by the above calculation that shows that, for $t \geq \theta$, the exponential moments are infinite.

(iii) Let X be log-normally distributed (see Example 15.5). Then $\mathbf{E}[X^n] = e^{n^2/2}$. In particular, here $\alpha = \infty$. In fact, in Example 15.5, we saw that here the moments do not determine the distribution of X.

(iv) If X takes values in $\mathbb{N}_0$ and if $\beta := \limsup_{n\to\infty} \mathbf{E}[X^n]^{1/n} < 1$, then by Hadamard's criterion $\psi_X(z) := \sum_{k=1}^\infty \mathbf{P}[X = k] z^k < \infty$ for $|z| < 1/\beta$. In particular, the probability generating function X is characterized by its derivatives $\psi_X^{(n)}(1)$, $n \in \mathbb{N}$, and thus by the moments of X. Compare Theorem 3.2(iii). ◊

Theorem 15.34 *Let X be a real random variable and let φ be its characteristic function. Let $n \in \mathbb{N}$, and assume that φ is $2n$-times differentiable at 0 with derivative $\varphi^{(2n)}(0)$. Then $\mathbf{E}[X^{2n}] = (-1)^n \varphi^{(2n)}(0) < \infty$.*

Proof We carry out the proof by induction on $n \in \mathbb{N}_0$. For $n = 0$, the claim is trivially true. Now, let $n \in \mathbb{N}$, and assume φ is $2n$-times (not necessarily continuously) differentiable at 0. Define $u(t) = \mathrm{Re}(\varphi(t))$. Then u is also $2n$-times differentiable at 0 and $u^{(2k-1)}(0) = 0$ for $k = 1, \ldots, n$ since u is even. Since $\varphi^{(2n)}(0)$ exists, $\varphi^{(2n-1)}$ is continuous at 0 and $\varphi^{(2n-1)}(t)$ exists for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. Furthermore, $\varphi^{(k)}$ exists in $(-\varepsilon, \varepsilon)$ and is continuous on $(-\varepsilon, \varepsilon)$ for any $k = 0, \ldots, 2n-2$. By Taylor's formula, for every $t \in (-\varepsilon, \varepsilon)$,

$$\left| u(t) - \sum_{k=0}^{n-1} u^{(2k)}(0) \frac{t^{2k}}{(2k)!} \right| \leq \frac{|t|^{2n-1}}{(2n-1)!} \sup_{\theta \in (0,1]} \left| u^{(2n-1)}(\theta t) \right|. \tag{15.5}$$

Define a continuous function $f_n : \mathbb{R} \to [0, \infty)$ by $f_n(0) = 1$ and

$$f_n(x) = (-1)^n (2n)! x^{-2n} \left[\cos(x) - \sum_{k=0}^{n-1} (-1)^k \frac{x^{2k}}{(2k)!} \right] \quad \text{for } x \neq 0.$$

By the induction hypothesis, $\mathbf{E}[X^{2k}] = u^{(2k)}(0)$ for all $k = 1, \ldots, n-1$. Using (15.5), we infer

$$\mathbf{E}\big[f_n(tX)X^{2n}\big] \le \frac{2n}{|t|} \sup_{\theta\in(0,1]} \big|u^{(2n-1)}(\theta t)\big| \le g_n(t)$$
$$:= 2n \sup_{\theta\in(0,1]} \frac{|u^{(2n-1)}(\theta t)|}{\theta|t|}.$$

Now Fatou's lemma implies

$$\mathbf{E}\big[X^{2n}\big] = \mathbf{E}\big[f_n(0)X^{2n}\big] \le \liminf_{t\to 0} \mathbf{E}\big[f_n(tX)X^{2n}\big]$$
$$\le \liminf_{t\to 0} g_n(t) = 2n\big|u^{(2n)}(0)\big| < \infty.$$

By Theorem 15.31, this implies $\mathbf{E}[X^{2n}] = (-1)^n u^{(2n)}(0) = (-1)^n \varphi^{(2n)}(0)$. □

Remark 15.35 For odd moments, the statement of the theorem may fail (see, e.g., Exercise 15.4.4 for the first moment). Indeed, φ is differentiable at 0 with derivative im for some $m \in \mathbb{R}$ if and only if $x\mathbf{P}[|X| > x] \overset{x\to\infty}{\longrightarrow} 0$ and $\mathbf{E}[X \mathbb{1}_{\{|X|\le x\}}] \overset{x\to\infty}{\longrightarrow} m$. (See [54, Chapter XVII.2a, p. 565].) ◊

Exercise 15.4.1 Let X and Y be nonnegative random variables with

$$\limsup_{n\to\infty} \frac{1}{n}\mathbf{E}\big[|X|^n\big]^{1/n} < \infty, \qquad \limsup_{n\to\infty} \frac{1}{n}\mathbf{E}\big[|Y|^n\big]^{1/n} < \infty,$$

and

$$\mathbf{E}\big[X^m Y^n\big] = \mathbf{E}\big[X^m\big]\mathbf{E}\big[Y^n\big] \quad \text{for all } m, n \in \mathbb{N}_0.$$

Show that X and Y are independent.

Hint: Consider the random variable Y with respect to the probability measure $X^m\mathbf{P}[\,\cdot\,]/\mathbf{E}[X^m]$, and use Corollary 15.32 to show that

$$\mathbf{E}\big[X^m \mathbb{1}_A(Y)\big]/\mathbf{E}\big[X^m\big] = \mathbf{P}[Y \in A] \quad \text{for all } A \in \mathcal{B}(R) \text{ and } m \in \mathbb{N}_0.$$

Now apply Corollary 15.32 to the random variable X with respect to the probability measure $\mathbf{P}[\,\cdot \mid Y \in A]$.

Exercise 15.4.2 Let $r, s > 0$ and let $Z \sim \Gamma_{1,r+s}$ and $B \sim \beta_{r,s}$ be independent (see Example 1.107). Use Exercise 15.4.1 to show that the random variables $X := BZ$ and $Y := (1-B)Z$ are independent with $X \sim \Gamma_{1,r}$ and $Y \sim \Gamma_{1,s}$.

Exercise 15.4.3 Show that, for $\alpha > 2$, the function $\phi_\alpha(t) = e^{-|t|^\alpha}$ is not a characteristic function.

Hint: Assume the contrary and show that the corresponding random variable would have variance zero.

Exercise 15.4.4 Let $X_1, X_2, \dots$ be i.i.d. real random variables with characteristic function φ. Show the following.

(i) If φ is differentiable at 0, then $\varphi'(0) = im$ for some $m \in \mathbb{R}$.
(ii) φ is differentiable at 0 with $\varphi'(0) = im$ if and only if $(X_1 + \dots + X_n)/n \overset{n\to\infty}{\longrightarrow} m$ in probability.
(iii) Assume that φ is differentiable at 0 and that $X_1 \geq 0$ almost surely. Then $\mathbf{E}[X_1] = -i\varphi'(0) < \infty$.
Hint: Use (ii) and the law of large numbers.
(iv) The distribution of X_1 can be chosen such that φ is differentiable at 0 but $\mathbf{E}[|X_1|] = \infty$.

Exercise 15.4.5 Let $X_1, X_2, \dots$ be real random variables. For $r > 0$ let $M_r(X_n) = \mathbf{E}[|X_n|^r]$ be the rth absolute moment. For $k \in \mathbb{N}$ let $m_k(X_n) = \mathbf{E}[X_n^k]$ be the kth moment if $M_k(X_n) < \infty$.

(i) Assume that X is a real random variable and that $(X_{n_l})_{l\in\mathbb{N}}$ is a subsequence such that

$$\mathbf{P}_{X_{n_l}} \overset{l\to\infty}{\longrightarrow} \mathbf{P}_X \quad \text{weakly.}$$

Assume further that there is an $r > 0$ such that $\sup_{n\in\mathbb{N}} M_r(X_n) < \infty$. Show that for any $k \in \mathbb{N} \cap (0, r)$ and $s \in (0, r)$ we have $M_s(X) < \infty$ as well as

$$M_s(X_{n_l}) \overset{l\to\infty}{\longrightarrow} M_s(X) \quad \text{and} \quad m_k(X_{n_l}) \overset{l\to\infty}{\longrightarrow} m_k(X).$$

(ii) Assume that for any $k \in \mathbb{N}$ the limit

$$m_k := \lim_{n\to\infty} m_k(X_n)$$

exists and is finite (note that finitely many of the $m_k(X_n)$ may be undefined for any k). Show that there exists a real random variable X with $m_k = m_k(X)$ for all $k \in \mathbb{N}$ and a subsequence $(X_{n_l})_{l\in\mathbb{N}}$ such that

$$\mathbf{P}_{X_{n_l}} \overset{l\to\infty}{\longrightarrow} \mathbf{P}_X \quad \text{weakly.}$$

(iii) Show the theorem of Fréchet–Shohat: If in (ii) the distribution of X is determined by its moments $m_k(X)$, $k \in \mathbb{N}$ (see Corollary 15.32), then

$$\mathbf{P}_{X_n} \overset{n\to\infty}{\longrightarrow} \mathbf{P}_X \quad \text{weakly.}$$

Exercise 15.4.6 Let $X_1, X_2, \dots$ be i.i.d. real random variables with $\mathbf{E}[X_1] = 0$ and $\mathbf{E}[|X_1|^k] < \infty$ for all $k \in \mathbb{N}$.

(i) Show that there exist finite numbers $(d_k)_{k\in\mathbb{N}}$ (depending on the distribution $\mathbf{P}_{X_1}$) such that for any $k, n \in \mathbb{N}$ we have

$$\left|\mathbf{E}\big[(X_1 + \dots + X_n)^{2k-1}\big]\right| \leq d_{2k-1} n^{k-1}$$

and

$$\left|\mathbf{E}\left[(X_1+\ldots+X_n)^{2k}\right]-\frac{(2k)!}{2^k k!}\mathbf{E}\left[X_1^2\right]^k n^k\right| \le d_{2k}n^{k-1}.$$

Hint: Expand the bracket expression, sort the terms by the different mixed moments and compute by combinatorial means the number of each type of summand. The number of summands of the type $\mathbf{E}[X_{l_1}^2 \ldots X_{l_k}^2]$ (for different $l_1, \ldots, l_k$) is of particular importance.

(ii) Let $Y \sim \mathcal{N}_{0,1}$. Use Theorem 15.31(i) to show that for any $k \in \mathbb{N}$ we have

$$\mathbf{E}\left[Y^{2k-1}\right]=0 \quad \text{and} \quad \mathbf{E}\left[Y^{2k}\right]=\frac{(2k)!}{2^k k!}.$$

(iii) Let $S_n^* = (X_1+\ldots+X_n)/\sqrt{n\mathbf{Var}[X_1]}$. Use Exercise 15.4.5 to infer the statement of the central limit theorem (compare Theorem 15.37)

$$\mathbf{P}_{S_n^*} \overset{n\to\infty}{\longrightarrow} \mathcal{N}_{0,1} \quad \text{weakly.}$$

15.5 The Central Limit Theorem

In the strong law of large numbers, we saw that, for large n, the order of magnitude of the sum $S_n = X_1+\ldots+X_n$ of i.i.d. integrable random variables is $n\cdot\mathbf{E}[X_1]$. Of course, for any n, the actual value of S_n will sometimes be smaller than $n\cdot\mathbf{E}[X_1]$ and sometimes larger. In the central limit theorem (CLT), we study the size and shape of the *typical fluctuations* around $n\cdot\mathbf{E}[X_1]$ in the case where the X_i have a finite variance.

We prepare for the proof of the CLT with a lemma.

Lemma 15.36 *Let $X_1, X_2, \ldots$ be i.i.d. real random variables with $\mathbf{E}[X_1]=\mu$ and $\mathbf{Var}[X_1]=\sigma^2 \in (0,\infty)$. Let*

$$S_n^* := \frac{1}{\sqrt{n\sigma^2}}\sum_{k=1}^n (X_k-\mu)$$

be the normalized nth partial sum. Then

$$\lim_{n\to\infty}\varphi_{S_n^*}(t)=e^{-t^2/2} \quad \textit{for all } t\in\mathbb{R}.$$

Proof Let $\varphi=\varphi_{X_k-\mu}$. Then, by Theorem 15.31(ii),

$$\varphi(t)=1-\frac{\sigma^2}{2}t^2+\varepsilon(t)t^2,$$

where the error term $\varepsilon(t)$ goes to 0 if $t\to 0$. By Lemma 15.11(iv) and (ii),

$$\varphi_{S_n^*}(t)=\varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)^n.$$

Now $(1-\frac{t^2}{2n})^n \overset{n\to\infty}{\longrightarrow} e^{-t^2/2}$ and

$$\left|\left(1-\frac{t^2}{2n}\right)^n - \varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)^n\right| \le n\left|1-\frac{t^2}{2n}-\varphi\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right|$$

$$\le n\frac{t^2}{n\sigma^2}\left|\varepsilon\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right| \overset{n\to\infty}{\longrightarrow} 0.$$

(Note that $|u^n - v^n| \le |u-v| \cdot n \cdot \max(|u|,|v|)^{n-1}$ for all $u, v \in \mathbb{C}$.) □

Theorem 15.37 (Central limit theorem (CLT)) *Let $X_1, X_2, \ldots$ be i.i.d. real random variables with $\mu := \mathbf{E}[X_1] \in \mathbb{R}$ and $\sigma^2 := \mathbf{Var}[X_1] \in (0,\infty)$. For $n \in \mathbb{N}$, let $S_n^* := \frac{1}{\sqrt{\sigma^2 n}} \sum_{i=1}^n (X_i - \mu)$. Then*

$$\mathbf{P}_{S_n^*} \overset{n\to\infty}{\longrightarrow} \mathcal{N}_{0,1} \quad \textit{weakly}.$$

For $-\infty \le a < b \le +\infty$, we have $\lim_{n\to\infty} \mathbf{P}[S_n^* \in [a,b]] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx$.

Proof By Lemma 15.36 and Lévy's continuity theorem (Theorem 15.23), $(\mathbf{P}_{S_n^*})$ converges to the distribution with characteristic function $\varphi(t) = e^{-t^2/2}$. By Theorem 15.12(i), this is $\mathcal{N}_{0,1}$. The additional claim follows by the Portemanteau theorem (Theorem 13.16) since $\mathcal{N}_{0,1}$ has a density; hence $\mathcal{N}_{0,1}(\partial[a,b]) = 0$. □

Remark 15.38 If we prefer to avoid the continuity theorem, we could argue as follows: For every $K > 0$ and $n \in \mathbb{N}$, we have $\mathbf{P}[|S_n^*| > K] \le \mathbf{Var}[S_n^*]/K^2 = 1/K^2$; hence the sequence $(\mathbf{P}_{S_n^*})$ is tight. As characteristic functions determine distributions, the claim follows by Theorem 13.34. ◊

We want to weaken the assumption in Theorem 15.37 that the random variables are identically distributed. In fact, we can even take a different set of summands for every n. The essential assumptions are that the summands are independent, each summand contributes only a little to the sum and the sum is centered and has variance 1.

Definition 15.39 For every $n \in \mathbb{N}$, let $k_n \in \mathbb{N}$ and let $X_{n,1}, \ldots, X_{n,k_n}$ be real random variables. We say that $(X_{n,l}) = (X_{n,l}, l = 1, \ldots, k_n, n \in \mathbb{N})$ is an *array of random variables*. Its row sum is denoted by $S_n = X_{n,1} + \ldots + X_{n,k_n}$. The array is called

- *independent* if, for every $n \in \mathbb{N}$, the family $(X_{n,l})_{l=1,\ldots,k_n}$ is independent,
- *centered* if $X_{n,l} \in \mathcal{L}^1(\mathbf{P})$ and $\mathbf{E}[X_{n,l}] = 0$ for all n and l, and
- *normed* if $X_{n,l} \in \mathcal{L}^2(\mathbf{P})$ and $\sum_{l=1}^{k_n} \mathbf{Var}[X_{n,l}] = 1$ for all $n \in \mathbb{N}$.

A centered array is called a *null array* if its individual components are asymptotically negligible in the sense that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \max_{1\le l\le k_n} \mathbf{P}\big[|X_{n,l}| > \varepsilon\big] = 0.$$

Definition 15.40 A centered array of random variables $(X_{n,l})$ with $X_{n,l} \in \mathcal{L}^2(\mathbf{P})$ for every $n \in \mathbb{N}$ and $l = 1, \ldots, k_n$ is said to satisfy the *Lindeberg condition* if, for all $\varepsilon > 0$,

$$L_n(\varepsilon) := \frac{1}{\mathbf{Var}[S_n]} \sum_{l=1}^{k_n} \mathbf{E}\big[X_{n,l}^2 \mathbb{1}_{\{X_{n,l}^2 > \varepsilon^2 \mathbf{Var}[S_n]\}}\big] \overset{n\to\infty}{\longrightarrow} 0. \qquad (15.6)$$

The array fulfills the *Lyapunov condition* if there exists a $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{\mathbf{Var}[S_n]^{1+(\delta/2)}} \sum_{l=1}^{k_n} \mathbf{E}\big[|X_{n,l}|^{2+\delta}\big] = 0. \qquad (15.7)$$

Lemma 15.41 *The Lyapunov condition implies the Lindeberg condition.*

Proof For $x \in \mathbb{R}$, we have $x^2 \mathbb{1}_{\{|x|>\varepsilon'\}} \le (\varepsilon')^{-\delta}|x|^{2+\delta}\mathbb{1}_{\{|x|>\varepsilon'\}} \le (\varepsilon')^{-\delta}|x|^{2+\delta}$. Letting $\varepsilon' := \varepsilon\sqrt{\mathbf{Var}[S_n]}$, we get

$$L_n(\varepsilon) \le \varepsilon^{-\delta} \frac{1}{\mathbf{Var}[S_n]^{1+(\delta/2)}} \sum_{l=1}^{k_n} \mathbf{E}\big[|X_{n,l}|^{2+\delta}\big].$$

□

Example 15.42 Let $(Y_n)_{n\in\mathbb{N}}$ be i.i.d. with $\mathbf{E}[Y_n] = 0$ and $\mathbf{Var}[Y_n] = 1$. Let $k_n = n$ and $X_{n,l} = \frac{Y_l}{\sqrt{n}}$. Then $(X_{n,l})$ is independent, centered and normed. Clearly, $\mathbf{P}[|X_{n,l}| > \varepsilon] = \mathbf{P}[|Y_1| > \sqrt{\varepsilon n}] \overset{n\to\infty}{\longrightarrow} 0$; hence $(X_{n,l})$ is a null array. Furthermore, $L_n(\varepsilon) = \mathbf{E}[Y_1^2 \mathbb{1}_{\{|Y_1|>\varepsilon\sqrt{n}\}}] \overset{n\to\infty}{\longrightarrow} 0$; hence $(X_{n,l})$ satisfies the Lindeberg condition.

If $Y_1 \in \mathcal{L}^{2+\delta}(\mathbf{P})$ for some $\delta > 0$, then

$$\sum_{l=1}^{n} \mathbf{E}\big[|X_{n,l}|^{2+\delta}\big] = n^{-(\delta/2)}\mathbf{E}\big[|Y_1|^{2+\delta}\big] \overset{n\to\infty}{\longrightarrow} 0.$$

In this case, $(X_{n,l})$ also satisfies the Lyapunov condition. ◊

The following theorem is due to Lindeberg (1922, see [108]) for the implication (i)⟹(ii) and is attributed to Feller (1935 and 1937, see [51, 52]) for the converse implication (ii)⟹(i). As most applications only need (i)⟹(ii), we only prove that implication. For a proof of (ii)⟹(i) see, e.g., [154, Theorem III.4.3].

Theorem 15.43 (Central limit theorem of Lindeberg–Feller) *Let $(X_{n,l})$ be an independent centered and normed array of real random variables. For every $n \in \mathbb{N}$, let $S_n = X_{n,1} + \ldots + X_{n,k_n}$. Then the following are equivalent.*

(i) *The Lindeberg condition holds.*
(ii) *$(X_{n,l})$ is a null array and $\mathbf{P}_{S_n} \overset{n\to\infty}{\longrightarrow} \mathcal{N}_{0,1}$.*

We prepare for the proof of Lindeberg's theorem with a couple of lemmas.

Lemma 15.44 *If* (i) *of Theorem* 15.43 *holds, then* $(X_{n,l})$ *is a null array.*

Proof For $\varepsilon > 0$, by Chebyshev's inequality,

$$\sum_{l=1}^{k_n} \mathbf{P}[|X_{n,l}| > \varepsilon] \le \varepsilon^{-2} \sum_{l=1}^{k_n} \mathbf{E}\big[X_{n,l}^2 \mathbb{1}_{\{|X_{n,l}|>\varepsilon\}}\big] = L_n(\varepsilon) \stackrel{n\to\infty}{\longrightarrow} 0.$$

□

In the following, $\varphi_{n,l}$ and φ_n will always denote the characteristic functions of $X_{n,l}$ and S_n.

Lemma 15.45 *For every* $n \in \mathbb{N}$ *and* $t \in \mathbb{R}$, *we have*

$$\sum_{l=1}^{k_n} \big|1 - \varphi_{n,l}(t)\big| \le \frac{t^2}{2}.$$

Proof For every $x \in \mathbb{R}$, we have $|e^{itx} - 1 - itx| \le \frac{t^2x^2}{2}$. Since $\mathbf{E}[X_{n,l}] = 0$,

$$\begin{aligned}
\sum_{l=1}^{k_n} \big|\varphi_{n,l}(t) - 1\big| &= \sum_{l=1}^{k_n} \big|\mathbf{E}\big[e^{itX_{n,l}} - 1\big]\big| \\
&\le \sum_{l=1}^{k_n} \mathbf{E}\big[\big|e^{itX_{n,l}} - itX_{n,l} - 1\big|\big] + \big|\mathbf{E}[itX_{n,l}]\big| \\
&\le \sum_{l=1}^{k_n} \frac{t^2}{2} \mathbf{E}\big[X_{n,l}^2\big] = \frac{t^2}{2}.
\end{aligned}$$

□

Lemma 15.46 *If* (i) *of Theorem* 15.43 *holds, then*

$$\lim_{n\to\infty} \left| \log \varphi_n(t) - \sum_{l=1}^{k_n} \mathbf{E}\big[e^{itX_{n,l}} - 1\big] \right| = 0.$$

Proof Let $m_n := \max_{l=1,\ldots,k_n} |\varphi_{n,l}(t) - 1|$. Note that, for all $\varepsilon > 0$,

$$\big|e^{itx} - 1\big| \le \begin{cases} 2x^2/\varepsilon^2, & \text{if } |x| > \varepsilon, \\ \varepsilon t, & \text{if } |x| \le \varepsilon. \end{cases}$$

This implies

$$\begin{aligned}
\big|\varphi_{n,l}(t) - 1\big| &\le \mathbf{E}\big[\big|e^{itX_{n,l}} - 1\big| \mathbb{1}_{\{|X_{n,l}|\le\varepsilon\}}\big] + \mathbf{E}\big[\big|e^{itX_{n,l}} - 1\big| \mathbb{1}_{\{|X_{n,l}|>\varepsilon\}}\big] \\
&\le \varepsilon t + 2\varepsilon^{-2} \mathbf{E}\big[X_{n,l}^2 \mathbb{1}_{\{|X_{n,l}|>\varepsilon\}}\big].
\end{aligned}$$

Hence, for all $\varepsilon > 0$,

$$\limsup_{n\to\infty} m_n \le \limsup_{n\to\infty}\big(\varepsilon t + 2\varepsilon^{-2} L_n(\varepsilon)\big) = \varepsilon t,$$

and thus $\lim_{n\to\infty} m_n = 0$. Now $|\log(1+x) - x| \le x^2$ for all $x \in \mathbb{C}$ with $|x| \le \frac{1}{2}$. If n is sufficiently large that $m_n < \frac{1}{2}$, then

$$\begin{aligned}\left|\log\varphi_n(t) - \sum_{l=1}^{k_n}\mathbf{E}\big[e^{itX_{n,l}} - 1\big]\right| &= \left|\sum_{l=1}^{k_n}\log\big(\varphi_{n,l}(t)\big) - \mathbf{E}\big[e^{itX_{n,l}} - 1\big]\right| \\ &\le \sum_{l=1}^{k_n}\big(\varphi_{n,l}(t) - 1\big)^2 \\ &\le m_n \sum_{l=1}^{k_n}\big|\varphi_{n,l}(t) - 1\big| \\ &\le \frac{1}{2} m_n t^2 \quad \text{(by Lemma 15.45)} \\ &\longrightarrow 0 \quad \text{for } n \to \infty. \end{aligned}$$
□

The fundamental trick of the proof, which is worth remembering, consists in the introduction of the function

$$f_t(x) := \begin{cases} \frac{1+x^2}{x^2}\big(e^{itx} - 1 - \frac{itx}{1+x^2}\big), & \text{if } x \ne 0, \\ -\frac{t^2}{2}, & \text{if } x = 0, \end{cases} \tag{15.8}$$

and the measures $\mu_n, \nu_n \in \mathcal{M}_f(\mathbb{R})$, $n \in \mathbb{N}$,

$$\nu_n(dx) := \sum_{l=1}^{k_n} x^2 \mathbf{P}_{X_{n,l}}(dx) \quad \text{and} \quad \mu_n(dx) := \sum_{l=1}^{k_n} \frac{x^2}{1+x^2}\mathbf{P}_{X_{n,l}}(dx).$$

Lemma 15.47 *For every $t \in \mathbb{R}$, we have $f_t \in C_b(\mathbb{R})$.*

Proof For all $|x| \ge 1$, we have $\frac{1+x^2}{x^2} \le 2$; hence

$$\big|f_t(x)\big| \le 2\left(\big|e^{itx}\big| + 1 + \frac{tx}{1+x^2}\right) \le 4 + 2|t|.$$

We have to show that f_t is continuous at 0. By Taylor's formula (Lemma 15.30), we get

$$e^{itx} = 1 + itx - \frac{t^2x^2}{2} + R(tx),$$

where the error term is bounded by $|R(tx)| \le \frac{1}{6}|tx|^3$. Hence, for fixed t,

$$\lim_{0\neq x\to 0} f_t(x) = \lim_{0\neq x\to 0} \frac{1}{x^2}\left(itx\left(1-\frac{1}{1+x^2}\right) - \frac{t^2x^2}{2} + R(tx)\right) = -\frac{t^2}{2}. \qquad \square$$

Lemma 15.48 *If* (i) *of Theorem* 15.43 *holds, then* $\nu_n \overset{n\to\infty}{\longrightarrow} \delta_0$ *weakly.*

Proof For every $n \in \mathbb{N}$, we have $\nu_n \in \mathcal{M}_1(\mathbb{R})$ since

$$\nu_n(\mathbb{R}) = \sum_{l=1}^{k_n} \int x^2 \mathbf{P}_{X_{n,l}}(dx) = \sum_{l=1}^{k_n} \mathbf{Var}[X_{n,l}] = 1.$$

However, for $\varepsilon > 0$, we have $\nu_n((-\varepsilon,\varepsilon)^c) = L_n(\varepsilon) \overset{n\to\infty}{\longrightarrow} 0$; hence $\nu_n \overset{n\to\infty}{\longrightarrow} \delta_0$. $\square$

Lemma 15.49 *If* (i) *of Theorem* 15.43 *holds, then*

$$\int f_t(x)\mu_n(dx) + it\int \frac{1}{x}\mu_n(dx) \overset{n\to\infty}{\longrightarrow} -\frac{t^2}{2}.$$

Proof Since $(x \mapsto f_t(x)/(1+x^2)) \in C_b(\mathbb{R})$, by Lemma 15.48,

$$\int f_t(x)\mu_n(dx) = \int f_t(x)\frac{1}{1+x^2}\nu_n(dx) \overset{n\to\infty}{\longrightarrow} f_t(0) = -\frac{t^2}{2}.$$

Now $(x \mapsto x/(1+x^2)) \in C_b(\mathbb{R})$ and $\mathbf{E}[X_{n,l}] = 0$ for all n and l; hence

$$\begin{aligned}\int \frac{1}{x}\mu_n(dx) &= \sum_{l=1}^{k_n} \mathbf{E}\left[\frac{X_{n,l}}{1+X_{n,l}^2}\right] = \sum_{l=1}^{k_n} \mathbf{E}\left[\frac{X_{n,l}}{1+X_{n,l}^2} - X_{n,l}\right] \\ &= -\sum_{l=1}^{k_n} \mathbf{E}\left[X_{n,l}^2 \cdot \frac{X_{n,l}}{1+X_{n,l}^2}\right] \\ &= -\int \frac{x}{1+x^2}\nu_n(dx) \overset{n\to\infty}{\longrightarrow} 0. \qquad \square\end{aligned}$$

Proof of Theorem 15.43 "(i)$\Longrightarrow$(ii)" We have to show that $\lim_{n\to\infty} \log\varphi_n(t) = -\frac{t^2}{2}$ for every $t \in \mathbb{R}$. By Lemma 15.46, this is equivalent to

$$\lim_{n\to\infty} \sum_{l=1}^{k_n} \big(\varphi_{n,l}(t) - 1\big) = -\frac{t^2}{2}.$$

Now $f_t(x)\frac{x^2}{1+x^2} = e^{itx} - 1 - \frac{itx}{1+x^2}$. Hence

$$\sum_{l=1}^{k_n}\big(\varphi_{n,l}(t)-1\big) = \sum_{l=1}^{k_n}\int\left(f_t(x)\frac{x^2}{1+x^2}+\frac{itx}{1+x^2}\right)\mathbf{P}_{X_{n,l}}(dx)$$
$$= \int f_t\,d\mu_n + it\int \frac{1}{x}\mu_n(dx)$$
$$\overset{n\to\infty}{\longrightarrow} -\frac{t^2}{2} \quad \text{(by Lemma 15.49).} \qquad \square$$

As an application of the Lindeberg–Feller theorem, we give the so-called *three-series theorem*, which is due to Kolmogorov.

Theorem 15.50 (Kolmogorov's three-series theorem) *Let $X_1, X_2, \ldots$ be independent real random variables. Let $K > 0$ and $Y_n := X_n \mathbb{1}_{\{|X_n|\le K\}}$ for all $n \in \mathbb{N}$. The series $\sum_{n=1}^{\infty} X_n$ converges almost surely if and only if each of the following three conditions holds*:

(i) $\sum_{n=1}^{\infty} \mathbf{P}[|X_n| > K] < \infty$.
(ii) $\sum_{n=1}^{\infty} \mathbf{E}[Y_n]$ *converges*.
(iii) $\sum_{n=1}^{\infty} \mathbf{Var}[Y_n] < \infty$.

Proof "$\Longleftarrow$" Assume that (i), (ii) and (iii) hold. By Exercise 7.1.1, since (iii) holds, the series $\sum_{n=1}^{\infty}(Y_n - \mathbf{E}[Y_n])$ converges a.s. As (ii) holds, $\sum_{n=1}^{\infty} Y_n$ converges almost surely. By the Borel–Cantelli lemma, there exists an $N = N(\omega)$ such that $|X_n| < K$; hence $X_n = Y_n$ for all $n \ge N$. Hence $\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{N-1} X_n + \sum_{n=N}^{\infty} Y_n$ converges a.s.

"$\Longrightarrow$" Assume that $\sum_{n=1}^{\infty} X_n$ converges a.s. Clearly, this implies (i) (otherwise, by the Borel–Cantelli lemma, $|X_n| > K$ infinitely often, contradicting the assumption).

We assume that (iii) does not hold and produce a contradiction. To this end, let $\sigma_n^2 = \sum_{k=1}^{n} \mathbf{Var}[Y_k]$ and define an array $(X_{n,l};\, l = 1,\ldots,n,\, n \in \mathbb{N})$ by $X_{n,l} = (Y_l - \mathbf{E}[Y_l])/\sigma_n$. This array is centered and normed. Since $\sigma_n^2 \overset{n\to\infty}{\longrightarrow} \infty$, for every $\varepsilon > 0$ and for sufficiently large $n \in \mathbb{N}$, we have $2K < \varepsilon\sigma_n$; thus $|X_{n,l}| \le \varepsilon$ for all $l = 1,\ldots,n$. This implies $L_n(\varepsilon) \overset{n\to\infty}{\longrightarrow} 0$, where $L_n(\varepsilon) = \sum_{l=1}^{n} \mathbf{E}[X_{n,l}^2 \mathbb{1}_{\{|X_{n,l}|\ge\varepsilon\}}]$ is the quantity of the Lindeberg condition (see (15.6)). By the Lindeberg–Feller theorem, we then get $S_n := X_{n,1} + \ldots + X_{n,n} \overset{n\to\infty}{\Longrightarrow} \mathcal{N}_{0,1}$. As shown in the first part of this proof, almost sure convergence of $\sum_{n=1}^{\infty} X_n$ and (i) imply that

$$\sum_{n=1}^{\infty} Y_n \quad \text{converges almost surely.} \tag{15.9}$$

In particular, $T_n := (Y_1 + \ldots + Y_n)/\sigma_n \overset{n\to\infty}{\Longrightarrow} 0$. Thus, by Slutzky's theorem, we also have $(S_n - T_n) \overset{n\to\infty}{\Longrightarrow} \mathcal{N}_{0,1}$. On the other hand, for all $n \in \mathbb{N}$, the difference $S_n - T_n$ is deterministic, contradicting the assumption that (iii) does not hold.

Now that we have established (iii), by Exercise 7.1.1, we see that $\sum_{n=1}^{\infty}(Y_n - \mathbf{E}[Y_n])$ converges almost surely. Together with (15.9), we conclude (ii). □

As a supplement, we cite a statement about the speed of convergence in the central limit theorem (see, e.g., [154, Chapter III, Section 11] for a proof). With different bounds (instead of 0.8), the statement was found independently by Berry [10] and Esseen [46].

Theorem 15.51 (Berry–Esseen) *Let $X_1, X_2, \ldots$ be independent and identically distributed with $\mathbf{E}[X_1] = 0$, $\mathbf{E}[X_1^2] = \sigma^2 \in (0, \infty)$ and $\gamma := \mathbf{E}[|X_1|^3] < \infty$. Let $S_n^* := \frac{1}{\sqrt{n\sigma^2}}(X_1 + \ldots + X_n)$ and let $\Phi : x \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$ be the distribution function of the standard normal distribution. Then, for all $n \in \mathbb{N}$,*

$$\sup_{x \in \mathbb{R}} \left| \mathbf{P}\left[S_n^* \le x\right] - \Phi(x) \right| \le \frac{0.8\gamma}{\sigma^3 \sqrt{n}}.$$

Example 15.52 Let $\alpha \in (0, 1)$. Consider the distribution μ_α on $\mathbb{R}$ with density

$$f_\alpha(x) = \frac{1}{2\alpha} |x|^{-1-1/\alpha} \mathbb{1}_{\{|x| \ge 1\}}.$$

Let $X_1, X_2, \ldots,$ be i.i.d. random variables with distribution μ_α. Then $\mathbf{E}[X_1] = 0$ and $\sigma^2 := \mathbf{Var}[X_1] = 1/(1 - 2\alpha) < \infty$ if $\alpha < 1/2$. Let F_n denote the distribution function of S_n^* and F_Φ the distribution function of the standard normal distribution.

The closer F_n and F_Φ are, the closer lie the points $(F_\Phi^{-1}(t), F_n^{-1}(t))$ on the diagonal $\{(x, x) : x \in \mathbb{R}\}$. A graphical representation of the points $(F_\Phi^{-1}(t), F_n^{-1}(t))$, $t \in \mathbb{R}$ is called *Q-Q-plot* or quantile-quantile-plot.

As α approaches 1/2, the distribution μ_α has less and less moments. Hence we expect the convergence in the central limit theorem to be slower. For fixed n, we expect the deviation of F_n from F_Φ to be larger for larger α. The graphs in Fig. 15.2 illustrate this. ◊

Exercise 15.5.1 The argument of Remark 15.38 is more direct than the argument with Lévy's continuity theorem but is less robust: Give a sequence $X_1, X_2, \ldots$ of independent real random variables with $\mathbf{E}[|X_n|] = \infty$ for all $n \in \mathbb{N}$ but such that

$$\frac{X_1 + \ldots + X_n}{\sqrt{n}} \overset{n \to \infty}{\Longrightarrow} \mathcal{N}_{0,1}.$$

Exercise 15.5.2 Let $Y_1, Y_2, \ldots$ be i.i.d. with $\mathbf{E}[Y_i] = 0$ and $\mathbf{E}[Y_i^2] = 1$. Let $Z_1, Z_2, \ldots$ be independent random variables (and independent of $Y_1, Y_2, \ldots$) with

$$\mathbf{P}[Z_i = i] = \mathbf{P}[Z_i = -i] = \frac{1}{2}\big(1 - \mathbf{P}[Z_i = 0]\big) = \frac{1}{2}\frac{1}{i^2}.$$

For $i, n \in \mathbb{N}$, define $X_i := Y_i + Z_i$ and $S_n = X_1 + \ldots + X_n$.

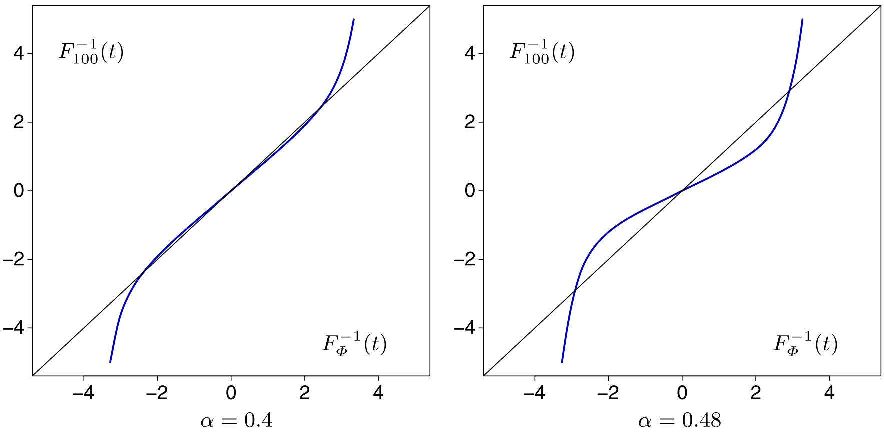

Fig. 15.2 Q-Q-plots for S^*_{100} from Example 15.52 with $\alpha = 0.4$ (*left*) and $\alpha = 0.48$ (*right*). The *abscissa* shows the quantiles of the standard normal distribution. For convenience, also the *diagonal* is drawn

Show that $n^{-1/2} S_n \overset{n\to\infty}{\Longrightarrow} \mathcal{N}_{0,1}$ but that $(X_i)_{i\in\mathbb{N}}$ does not satisfy the Lindeberg condition.

Hint: Do not try a direct computation!

Exercise 15.5.3 Let $X_1, X_2, \ldots$ be i.i.d. random variables with density

$$f(x) = \frac{1}{|x|^3} \mathbb{1}_{\mathbb{R}\setminus[-1,1]}(x).$$

Then $\mathbf{E}[X_1^2] = \infty$ but there are numbers $A_1, A_2, \ldots$, such that

$$\frac{X_1 + \ldots + X_n}{A_n} \quad \overset{n\to\infty}{\Longrightarrow} \quad \mathcal{N}_{0,1}.$$

Determine one such sequence $(A_n)_{n\in\mathbb{N}}$ explicitly.

15.6 Multidimensional Central Limit Theorem

We come to a multidimensional version of the CLT.

Definition 15.53 Let C be a (strictly) positive definite symmetric real $d \times d$ matrix and let $\mu \in \mathbb{R}^d$. A random vector $X = (X_1, \ldots, X_d)^T$ is called *d-dimensional normally distributed* with expectation μ and covariance matrix C if X has the density

$$f_{\mu,C}(x) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\left(-\frac{1}{2}\left\langle x-\mu, C^{-1}(x-\mu)\right\rangle\right) \qquad (15.10)$$

for $x \in \mathbb{R}^d$. In this case, we write $X \sim \mathcal{N}_{\mu,C}$.

Theorem 15.54 *Let $\mu \in \mathbb{R}^d$ and let C be a real positive definite symmetric $d \times d$ matrix. If $X \sim \mathcal{N}_{\mu,C}$, then the following statements hold.*

(i) $\mathbf{E}[X_i] = \mu_i$ *for all* $i = 1, \ldots, d$.
(ii) $\mathbf{Cov}[X_i, X_j] = C_{i,j}$ *for all* $i, j = 1, \ldots, d$.
(iii) $\langle \lambda, X\rangle \sim \mathcal{N}_{\langle \lambda, \mu\rangle, \langle \lambda, C\lambda\rangle}$ *for every* $\lambda \in \mathbb{R}^d$.
(iv) $\varphi(t) := \mathbf{E}[e^{i\langle t, X\rangle}] = e^{i\langle t,\mu\rangle} e^{-\frac{1}{2}\langle t, Ct\rangle}$ *for every* $t \in \mathbb{R}^d$.

Moreover, $X \sim \mathcal{N}_{\mu,C} \iff$ (iii) $\iff$ (iv).

Proof (i) and (ii) follow by simple computations. The same is true for (iii) and (iv). The implication (iii)$\Longrightarrow$(iv) is straightforward. The family

$$\left\{ f_t : x \mapsto e^{i\langle t, x\rangle}, t \in \mathbb{R}^d \right\}$$

is a separating class for $\mathcal{M}_1(\mathbb{R}^d)$ by the Stone–Weierstraß theorem. Hence φ determines the distribution of X uniquely. □

Remark 15.55 For one-dimensional normal distributions, it is natural to define the degenerate normal distribution by $\mathcal{N}_{\mu,0} := \delta_\mu$. For the multidimensional situation, there are various possibilities for degeneracy depending on the size of the kernel of C. If C is only positive semidefinite (and symmetric, of course), we define $\mathcal{N}_{\mu,C}$ as that distribution on $\mathbb{R}^n$ with characteristic function $\varphi(t) = e^{i\langle t,\mu\rangle} e^{-\frac{1}{2}\langle t, Ct\rangle}$. ◊

Theorem 15.56 (Cramér–Wold device) *Let $X_n = (X_{n,1}, \ldots, X_{n,d})^T \in \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors. Then, the following are equivalent*:

(i) *There is a random vector X such that* $X_n \overset{n\to\infty}{\Longrightarrow} X$.
(ii) *For any $\lambda \in \mathbb{R}^d$, there is a random variable X^λ such that* $\langle \lambda, X_n\rangle \overset{n\to\infty}{\Longrightarrow} X^\lambda$.

If (i) *and* (ii) *hold, then* $X^\lambda \overset{\mathcal{D}}{=} \langle \lambda, X\rangle$ *for all* $\lambda \in \mathbb{R}^d$.

Proof Assume (i). Let $\lambda \in \mathbb{R}^d$ and $s \in \mathbb{R}$. The map $\mathbb{R}^d \to \mathbb{C}$, $x \mapsto e^{is\langle \lambda, x\rangle}$ is continuous and bounded; hence we have $\mathbf{E}[e^{is\langle\lambda, X_n\rangle}] \overset{n\to\infty}{\longrightarrow} \mathbf{E}[e^{is\langle \lambda, X_\infty\rangle}]$. Thus (ii) holds with $X^\lambda := \langle \lambda, X\rangle$.

Now assume (ii). Then $(\mathbf{P}_{X_{n,l}})_{n\in\mathbb{N}}$ is tight for every $l = 1, \ldots, d$. Hence $(\mathbf{P}_{X_n})_{n\in\mathbb{N}}$ is tight and thus relatively sequentially compact (Prohorov's theorem). For any weak limit point Q for $(\mathbf{P}_{X_n})_{n\in\mathbb{N}}$ and for any $\lambda \in \mathbb{R}^d$, we have

$$\int Q(dx) e^{i\langle \lambda, x\rangle} = \mathbf{E}\big[e^{iX^\lambda}\big].$$

Hence the limit point Q is unique and thus $(\mathbf{P}_{X_n})_{n\in\mathbb{N}}$ converges weakly to Q. That is, (i) holds.

If (ii) holds, then the distributions of the limiting random variables X^λ are uniquely determined and by what we have shown already, $X^\lambda = \langle \lambda, X\rangle$ is one possible choice. Thus $X^\lambda \stackrel{\mathcal{D}}{=} \langle \lambda, X\rangle$. □

Theorem 15.57 (Central limit theorem in $\mathbb{R}^d$) *Let* $(X_n)_{n\in\mathbb{N}}$ *be i.i.d. random vectors with* $\mathbf{E}[X_{n,i}] = 0$ *and* $\mathbf{E}[X_{n,i}X_{n,j}] = C_{ij}$, $i, j = 1, \ldots, d$. *Let* $S_n^* := \frac{X_1+\ldots+X_n}{\sqrt{n}}$. *Then*

$$\mathbf{P}_{S_n^*} \stackrel{n\to\infty}{\longrightarrow} \mathcal{N}_{0,C} \quad \textit{weakly.}$$

Proof Let $\lambda \in \mathbb{R}^d$. Define $X_n^\lambda = \langle \lambda, X_n\rangle$, $S_n^\lambda = \langle \lambda, S_n^*\rangle$ and $S_\infty \sim \mathcal{N}_{0,C}$. Then $\mathbf{E}[X_n^\lambda] = 0$ and $\mathbf{Var}[X_n^\lambda] = \langle \lambda, C\lambda\rangle$. By the one-dimensional central limit theorem, we have $\mathbf{P}_{S_n^\lambda} \stackrel{n\to\infty}{\longrightarrow} \mathcal{N}_{0,\langle\lambda,C\lambda\rangle} = \mathbf{P}_{\langle\lambda,S_\infty\rangle}$. By Theorem 15.56, this yields the claim. □

Exercise 15.6.1 Let $\mu \in \mathbb{R}^d$, let C be a symmetric positive semidefinite real $d \times d$ matrix and let $X \sim \mathcal{N}_{\mu,C}$ (in the sense of Remark 15.55). Show that $AX \sim \mathcal{N}_{A\mu,ACA^T}$ for every $m \in \mathbb{N}$ and every real $m \times d$ matrix A.

Exercise 15.6.2 (Cholesky factorization) Let C be a positive definite symmetric real $d \times d$ matrix. Then there exists a real $d \times d$ matrix $A = (a_{kl})$ with $A \cdot A^T = C$. The matrix A can be chosen to be lower triangular. Let $W := (W_1, \ldots, W_d)^T$, where $W_1, \ldots, W_d$ are independent and $\mathcal{N}_{0,1}$-distributed. Define $X := AW + \mu$. Show that $X \sim \mathcal{N}_{\mu,C}$.