

Chapter 2

Fourier Theory of Probability Distributions

Characteristic functions, that is Fourier transforms of probability measures, play a major role in Probability theory, in particular in the Fourier theory of wide-sense stationary stochastic processes, whose starting point is the notion of power spectral measure. It turns out that the existence of such a measure is a direct consequence of Bochner's theorem of characterization of characteristic functions, and that the proof of its unicity is a rephrasing of Paul Lévy's inversion theorem. Another result of Paul Lévy, characterizing convergence in distribution in terms of characteristic functions, intervenes in an essential way in the proof of Bochner's theorem. In fact, characteristic functions are the link between the Fourier theory of deterministic functions and that of stochastic processes. This chapter could have been entitled "Convergence in distribution of random sequences", a classical topic of probability theory. However, we shall need to go slightly beyond this and give the extension of Paul Lévy's convergence theorem to sequences of finite measures (instead of probability distributions) as this is needed in Chap. 5 for the proof of existence of the Bartlett spectral measure.

2.1 Characteristic Functions

2.1.1 Basic Facts

Denote by $M^+(\mathbb{R}^d)$ the collection of *finite measures* on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Definition 2.1.1 The Fourier transform of a measure $\mu \in M^+(\mathbb{R}^d)$ is the function $\widehat{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$ defined by

$$\widehat{\mu}(v) = \int_{\mathbb{R}^d} e^{-2i\pi \langle v, x \rangle} \mu(dx),$$

where $\langle v, x \rangle := \sum_{j=1}^d v_j x_j$.

Theorem 2.1.1 *The Fourier transform of a measure $\mu \in M^+(\mathbb{R}^d)$ is bounded and uniformly continuous.*

Proof The proof is similar to that of Theorem 1.1.1. From the definition, we have that

$$\begin{aligned} |\widehat{\mu}(v)| &\leq \int_{\mathbb{R}^d} |e^{-2i\pi\langle v, x \rangle}| \mu(dx) \\ &= \int_{\mathbb{R}^d} \mu(dx) = \mu(\mathbb{R}^d), \end{aligned}$$

where the last term does not depend on v and is finite. Also, for all $h \in \mathbb{R}^d$,

$$\begin{aligned} |\widehat{\mu}(v+h) - \widehat{\mu}(v)| &\leq \int_{\mathbb{R}^d} \left| e^{-2i\pi\langle v, x+h \rangle} - e^{-2i\pi\langle v, x \rangle} \right| \mu(dx) \\ &= \int_{\mathbb{R}^d} \left| e^{-2i\pi\langle h, x \rangle} - 1 \right| \mu(dx). \end{aligned}$$

The last term is independent of v and tends to 0 as $h \rightarrow 0$ by dominated convergence (recall that μ is finite). \square

For future reference, we shall quote the following:

Theorem 2.1.2 *Let $\mu \in M^+(\mathbb{R}^d)$ and let \widehat{f} be the Fourier transform of $f \in L^1_{\mathbb{C}}(\mathbb{R}^d)$. Then*

$$\int_{\mathbb{R}^d} \widehat{f} d\mu = \int_{\mathbb{R}^d} f \widehat{\mu} dx.$$

Proof This follows from Fubini's theorem. In fact,

$$\begin{aligned} \int_{\mathbb{R}^d} \widehat{f} d\mu &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x) e^{-2i\pi\langle v, x \rangle} dx \right) \mu(dv) \\ &= \int_{\mathbb{R}^d} f(x) \left(\int_{\mathbb{R}^d} e^{-2i\pi\langle v, x \rangle} \mu(dv) \right) dx. \end{aligned}$$

(Interversion of the order of integration is justified by the fact that the function $(x, v) \rightarrow |f(x) e^{-2i\pi\langle v, x \rangle}| = |f(x)|$ is integrable with respect to the product measure $dx \times \mu(dv)$. Recall that μ is finite.) \square

Sampling Theorem for the FT of a Finite Measure

The following approach to sampling acknowledges the fact that a function is a combination of complex sinusoids, and therefore starts by obtaining the sampling theorem for this type of elementary functions.

Consider the function

$$f(t) = e^{2i\pi vt},$$

where $v \in \mathbb{R}$. This function is not integrable and therefore does not fit into the framework of Shannon's sampling theorem. However, the Shannon–Nyquist formula remains essentially true.

Theorem 2.1.3 *For all $t \in \mathbb{R}$ and all $v \in (-W, +W)$:*

$$e^{2i\pi vt} = \sum_{n \in \mathbb{Z}} e^{2i\pi n 2Wv} \frac{\sin(2\pi W(t - \frac{n}{2W}))}{2\pi W(t - \frac{n}{2W})}. \quad (2.1)$$

For all $B < W$, the convergence is uniform in $v \in [-B, +B]$.

Proof We first prove that for all $v \in \mathbb{R}$ and all $t \in (-W, +W)$

$$e^{2i\pi vt} = \sum_{n \in \mathbb{Z}} e^{2i\pi n 2Wt} \frac{\sin(2\pi W(v - \frac{n}{2W}))}{2\pi W(v - \frac{n}{2W})}, \quad (\dagger)$$

where the series converges uniformly for all $t \in [-B, +B]$ for any $B < W$. The result then follows by exchanging the roles of t and v .

Let $g : \mathbb{R} \rightarrow \mathbb{C}$, whose argument is denoted by t , be the $2W$ -periodic function which is equal to $e^{2i\pi vt}$ on $(-W, +W]$. The series in (\dagger) is the Fourier series of g . We must therefore show uniform pointwise convergence of this Fourier series to the original function, which was done in Example 1.1.11. \square

In particular, if f is a finite linear combination of complex trigonometric functions, that is

$$f(t) = \sum_{k=1}^M \gamma_k e^{2i\pi v_k t},$$

where $\gamma_k \in \mathbb{C}$, $v_k \in \mathbb{R}$, and if W satisfies the condition

$$W > \sup\{|v_k| : 1 \leq k \leq M\}, \quad (2.2)$$

we have the Shannon–Nyquist reconstruction formula, with $T = \frac{1}{2W}$,

$$f(t) = \sum_{n \in \mathbb{Z}} f(nT) \frac{\sin(\frac{\pi}{T}(t - nT))}{\frac{\pi}{T}(t - nT)}. \quad (2.3)$$

Exercise 2.4.1 shows that one really needs the strict inequality in (2.2).

The following extension of the sampling theorem for sinusoids is now straightforward:

Theorem 2.1.4 *Let μ be a finite measure on $[-B, +B]$, where $B \in \mathbb{R}_+ \setminus \{0\}$, and define the function*

$$f(t) = \int_{[-B, +B]} e^{2i\pi vt} \mu(dv).$$

Then for any $T < 1/2B$ and for all $t \in \mathbb{R}$, we have the Shannon–Nyquist reconstruction formula (2.3).

Proof Since μ is finite and the convergence in (2.1) is uniform in $v \in [-B, +B]$,

$$\begin{aligned} f(t) &= \int_{[-B, +B]} \left\{ \sum_{n \in \mathbb{Z}} e^{2i\pi vnT} \frac{\sin\left(\frac{\pi}{T}(t - nT)\right)}{\frac{\pi}{T}(t - nT)} \right\} \mu(dv) \\ &= \sum_{n \in \mathbb{Z}} \left\{ \int_{[-B, +B]} e^{2i\pi vnT} \frac{\sin\left(\frac{\pi}{T}(t - nT)\right)}{\frac{\pi}{T}(t - nT)} \mu(dv) \right\} \\ &= \sum_{n \in \mathbb{Z}} \left\{ \int_{[-B, +B]} e^{2i\pi vnT} \mu(dv) \right\} \frac{\sin\left(\frac{\pi}{T}(t - nT)\right)}{\frac{\pi}{T}(t - nT)}. \quad \square \end{aligned}$$

2.1.2 Inversion Formula

The *characteristic function* of a random vector $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ is the function $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ defined by

$$\varphi_X(u) = E[e^{i\langle u, X \rangle}].$$

In other words, φ_X is the Fourier transform of the probability distribution of X .

It turns out that the characteristic function of a random vector determines uniquely its probability distribution. This result will be obtained as a consequence of *Paul Lévy's inversion formula*:

Theorem 2.1.5 *Let $X \in \mathbb{R}^d$ be a random vector with characteristic function φ . Then for all $1 \leq j \leq d$, all $a_j, b_j \in \mathbb{R}^d$ such that $a_j < b_j$,*

$$\begin{aligned}
& \lim_{c \uparrow +\infty} \frac{1}{(2\pi)^d} \int_{-c}^{+c} \cdots \int_{-c}^{+c} \left(\prod_{j=1}^d \frac{e^{-iu_j a_j} - e^{-iu_j b_j}}{iu_j} \right) \varphi(u_1, \dots, u_d) du_1 \dots du_d \\
&= E \left[\prod_{j=1}^d \left(\frac{1}{2} 1_{\{X_j = a_j \text{ or } b_j\}} + 1_{\{a_j < X_j < b_j\}} \right) \right].
\end{aligned}$$

Before proving the theorem, we give its main consequence:

Corollary 2.1.1 *The distribution of a random vector of \mathbb{R}^d is uniquely determined by its characteristic function.*

Proof Let X and Y be two vectors of \mathbb{R}^d with the same characteristic function φ . Lévy's inversion formula shows that the distributions of X and Y agree on \mathcal{A} , the class of rectangles $A = \prod_{j=1}^d (a_j, b_j]$ whose boundary has a null measure with respect to the distributions of both X and Y . Since there are at most a countable number of rectangles whose boundary has positive measure with respect to the distributions of both X and Y , \mathcal{A} generates $\mathcal{B}(\mathbb{R}^d)$. Moreover, \mathcal{A} is a π -system and therefore, by Theorem A.2.1, the two distributions coincide. \square

Corollary 2.1.2 *A necessary and sufficient condition for the random variables X_1, \dots, X_d to be independent is that the characteristic function φ_X of the vector $X = (X_1, \dots, X_d)$ factorizes as*

$$\varphi_X(u_1, \dots, u_d) = \prod_{j=1}^d \varphi_j(u_j),$$

where for all $1 \leq j \leq d$, φ_j is a characteristic function. In this case, for all $1 \leq j \leq d$, $\varphi_j = \varphi_{X_j}$, the characteristic function of X_j .

Proof Necessity. Write

$$\begin{aligned}
\varphi_X(u) &= E \left[e^{i \sum_{j=1}^d u_j X_j} \right] \\
&= E \left[\prod_{j=1}^d e^{iu_j X_j} \right] = \prod_{j=1}^d E \left[e^{iu_j X_j} \right] = \prod_{j=1}^d \varphi_{X_j}(u_j),
\end{aligned}$$

by the product formula for expectations (Theorem A.2.8).

Sufficiency. Let $X' := (X'_1, \dots, X'_d) \in \mathbb{R}^d$ be a random vector whose *independent* coordinate random variables X'_1, \dots, X'_d have the respective characteristic functions $\varphi_1, \dots, \varphi_d$. The characteristic function of X' is $\prod_{j=1}^d \varphi_j(u_j)$, and therefore X and X' have the same distribution. In particular X_1, \dots, X_d are independent random variables with respective characteristic functions $\varphi_1, \dots, \varphi_d$. \square

We now return to the proof of Theorem 2.1.5.

Proof We do the proof in the univariate case for the sake of notational ease. The multivariate case is a straightforward adaptation of it. Let X be a real-valued random variable with cumulative distribution function F and characteristic function φ . We have to show that for any pair of points a, b with $(a < b)$,

$$\lim_{c \uparrow +\infty} \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) du = E \left[\left(\frac{1}{2} 1_{\{X=a \text{ or } b\}} + 1_{\{a < X < b\}} \right) \right]. \quad (*)$$

For this, write

$$\begin{aligned} \Phi_c &:= \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) du \\ &= \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} \left(\int_{-\infty}^{+\infty} e^{iux} dF(x) \right) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} e^{iux} du \right) dF(x) = \int_{-\infty}^{+\infty} \Psi_c(x) dF(x), \end{aligned}$$

where

$$\Psi_c(x) := \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} e^{-iux} du.$$

The above computations make use of Fubini's theorem. This is allowed since, observing that

$$\left| \frac{e^{-iua} - e^{-iub}}{iu} \right| = \left| \int_a^b e^{-iux} dx \right| \leq (b - a),$$

we have

$$\begin{aligned} \int_{-c}^{+c} \int_{-\infty}^{+\infty} \left| \frac{e^{-iua} - e^{-iub}}{iu} e^{iux} \right| dF(x) du &= \int_{-c}^{+c} \int_{-\infty}^{+\infty} \left| \frac{e^{-iua} - e^{-iub}}{iu} \right| dF(x) du \\ &\leq \int_{-c}^{+c} \int_{-\infty}^{+\infty} (b - a) dF(x) du \\ &= 2c(b - a) < \infty. \end{aligned}$$

Since the function $u \rightarrow \frac{\cos(au)}{u}$ is antisymmetric, $\int_{-c}^{+c} \frac{\cos(au)}{u} du = 0$, and therefore

$$\begin{aligned}\Psi_c(x) &= \frac{1}{2\pi} \int_{-c}^{+c} \frac{\sin u(x-a) - \sin u(x-b)}{u} du \\ &= \frac{1}{2\pi} \int_{-c(x-a)}^{+c(x-a)} \frac{\sin u}{u} du - \frac{1}{2\pi} \int_{-c(x-b)}^{+c(x-b)} \frac{\sin u}{u} du.\end{aligned}$$

The function $c \rightarrow \int_0^c \frac{\sin u}{u} du = \int_{-c}^0 \frac{\sin u}{u} du$ is uniformly continuous in c and tends to $\int_0^{+\infty} \frac{\sin u}{u} du = \frac{1}{2}\pi$ as $c \uparrow +\infty$. Therefore the function $(c, x) \rightarrow \Psi_c(x)$ is uniformly bounded. Moreover, in view of the above expression for Ψ_c ,

$$\begin{aligned}\lim_{c \uparrow \infty} \Psi_c(x) &:= \Psi(x) = 0 \quad \text{if } x < a \text{ or } x > b \\ &= \frac{1}{2} \quad \text{if } x = a \text{ or } x = b \\ &= 1 \quad \text{if } a < x < b.\end{aligned}$$

Therefore, by dominated convergence,

$$\begin{aligned}\lim_{c \uparrow \infty} \Phi_c &= \int_{-\infty}^{+\infty} \lim_{c \uparrow \infty} \Psi_c(x) dF(x) \\ &= \int_{-\infty}^{+\infty} \Psi(x) dF(x) = E \left[\left(\frac{1}{2} 1_{\{X=a \text{ or } b\}} + 1_{\{a < X < b\}} \right) \right].\end{aligned}$$

□

Note that, in the univariate case, denoting by F the cumulative distribution function of the random variable X ,

$$E \left[\left(\frac{1}{2} 1_{\{X=a \text{ or } b\}} + 1_{\{a < X < b\}} \right) \right] = \frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2},$$

so that formula (*) takes the perhaps more familiar form

$$\frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2} = \lim_{c \uparrow +\infty} \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) du.$$

Corollary 2.1.3 *If moreover φ is integrable, then X admits a probability density f given by*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}^d} \varphi(u) e^{-i\langle u, x \rangle} du.$$

Proof We do the proof in the univariate case (the extension to the multivariate case follows the same lines of argument). The function f is well-defined, continuous and bounded, and in particular integrable on finite intervals. We have (Fubini)

$$\begin{aligned} \int_a^b f(x) dx &= \int_a^b \left(\frac{1}{2\pi} \int_{\mathbb{R}} \varphi(u) e^{-iux} du \right) dx \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(u) \left[\int_a^b e^{-iux} dx \right] du \\ &= \lim_{c \uparrow \infty} \frac{1}{2\pi} \int_{-c}^{+c} \varphi(u) \left[\int_a^b e^{-iux} dx \right] du \\ &= \lim_{c \uparrow \infty} \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) du = F(b) - F(a) \end{aligned}$$

for all $a < b$ that are points of continuity of F , from which the result easily follows for any interval $[a, b]$. \square

Since a finite probability measure of $M^+(\mathbb{R}^d)$ is a multiple of a probability measure on \mathbb{R}^d , which is in turn the probability distribution of some random vector $X \in \mathbb{R}^d$, and since $\hat{\mu}(0)$ enables us to recover the total mass, the uniqueness theorem concerning characteristic functions gives:

Corollary 2.1.4 *A measure of $M^+(\mathbb{R}^d)$ is characterized by its Fourier transform.*

2.1.3 Gaussian Vectors

Gaussian variables and Gaussian vectors play an important role in probabilistic modeling. We decide to include them at this place in the book because they are defined in terms of characteristic functions, and their properties are best studied in these terms.

A Gaussian random variable with mean m and variance $\sigma^2 > 0$ is defined by its probability density

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

Its characteristic function, which can be derived from the result of Example 1.1.5, is:

$$\varphi_X(u) = \exp \left\{ i m u - \frac{1}{2} \sigma^2 u^2 \right\}. \quad (\dagger)$$

Definition 2.1.2 An *extended Gaussian variable* X is any real random variable with a characteristic function of the form (\dagger) , where $m \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$. A standard Gaussian variable is one for which $m = 0$ and $\sigma^2 = 1$.

Note that a null variance ($\sigma^2 = 0$) is allowed, and in this case, the random variable X is the constant m . This is precisely the extension we need.

Definition 2.1.3 A random vector $X \in \mathbb{R}^d$ is called a *Gaussian random vector* if for all $\alpha \in \mathbb{R}^d$, the random variable $\langle \alpha, X \rangle$ is an extended Gaussian random variable.

We now connect this definition with another one in terms of characteristic functions.

Theorem 2.1.6 For a random vector $X \in \mathbb{R}^d$ to be a Gaussian vector, it is necessary and sufficient that its characteristic function ϕ_X be of the form:

$$\varphi_X(u) = \exp \left\{ i \langle u, m \rangle - \frac{1}{2} \langle u, \Gamma u \rangle \right\}. \quad (2.4)$$

where $m \in \mathbb{R}^n$ and where Γ is a symmetric and non-negative definite $d \times d$ matrix.

In this case, we write $X \stackrel{\mathcal{D}}{\sim} \mathcal{N}(m, \Gamma)$.

Proof Necessity: The characteristic function of a Gaussian vector as defined in Definition 2.1.3 can be expressed as

$$E[e^{i \langle u, X \rangle}] = \varphi_Z(1),$$

where φ_Z is the characteristic function of $Z = \langle u, X \rangle$. The random variable Z being an extended Gaussian variable,

$$\varphi_Z(1) = \exp \left\{ i m_Z - \frac{1}{2} \sigma_Z^2 \right\},$$

where

$$m_Z := E[Z] = \langle u, E[X] \rangle = \langle u, m \rangle$$

and

$$\begin{aligned} \sigma_Z^2 &= E[\langle u, X - m \rangle \langle u, X - m \rangle] \\ &= \langle u, E[(X - m)(X - m)^T] u \rangle = \langle u, \Gamma u \rangle. \end{aligned}$$

Therefore, finally,

$$\varphi_X(u) = \exp \left\{ i \langle u, m \rangle - \frac{1}{2} \langle u, \Gamma u \rangle \right\}.$$

Sufficiency: Let X be a random vector with characteristic function given by (2.4). Let $Z = \langle \alpha, X \rangle$, where $\alpha \in \mathbb{R}^d$. The characteristic function of the random variable Z is

$$\begin{aligned} \varphi_Z(v) &= E[\exp\{i v Z\}] \\ &= E[\exp\{i v \langle \alpha, X \rangle\}] \\ &= \exp \left\{ i v (\langle \alpha, m \rangle) - \frac{1}{2} v^2 (\langle \alpha, \Gamma \alpha \rangle) \right\}. \end{aligned}$$

Therefore Z is an extended Gaussian random variable. \square

It remains to prove the existence of a random vector with the characteristic function (2.4):

Theorem 2.1.7 *Let Γ be a non-negative definite matrix $d \times d$ matrix and let $m \in \mathbb{R}^d$. There exists a vector $X \in \mathbb{R}^d$ with characteristic functions (2.4).*

Proof Since Γ is non-negative definite, there exists a matrix A of the same dimension and such that $\Gamma = A A^T$. Let $X = m + A Z$ where Z is a vector of independent standard Gaussian variables Z_1, \dots, Z_d . Then X has the required characteristic functions. In fact, the characteristic function of Z is

$$\begin{aligned} \varphi_Z(u_1, \dots, u_d) &= E \left[e^{i \sum_{j=1}^d u_j Z_j} \right] = \prod_{j=1}^d E \left[e^{i u_j Z_j} \right] \\ &= \prod_{j=1}^d \varphi_{Z_j}(u_j) = e^{-\frac{1}{2} \sum_{j=1}^d u_j^2} = e^{-\frac{1}{2} \|u\|^2}, \end{aligned}$$

and therefore

$$\begin{aligned} \varphi_X(u) &= E \left[e^{i \langle u, m + A Z \rangle} \right] = e^{i \langle u, m \rangle} \varphi_Z(u^T A) \\ &= e^{i \langle u, m \rangle} e^{-\frac{1}{2} \|u^T A\|^2} = e^{i \langle u, m \rangle - \frac{1}{2} \langle u, \Gamma u \rangle}. \end{aligned}$$

\square

It is clear from the above proof that the parameters m and Γ in (2.4) are respectively the mean and the covariance matrix of X .

In general, non-correlation does *not* imply independence. However, this is nearly true for Gaussian vectors. We start with a definition that is needed for a precise statement of the corresponding result.

Definition 2.1.4 Two real random vectors X and Y are said to be *jointly Gaussian* if the vector $Z = (X, Y)$ is Gaussian.

Theorem 2.1.8 *Two jointly Gaussian random vectors X and Y of arbitrary dimensions are independent if and only if they are uncorrelated (that is, if $\Gamma_{XY} := E[(X - m_X)(Y - m_Y)^T] = 0$).*

Proof Necessity: If X and Y are independent, then, by the product formula for expectations,

$$E[(X - m_X)(Y - m_Y)^T] = E[X - m_X]E[Y - m_Y]^T = 0.$$

Sufficiency: The random vector $Z = (X, Y)$ has the mean

$$m_Z = \begin{pmatrix} m_X \\ m_Y \end{pmatrix},$$

and if X and Y are uncorrelated, its covariance matrix is of the form

$$\Gamma_Z = \begin{pmatrix} \Gamma_X & 0 \\ 0 & \Gamma_Y \end{pmatrix}.$$

Vector Z is Gaussian by hypothesis and therefore, with $w = (u, v)$, u and v being real vectors of appropriate dimensions,

$$\begin{aligned} E[\exp\{i(u^T X + v^T Y)\}] &= E[\exp\{i w^T Z\}] \\ &= \exp\left\{i w^T m_Z - \frac{1}{2} w^T \Gamma_Z w\right\} \\ &= \exp\left\{i(u^T m_X + v^T m_Y) - \frac{1}{2} u^T \Gamma_X u - \frac{1}{2} v^T \Gamma_Y v\right\} \\ &= E[\exp\{i u^T X\}] E[\exp\{i v^T Y\}], \end{aligned}$$

and the conclusion follows from the characteristic function independence criterion □

Theorem 2.1.9 *Let X be a d -dimensional Gaussian vector with mean m and covariance matrix Γ , and assume that it is non-degenerate, that is:*

$$\langle u, \Gamma u \rangle = u^T \Gamma u = 0 \Rightarrow u = 0.$$

Then X admits the probability distribution function

$$f_X(x) = \frac{1}{(2\pi)^{d/2}(\det \Gamma)^{1/2}} \exp \left\{ -\frac{1}{2} \langle x - m, \Gamma^{-1}(x - m) \rangle \right\}. \quad (2.5)$$

Proof Since $\Gamma > 0$, there exists a non-singular matrix A of the same dimension as Γ and such that $\Gamma = AA^T$. Define the n -vector $Z = A^{-1}(X - m)$. According to Definition 2.1.3, it is a Gaussian vector, and furthermore $E[Z] = 0$ and

$$\Gamma_Z = A^{-1}\Gamma A^{-T} = A^{-1}AA^T A^{-T} = I,$$

where I is the identity matrix. Therefore

$$E[\exp\{iu^T Z\}] = \exp \left\{ -\frac{1}{2} \sum_{i=1}^d u_i^2 \right\}.$$

Since this is the characteristic function of a vector of independent standard (centered, variance 1) Gaussian variables, we can assert that Z_1, \dots, Z_d are independent standard Gaussian random variables Corollary 2.1.1. In particular, Z admits the probability distribution function

$$\begin{aligned} f_Z(z) &= \prod_{i=1}^d \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2/2} \\ &= \frac{1}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} \|z\|^2 \right\}. \end{aligned}$$

Now, $X = AZ + m$, and therefore, by the formula of smooth change of variables,

$$\begin{aligned} f_X(x) &= \frac{1}{|\det A|} f_Y(A^{-1}(x - m)) \\ &= \frac{1}{(\det \Gamma)^{1/2}} \frac{1}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} \|A^{-1}(x - m)\|^2 \right\}, \end{aligned}$$

and this is precisely the announced result since

$$\begin{aligned} \|A^{-1}(x - m)\|^2 &= \langle A^{-1}(x - m), A^{-1}(x - m) \rangle \\ &= \langle x - m, A^{-T} A^{-1}(x - m) \rangle \\ &= \langle x - m, \Gamma^{-1}(x - m) \rangle. \end{aligned}$$

□

2.2 Convergence in Distribution

2.2.1 Paul Lévy's Theorem

Recall that $C_b(\mathbb{R}^d)$ is the collection of bounded continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

Definition 2.2.1 (a) The sequence $\{\mu_n\}_{n \geq 1}$ in $M^+(\mathbb{R}^d)$ is said to converge *weakly* to μ if, for all $f \in C_b(\mathbb{R}^d)$, $\lim_{n \uparrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$. This is denoted by

$$\mu_n \xrightarrow{w} \mu.$$

(b) The sequence of random vectors $\{X_n\}_{n \geq 1}$ of \mathbb{R}^d , with respective probability distributions $\{Q_{X_n}\}_{n \geq 1}$, is said to converge *in distribution* to the random vector $X \in \mathbb{R}^d$ with distribution Q_X if $Q_{X_n} \xrightarrow{w} Q_X$. This is also denoted by

$$X_n \xrightarrow{\mathcal{D}} X.$$

Observe that the vectors X and X_n 's need not be defined on the same probability space. Convergence in distribution concerns only probability distributions. As a matter of fact, very often the X_n 's are defined on the same probability space but there is no “visible” (that is, defined on the same probability space) limit random vector X . Therefore we sometimes denote convergence in distribution as follows: $X_n \xrightarrow{\mathcal{D}} Q$, where Q is a probability distribution on \mathbb{R}^d . If Q is a “famous” probability distribution, for instance a standard Gaussian variable, we then say, that “ $\{X_n\}_{n \geq 1}$ converges in distribution to a standard Gaussian distribution”. We would then denote this by $X_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$.

Denote by B° and B^c respectively the interior and the closure of the set $B \in \mathbb{R}^d$, and by ∂B its boundary ($:= B^c \setminus B^\circ$).

Theorem 2.2.1 Let $\{\mu_n\}_{n \geq 1}$ and μ be probability distributions on \mathbb{R}^d . The following conditions are equivalent:

- (i) $\mu_n \xrightarrow{w} \mu$.
- (ii) For any open set $G \subseteq \mathbb{R}^d$, $\liminf_n \mu_n(G) \geq \mu(G)$.
- (iii) For any closed set $F \subseteq \mathbb{R}^d$, $\limsup_n \mu_n(F) \leq \mu(F)$.
- (iv) For any measurable set $B \subseteq \mathbb{R}^d$ such that $\mu(\partial B) = 0$, $\lim_n \mu_n(B) = \mu(B)$.

Proof (i) \Rightarrow (ii). For any open set $G \in \mathbb{R}^d$ there exists a non-decreasing sequence $\{\varphi_k\}_{k \geq 1}$ of non-negative functions of $C_b(\mathbb{R}^d)$ such that $0 \leq \varphi_k \leq 1$ and $\varphi_k \uparrow 1_G$ (for instance, $\varphi_k(x) = 1 - e^{-kd(x, \overline{G})}$). For all $k \geq 1$, $\int 1_G d\mu_n \geq \int \varphi_k d\mu_n$, and therefore

$$\liminf_n \mu_n(G) = \liminf_n \int 1_G d\mu_n \geq \liminf_n \int \varphi_k d\mu_n.$$

Since this is true for all $k \geq 1$,

$$\begin{aligned} \liminf_n \mu_n(G) &\geq \sup_k \left(\liminf_n \int \varphi_k d\mu_n \right) \\ &= \sup_k \left(\lim_n \int \varphi_k d\mu_n \right) \\ &= \sup_k \int \varphi_k d\mu = \mu(G). \end{aligned}$$

(ii) \Leftrightarrow (iii). Take complements.

(ii) + (iii) \Rightarrow (iv). Indeed, by (ii) and (iii),

$$\limsup_n \mu_n(B) \leq \limsup_n \mu_n(B^c) \leq \mu(B^c)$$

and

$$\liminf_n \mu_n(B) \geq \liminf_n \mu_n(B^o) \geq \mu(B^o).$$

But, since $\mu(\partial B) = 0$, $\mu(B^o) = \mu(B^c) = \mu(B)$, and therefore (iv) is verified.

(iv) \Rightarrow (i). Let $f \in C_b(\mathbb{R}^d)$. We must show that $\lim_{n \uparrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$. It is enough to show this for $f \geq 0$. Let $K < \infty$ be a bound of f . Write, using Fubini,

$$\begin{aligned} \int_{\mathbb{R}^d} f(x) d\mu(x) &= \int_{\mathbb{R}^d} \left(\int_0^K 1_{\{t \leq f(x)\}} dt \right) d\mu(x) \\ &= \int_0^K \mu(\{x; t \leq f(x)\}) dt = \int_0^K \mu(D_t^f) dt, \end{aligned}$$

where $D_t^f := \{x; t \leq f(x)\}$. Observe that $\partial D_t^f \subseteq \{x; t = f(x)\}$ and that the collection of positive t such that $\mu(\{x; t = f(x)\}) > 0$ is at most denumerable (for each positive integer k there are at most k values of t such that $\mu(\{x; t = f(x)\}) \geq \frac{1}{k}$). Therefore, by (iv), for almost all t (with respect to the Lebesgue measure),

$$\lim_n \mu_n(D_t^f) = \mu(D_t^f)$$

and, by dominated convergence,

$$\begin{aligned}\lim_n \int f d\mu_n &= \lim_n \int_0^K \mu_n(D_t^f) dt \\ &= \int_0^K \mu(D_t^f) dt = \int f d\mu.\end{aligned}$$

□

We now state *Lévy's characterization of convergence in distribution*. Its proof will be given later in Theorem 2.3.5.

Theorem 2.2.2 *A necessary and sufficient condition for the sequence $\{X_n\}_{n \geq 1}$ of random vectors of \mathbb{R}^d to converge in distribution is that the sequence of their characteristic functions $\{\varphi_n\}_{n \geq 1}$ converges to some function φ that is continuous at 0. In such a case, φ is the characteristic function of the limit probability distribution.*

The following equivalent formulations of convergence in distribution are then a consequence of Theorems 2.2.2 and 2.2.1.

Corollary 2.2.1 *Let $\{X_n\}_{n \geq 1}$ and X be random vectors of \mathbb{R}^d with respective characteristic functions $\{\varphi_n\}_{n \geq 1}$ and φ . The following three statements are equivalent.*

- (A) $X_n \xrightarrow{\mathcal{D}} X$.
- (B) For all continuous and bounded functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\lim_{n \uparrow \infty} E[f(X_n)] = E[f(X)].$$

- (C) $\lim_{n \uparrow \infty} \varphi_n = \varphi$.

Corollary 2.2.2 *In the univariate case, denote by F_n and F the cumulative distribution functions of X_n and X respectively. Call a point $x \in \mathbb{R}$ a continuity point of F if $F(x) = F(x_-)$. Then $X_n \xrightarrow{\mathcal{D}} X$ if and only*

$$\lim_n F_n(x) = F(x) \quad \text{for all continuity points } x \text{ of } F.$$

Proof Necessity. Let Q_X be the probability distribution of X . If x is a continuity point of F , the boundary of $C := (-\infty, x]$ is $\{x\}$ of null Q_X -measure. Therefore by (iv), $\lim_n Q_{X_n}((-\infty, x]) = Q_X((-\infty, x])$, that is $\lim_n F_n(x) = F(x)$.

Sufficiency. Let $f \in C_b(\mathbb{R})$ and let $M < \infty$ be an upper bound of f . For arbitrary $\varepsilon > 0$, there exists a subdivision $-\infty < a = x_0 < x_1 < \dots < x_k = b < +\infty$ formed by continuity points of F , such that $F(a) < \varepsilon$, $F(b) > 1 - \varepsilon$ and

$|f(x) - f(x_i)| < \varepsilon$ on $[x_{i-1}, x_i]$. By hypothesis,

$$S_n := \sum_{i=1}^k f(x_i)(F_n(x_i) - F_n(x_{i-1})) \rightarrow S := \sum_{i=1}^k f(x_i)(F(x_i) - F(x_{i-1}))$$

Also

$$|E[f(X)] - S| \leq \varepsilon + MF(a) + M(1 - F(b)) \leq (2M + 1)\varepsilon$$

and

$$\begin{aligned} |E[f(X_n)] - S_n| &\leq \varepsilon + MF_n(a) + M(1 - F_n(b)) \rightarrow \varepsilon + MF(a) + M(1 - F(b)) \\ &\leq (2M + 1)\varepsilon. \end{aligned} \quad (2.6)$$

Therefore,

$$\begin{aligned} \limsup_n |E[f(X_n)] - E[f(X)]| &\leq \limsup_n |E[f(X_n)] - S_n| + \limsup_n |S_n - S| + |E[f(X)] - S| \\ &\leq (4M + 2)\varepsilon. \end{aligned}$$

Since ε is arbitrary, $\lim_n |E[f(X_n)] - E[f(X)]| = 0$. □

Theorem 2.2.3 *Let $\{X_n\}_{n \geq 1}$ and $\{Y_n\}_{n \geq 1}$ be sequences of random vectors of \mathbb{R}^d such that $X_n \xrightarrow{\mathcal{D}} X$ and $d(X_n, Y_n) \xrightarrow{Pr.} 0$ where d denotes the euclidean distance. Then $Y_n \xrightarrow{\mathcal{D}} X$.*

Proof By Corollary 2.2.1, it suffices to show that for all closed sets F , $\limsup_n P(Y_n \in F) \leq P(X \in F)$. For all $\varepsilon > 0$, define the closed set $F_\varepsilon = \{x \in \mathbb{R}^d; d(x, F) \leq \varepsilon\}$. We have

$$P(Y_n \in F) \leq P(d(X_n, F) \geq \varepsilon) + P(X_n \in F_\varepsilon),$$

and therefore, by Corollary 2.2.1,

$$\begin{aligned} \limsup_n P(Y_n \in F) &\leq \limsup_n P(d(X_n, F) \geq \varepsilon) + \limsup_n P(X_n \in F_\varepsilon) \\ &= \limsup_n P(X_n \in F_\varepsilon) \leq P(X \in F_\varepsilon). \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary and $\lim_{\varepsilon \downarrow 0} P(X \in F_\varepsilon) = P(X \in F)$, $\limsup_n P(Y_n \in F) \leq P(X \in F)$. □

Links with Other Types of Convergence

Convergences in probability, almost-sure, and in quadratic mean (in $L^2_{\mathbb{C}}(P)$) are linked to convergence in distribution as the following results show.

Theorem 2.2.4 *If the sequence $\{X_n\}_{n \geq 1}$ of random vectors of \mathbb{R}^d converges almost surely to some random vector X , it also converges in distribution to the same vector X .*

Proof By dominated convergence, for all $u \in \mathbb{R}$,

$$\lim_{n \uparrow \infty} E \left[e^{i \langle u, X_n \rangle} \right] = E \left[e^{i \langle u, X \rangle} \right]$$

which implies, by Theorem 2.2.1 that $\{X_n\}_{n \geq 1}$ converges in distribution to X . \square

In fact, we have the stronger result:

Theorem 2.2.5 *If the sequence $\{X_n\}_{n \geq 1}$ of random vectors of \mathbb{R}^d converges in probability to some random vector X , it also converges in distribution to X .*

Proof If this were not the case, one could find a function $f \in C_b(\mathbb{R}^d)$ such that $E[f(X_n)]$ does not converge to $E[f(X)]$. In particular, there would exist a subsequence n_k and some $\varepsilon > 0$ such that $|E[f(X_{n_k})] - E[f(X)]| \geq \varepsilon$ for all k . As $\{X_{n_k}\}_{k \geq 1}$ converges in probability to X , one can extract from it a subsequence $\{X_{n_{k_\ell}}\}_{\ell \geq 1}$ converging almost-surely to X . In particular, since f is bounded and continuous, $\lim_{\ell} E[f(X_{n_{k_\ell}})] = E[f(X)]$ by dominated convergence, a contradiction. \square

Theorem 2.2.6 *If the sequence of real random variables $\{Z_n\}_{n \geq 1}$ converges in quadratic mean to some real random variable Z , it also converges in probability to the same random variable Z .*

Proof This follows immediately from Chebyshev's inequality

$$P(|Z_n - Z| \geq \varepsilon) \leq \frac{E[|Z_n - Z|^2]}{\varepsilon^2}.$$

\square

Pasting Theorems A.2.17 and 2.2.5, we have:

Theorem 2.2.7 *If the sequence $\{Z_n\}_{n \geq 1}$ converges in quadratic mean to some random variable Z , it also converges in distribution to the same random variable Z .*

Example 2.2.1 (A stability property of the Gaussian distribution)

Let $\{Z_n\}_{n \geq 1}$, where $Z_n = (Z_n^{(1)}, \dots, Z_n^{(m)})$, be a sequence of Gaussian random vectors of fixed dimension m that converges componentwise in quadratic mean to some vector $Z = (Z^{(1)}, \dots, Z^{(m)})$. Then the latter vector is Gaussian. In fact, by

continuity of the inner product in $L^2_{\mathbb{R}}(P)$, for all $1 \leq i, j \leq m$, $\lim_{n \uparrow \infty} E[Z_n^{(i)} Z_n^{(j)}] = E[Z^{(i)} Z^{(j)}]$ and $\lim_{n \uparrow \infty} E[Z_n^{(i)}] = E[Z^{(i)}]$, that is

$$\lim_{n \uparrow \infty} m_{Z_n} = m_Z, \quad \lim_{n \uparrow \infty} \Gamma_{Z_n} = \Gamma_Z$$

and in particular, for all $u \in \mathbb{R}^m$,

$$\begin{aligned} \lim_{n \uparrow \infty} E \left[e^{iu^T Z_n} \right] &= \lim_{n \uparrow \infty} e^{iu^T \mu_{Z_n} - \frac{i}{2} u^T \Gamma_{Z_n} u} \\ &= e^{iu^T \mu_Z - \frac{i}{2} u^T \Gamma_Z u}. \end{aligned}$$

The sequence $\{u^T Z_n\}_{n \geq 1}$ converges in quadratic mean to $u^T Z$, and therefore it also converges in distribution to $u^T Z$. Therefore, $\lim_{n \uparrow \infty} E \left[e^{iu^T Z_n} \right] = E[e^{iu^T Z}]$, and finally

$$E[e^{iu^T Z}] = e^{iu^T \mu_Z - \frac{i}{2} u^T \Gamma_Z u}$$

for all $u \in \mathbb{R}^m$. This shows that Z is a Gaussian vector.

Therefore, limits in the quadratic mean preserve the Gaussian nature of random vectors. This is the stability property referred to in the title of this example. Note that the Gaussian nature of random vectors is also preserved by linear transformations as we already know.

2.2.2 Bochner's Theorem

The result of this subsection links the classical Fourier theory and the Fourier theory of stochastic processes.

The characteristic function φ of a real random variable X has the following properties:

- (A) It is hermitian symmetric, that is $\varphi(-u) = \varphi(u)^*$, and it is uniformly bounded: $|\varphi(u)| \leq \varphi(0)$.
- (B) It is uniformly continuous on \mathbb{R} , and
- (C) It is definite non-negative, in the sense that for all integers n , all $u_1, \dots, u_n \in \mathbb{R}$, and all $z_1, \dots, z_n \in \mathbb{C}$,

$$\sum_{j=1}^n \sum_{k=1}^n \varphi(u_j - u_k) z_j z_k^* \geq 0$$

(just observe that the left-hand side equals $E \left[\left| \sum_{j=1}^n z_j e^{iu_j X} \right|^2 \right]$).

It turns out that Properties A, B and C characterize characteristic functions (up to a multiplicative constant). This is *Bochner's theorem*:

Theorem 2.2.8 *Let $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ be a function satisfying properties A, B and C. Then there exists a constant $0 \leq \beta < \infty$ and a real random variable X such that for all $u \in \mathbb{R}$,*

$$\varphi(u) = \beta E \left[e^{iuX} \right].$$

Proof We henceforth eliminate the trivial case where $\varphi(0) = 0$ (implying, in view of condition A, that φ is the null function). For any continuous function $z : \mathbb{R} \rightarrow \mathbb{C}$ and any $A \geq 0$,

$$\int_0^A \int_0^A \varphi(u-v) z(u) z^*(v) du dv \geq 0. \quad (*)$$

Indeed, as the integrand is continuous, the integral is the limit as $n \uparrow \infty$ of

$$\frac{A^2}{4^n} \sum_{j=1}^{2^n} \sum_{k=1}^{2^n} \varphi \left(\frac{A(j-k)}{2^n} \right) z \left(\frac{Aj}{2^n} \right) z \left(\frac{Ak}{2^n} \right)^*,$$

a non-negative quantity, by condition C. From (*) with $z = e^{-ixu}$, we have that

$$g(x, A) := \frac{1}{2\pi A} \int_0^A \int_0^A \varphi(u-v) e^{-ix(u-v)} du dv \geq 0.$$

Changing variables, we obtain for $g(x, A)$ the alternative expression

$$\begin{aligned} g(x, A) &:= \frac{1}{2\pi} \int_{-A}^A \left(1 - \frac{|u|}{A} \right) \varphi(u) e^{-iux} du \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} h \left(\frac{u}{A} \right) \varphi(u) e^{-iux} du, \end{aligned}$$

where

$$h(u) = (1 - |u|) 1_{\{|u| \leq 1\}}.$$

Let $M > 0$. We have

$$\begin{aligned} \int_{-\infty}^{+\infty} h\left(\frac{x}{2M}\right) g(x, A) dx &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} h\left(\frac{u}{A}\right) \varphi(u) \left(\int_{-\infty}^{+\infty} h\left(\frac{x}{2M}\right) e^{-iux} dx \right) du \\ &= \frac{1}{\pi} M \int_{-\infty}^{+\infty} h\left(\frac{u}{A}\right) \varphi(u) \left(\frac{\sin Mu}{Mu} \right)^2 du. \end{aligned}$$

Therefore

$$\begin{aligned} \int_{-\infty}^{+\infty} h\left(\frac{x}{2M}\right) g(x, A) dx &\leq \frac{1}{\pi} M \int_{-\infty}^{+\infty} h\left(\frac{u}{A}\right) |\varphi(u)| \left(\frac{\sin Mu}{Mu} \right)^2 du \\ &\leq \frac{1}{\pi} \varphi(0) \int_{-\infty}^{+\infty} \left(\frac{\sin u}{u} \right)^2 du = \varphi(0). \end{aligned}$$

By monotone convergence,

$$\lim_{M \uparrow \infty} \int_{-\infty}^{+\infty} h\left(\frac{x}{2M}\right) g(x, A) dx = \int_{-\infty}^{+\infty} g(x, A) dx,$$

and therefore

$$\int_{-\infty}^{+\infty} g(x, A) dx \leq \varphi(0).$$

The function $x \rightarrow g(x, A)$ is therefore integrable and it is the Fourier transform of the integrable and continuous function $u \rightarrow h\left(\frac{u}{A}\right) \varphi(u)$. Therefore, by the Fourier inversion formula:

$$h\left(\frac{u}{A}\right) \varphi(u) = \int_{-\infty}^{+\infty} g(x, A) e^{iux} dx.$$

In particular, with $u = 0$, $\int_{-\infty}^{+\infty} g(x, A) dx = \varphi(0)$. Therefore, $f(x, A) := \frac{g(x, A)}{\varphi(0)}$ is the probability density of some real random variable with characteristic function $h\left(\frac{u}{A}\right) \frac{\varphi(u)}{\varphi(0)}$. But

$$\lim_{A \uparrow \infty} h\left(\frac{u}{A}\right) \frac{\varphi(u)}{\varphi(0)} = \frac{\varphi(u)}{\varphi(0)}.$$

This limit of a sequence of characteristic functions is continuous at 0, and therefore it is a characteristic function (Paul Lévy's criterion, Theorem 2.2.2). \square

2.3 Weak Convergence of Finite Measures

2.3.1 Helly's Theorem

We are now left with the task of proving Paul Lévy's theorem. We shall need the following slight extension of Riesz's theorem (Theorem A.1.31), Part (ii) of the following:

- Theorem 2.3.1** (i) Let $\mu \in M^+(\mathbb{R}^d)$. The linear form $L : C_0(\mathbb{R}^d) \rightarrow \mathbb{R}$ defined by $L(f) := \int_{\mathbb{R}^d} f d\mu$ is positive ($L(f) \geq 0$ whenever $f \geq 0$) and continuous, and its norm is $\mu(\mathbb{R}^d)$.
- (ii) Let $L : C_0(\mathbb{R}^d) \rightarrow \mathbb{R}$ be a positive continuous linear form. There exists a unique measure $\mu \in M^+(\mathbb{R}^d)$ such that for all $f \in C_0(\mathbb{R}^d)$,

$$L(f) = \int_{\mathbb{R}^d} f d\mu.$$

Proof Part (i) is left as an exercise Exercise 2.4.13. We turn to the proof of (ii). The restriction of L to C_c is a positive Radon linear form, and therefore, according to Riesz's theorem A.1.31, there exists a locally finite μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that for all $f \in C_c(\mathbb{R}^d)$, $L(f) = \int_{\mathbb{R}^d} f d\mu$.

We show that μ is a *finite* (not just locally finite) measure. If not, there would exist a sequence $\{K_m\}_{m \geq 1}$ of compact subsets of \mathbb{R}^d such that $\mu(K_m) \geq 3^m$ for all $m \geq 1$. Let then $\{\varphi_m\}_{m \geq 1}$ be a sequence of non-negative functions in $C_c(\mathbb{R}^d)$ with values in $[0, 1]$ and such that for all $m \geq 1$, $\varphi_m(x) = 1$ for all $x \in K_m$. In particular, the function $\varphi := \sum_{m \geq 1} 2^{-m} \varphi_m$ is in $C_0(\mathbb{R}^d)$ and

$$\begin{aligned} L(\varphi) &\geq L\left(\sum_{m=1}^k 2^{-m} \varphi_m\right) = \sum_{m=1}^k 2^{-m} L(\varphi_m) \\ &= \sum_{m=1}^k 2^{-m} \int_{\mathbb{R}^d} \varphi_m d\mu \geq \sum_{m=1}^k 2^{-m} \mu(K_m) \geq \left(\frac{3}{2}\right)^k. \end{aligned}$$

Letting $k \uparrow \infty$ leads to $L(\varphi) = \infty$, a contradiction.

We show that L is continuous. Suppose it is not. We could then find a sequence $\{\varphi_m\}_{m \geq 1}$ of functions in $C_c(\mathbb{R}^d)$ such that $|\varphi_m| \leq 1$ and $L(\varphi_m) \geq 3^m$. The function

$\varphi := \sum_{m \geq 1} 2^{-m} \varphi_m$ is in $C_0(\mathbb{R}^d)$ and

$$\begin{aligned} L(\varphi) &\geq L\left(\sum_{m=1}^k 2^{-m} \varphi_m\right) = \sum_{m=1}^k 2^{-m} L(\varphi_m) \\ &= \sum_{m=1}^k 2^{-m} \int_{\mathbb{R}^d} \varphi_m d\mu \geq \left(\frac{3}{2}\right)^k, \end{aligned}$$

Letting $k \uparrow \infty$ again leads to $L(\varphi) = \infty$, a contradiction.

It remains to show that $L(f) = \int_{\mathbb{R}^d} f d\mu$ for all $f \in C_0(\mathbb{R}^d)$. For this we consider a sequence $\{f_m\}_{m \geq 1}$ of functions in $C_c(\mathbb{R}^d)$ converging uniformly to $f \in C_0(\mathbb{R}^d)$. We have, since L is continuous, $\lim_{m \uparrow \infty} L(f_m) = L(f)$ and, since μ is finite, $\lim_{m \uparrow \infty} \int f_m d\mu = \int f d\mu$ by dominated convergence. Therefore, since $L(f_m) = \int f_m d\mu$ for all $m \geq 1$, $L(f) = \int f d\mu$. \square

We now introduce the notion of *vague convergence*.

Definition 2.3.1 The sequence $\{\mu_n\}_{n \geq 1}$ in $M^+(\mathbb{R}^d)$ is said to converge *vaguely* to μ if, for all $f \in C_0(\mathbb{R}^d)$, $\lim_{n \uparrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$.

Theorem 2.3.2 The sequence $\{\mu_n\}_{n \geq 1}$ in $M^+(\mathbb{R}^d)$ converges vaguely if and only if

- (a) $\sup_n \mu_n(\mathbb{R}^d) < \infty$, and
- (b) There exists a dense subset \mathcal{E} in $C_0(\mathbb{R}^d)$ such that for all $f \in \mathcal{E}$, there exists $\lim_{n \uparrow \infty} \int_{\mathbb{R}^d} f d\mu_n$.

Proof Necessity. If the sequence converges vaguely, it obviously satisfies (b). As for (a), it is a consequence of the Banach–Steinhaus theorem.¹ Indeed, $\mu_n(\mathbb{R}^d)$ is the norm of L_n , where L_n is the continuous linear form $f \rightarrow \int_{\mathbb{R}^d} f d\mu_n$ from the Banach space $C_0(\mathbb{R}^d)$ (with the sup norm) to \mathbb{R} , and for all $f \in C_0(\mathbb{R}^d)$, $\sup_n \left| \int_{\mathbb{R}^d} f d\mu_n \right| < \infty$.

Sufficiency. Suppose the sequence satisfies (a) and (b). Let $f \in C_0(\mathbb{R}^d)$. For all $\varphi \in \mathcal{E}$,

$$\begin{aligned} &\left| \int f d\mu_m - \int f d\mu_n \right| \\ &\leq \left| \int \varphi d\mu_m - \int \varphi d\mu_n \right| + \left| \int f d\mu_m - \int \varphi d\mu_m \right| + \left| \int f d\mu_n - \int \varphi d\mu_n \right| \\ &\leq \left| \int \varphi d\mu_m - \int \varphi d\mu_n \right| + \sup_{x \in \mathbb{R}^d} |f(x) - \varphi(x)| \times \sup_n \mu_n(\mathbb{R}^d). \end{aligned}$$

¹ Let E be a Banach space and F be a normed vector space. Let $\{L_i\}_{i \in I}$ be a family of continuous linear mappings from E to F such that, $\sup_{i \in I} \|L_i(x)\| < \infty$ for all $x \in E$. Then $\sup_{i \in I} \|L_i\| < \infty$. See for instance Rudin (1986), Theorem 5.8.

Since $\sup_{x \in \mathbb{R}^d} |f(x) - \varphi(x)|$ can be made arbitrarily small by a proper choice of φ , this shows that the sequence $\{\int f d\mu_n\}_{n \geq 1}$ is a Cauchy sequence. It therefore converges to some $L(f)$, and L so defined is a positive linear form on $C_0(\mathbb{R}^d)$. Therefore, there exists $\mu \in M^+(\mathbb{R}^d)$ such that $L(f) = \int_{\mathbb{R}^d} f d\mu$ and $\{\mu_n\}_{n \geq 1}$ converges vaguely to μ . \square

We now give *Helly's theorem*.

Theorem 2.3.3 *From any bounded sequence of $M^+(\mathbb{R}^d)$, one can extract a vaguely convergent subsequence.*

Proof Let $\{\mu_n\}_{n \geq 1}$ be a bounded sequence of $M^+(\mathbb{R}^d)$. Let $\{f_n\}_{n \geq 1}$ be a dense sequence of $C_0(\mathbb{R}^d)$.

Since the sequence $\{\int f_1 d\mu_n\}_{n \geq 1}$ is bounded, one can extract from it a convergent subsequence $\{\int f_1 d\mu_{1,n}\}_{n \geq 1}$. Since the sequence $\{\int f_2 d\mu_{1,n}\}_{n \geq 1}$ is bounded, one can extract from it a convergent subsequence $\{\int f_2 d\mu_{2,n}\}_{n \geq 1}$. This diagonal selection process is continued. At step k , since the sequence $\{\int f_{k+1} d\mu_{k,n}\}_{n \geq 1}$ is bounded, one can extract from it a convergent subsequence $\{\int f_{k+1} d\mu_{k+1,n}\}_{n \geq 1}$. The sequence $\{\nu_k\}_{k \geq 1}$ where $\nu_k = \mu_{k,k}$ (the “diagonal” sequence) is extracted from the original sequence and for all f_n , the sequence $\{\int f_n d\nu_k\}_{k \geq 1}$ converges. The conclusion follows from Theorem 2.3.2. \square

2.3.2 Proof of Paul Lévy's Theorem

For the next definition, recall that $C_b(\mathbb{R}^d)$ denotes the collection of uniformly bounded and continuous functions from \mathbb{R}^d to \mathbb{R} .

Theorem 2.3.4 *The sequence $\{\mu_n\}_{n \geq 1}$ in $M^+(\mathbb{R}^d)$ converges weakly to μ if and only if*

- (i) *It converges vaguely to μ , and*
- (ii) $\lim_{n \uparrow \infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d)$.

Proof The necessity of (i) immediately follows from the observation that $C_0(\mathbb{R}^d) \subset C_b(\mathbb{R}^d)$. The necessity of (ii) follows from the fact that the function that is the constant 1 is in $C_b(\mathbb{R}^d)$ and therefore $\int 1 d\mu_n = \mu_n(\mathbb{R}^d)$ tends to $\int 1 d\mu = \mu(\mathbb{R}^d)$ as $n \uparrow \infty$.

Sufficiency. Suppose that (i) and (ii) are satisfied. To prove weak convergence, it suffices to prove that $\lim_{n \uparrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$ for any non-negative function $f \in C_b(\mathbb{R}^d)$.

Since the measure μ is of finite total mass, for any $\varepsilon > 0$ one can find a compact set $K_\varepsilon = K$ such that $\mu(\bar{K}) \leq \varepsilon$. Choose a continuous function with compact support φ with values in $[0, 1]$ and such that $\varphi \geq 1_K$. Since $|f - f\varphi| \leq \|f\|(1 - \varphi)$

(where $\|f\| = \sup_{x \in \mathbb{R}^d} |f(x)|$),

$$\begin{aligned} \limsup_{n \uparrow \infty} \left| \int f d\mu_n - \int f \varphi d\mu_n \right| &\leq \limsup_{n \uparrow \infty} \|f\| \int (1 - \varphi) d\mu_n \\ &= \|f\| \left(\lim_{n \uparrow \infty} \int d\mu_n - \lim_{n \uparrow \infty} \int \varphi d\mu_n \right) \\ &= \|f\| \int (1 - \varphi) d\mu \leq \varepsilon \|f\|. \end{aligned}$$

Similarly, $\left| \int f d\mu - \int f \varphi d\mu \right| \leq \varepsilon \|f\|$. Therefore, for all $\varepsilon > 0$,

$$\limsup_{n \uparrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \leq 2\varepsilon \|f\|,$$

and this completes the proof. \square

We will be ready for the generalization of Paul Lévy's criterion of convergence in distribution after a few preliminaries.

Definition 2.3.2 A family $\{\alpha_t\}_{t>0}$ of functions $\alpha_t : \mathbb{R}^d \rightarrow \mathbb{C}$ in $L^1_{\mathbb{C}}(\mathbb{R}^d)$ is called an *approximation of the Dirac distribution* in \mathbb{R}^d if it satisfies the following three conditions:

- (i) $\int_{\mathbb{R}^d} \alpha_t(x) dx = 1$,
- (ii) $\sup_{t>0} \int_{\mathbb{R}^d} |\alpha_t(x)| dx := M < \infty$, and
- (iii) For any compact neighborhood V of $0 \in \mathbb{R}^d$, $\lim_{t \downarrow 0} \int_V |\alpha_t(x)| dx = 0$.

Lemma 2.3.1 Let $\{\alpha_t\}_{t>0}$ be an approximation of the Dirac distribution in \mathbb{R}^d . Let $f : \mathbb{R}^d \rightarrow \mathbb{C}$ be a bounded function continuous at all points of a compact $K \subset \mathbb{R}^d$. Then $\lim_{t \downarrow 0} f * \alpha_t$ uniformly in K .

Proof We will show later that

$$\lim_{y \rightarrow 0} \sup_{x \in K} |f(x - y) - f(x)| \rightarrow 0. \quad (*)$$

V being a compact neighborhood of 0 , we have that

$$\begin{aligned} \sup_{x \in K} |f(x) - (f * \alpha_t)(x)| &\leq M \sup_{y \in V} \sup_{x \in K} |f(x - y) - f(x)| \\ &\quad + 2 \sup_{x \in \mathbb{R}^d} |f(x)| \int_V |\alpha_t(y)| dy. \end{aligned}$$

This quantity can be made smaller than an arbitrary $\varepsilon > 0$ by choosing V such that the first term be $< \frac{1}{2}\varepsilon$ (uniform continuity of f on compact sets) and the second term can then be made $< \frac{1}{2}\varepsilon$ by letting $t \downarrow$ (condition (iii) of Definition 2.3.2).

Proof of (*). Let $\varepsilon > 0$ be given. For all $x \in K$, there exists an open and symmetric neighborhood V_x of 0 such that for all $y \in V_x$, $|f(x-y) - f(x)| \leq \frac{1}{2}\varepsilon$. Also, one can find an open and symmetric neighborhood W_x of 0 such that $W_x + W_x \subset V_x$. The union of open sets $\cup_{x \in K} \{x + W_x\}$ obviously covers K , and since the latter is a compact set, one can extract a finite covering of K : $\cup_{j=1}^m (x_j + W_{x_j})$. Define $W = \cap_{j=1}^m W_{x_j}$, open neighborhood of 0.

Let $y \in W$. Any $x \in K$ belongs to some $x_j + W_{x_j}$, and for such j , and

$$\begin{aligned} |f(x-y) - f(x)| &\leq |f(x_j) - f(x_j - (x_j - x))| \\ &\quad + |f(x_j) - f(x_j - (x_j - x + y))|. \end{aligned}$$

But $x_j - x \in W_{x_j}$ and $x_j - x + y \in W_{x_j} + W \subset V_{x_j}$. Therefore

$$|f(x-y) - f(x)| \leq \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon.$$

□

Theorem 2.3.5 *Let $\{\mu_n\}_{n \geq 1}$ be a sequence of $M^+(\mathbb{R}^d)$ such that for all $v \in \mathbb{R}^d$, there exists $\lim_{n \uparrow \infty} \widehat{\mu}_n(v) = \varphi(v)$ for some function φ that is continuous at 0. Then $\{\mu_n\}_{n \geq 1}$ converges weakly to a finite measure μ whose Fourier transform is φ .*

Proof The sequence $\{\mu_n\}_{n \geq 1}$ is bounded (that is $\sup_n \mu_n(\mathbb{R}^d) < \infty$) since $\mu_n(\mathbb{R}^d) = \widehat{\mu}_n(1)$ has a limit as $n \uparrow \infty$. In particular

$$\widehat{\mu}_n(v) \leq \mu_n(\mathbb{R}^d) \leq \sup_n \mu_n(\mathbb{R}^d) < \infty. \quad (\dagger)$$

If $f \in L^1_{\mathbb{R}}(\mathbb{R}^d)$, then by Theorem 2.1.2, $\int \widehat{f} d\mu_n = \int f \widehat{\mu}_n dx$. By dominated convergence (using (\dagger)), $\lim_{n \uparrow \infty} \int f \widehat{\mu}_n dx = \int f \varphi dx$. Therefore

$$\lim_{n \uparrow \infty} \int \widehat{f} d\mu_n = \int f \varphi dx.$$

One can replace in the above equality \widehat{f} by any function in $\mathcal{D}(\mathbb{R}^d)$, since such function is always the Fourier transform of some integrable function (Corollary 1.1.5).

Therefore, by Theorem 2.3.2, $\{\mu_n\}_{n \geq 1}$ converges vaguely to some finite measure μ .

We now show that it converges *weakly* to μ . Let f be an integrable function of integral 1 such that $f(x) = f(-x)$ and $\widehat{f} \in \mathcal{D}(\mathbb{R}^d)$. For $t > 0$, define $f_t(x) := t^{-d} f(t^{-1}x)$. Using Theorem 2.1.2, we have

$$\int \widehat{f}(tx) \mu_n(dx) = \int f_t(x) \widehat{\mu}_n(x) dx = (f_t * \widehat{\mu}_n)(0).$$

By dominated convergence,

$$\lim_{n \uparrow \infty} \int \widehat{f}(tx) \mu_n(dx) = (f_t * \varphi)(0),$$

and by vague convergence,

$$\lim_{n \uparrow \infty} \int \widehat{f}(tx) \mu_n(dx) = \int \widehat{f}(tx) \mu(dx).$$

Therefore, for all $t > 0$, $\int \widehat{f}(tx) \mu(dx) = (f_t * \varphi)(0)$.

Since the function φ is bounded and continuous at the origin, by Lemma 2.3.1, $\lim_{t \downarrow 0} (f_t * \varphi)(0) = \varphi(0)$. Also, by dominated convergence $\lim_{t \downarrow 0} \int \widehat{f}(tx) \mu(dx) = \mu(\mathbb{R}^d)$. Therefore,

$$\mu(\mathbb{R}^d) = \varphi(0) = \lim_{n \uparrow \infty} \mu_n(\mathbb{R}^d).$$

Therefore, by Theorem 2.3.4, $\{\mu_n\}_{n \geq 1}$ converges weakly to μ . Since the function $x \rightarrow e^{-2i\pi \langle v, x \rangle}$ is continuous and bounded,

$$\widehat{\mu}(v) = \int e^{-2i\pi \langle v, x \rangle} \mu(dx) = \lim_{n \uparrow \infty} \int e^{-2i\pi \langle v, x \rangle} \mu_n(dx) = \varphi(v).$$

□

2.4 Exercises

Exercise 2.4.1 (*Sampling a sinusoid*) Check with a single sinusoid that you really need the strict inequality in (2.2).

Exercise 2.4.2 (*Characteristic function of the residual time*) Let $G : \mathbb{R}_+ \rightarrow \mathbb{C}$ be the primitive function of the locally integrable function $g : \mathbb{R}_+ \rightarrow \mathbb{C}$, that is, for all $x \geq 0$,

$$G(x) = G(0) + \int_0^x g(u) du.$$

(a) Let X be a non-negative random variable with finite mean μ and such that $E[|G(X)|] < \infty$. Show that

$$E[G(X)] = G(0) + \int_0^\infty g(x) P(X \geq x) dx.$$

- (b) Let X be as in (a), and let \bar{X} be a non-negative random variable with the probability density $\mu^{-1}P(X \geq x)$. Compute the characteristic function of \bar{X} .

Exercise 2.4.3 (*Characteristic functions of lattice distributions*) Let X be a real random variable whose characteristic function ψ is such that $|\psi(t_0)| = 1$ for some $t_0 \neq 0$. Show that there exists some $a \in \mathbb{R}$ such that $\sum_{n \in \mathbb{Z}} P(X = a + n \frac{2\pi}{t_0}) = 1$.

Exercise 2.4.4 (*Probability of the positive quadrant*) Let (X, Y) be a 2-dimensional Gaussian vector with probability density

$$f(x, y) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right\}$$

where $|\rho| < 1$.

- (a) Show that X and $(Y - \rho X) / (1 - \rho^2)^{1/2}$ are independent Gaussian random variable with mean 0 and variance 1.
 (b) Prove that

$$P(X > 0, Y > 0) = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}(\rho).$$

Exercise 2.4.5 (*Discrete uniform variable and convergence in distribution*) Let Z_n be a random variable uniformly distributed on $\{\frac{1}{2^n}, 2\frac{1}{2^n}, \dots, 2^n \frac{1}{2^n} = 1\}$. Does it converge in distribution?

Exercise 2.4.6 (*Convergence in distribution for integer-valued variables*) Show that if the random variables X_n 's and X take integer values, $X_n \xrightarrow{\mathcal{D}} X$ if and only if for all $k \geq 0$,

$$\lim_{n \uparrow \infty} P(X_n = k) = P(X = k).$$

Exercise 2.4.7 (*Poisson's law of rare events in the plane*) Let Z_1, \dots, Z_M be M bidimensional IID random vectors uniformly distributed on the square $[0, A] \times [0, A] = \Gamma_A$. Define for any measurable set $C \subseteq \Gamma_A$, $N(C)$ to be the number of random vectors Z_i that fall in C . Let C_1, \dots, C_K be measurable *disjoint* subsets of Γ_A .

- (i) Give the characteristic function of the vectors $(N(C_1), \dots, N(C_K))$.
 (ii) We now let M be a function of $A \in \mathbb{R}_+$ such that $\frac{M(A)}{A^2} = \lambda > 0$. Show that, as $A \uparrow \infty$, $(N(C_1), \dots, N(C_K))$ converges in distribution. Identify the limit distribution.

Exercise 2.4.8 (*Cauchy random variable*) A random variable X with PDF

$$f(x) = \frac{1}{\pi(1+x^2)}$$

is called a *Cauchy* random variable. Compute the Fourier transform of $u \rightarrow e^{-|u|}$ and deduce from the result that the CF of the Cauchy random variable is $\psi(u) = e^{-|u|}$.

Exercise 2.4.9 (*Cauchy random variables and convergences*) Let $\{X_n\}_{n \geq 1}$ be a sequence of IID Cauchy random variables.

- (A.) What is the limit in distribution of $\frac{X_1 + \dots + X_n}{n}$?
- (B.) Does $\frac{X_1 + \dots + X_n}{n^2}$ converge in distribution?
- (C.) Does $\frac{X_1 + \dots + X_n}{n}$ converge almost surely to a deterministic constant?

Exercise 2.4.10 (*A simple counter example*) Let Z be a random variable with a symmetric distribution (that is, Z and $-Z$ have the same distribution). Define the sequence $\{Z_n\}_{n \geq 1}$ as follows: $Z_n = Z$ if n is odd, $Z_n = -Z$ if n is even. In particular, $\{Z_n\}_{n \geq 1}$ converges in *distribution* to Z . Show that if Z is nondegenerate, then $\{Z_n\}_{n \geq 1}$ does NOT converge to Z in *probability*.

Exercise 2.4.11 (*Infimum of IID uniform variables*) Let $\{Y_n\}_{n \geq 1}$ be a sequence of IID random variables uniformly distributed on $[0, 1]$. Show that

$$X_n = n \min(Y_1, \dots, Y_n) \xrightarrow{\mathcal{D}} \mathcal{E}(1),$$

(the exponential distribution with mean 1).

Exercise 2.4.12 (*A stability property of the Gaussian distribution*) Prove the statement (ii) of Example 2.2.1.

Exercise 2.4.13 (*Riesz's theorem*) Prove Part (i) of Theorem 2.3.1.

Exercise 2.4.14 (*Dirac measures and vague convergence*)

- (i) Show that the sequence of Dirac measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\{\varepsilon_{a_n}\}_{n \geq 1}$ converges vaguely to ε_a if and only if $\lim_{n \uparrow \infty} a_n = a$.
- (ii) Show that ε_a converges vaguely to the null measure when $a \rightarrow \infty$.
- (iii) Show that the measure $\mu_a = \frac{1}{2B} 1_{[-B, +B]} \ell$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ converges vaguely to the null measure as $B \rightarrow \infty$.
- (iv) Let $f \in L^1_{\mathbb{C}}(\mathbb{R}^d)$ be non-negative and such that $\int_{\mathbb{R}^d} f(x) dx = 1$. Let $f_t(x) := t^{-d} f(t^{-1}x)$. Show that, as $n \uparrow \infty$, the measures on \mathbb{R}^d with densities $f_{\frac{1}{n}}$ and f_n converge vaguely to the Dirac measure ε_0 and the null measure respectively.

Exercise 2.4.15 (*Polya's theorem*) Prove the following result:

A non-negative symmetric continuous function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ that is convex on $[0, +\infty)$ and such that $\varphi(0) = 1$ and $\lim_{|u| \uparrow \infty} \varphi(u) = 0$ is a characteristic function.

Hint: first show that this is true if φ is in addition piecewise linear.

Use the result to prove that the fact that two real random variables whose characteristic functions coincide on an interval of \mathbb{R} do not necessarily have the same distribution.

Fourier Analysis and Stochastic Processes

Brémaud, P.

2014, XIII, 385 p. 2 illus., Softcover

ISBN: 978-3-319-09589-9