7 Fourier Integral and Sobolev Space H^s

The Fourier integral is a useful construct in analysis which is based on an idea of **J. Fourier** for resolving functions into basic harmonics in his treatment of conduction of heat. When functions are periodic, say of period 2π, they are resolved as Fourier series (see Section 5.9). For nonperiodic functions on $\mathbb{R}$, the idea leads to a Fourier integral. The Fourier integral for L^1 functions on $\mathbb{R}^n$ can be defined straight away, and is treated in Section 7.1. Since L^2 is a Hilbert space, it is desirable to define a Fourier integral for L^2 functions; but a straightforward definition for L^2 functions is lacking; some variation is therefore necessary for the purpose. We shall get around this through the Fourier integral for rapidly decreasing functions, introduced in Section 7.2. Applications to Sobolev spaces H^s and to partial differential equations are provided in later sections of the chapter. The Fourier integral of probability distributions is introduced in Section 7.5, and is applied to prove the central limit theorem of probability theory.

A Fourier integral is also called a Fourier transform.

For the convenience of expressing certain functions defined on $\mathbb{R}^n$, the function $x \mapsto f(x)$ will sometimes be expressed by $f(x)$. For example, $x \mapsto x^\alpha$ is simply denoted by x^α, and if f is a function on $\mathbb{R}^n$, the function $x \mapsto x^\alpha f(x)$ is denoted by $x^\alpha f$.

7.1 Fourier integral for L^1 functions

For $f \in L^1 := L^1(\mathbb{R}^n)$, define the **Fourier integral** Ff of f by

$$(Ff)(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} f(x) e^{-i\xi \cdot x} dx, \quad \xi \in \mathbb{R}^n.$$

Since $|f(x)e^{-i\xi \cdot x}| = |f(x)|$, Ff is defined and is finite for every $\xi \in \mathbb{R}^n$. One verifies readily that

(1) $\|Ff\|_\infty \le \frac{1}{(2\pi)^{\frac{n}{2}}}\|f\|_1$;

(2) Ff is uniformly continuous on $\mathbb{R}^n$ (note that this follows from LDCT).

Exercise 7.1.1 Let $f(x) = e^{-|x|}$, $x \in \mathbb{R}$. Find Ff.

Exercise 7.1.2 Suppose that $f_1, \ldots, f_n$ are in $L^1(\mathbb{R})$ and let $f(x) = \prod_{j=1}^n f_j(x_j)$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Show that $f \in L^1(\mathbb{R}^n)$ and $(Ff)(\xi) = \prod_{j=1}^n (Ff_j)(\xi_j)$ for $\xi = (\xi_1, \ldots, \xi_n)$.

Example 7.1.1

(i) For $\alpha > 0$, consider the function $f = I_{[-\alpha,\alpha]}$ on $\mathbb{R}$; then $(Ff)(\xi) = \frac{2\sin\alpha\xi}{\sqrt{2\pi}\,\xi}$, $\xi \in \mathbb{R}$. For $n > 1$, $f = I_{[-\alpha,\alpha]\times\cdots\times[-\alpha,\alpha]}$, $(Ff)(\xi) = \frac{2^n}{(2\pi)^{\frac{n}{2}}}\prod_{j=1}^n \frac{\sin\alpha\xi_j}{\xi_j}$.

This follows from Exercise 7.1.2.

(ii) For $n = 1$, consider the function $f(x) = e^{-\frac{x^2}{2}}$. We have

$$(Ff)(\xi) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-i\xi x}dx = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\cos\xi x dx,$$

and

$$\begin{aligned}(Ff)'(\xi) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}(-x\sin\xi x)dx \\ &= \frac{1}{\sqrt{2\pi}}\xi\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\cos\xi x dx = \xi(Ff)(\xi).\end{aligned}$$

The first equality follows by LDCT and the second by integration by parts. Then $(Ff)(\xi) = Ce^{-\frac{\xi^2}{2}}$ with C being a constant. But $(Ff)(0) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{x^2}{2}}dx$ $= 1 = C$. Thus $(Ff)(\xi) = e^{-\frac{\xi^2}{2}}$. For $n > 1$, if $f(x) = e^{-\frac{|x|^2}{2}}$, then $(Ff)(\xi) = e^{-\frac{|\xi|^2}{2}}$.

Exercise 7.1.3 Consider the function $f(x) = e^{-\frac{1}{2}x^2}$ in Example 7.1.1 (ii). Use a contour integral to show that

$$\int_{\mathbb{R}} e^{-\frac{1}{2}(x+i\xi)^2}dx = \int_{\mathbb{R}} e^{-\frac{1}{2}x^2}dx = \sqrt{2\pi}$$

and give a direct verification that

$$Ff(\xi) = e^{-\frac{1}{2}\xi^2}.$$

Theorem 7.1.1 *If $f, g \in L^1(\mathbb{R}^n)$, $(F\{f * g\})(\xi) = (2\pi)^{\frac{n}{2}}(Ff)(\xi)(Fg)(\xi)$.*

Proof Observe first that $f * g \in L^1(\mathbb{R}^n)$, by the Young inequality (Theorem 6.5.1). Then,

$$\begin{aligned}(F\{f * g\})(\xi) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)g(y)dy \right) e^{-i\xi\cdot x}dx \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x-y)e^{-i\xi\cdot(x-y)}dx \right) g(y)e^{-i\xi\cdot y}dy \\ &= \int_{\mathbb{R}^n} (Ff)(\xi)g(y)e^{-i\xi\cdot y}dy = (2\pi)^{\frac{n}{2}}(Ff)(\xi)(Fg)(\xi).\end{aligned}$$

It is to be noted that since $f(x-y)g(y)e^{-i\xi\cdot x}$ is an integrable function of (x, y) in $\mathbb{R}^{2n}$, it is legitimate to use the Fubini theorem in the above argument. ∎

Example 7.1.2 Let $\alpha > 0$. It is readily verified that $\frac{1}{\alpha}I_{[-\frac{\alpha}{2},\frac{\alpha}{2}]} * I_{[-\frac{\alpha}{2},\frac{\alpha}{2}]}(x) = \left(1 - \frac{|x|}{\alpha}\right)^+$ (cf. Exercise 6.5.2), it then follows than

$$\left(F\left(1 - \frac{|x|}{\alpha}\right)^+\right)(\xi) = \frac{\sqrt{2\pi}}{\alpha}\left(\frac{2\sin\frac{\alpha}{2}\xi}{\sqrt{2\pi}\xi}\right)^2 = \frac{1}{\alpha}\cdot\frac{2(1-\cos\alpha\xi)}{\sqrt{2\pi}\xi^2}.$$

For $f \in L^1(\mathbb{R}^n)$, the **inverse Fourier integral** $\check{F}f$ of f is defined by

$$(\check{F}f)(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} f(x)e^{i\xi\cdot x}dx, \quad \xi \in \mathbb{R}^n.$$

For $f \in L^1$, Ff and $\check{F}f$ are often denoted by $\hat{f}$ and $\check{f}$ respectively.

Exercise 7.1.4 Recall that for $a \in \mathbb{R}^n, \sigma > 0$, and a function f on $\mathbb{R}^n$, $f^a(x) = f(x-a)$, $f_\sigma(x) = \sigma^{-n}f(\frac{x}{\sigma})$ for $x \in \mathbb{R}^n$.

(i) Show that $\widehat{f^a}(\xi) = e^{-i\xi\cdot a}\hat{f}(\xi)$ for $f \in L^1$.

(ii) Show that $\widehat{f_\sigma}(\xi) = \hat{f}(\sigma\xi)$.

Exercise 7.1.5 Let f, g be in L^1. Show that

$$\int_{\mathbb{R}^n} f\hat{g}d\lambda^n = \int_{\mathbb{R}^n} \hat{f}gd\lambda^n.$$

Theorem 7.1.2 (Riemann–Lebesgue) *If $f \in L^1$, then $\hat{f} \in C_0(\mathbb{R}^n)$.*

Proof If f is the function considered in Example 7.1.1 (i), then $\lim_{|\xi|\to\infty}\hat{f}(\xi) = 0$; hence the theorem holds for indicator functions of cubes, by Exercise 7.1.4 (i); as a consequence the theorem holds for finite linear combinations of indicator functions of cubes. But, as $C_c(\mathbb{R}^n)$ is dense in L^1, one verifies easily that the family of all finite linear combinations of indicator functions of cubes is dense in L^1. Thus for $f \in L^1$ and $\varepsilon > 0$, there is a finite linear combination φ of indicator functions of cubes such that $\|f-\varphi\|_1 < \frac{\varepsilon}{2}$, then $|\hat{f}(\xi)| \le |(\widehat{f-\varphi})(\xi)| + |\hat{\varphi}(\xi)| < \frac{\varepsilon}{2} + |\hat{\varphi}(\xi)|$, from which it follows that $|\hat{f}(\xi)| < \varepsilon$ if $|\xi|$ is large enough, because $\hat{\varphi} \in C_0(\mathbb{R}^n)$. ■

Theorem 7.1.3 *For $f \in L^1$, f is uniquely determined by $\hat{f}$; in other words, the map $f \mapsto \hat{f}$ is injective on L^1.*

Proof Take h to be the function defined on $\mathbb{R}^n$ by $h(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$ (note $\int h d\lambda^n = 1$), then $\{h_\sigma\}_{\sigma>0}$ is an approximate identity for L^1. Put $m_\sigma f = f * h_\sigma$. Then, since $\hat{h}(\xi) = h(\xi)$, as in Example 7.1.1 (ii), we have

$$\begin{aligned}
(m_\sigma f)(x) &= \int_{\mathbb{R}^n} f(y) h_\sigma(x-y) dy = \int_{\mathbb{R}^n} f(y) h_\sigma(y-x) dy \\
&= \sigma^{-n} \int_{\mathbb{R}^n} f(y) h\left(\frac{y-x}{\sigma}\right) dy \\
&= \frac{\sigma^{-n}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(f(y) \int_{\mathbb{R}^n} e^{-i\frac{y-x}{\sigma}\cdot z} h(z) dz\right) dy \\
&= \frac{\sigma^{-n}}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(f(y) \int_{\mathbb{R}^n} e^{-i(y-x)\cdot\frac{z}{\sigma}} h(z) dz\right) dy \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} \left(f(y) \int_{\mathbb{R}^n} e^{-i(y-x)\cdot z} h(\sigma z) dz\right) dy \\
&= \int_{\mathbb{R}^n} e^{ix\cdot z} \hat{f}(z) h(\sigma z) dz;
\end{aligned}$$

this means that the function $m_\sigma f$ is uniquely determined by $\hat{f}$. But as $m_\sigma f \to f$ in L^1 as $\sigma \to 0$, f is uniquely determined by $\hat{f}$. ■

Theorem 7.1.4 (L^1 inversion theorem) *If both f and $\hat{f}$ are in L^1, then $f = (\hat{f})^\vee$, i.e. f is the inverse Fourier integral of $\hat{f}$.*

Proof Let h and $\{m_\sigma\}_{\sigma>0}$ be as in the proof of Theorem 7.1.3. There we have shown that

$$(m_\sigma f)(x) = \int_{\mathbb{R}^n} e^{ix\cdot\xi} \hat{f}(\xi) h(\sigma\xi) d\xi;$$

since $|e^{ix\cdot\xi}\hat{f}(\xi)h(\sigma\xi)| \le \frac{1}{(2\pi)^{n/2}} |\hat{f}(\xi)|$ and $\lim_{\sigma\to 0} e^{ix\cdot\xi}\hat{f}(\xi)h(\sigma\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{ix\cdot\xi}\hat{f}(\xi)$, it follows from LDCT that $\lim_{\sigma\to 0}(m_\sigma f)(x) = (\hat{f})^\vee(x)$ for each $x \in \mathbb{R}^n$. Now, $|(m_\sigma f)(x)| \le \frac{1}{(2\pi)^{\frac{n}{2}}} \|\hat{f}\|_1$ implies that $\lim_{\sigma\to 0} \int_{B_R(0)} |m_\sigma f - (\hat{f})^\vee| d\lambda^n = 0$ for any $R > 0$, again by LDCT; this, together with $\lim_{\sigma\to 0} \int_{B_R(0)} |m_\sigma f - f| d\lambda^n = 0$ (cf. Theorem 6.5.3), shows that $f = (\hat{f})^\vee$ a.e. on $B_R(0)$ for any $R > 0$, and consequently $f = (\hat{f})^\vee$ a.e. on $\mathbb{R}^n$. ■

As an application of the L^1 inversion theorem, we establish the fact that the family $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ of normalized Hermite functions introduced in Example 5.8.1 is an orthonormal basis for $L^2(\mathbb{R})$, or equivalently, that the family $\{h_0, h_1, h_2, \ldots\}$ of normalized Hermite polynomials is an orthonormal basis for $L^2_w(\mathbb{R})$ where $w(x) = e^{-x^2}$.

Corollary 7.1.1 *The family $\{\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2, \ldots\}$ of normalized Hermite functions is an orthonormal basis for $L^2(\mathbb{R})$.*

Proof By Theorem 5.8.3, we need to show that if $f \in L^2(\mathbb{R})$ is such that

$$\text{(A)} \qquad \int_{-\infty}^{\infty} f(x)\mathcal{E}_n(x)dx = 0, \quad n = 0, 1, 2, \ldots,$$

then $f = 0$ a.e. Recall from Example 5.8.1 that $\mathcal{E}_n(x) = e^{-\frac{x^2}{2}} h_n(x)$, where $h_n(x)$ is a polynomial of degree n and that each monomial x^n is a linear combination of $h_0(x), \ldots, h_n(x)$; hence if $f \in L^2(\mathbb{R})$ satisfies the condition (A), then it satisfies the condition

$$\text{(B)} \qquad \int_{-\infty}^{\infty} f(x)e^{-\frac{x^2}{2}} x^n dx = 0, \quad n = 0, 1, 2, \ldots.$$

Therefore, it suffices to show that if $f \in L^2(\mathbb{R})$ satisfies the condition (B), then $f = 0$ a.e. Now let $f \in L^2(\mathbb{R})$ satisfy the condition (B). Put $g(x) = f(x)e^{-\frac{x^2}{2}}$, then $g \in L^1(\mathbb{R})$, by the Schwarz inequality and

$$\begin{aligned} \hat{g}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} f(x)e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} \frac{(-itx)^n}{n!} \right) f(x)e^{-\frac{x^2}{2}} dx; \end{aligned}$$

but for $N \in \mathbb{N}$,

$$\left| \sum_{n=0}^{N} \frac{(-itx)^n}{n!} f(x)e^{-\frac{x^2}{2}} \right| \leq |f(x)|e^{|tx|}e^{-\frac{x^2}{2}},$$

of which the function on the right-hand side is integrable because

$$\int_{-\infty}^{\infty} |f(x)|e^{|tx|}e^{-\frac{x^2}{2}} dx \leq \|f\|_2 \left\{ \int_{-\infty}^{\infty} e^{2|tx|}e^{-x^2} dx \right\}^{\frac{1}{2}} < \infty.$$

It follows then from LDCT that

$$\begin{aligned}\hat{g}(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\lim_{N\to\infty}\left(\sum_{n=0}^{N}\frac{(-itx)^n}{n!}f(x)e^{-\frac{x^2}{2}}\right)dx\\ &= \lim_{N\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(\sum_{n=0}^{N}\frac{(-it)^n}{n!}x^n f(x)e^{-\frac{x^2}{2}}\right)dx\\ &= \lim_{N\to\infty}\frac{1}{\sqrt{2\pi}}\sum_{n=0}^{N}\frac{(-it)^n}{n!}\int_{-\infty}^{\infty}f(x)e^{-\frac{x^2}{2}}x^n\,dx = 0\end{aligned}$$

by condition (B). Thus, $\hat{g} = 0$ and by Theorem 7.1.4, $g(t) = (\hat{g})^{\vee}(t) = 0$ a.e. and hence $f = 0$ a.e. ■

A remarkable application of the Fourier integral is the **Poisson summation formula**, which states that

$$\sum_{n=-\infty}^{\infty} f(2n\pi) = \frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty}\hat{f}(n)$$

for integrable functions f satisfying certain condition. Usually, the Poisson summation formula is established for $f \in C^2(\mathbb{R})$ such that

$$|f(x)| + |f'(x)| + |f''(x)| \le C(1+x^2)^{-1}, \quad x \in \mathbb{R},$$

for some constant $C > 0$. We shall prove the formula under weaker conditions. For an integrable function f on $\mathbb{R}$ and $n \in \mathbb{Z}$, let

$$f_n(x) = f(x + 2\pi n), \quad x \in \mathbb{R}.$$

We first claim that $\{f_n(x)\} = \{f_n(x)\}_{n\in\mathbb{Z}}$ is summable for a.e. x in $\mathbb{R}$. For this purpose it is sufficient to show that $\{f_n(x)\}$ is summable for a.e. x in $[-\pi, \pi]$, because if $\{f_n(x)\}$ is summable, then $\{f_n(x + 2\pi m)\} = \{f_{n+m}(x)\} = \{f_n(x)\}$ for any $m \in \mathbb{Z}$, and hence $\sum_{n\in\mathbb{Z}} f_n(x + 2\pi m) = \sum_{n\in\mathbb{Z}} f_n(x)$. Now,

$$\int_{-\pi}^{\pi}\sum_{n\in\mathbb{Z}}|f_n(x)|dx = \sum_{n\in\mathbb{Z}}\int_{-\pi}^{\pi}|f_n(x)|dx = \int_{\mathbb{R}}|f|d\lambda < \infty$$

implies that $\sum_{n\in\mathbb{Z}}|f_n(x)| < \infty$ for a.e. x in $[-\pi, \pi]$. Hence $\{f_n(x)\}$ is summable for a.e. x in $\mathbb{R}$ and if we put $[f](x) = \sum_{n\in\mathbb{Z}} f_n(x)$, if $\{f_n(x)\}$ is summable and $[f](x) = 0$ otherwise, $[f]$ is defined on $\mathbb{R}$ and periodic with period 2π. Furthermore, $[f]$ is integrable on $[-\pi, \pi]$. The function $[f]$ is called the **stacked function** of f. If we define for $j \in \mathbb{N}$ the function $[f]_j$ on $\mathbb{R}$ by

$$[f]_j(x) = \sum_{|n|\le j} f_n(x),$$

then $[f]_j \to [f]$ a.e. and $|[f]_j| \le [|f|]$ a.e. Since $[|f|]$ is integrable on $[-\pi,\pi]$, it follows from LDCT that $[f]_j \to [f]$ in $L^1[-\pi,\pi]$. We have proved the following lemma (7.1.1).

Lemma 7.1.1 *Suppose that $f \in L^1 = L^1(\mathbb{R})$. Then $[f]_j \to [f]$ a.e. as well as in $L^1[-\pi,\pi]$ as $j \to \infty$.*

In the immediate following, for $f \in W^{1,1}(\mathbb{R})$ we always take a version of f which is AC on every finite closed interval of $\mathbb{R}$ (note that since $W^{1,1}(\mathbb{R}) = \mathring{W}^{1,1}(\mathbb{R})$, $f(x) = \int_{-\infty}^{x} f'(x)dx$ for a.e. x).

Lemma 7.1.2 *If $f \in W^{1,1}(\mathbb{R})$, then $[f]$ is an AC function on $[-\pi,\pi]$ and satisfies $[f](-\pi) = [f](\pi)$. Furthermore, $[f]' = [f']$ a.e.*

Proof We take a version of f which is AC on every finite closed interval of $\mathbb{R}$. Then f' exists a.e. and is integrable on $\mathbb{R}$; and since

$$
\begin{aligned}
f_n(x) &= f(x+2n\pi) = f(-\pi+2n\pi) + \int_{-\pi+2n\pi}^{x+2n\pi} f'(s)ds \\
&= f_n(-\pi) + \int_{-\pi}^{x} f_n'(s)ds
\end{aligned}
\tag{7.1}
$$

for $x \in [-\pi,\pi]$, we have

$$
[f]_j(x) = [f]_j(-\pi) + \int_{-\pi}^{x} [f']_j(s)ds, \quad x \in [-\pi,\pi]. \tag{7.2}
$$

As $[f']_j \to [f']$ in $L^1[-\pi,\pi]$ as $j \to \infty$, by Lemma 7.1.1, $\lim_{j\to\infty} \int_{-\pi}^{x} [f']_j(s)ds = \int_{-\pi}^{x} [f'](s)ds$ for $x \in [-\pi,\pi]$; we conclude that $\lim_{j\to\infty}[f]_j(-\pi)$ exists and is finite by letting $j \to \infty$ in (7.2) for x such that $\lim_{j\to\infty}[f]_j(x) = [f](x)$. Because $|f| \in W^{1,1}(\mathbb{R})$, we also know that $\lim_{j\to\infty}[|f|]_j(-\pi)$ exists and is finite, from which follows that $\{f_n(-\pi)\}$ is summable and hence $[f](-\pi) = \lim_{j\to\infty}[f]_j(-\pi)$. Now for any finite subset F of $\mathbb{Z}$ and $x \in [-\pi,\pi]$,

$$
\sum_{n\in F} \left| \int_{-\pi}^{x} f_n'(s)ds \right| \le \int_{-\pi}^{x} \sum_{n\in F} |f_n'(s)|ds \le \int_{-\pi}^{x} [|f'|](s)ds < \infty,
$$

implying that $\{\int_{-\pi}^{x} f_n'(s)ds\}$ is summable for each $x \in [-\pi,\pi]$. We then infer from (7.1) that $\{f_n(x)\}$ is summable and $[f](x) = \lim_{j\to\infty}[f]_j(x)$ for each $x \in [-\pi,\pi]$. Now let $j \to \infty$ in (7.2); we have

$$
[f](x) = [f](-\pi) + \int_{-\pi}^{x} [f'](s)ds, \quad x \in [-\pi,\pi];
$$

consequently $[f]$ is AC on $[-\pi,\pi]$ and $[f]' = [f']$ a.e. on $[-\pi,\pi]$. Finally, $[f](-\pi) = \sum_{n\in\mathbb{Z}} f_n(-\pi) = \sum_{n\in\mathbb{Z}} f_{n+1}(-\pi) = \sum_{n\in\mathbb{Z}} f_n(\pi) = [f](\pi)$. ■

Lemma 7.1.3 *If $f \in W^{1,1}(\mathbb{R})$, then*

$$\sum_{k\in\mathbb{Z}} f(2\pi k) = \lim_{j\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{|n|\leq j} \hat{f}(n).$$

Proof Since $[f]$ is an AC function on $[-\pi,\pi]$, by Lemma 7.1.2, from Theorem 5.9.6 we know that

$$[f](0) = \lim_{j\to\infty} S_j([f],0) = \lim_{j\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{|k|\leq j} \widehat{[f]}(k),$$

where $\widehat{[f]}(k)$, $k \in \mathbb{Z}$, are the Fourier coefficients of $[f]$; but

$$\begin{aligned}\widehat{[f]}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} [f](x)e^{-ikx}dx = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \sum_{n\in\mathbb{Z}} f(x+2\pi n)e^{-ikx}dx \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} f(x+2\pi n)e^{-ikx}dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)e^{-ikx}dx = \hat{f}(k),\end{aligned}$$

hence, $[f](0) = \sum_{n\in\mathbb{Z}} f(2\pi n) = \lim_{j\to\infty} \frac{1}{\sqrt{2\pi}} \sum_{|k|\leq j} \hat{f}(k)$. ■

Theorem 7.1.5 (Poisson summation formula) *If $f \in W^{2,1}(\mathbb{R})$, then $\{\hat{f}(n)\}$ is summable and*

$$\sum_{n\in\mathbb{Z}} f(2\pi n) = \frac{1}{\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \hat{f}(n). \tag{7.3}$$

Proof In view of Lemma 7.1.3, it is sufficient to show that $\{\hat{f}(n)\}$ is summable.

Since $f \in W^{2,1}(\mathbb{R})$, $f' \in W^{1,1}(\mathbb{R})$. Then $[f'] = [f]'$ is AC and is therefore in $L^2[-\pi,\pi]$. Now,

$$\widehat{[f]}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} [f](x)e^{-ikx} = \frac{i}{k} \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} [f]'(x)e^{-ikx}dx = \frac{i}{k}\widehat{[f]'}(k),$$

if $k \neq 0$ (note $[f](-\pi) = [f](\pi)$), hence,

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \left|\widehat{[f]}(k)\right| &= \left|\widehat{[f]}(0)\right| + \sum_{k\neq 0} \frac{1}{|k|} \left|\widehat{[f]'}(k)\right| \\ &\leq \left|\widehat{[f]}(0)\right| + \left(\sum_{k\neq 0} \frac{1}{k^2}\right)^{\frac{1}{2}} \left(\sum_{k\in\mathbb{Z}} \left|\widehat{[f]'}(k)\right|^2\right)^{\frac{1}{2}} < \infty,\end{aligned}$$

because $\sum_{k\in\mathbb{Z}} |\widehat{[f]'}(k)|^2 = \|[f]'\|_2^2$. Thus $\{\widehat{[f]}(n)\}$ is summable. We have shown in the proof of Lemma 7.1.3 that $\hat{f}(n) = \widehat{[f]}(n)$ for $n \in \mathbb{Z}$, hence $\{\hat{f}(n)\}$ is summable. ■

Example 7.1.3 Let $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, and for $t > 0$ let $g_t(x) = \frac{1}{\sqrt{2\pi}t} e^{-\frac{x^2}{2t^2}}$, $x \in \mathbb{R}$. The family $\{g_t\}$ is called the **Gauss kernel**. From Example 7.1.1 (ii) and Exercise 7.1.4 (ii), $\hat{g}_t(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2\xi^2}{2}}$. Using (7.3), we conclude that

$$\frac{1}{\sqrt{2\pi}\,t} \sum_{n\in\mathbb{Z}} e^{-\frac{(2\pi n)^2}{2t^2}} = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} e^{-\frac{n^2t^2}{2}},$$

from which on replacing t by $2\pi\sqrt{t}$, we have

$$\frac{1}{\sqrt{2\pi t}} \sum_{n\in\mathbb{Z}} e^{-\frac{n^2}{2t}} = \sum_{n\in\mathbb{Z}} e^{-2\pi^2 n^2 t}. \tag{7.4}$$

The relation (7.4) is Jacobi's identity for the theta function θ,

$$\theta(t) = t^{-\frac{1}{2}} \theta\left(\frac{1}{t}\right), \quad t > 0, \tag{7.5}$$

where $\theta(t) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t}$.

Example 7.1.4 Consider the **Poisson kernel** $P_t(x) = \frac{1}{\pi}\frac{t}{t^2+x^2}$, $t > 0$, $x \in \mathbb{R}$. From the Cauchy integral formula, if $\xi > 0$ and $y < -t$,

$$\int_{\mathbb{R}} \frac{e^{-i\xi x}dx}{(x-it)(x+it)} - \int_{\mathbb{R}} \frac{e^{-i\xi(x+iy)}dx}{(x+iy-it)(x+iy+it)} = 2\pi i\left(\frac{e^{-\xi t}}{-2it}\right) = \frac{\pi}{t}e^{-\xi t},$$

where $\frac{e^{-\xi t}}{-2it}$ is the value of the function $\frac{e^{-i\xi z}}{z-it}$ at $z = -it$. But,

$$\int_{\mathbb{R}} \frac{e^{-i\xi(x+iy)}dx}{(x+iy)^2+t^2} = e^{\xi y}\int_{\mathbb{R}} \frac{e^{-i\xi x}}{(x+iy)^2+t^2}dx \to 0$$

as $y \to -\infty$. Hence $\int_{\mathbb{R}} \frac{e^{-i\xi x}}{x^2+t^2}dx = \frac{\pi}{t}e^{-\xi t}$ if $\xi > 0$.

If $\xi < 0$, take $y > t$ and then let $y \to \infty$; we obtain $\int_{\mathbb{R}} \frac{e^{-i\xi x}}{x^2+t^2}dx = \frac{\pi}{t}e^{\xi t}$ by the same argument. Thus $\widehat{P}_t(\xi) = \frac{1}{\sqrt{2\pi}}\frac{t}{\pi}\int_{\mathbb{R}} \frac{e^{-i\xi x}}{x^2+t^2}dx = \frac{1}{\sqrt{2\pi}}e^{-|\xi|t}$. Apply (7.3); we have

$$\sum_{n\in\mathbb{Z}} P_t(2n\pi) = \frac{t}{\pi}\sum_{n\in\mathbb{Z}} \frac{1}{t^2+(2n\pi)^2} = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} e^{-|n|t} = \frac{1}{2\pi}\frac{1+e^{-t}}{1-e^{-t}}, \tag{7.6}$$

or

$$\sum_{n\in\mathbb{Z}} \frac{1}{t^2+n^2} = \frac{\pi}{t}\frac{1+e^{-2\pi t}}{1-e^{-2\pi t}}$$

on replacing t by $2\pi t$. When $t \to 0+$, (7.6) becomes $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

Exercise 7.1.6 Show that $\int_{\mathbb{R}} e^{-|\xi|t}e^{i\xi x}d\xi = 2\pi P_t(x)$ and verify that $\widehat{P}_t(\xi) = \frac{1}{\sqrt{2\pi}}e^{-|\xi|t}$.

7.2 Fourier integral on L^2

The Fourier integral for L^2 functions will be defined by using properties of the Fourier integral operator on the space of rapidly decreasing functions.

Denote by $\mathcal{S}$ the space of all complex-valued functions f in $C^\infty(\mathbb{R}^n)$ such that for all multi-indices α and β

$$P_{\alpha\beta}(f) := \sup_{x\in\mathbb{R}^n} |x^\alpha \partial^\beta f(x)| < \infty,$$

where $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $\partial^\beta f(x) = \frac{\partial^{|\beta|} f}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}(x)$.

$\mathcal{S}$ is called the **Schwartz space** in $\mathbb{R}^n$, and functions in $\mathcal{S}$ are usually referred to as **rapidly decreasing** functions. For each pair α, β of multi-indices, $P_{\alpha\beta}(\cdot)$ is a semi-norm on $\mathcal{S}$. Note that (i) $\mathcal{D} := C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}$; and (ii) the function $e^{-\frac{|x|^2}{2}}$ is in $\mathcal{S}$.

Define a metric ρ on $\mathcal{S}$ by

$$\rho(f,g) = \sum_{\alpha,\beta} \frac{1}{e^{|\alpha|}e^{|\beta|}} \cdot \frac{P_{\alpha\beta}(f-g)}{1+P_{\alpha\beta}(f-g)} \cdot \frac{1}{n^{|\alpha|}n^{|\beta|}}. \tag{7.7}$$

Since $\{\frac{1}{e^{|\alpha|}e^{|\beta|}} \cdot \frac{P_{\alpha\beta}(f-g)}{1+P_{\alpha\beta}(f-g)} \cdot \frac{1}{n^{|\alpha|}n^{|\beta|}}\}_{\alpha,\beta}$ is summable with sum $\le \sum_{j,k\ge 0} \frac{1}{e^j e^k}$, $\rho(f,g)$ is a nonnegative finite number.

Exercise 7.2.1 Show that ρ is actually a metric on $\mathcal{S}$.

We observe first the following elementary inequalities:

$$\begin{aligned} (1+|x|)^N &\le 2^N(1+|x|^N), \quad N \ge 0,\ x \in \mathbb{R}^n; \\ |x|^N &\le \delta^{-1}\sum_{j=1}^n |x_j|^N, \quad N \ge 0,\ x \in \mathbb{R}^n, \end{aligned} \tag{7.8}$$

where $\delta = \min_{|x|=1} \sum_{j=1}^n |x_j|^N$. For the first one, we may assume that $|x| > 1$, then $(1+|x|)^N \le (2|x|)^N < 2^N(1+|x|^N)$; while the second inequality follows by first considering the case $|x| = 1$ and then reducing the general case to this particular case.

Proposition 7.2.1 *For $f \in \mathcal{S}$, $x^\alpha\partial^\beta f \in L^1$ for any multi-indices α and β.*

Proof

$$\begin{aligned} \int_{\mathbb{R}^n} |x^\alpha \partial^\beta f(x)|dx &= \int_{\mathbb{R}^n} |x^\alpha|(1+|x|^{n+1})|\partial^\beta f(x)|\frac{1}{1+|x|^{n+1}}dx \\ &\le \int_{\mathbb{R}^n} |x^\alpha|\left(1+\delta^{-1}\sum_{j=1}^n |x_j|^{n+1}\right)|\partial^\beta f(x)|\frac{1}{1+|x|^{n+1}}dx \\ &\le M\int_{\mathbb{R}^n} \frac{1}{1+|x|^{n+1}}dx < \infty, \end{aligned}$$

for some $M > 0$, where $\delta = \min_{|x|=1}\sum_{j=1}^n |x_j|^{n+1}$ (cf. (7.8)). ∎

Now let $f \in \mathcal{S}$, then $f \in L^1$ by Proposition 7.2.1, and $\hat{f}$ is defined. We show the existence of $\frac{\partial}{\partial \xi_j}\hat{f}(\xi)$ as follows. Consider for $h \neq 0$ the difference quotient

$$\frac{\hat{f}(\xi_1, \ldots, \xi_j + h, \ldots, \xi_n) - \hat{f}(\xi)}{h} = \frac{1}{(2\pi)^{n/2}} \int f(x) e^{-i\xi \cdot x} \frac{(e^{-ihx_j} - 1)}{h} dx.$$

Since $|\frac{e^{-ihx_j}-1}{h}| \leq |x_j|$ and $|x_j||f| \in L^1$, by Proposition 7.2.1, it follows from LDCT that $\frac{\partial}{\partial \xi_j}\hat{f}(\xi)$ exists and

$$\frac{\partial}{\partial \xi_j}\hat{f}(\xi) = (-i)\widehat{x_j f}(\xi).$$

By Proposition 7.2.1, we can repeat the above argument with f replaced by $x_j f$, and obtain for any multi-index α the following formula:

$$\partial_\xi^\alpha \hat{f}(\xi) = (-i)^{|\alpha|}\widehat{x^\alpha f}(\xi). \tag{7.9}$$

Since $x^\alpha f \in L^1$ and $\widehat{x^\alpha f}$ is uniformly continuous, Proposition 7.2.2 then follows from (7.9).

Proposition 7.2.2 *If $f \in \mathcal{S}$, then $\hat{f} \in C^\infty(\mathbb{R}^n)$.*

Using the Fubini theorem and integration by parts, one asserts

$$\widehat{\partial^\beta f}(\xi) = (i)^{|\beta|}\xi^\beta \hat{f}(\xi) \tag{7.10}$$

for any multi-index β. Combining (7.9) and (7.10), one obtains

$$(i)^{|\alpha+\beta|}\xi^\beta \partial_\xi^\alpha \hat{f}(\xi) = \widehat{\partial_x^\beta (x^\alpha f)}(\xi) \tag{7.11}$$

for any multi-indices α and β.

Theorem 7.2.1 *$F\mathcal{S} \subset \mathcal{S}$, and F is a continuous map with respect to the metric ρ on $\mathcal{S}$ defined by* (7.7).

Proof That $f \in \mathcal{S}$ implies that $\hat{f} \in \mathcal{S}$ follows directly from (7.11):

$$\sup_{\xi \in \mathbb{R}^n} |\xi^\beta \partial_\xi^\alpha \hat{f}(\xi)| \leq \|\widehat{\partial_x^\beta (x^\alpha f)}\|_\infty \leq \|\partial_x^\beta (x^\alpha f)\|_1 < \infty.$$

To see that F is continuous, first observe that a sequence $\{f_k\} \subset \mathcal{S}$ converges to $f \in \mathcal{S}$ in the metric ρ defined by (7.7) if and only if $\lim_{k\to\infty} P_{\alpha\beta}(f_k - f) = 0$ for each pair α, β of multi-indices. Now from (7.11),

$$P_{\beta\alpha}(\hat{f}_k - \hat{f}) \leq \|\widehat{\partial_x^\beta [x^\alpha (f_k - f)]}\|_\infty \leq \|\partial_x^\beta [x^\alpha (f_k - f)]\|_1;$$

observe that if $\rho(f_k,f) \to 0$, then $\partial_x^\beta[x^\alpha(f_k(x)-f(x))] \to 0$ uniformly on $\mathbb{R}^n$ and

$$|\partial_x^\beta[x^\alpha(f_k(x)-f(x))]| \le |(1+|x|^{n+1})\partial_x^\beta[x^\alpha(f_k(x)-f(x))]|\frac{1}{1+|x|^{n+1}}$$
$$\le M\frac{1}{1+|x|^{n+1}}.$$

LDCT can be applied to obtain $\lim_{k\to\infty}\|\partial_x^\beta[x^\alpha(f_k-f)]\|_1 = 0$, implying that $\lim_{k\to\infty} P_{\beta\alpha}(\hat{f}_k-\hat{f}) = 0$ and consequently $\rho(\hat{f}_k,\hat{f}) \to 0$. ■

Since $\check{F}f = F\tilde{f}$, where $\tilde{f}(x) = f(-x)$, $\check{F}$ is also a continuous map from $\mathcal{S}$ to $\mathcal{S}$ w.r.t. the metric defined by (7.7).

Taking into account Theorem 7.1.4 and the fact that $\mathcal{S} \subset L^1$, we conclude that Theorem 7.2.2 holds.

Theorem 7.2.2 (Fourier inversion theorem) *Both F and $\check{F}$ are continuous and bijective from $\mathcal{S}$ to $\mathcal{S}$ and $\check{F}(Ff) = f = F(\check{F}f)$ for $f \in \mathcal{S}$.*

Theorem 7.2.3 (Parseval relations) *For f, g in $\mathcal{S}$ the following relations hold:*

(i) $\int \hat{f}g d\lambda^n = \int f\hat{g} d\lambda^n$;

(ii) $\int f\bar{g} d\lambda^n = \int \hat{f}\bar{\hat{g}} d\lambda^n$.

Proof (i) is the conclusion of Exercise 7.1.5; (ii) follows from (i) by replacing f and g by $\check{f}$ and $\bar{g}$ respectively. ■

Exercise 7.2.2 Let $(f,g) = \int f\bar{g}d\lambda^n$ be the L^2 inner product of f and g in $\mathcal{S}$. Show that (i) and (ii) in Theorem 7.2.3 are equivalent and are equivalent to any of the following relations:

(a) $(\hat{f},g) = (f,\check{g})$;

(b) $(f,g) = (\hat{f},\hat{g})$.

We are ready to define the Fourier integral for functions in L^2. Since $C_c^\infty(\mathbb{R}^n)$ is dense in L^p, $1 \le p < \infty$, and $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}$, $\mathcal{S}$ is dense in L^2. For $f \in L^2$, there is a sequence $\{f_k\}$ in $\mathcal{S}$ such that $\lim_{k\to\infty}\|f_k-f\|_2 = 0$; a fortiori, $\{f_k\}$ is a Cauchy sequence in L^2. By relation (b) in Exercise 7.2.2, $\|f_k-f_l\|_2^2 = \|\hat{f}_k-\hat{f}_l\|_2^2$ for all k, l in $\mathbb{N}$, therefore $\{\hat{f}_k\}$ is a Cauchy sequence in L^2 and converges in L^2 to $g \in L^2$. We claim that g is independent of the sequence $\{f_k\}$ in $\mathcal{S}$, which converges to f in L^2. Suppose that $\{g_k\}$ is another sequence in $\mathcal{S}$ that converges to f in L^2; then $\lim_{k\to\infty}\|f_k-g_k\|_2 = 0$, but $\|\hat{f}_k-\hat{g}_k\|_2 = \|f_k-g_k\|_2$ implies that $\lim_{k\to\infty}\hat{g}_k = \lim_{k\to\infty}\hat{f}_k = g$ in L^2. Thus g is uniquely determined by f in the way we specify; we then denote g by $\hat{f}'$ for the moment. From the definition, one verifies readily that $(f,g) = (\hat{f}',\hat{g}')$ for f, g in L^2.

Lemma 7.2.1 *If $f \in L^1 \cap L^2$, then $\hat{f} = \hat{f}'$.*

Proof Fix a Friederich mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ constructed from a mollifying function $\varphi \geq 0$. For $\varepsilon > 0$, let $f_\varepsilon = f I_{B_{1/\varepsilon}(0)}$. Then $J_\varepsilon f_\varepsilon \in C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}$. We claim that $J_\varepsilon f_\varepsilon \to f$ in both L^1 and L^2. Actually for $p = 1$ or 2, we have

$$\begin{aligned}\|J_\varepsilon f_\varepsilon - f\|_p &\leq \|J_\varepsilon(f_\varepsilon - f)\|_p + \|J_\varepsilon f - f\|_p \\ &\leq \|f_\varepsilon - f\|_p + \|J_\varepsilon f - f\|_p \to 0\end{aligned}$$

as $\varepsilon \to 0$. From $\|J_\varepsilon f_\varepsilon - f\|_1 \to 0$, as $\varepsilon \to 0$, we infer that $\widehat{J_\varepsilon f_\varepsilon} \to \hat{f}$ uniformly on $\mathbb{R}^n$; while from $\|J_\varepsilon f_\varepsilon - f\|_2 \to 0$, as $\varepsilon \to 0$, we conclude that $\widehat{J_\varepsilon f_\varepsilon} \to \hat{f}'$ in L^2 and, consequently, there is a sequence of ε tending to zero such that $\widehat{J_\varepsilon f_\varepsilon} \to \hat{f}'$ a.e. on $\mathbb{R}^n$. Hence $\hat{f} = \hat{f}'$ a.e. ■

Because of Lemma 7.2.1, it is natural to call $\hat{f}'$ the Fourier integral of f in L^2 and also denote $\hat{f}'$ by $\hat{f}$. Similarly $\check{f}$ is also defined for $f \in L^2$. We shall also use F and $\check{F}$ to denote the maps $f \mapsto \hat{f}$ and $f \mapsto \check{f}$ respectively from L^2 onto L^2. Note that $(f, g) = (\hat{f}, \hat{g}) = (\check{f}, \check{g})$ for f, g in L^2.

Exercise 7.2.3 Show that both F and $\check{F}$ are linear bijective isometries from L^2 onto itself and $\check{F} = F^{-1}$.

Exercise 7.2.4 Suppose that $f \in W^{k,2}(\mathbb{R}^n)$, $k \in \mathbb{N}$. Show that $\widehat{\partial^\alpha f}(\xi) = (i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)$ for a.e. $\xi \in \mathbb{R}^n$ if $|\alpha| \leq k$. (Hint: $W^{k,2}(\mathbb{R}^n) = \overset{\circ}{W}{}^{k,2}(\mathbb{R}^n)$.)

7.3 The Sobolev space H^s

For each $s \in \mathbb{R}$, an inner product $(\cdot,\cdot)_s$ on $\mathcal{S}$ is defined by

$$(f, g)_s = \int (1 + |\xi|^2)^s \hat{f}(\xi)\overline{\hat{g}(\xi)}d\xi;$$

and the associated norm on $\mathcal{S}$ is denoted by $|\cdot|_s$. Thus,

$$|f|_s = \left(\int (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 d\xi\right)^{\frac{1}{2}}.$$

As usual, $(f, g) = \int f\bar{g}d\lambda^n$ is the inner product of f and g in L^2.

A few basic properties of inner products $(\cdot,\cdot)_s$ are now listed.

(1) $(f, g) = (f, g)_0$.

(2) $|(f, g)_0| \leq |f|_s|g|_{-s}$. This follows directly from

$$(f, g)_0 = \int (1 + |\xi|^2)^{s/2}\hat{f}(\xi)(1 + |\xi|^2)^{-s/2}\hat{g}(\xi)d\xi,$$

by Schwarz's inequality.

(3) $|f|_s = \max_{\substack{g\in\mathcal{S}\\ g\neq 0}} \frac{|(f,g)_0|}{|g|_{-s}}$. To see this, one observes first from (2) that

$$|f|_s \geq \sup_{\substack{g\in\mathcal{S}\\ g\neq 0}} \frac{|(f,g)_0|}{|g|_{-s}};$$

now, since $(1+|\xi|^2)^s\hat{f}(\xi) \in \mathcal{S}$, there is $h \in \mathcal{S}$ such that $\hat{h}(\xi) = (1+|\xi|^2)^s\hat{f}(\xi)$, and hence,

$$\begin{aligned} |h|_{-s}^2 &= \int (1+|\xi|^2)^{-s}(1+|\xi|^2)^{2s}|\hat{f}(\xi)|^2 d\xi \\ &= \int (1+|\xi|^2)^s|\hat{f}(\xi)|^2 d\xi = |f|_s^2; \\ (f,h)_0 &= \int (1+|\xi|^2)^s|\hat{f}(\xi)|^2 d\xi = |f|_s^2, \end{aligned}$$

resulting in $\frac{|(f,h)_0|}{|h|_{-s}} = |f|_s$.

(4) $|\partial^\alpha f|_s \leq |f|_{s+|\alpha|}$. This is obvious by (7.10).

The Sobolev space H^s is the completion of $\mathcal{S}$ under the norm $|\cdot|_s$. The Sobolev space H^s is a Hilbert space for each $s \in \mathbb{R}$. Observe that in the case $s \geq 0$, if $\{f_k\}$ is a Cauchy sequence in $\mathcal{S}$ in the norm $|\cdot|_s$, then it is a Cauchy sequence in L^2, hence it is legitimate to identify each element of H^s, $s \geq 0$, with an element of L^2. Those elements of L^2 which belong to H^s can be characterized as follows.

Theorem 7.3.1 *An element f of L^2 is in H^s, $s \geq 0$, if and only if there is a sequence $\{f_k\} \subset \mathcal{S}$ such that $\|f_k - f\|_2 \to 0$ as $k \to \infty$ and $\sup_k |f_k|_s < \infty$.*

Proof If $f \in H^s$, there is $\{f_k\} \subset \mathcal{S}$ such that $|f_k - f|_s \to 0$ as $k \to \infty$, a fortiori, $\|f_k - f\|_2 \to 0$ as $k \to \infty$ and $\sup_k |f_k|_s < \infty$.

Conversely, suppose that there is a sequence $\{f_k\} \subset \mathcal{S}$ such that $\|f_k - f\|_2 \to 0$ and $\sup_k |f_k|_s < \infty$. By the Banach–Saks theorem (Theorem 5.10.2), there is a subsequence $\{g_k\}$ of $\{f_k\}$ and g in H^s such that $|\frac{1}{N}\sum_{k=1}^N g_k - g|_s \to 0$ as $N \to \infty$, a fortiori, $\|\frac{1}{N}\sum_{k=1}^N g_k - g\|_2 \to 0$. But $\|g_k - f\|_2 \to 0$ implies that $\|\frac{1}{N}\sum_{k=1}^N g_k - f\|_2 \to 0$, and consequently $f = g$. Thus $f \in H^s$. ■

Exercise 7.3.1 Show that if k is a nonnegative integer, then $W^{k,2}(\mathbb{R}^n) = H^k$, in the sense that $W^{k,2}(\mathbb{R}^n) = H^k$ as set and the norms $\|\cdot\|_{k,2}$ and $|\cdot|_k$ are equivalent.

We will now show that in tempo with s becoming larger, elements of H^s become smoother. This is the content of the Sobolev lemma.

A preliminary lemma is shown first.

Lemma 7.3.1 *Suppose that $s \in \mathbb{R}$ and k is a nonnegative integer such that $s - k > \frac{n}{2}$; then there is $C > 0$ such that*

$$\max_{x\in\mathbb{R}^n} \sum_{|\alpha|\le k} |\partial^\alpha f(x)| \le C|f|_s$$

for $f \in \mathcal{S}$.

Proof Since $\widehat{\partial^\alpha f}(\xi) = (i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)$, $\widehat{\partial^\alpha f}$ is in $\mathcal{S}$; it follows from Fourier's Inversion theorem (Theorem 7.2.2) that

$$\begin{aligned}\partial^\alpha f(x) &= (2\pi)^{-\frac{n}{2}} \int e^{ix\cdot\xi}(i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)d\xi \\ &= (2\pi)^{-\frac{n}{2}}(i)^{|\alpha|} \int e^{ix\cdot\xi}\xi^\alpha(1+|\xi|^2)^{\frac{(s-k)}{2}}\hat{f}(\xi)(1+|\xi|^2)^{\frac{(k-s)}{2}}d\xi,\end{aligned}$$

and hence, when $|\alpha| \le k$,

$$\begin{aligned}|\partial^\alpha f(x)|^2 &\le (2\pi)^{-n}\int |\xi^\alpha|^2(1+|\xi|^2)^{s-k}|\hat{f}(\xi)|^2 d\xi \cdot \int (1+|\xi|^2)^{k-s}d\xi \\ &\le C' \int \frac{|\xi|^{2|\alpha|}}{(1+|\xi|^2)^k}(1+|\xi|^2)^s|\hat{f}(\xi)|^2 d\xi \\ &\le C|f|_s^2,\end{aligned}$$

where we have used the obvious fact that $\int (1+|\xi|^2)^{k-s}d\xi < \infty$. Thus,

$$\max_{x\in\mathbb{R}^n} \sum_{|\alpha|\le k} |\partial^\alpha f(x)| \le C|f|_s,$$

with $C > 0$ depending only on s, k, and n. ■

Theorem 7.3.2 (Sobolev lemma) *Suppose that $s \in \mathbb{R}$ and k is a nonnegative integer such that $s - k > \frac{n}{2}$; then $H^s \subset C^k(\mathbb{R}^n)$.*

Proof Consider f in H^s. There is a sequence $\{f_k\} \subset \mathcal{S}$ such that $|f_k - f|_s \to 0$ as $k \to \infty$; $\{f_k\}$ is therefore a Cauchy sequence in H^s. From Lemma 7.3.1, there is $C > 0$ such that

$$\max_{x\in\mathbb{R}^n} \sum_{|\alpha|\le k} |\partial^\alpha(f_m(x) - f_l(x))| \le C|f_m - f_l|_s \to 0$$

as $m, l \to \infty$, which means that $\{f_k\}$ converges uniformly on $\mathbb{R}^n$ to a function g in $C^k(\mathbb{R}^n)$. Then, $\int_{|x|\le R} |f_k(x) - g(x)|^2 dx \to 0$ as $k \to \infty$ for any $R > 0$;

but $\lim_{k\to\infty} |f_k - f|_s = 0$ implies that $\lim_{k\to\infty} \|f_k - f\|_2 = 0$, and therefore that $\int_{|x|\le R} |f_k(x) - f(x)|^2 dx \to 0$ as $k \to \infty$. Now,

$$\left\{\int_{|x|\le R} |f(x) - g(x)|^2 dx\right\}^{\frac{1}{2}} \le \left\{\int_{|x|\le R} |f(x) - f_k(x)|^2 dx\right\}^{\frac{1}{2}} + \left\{\int_{|x|\le R} |f_k(x) - g(x)|^2 dx\right\}^{\frac{1}{2}} \to 0$$

as $k \to \infty$, hence, $\int_{|x|\le R} |f(x) - g(x)|^2 dx = 0$ and consequently $f = g$ a.e. on $B_R(0)$. Since $R > 0$ is arbitrary, $f = g$ a.e. on $\mathbb{R}^n$. ■

7.4 Weak solutions of the Poisson equation

We illustrate the use of Sobolev space in this section by considering the existence and regularity of weak solutions of the Poisson equation,

$$\Delta u = f. \tag{7.12}$$

A **classical solution** u of (7.12) on an open domain Ω of $\mathbb{R}^n$ is a function u, defined on Ω such that $\Delta u(x) = f(x)$ for a.e. x of Ω. If f is continuous and u is a C^2 classical solution of (7.12) on Ω, then for any $v \in C_c^\infty(\Omega)$, we have

$$\int_\Omega f v d\lambda^n = \int v \Delta u d\lambda^n = \int u \Delta v d\lambda^n.$$

Therefore, when f is locally integrable on Ω, a locally integrable function u on Ω is called a **weak solution** of (7.12) if

$$\int_\Omega f v d\lambda^n = \int u \Delta v d\lambda^n$$

for all $v \in C_c^\infty(\Omega)$.

Exercise 7.4.1 Show that a C^2 function u on Ω is a classical solution of (7.12) if and only if it is a weak solution of (7.12).

We shall first prove the following regularity result for weak solutions of (7.12).

Theorem 7.4.1 *Suppose that $f \in C^\infty(\Omega)$. Then any locally L^2 weak solution of (7.12) is in $C^\infty(\Omega)$.*

The proof of Theorem 7.4.1 is preceded by some preliminaries relating to Friederich mollifiers. We fix a Friederich mollifier $\{J_\varepsilon\}_{\varepsilon>0}$ with a mollifying function φ which is nonnegative and satisfies the symmetry property: $\varphi(-x) = \varphi(x)$ for all x in $\mathbb{R}^n$. For example, we may take φ to be the function defined by $\varphi(x) = c\exp\{-\frac{1}{1-|x|^2}\}$ if $|x| < 1$ and $\varphi(x) = 0$ if $|x| \geq 1$, where c is a positive constant chosen so that $\int \varphi d\lambda^n = 1$.

Lemma 7.4.1 *Let $\{J_\varepsilon\}$ be a Friederich mollifier as previously specified.*

(i) $\|J_\varepsilon f\|_p \leq \|f\|_p$ *for* $f \in L^p$, $1 \leq p < \infty$.

(ii) $(J_\varepsilon f, g) = (f, J_\varepsilon g)$ *for* $f, g \in L^2$.

(iii) *If* $f \in C^1(\mathbb{R}^n)$, *then* $\frac{\partial}{\partial x_j} J_\varepsilon f(x) = J_\varepsilon \frac{\partial f}{\partial x_j}(x)$ *for all* $x \in \mathbb{R}^n$ *and* $j = 1, \ldots, n$.

(iv) *If* $f \in \mathcal{S}$, *then* $J_\varepsilon f \in \mathcal{S}$ *and* $|J_\varepsilon f|_s \leq |f|_s$.

Proof (i) is known in Section 4.10; (ii) follows directly from the definition of J_ε and the assumption that $\varphi(-x) = \varphi(x)$; while (iii) is a consequence of applying LDCT to the difference quotient involved in the definition of partial derivatives; it remains to show (iv). Since $J_\varepsilon f = f * \varphi_\varepsilon$, $\widehat{J_\varepsilon f} = (2\pi)^{\frac{n}{2}} \hat{f} \cdot \hat{\varphi}_\varepsilon$, which implies immediately that $\widehat{J_\varepsilon f} \in \mathcal{S}$, but by the Fourier inversion theorem, $J_\varepsilon f = (\widehat{J_\varepsilon f})^\vee$ and hence $J_\varepsilon f \in \mathcal{S}$. Now,

$$
\begin{aligned}
|J_\varepsilon f|_s^2 &= \int (1+|\xi|^2)^s |\widehat{J_\varepsilon f}(\xi)|^2 d\xi = (2\pi)^n \int (1+|\xi|^2)^s |\hat{f}(\xi)|^2 |\hat{\varphi}_\varepsilon(\xi)|^2 d\xi \\
&\leq \|\varphi_\varepsilon\|_1 \int (1+|\xi|^2)^s |\hat{f}(\xi)|^2 d\xi = |f|_s^2.
\end{aligned}
$$

Hence $|J_\varepsilon f|_s \leq |f|_s$. ■

Lemma 7.4.2 *There is a constant $C > 0$ such that*

$$|v|_s \leq C(|\Delta v|_{s-2} + |v|_{s-1})$$

for all $v \in \mathcal{S}$.

Proof For $\xi \in \mathbb{R}^n$, we have

$$(1+|\xi|^2)^2 = 1 + 2|\xi|^2 + |\xi|^4 < |\xi|^4 + 2(1+|\xi|^2) < 2\{|\xi|^4 + (1+|\xi|^2)\},$$

hence,

$$
\begin{aligned}
|v|_s^2 &= \int (1+|\xi|^2)^s |\hat{v}(\xi)|^2 d\xi \\
&< 2\int (1+|\xi|^2)^{s-2} \{|\xi|^4 + (1+|\xi|^2)\} |\hat{v}(\xi)|^2 d\xi \\
&= 2\left\{\int (1+|\xi|^2)^{s-2} |\widehat{\Delta v}(\xi)|^2 d\xi + \int (1+|\xi|^2)^{s-1} |\hat{v}(\xi)|^2 d\xi\right\} \\
&= 2(|\Delta v|_{s-2}^2 + |v|_{s-1}^2) \\
&\leq 2(|\Delta v|_{s-2} + |v|_{s-1})^2,
\end{aligned}
$$

and consequently,

$$|v|_s \leq \sqrt{2}(|\Delta v|_{s-2} + |v|_{s-1}).$$

■

Proof of Theorem 7.4.1 For $x \in \Omega$, there is $g \in C_c^\infty(\Omega)$, which takes a constant value in the neighborhood of x; it is therefore sufficient to prove that $gu \in C^\infty(\mathbb{R}^n)$ for each $g \in C_c^\infty(\Omega)$.

Consider now any $g \in C_c^\infty(\Omega)$. In order to show that $gu \in C^\infty(\mathbb{R}^n)$, it is sufficient to show that $gu \in H^s$ for all $s \in \mathbb{N}$, by the Sobolev lemma (Theorem 7.3.2); but since $\|J_\varepsilon(gu) - gu\|_2 \to 0$ as $\varepsilon \to 0$, from Theorem 7.3.1, it is sufficient to show that given $g \in C_c^\infty(\Omega)$, for each $s \in \mathbb{N}$, there is a constant $C_s > 0$ such that

$$|J_\varepsilon(gu)|_s \leq C_s, \quad \varepsilon > 0. \tag{7.13}$$

When $s = 0$, (7.13) is a consequence of $\|J_\varepsilon(gu)\|_2 \leq \|gu\|_2$ (cf. Lemma 7.4.1 (i)). Suppose that (7.13) holds for $s - 1$, we are going to show that (7.13) holds for s. Using the Fubini theorem and integration by parts, we have for $v \in \mathcal{S}$,

$$\begin{aligned}(\Delta(J_\varepsilon(gu)), v) &= (J_\varepsilon(gu), \Delta v) = (gu, \Delta J_\varepsilon v) = (u, g(\Delta J_\varepsilon v)) \\ &= (u, \Delta(gJ_\varepsilon v)) - \left(u, 2\sum_{j=1}^n \frac{\partial g}{\partial x_j}\frac{\partial J_\varepsilon v}{\partial x_j} + J_\varepsilon v \cdot \Delta g\right) \\ &= (f, gJ_\varepsilon v) - 2\sum_{j=1}^n \left(J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right), \frac{\partial v}{\partial x_j}\right) + (J_\varepsilon(u\Delta g), v),\end{aligned}$$

where Lemma 7.4.1 has been applied. Hence,

$$\begin{aligned}&|(\Delta J_\varepsilon(gu), v)| \\ &\leq \left\{|J_\varepsilon(gf)|_{s-2}|v|_{2-s} + 2\sum_{j=1}^n \left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} \cdot \left|\frac{\partial v}{\partial x_j}\right|_{1-s} + |J_\varepsilon(u\Delta g)|_{s-1}|v|_{1-s}\right\} \\ &\leq |v|_{2-s}\left\{|J_\varepsilon(gf)|_{s-2} + 2\sum_{j=1}^n \left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1}\right\},\end{aligned}$$

where (2) and (4) in Section 7.3 are used. Thus, by (3) in Section 7.3, we conclude that

$$|\Delta(J_\varepsilon(gu))|_{s-2} \leq |J_\varepsilon(gf)|_{s-2} + 2\sum_{j=1}^n \left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1}. \tag{7.14}$$

Now from Lemma 7.4.2,

$$|J_\varepsilon(gu)|_s \leq C(|\Delta J_\varepsilon(gu)|_{s-2} + |J_\varepsilon(gu)|_{s-1}). \tag{7.15}$$

Substitute (7.14) into (7.15); we have

$$|J_\varepsilon(gu)|_s \le C'\left(|J_\varepsilon(gf)|_{s-2} + 2\sum_{j=1}^{n}\left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1} + |J_\varepsilon(gu)|_{s-1}\right).$$

But $|J_\varepsilon(gf)|_{s-2} \le |gf|_{s-2}$, by Lemma 7.4.1 (iv), and

$$2\sum_{j=1}^{n}\left|J_\varepsilon\left(u\frac{\partial g}{\partial x_j}\right)\right|_{s-1} + |J_\varepsilon(u\Delta g)|_{s-1} + |J_\varepsilon(gu)|_{s-1} \le C'_s,$$

by the assumption that (7.13) holds for $(s-1)$. Therefore,

$$|J_\varepsilon(gu)|_s \le C_s = |gf|_{s-2} + C'_s. \qquad \blacksquare$$

Regarding the existence of weak solutions of the Poisson equation (7.12), we now establish the existence and uniqueness of a weak solution of (7.12) in $\overset{\circ}{W}{}^{1,2}(\Omega)$ when Ω is bounded and $f \in L^2(\Omega)$.

Theorem 7.4.2 *Suppose that Ω is bounded and $f \in L^2(\Omega)$, then there is a unique weak solution of* (7.12) *in* $\overset{\circ}{W}{}^{1,2}(\Omega)$.

Proof It is only necessary to consider the case that f and solutions to be sought are real-valued; therefore $\overset{\circ}{W}{}^{1,2}(\Omega)$ is assumed to consist of real-valued functions. By the Poincaré inequality (Theorem 6.6.3), $\overset{\circ}{W}{}^{1,2}(\Omega)$ can be considered as a Hilbert space with the inner product

$$(u,v)'_1 = \sum_{j=1}^{n}\int_\Omega \frac{\partial u}{\partial x_j}\frac{\partial v}{\partial x_j}\,d\lambda^n$$

for u, v in $\overset{\circ}{W}{}^{1,2}(\Omega)$. Since $|(f,v)| \le \|f\|_2\|v\|_2 \le \|f\|_2\|v\|_{1,2} \le C\|f\|_2|v|_{1,2}$ for all $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$, by (6.31), the linear functional $v \mapsto -\int_\Omega fv\,d\lambda^n$ is a bounded linear functional on $\overset{\circ}{W}{}^{1,2}(\Omega)$; it then follows from the Riesz representation theorem that there is $u \in \overset{\circ}{W}{}^{1,2}(\Omega)$, such that

$$-\int_\Omega fv\,d\lambda^n = (v,u)'_1 = \sum_{j=1}^{n}\int_\Omega \frac{\partial v}{\partial x_j}\frac{\partial u}{\partial x_j}\,d\lambda^n$$

for $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$ and therefore for $v \in C_c^\infty(\Omega)$ in particular. But if $v \in C_c^\infty(\Omega)$,

$$\int_\Omega \frac{\partial v}{\partial x_j}\frac{\partial u}{\partial x_j}\,d\lambda^n = -\int_\Omega \frac{\partial^2 v}{\partial x_j^2}u\,d\lambda^n$$

for each $j = 1, \ldots, n$; thus we have

$$\int_\Omega fv d\lambda^n = \int_\Omega u \Delta v d\lambda^n$$

for $v \in C_c^\infty(\Omega)$. Hence u is a weak solution of (7.12).

Suppose now that $w \in \overset{\circ}{W}{}^{1,2}(\Omega)$ is also a weak solution of (7.12). Then,

$$\int_\Omega (u - w)\Delta v d\lambda^n = -\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial v}{\partial x_j} d\lambda^n = 0$$

for all $v \in C_c^\infty(\Omega)$. We claim now that

$$\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial v}{\partial x_j} d\lambda^n = 0$$

for all $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$. Let $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$; choose a sequence $\{v_k\}$ in $C_c^\infty(\Omega)$ such that $\lim_{k\to\infty} |v - v_k|_{1,2} = 0$; then,

$$\begin{aligned}
&\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial v}{\partial x_j} d\lambda^n \\
&= \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial v_k}{\partial x_j} d\lambda^n + \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial(v-v_k)}{\partial x_j} d\lambda^n \\
&= \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial(v-v_k)}{\partial x_j} d\lambda^n,
\end{aligned}$$

and consequently from Schwarz inequality,

$$\left| \sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial v}{\partial x_j} d\lambda^n \right| = |(u-w, v-v_k)_1'| \le |u-w|_{1,2} \cdot |v-v_k|_{1,2} \to 0$$

as $k \to \infty$. Hence,

$$\sum_{j=1}^n \int_\Omega \frac{\partial(u-w)}{\partial x_j} \frac{\partial v}{\partial x_j} d\lambda^n = 0$$

for $v \in \overset{\circ}{W}{}^{1,2}(\Omega)$. Since $u - w \in \overset{\circ}{W}{}^{1,2}(\Omega)$, we have

$$0 = \sum_{j=1}^n \int_\Omega \left[\frac{\partial(u-w)}{\partial x_j} \right]^2 d\lambda^n = |u-w|_{1,2}^2,$$

implying that $u = w$. Therefore, (7.12) has a unique weak solution in $\overset{\circ}{W}{}^{1,2}(\Omega)$. ■

7.5 Fourier integral of probability distributions

The Fourier integral of probability distributions will be discussed in this section, with an application to the central limit theorem in probability theory. This is preceded by a very brief introduction of the necessary basic notions, terminology, and notations in the probability theory, as formulated by **A.N. Kolmogoroff**.

Kolmogoroff's formulation of probability theory is based on measure theory. A measure space (Ω, Σ, P) with $P(\Omega) = 1$ is called a **probability space**, of which Ω is called the **sample space**; and sets in the σ-algebra Σ are called **events** (more precisely, measurable events); and for $A \in \Sigma$, $P(A)$ is referred to as the probability of event A. A measurable function on the sample space Ω is called a **random variable** (often abbreviated as r.v.). Random variables are usually denoted by capital Roman letters, such as $X, Y, Z, \ldots$ etc. It should be noted that a probability space is usually a construct suggested by first observations of outcomes of experiments on a random phenomenon; these outcomes are referred to as sample points and form the sample space Ω. Such a construct provides a solid mathematical framework to discuss questions related to the random phenomenon; such questions are usually addressed in terms of random variables. Our construction of the Bernoulli sequence space, starting with Section 1.3 and through Examples 1.7.1, 2.1.1, and 3.4.6, illustrates revealingly the point we just made. Henceforth, random variables are assumed to take finite real value P-almost everywhere and hence for a random variable, we always consider a finite real-valued version (for a probability space, P-almost everywhere is expressed as P-**almost surely** and is abbreviated as a.s.). Suppose that X is a random variable; if $\int_\Omega X dP$ exists, it is called the **expectation** of X and is denoted by $E(X)$; if $E(X)$ is finite, $\int_\Omega |X - E(X)|^2 dP$ is called the **variance** of X and is denoted by $\mathrm{Var}(X)$. The σ-algebra $\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}\}$ is the smallest subσ-algebra of Σ relative to which X is measurable; as implied by Exercise 2.5.10 (ii), if $E(X)$ is finite, the family $\{\int_A X dP : A \in \sigma X\}$ characterizes the r.v. X, or intuitively, $\{\int_A X dP : A \in \sigma(X)\}$ is the information one obtains by observing the r.v. X. This suggests considering $\sigma(X)$ as where the information regarding X resides. Accordingly, the σ-algebra Σ is where information on all random variables resides. As we know, in Example 4.3.2, the Bernoulli sequence space and $([0,1], \mathcal{B}|[0,1], \lambda)$ are measure-theoretically the same space, hence the choice of probability space is for convenience, and not of primary importance.

The most simple but fundamental notion in probability theory is that of independence. We shall discuss independence at some length to give a touch of the flavor of a basic aspect of probabilistic argument; however the notion of conditioning, basic and fundamental as it is, will not be touched upon here.

In the following, random variables are in reference to a fixed probability space (Ω, Σ, P) and σ-algebras on Ω are always subσ-algebras of Σ. A finite family $\{\Sigma_1, \ldots, \Sigma_k\}$ of σ-algebras on Ω is said to be **independent** if for any choice of $A_j \in \Sigma_j$, $j = 1, \ldots, k$, $P(\bigcap_{j=1}^k A_j) = \prod_{j=1}^k P(A_j)$ holds. A family $\{\Sigma_\alpha\}$ of σ-algebras on Ω is said to be **independent** if all of its finite subfamilies are independent. If $\{\Sigma_\alpha\}$ is independent, then $\Sigma'_\alpha s$ are said to be independent. For a family $\{A_\alpha\}$ of events, the σ-algebra $\sigma(\{A_\alpha\})$ is abbreviated to $\sigma(A'_\alpha s)$; in particular, if $A \in \Sigma$, $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$. Events

A_α, $\alpha \in I$, are said to be independent if $\{\sigma(A_\alpha)\}_{\alpha\in I}$ is independent. It is readily verified that events $A'_\alpha s$ are independent if and only if for any finite set of indices $\alpha_1, \ldots, \alpha_k$, $P(\bigcap_{l=1}^k A_{\alpha_l}) = \prod_{l=1}^k P(A_{\alpha_l})$.

Given a sequence $A_1, A_2, \ldots$ of events, let $\mathcal{T} = \mathcal{T}(A_1, A_2, \ldots) = \bigcap_{n=1}^\infty \sigma(A_n, A_{n+1}, \ldots)$. Events in $\mathcal{T}$ are referred to as **tail events** of the sequence $\{A_n\}$. It is evident that $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$ are tail events of the sequence $\{A_n\}$. The following zero-one law of Kolmogoroff is a far-reaching consequence of the notion of independence.

Theorem 7.5.1 (Kolmogoroff's zero–one law) *If $A_1, A_2, A_3, \ldots$ are independent events, then every tail event of $\{A_n\}$ has probability zero or one.*

Proof Suppose that A is a tail event of the sequence $\{A_n\}$. For $n \geq 2$, let $\mathcal{L}$ be the family of all such $B \in \Sigma$, with the property that

$$P(B_1 \cap \cdots \cap B_{n-1} \cap B) = P(B_1) \cdots P(B_{n-1})P(B),$$

where for each $j = 1, \ldots, n-1$, $B_j = A_j$ or Ω; then $\mathcal{L}$ is a λ-system. Next, let $\mathcal{P}$ be the family of all finite intersections of $A_n, A_{n+1}, \ldots$; $\mathcal{P}$ is then a π-system and $\mathcal{P} \subset \mathcal{L}$. Hence $\sigma(\mathcal{P}) \subset \mathcal{L}$ by the (π-λ) theorem. But $A \in \sigma(A_n, A_{n+1}, \ldots) = \sigma(\mathcal{P}) \subset \mathcal{L}$; this means that $A, A_1, \ldots, A_{n-1}$ are independent.

We now claim that $P(A \cap B) = P(A)P(B)$ for $B \in \sigma(A_1, A_2, \ldots)$. For this purpose, let $\mathcal{L}' = \{B \in \Sigma : P(A \cap B) = P(A)P(B)\}$ and $\mathcal{P}'$ be the family of all finite intersections of $A_1, A_2, \ldots$; clearly, $\mathcal{L}'$ is a λ-system and $\mathcal{P}'$ a π-system. The fact that $A, A_1, \ldots, A_{n-1}$ are independent for each $n \geq 2$ implies that $\mathcal{P}' \subset \mathcal{L}'$. Thus, $\sigma(\mathcal{P}') = \sigma(A_1, A_2, \ldots) \subset \mathcal{L}'$, by the ($\pi$-$\lambda$) theorem, which means that $P(A \cap B) = P(A)P(B)$ for $B \in \sigma(A_1, A_2, \ldots)$; but since $A \in \sigma(A_1, A_2, \ldots)$, $P(A) = P(A)^2$. Hence $P(A) = 0$ or 1. ■

Exercise 7.5.1 Let $\mathcal{T} = \bigcap_n \sigma(A_n, A_{n+1}, \ldots)$, where $A_1, A_2, \ldots$ are independent events. Show that if X is a $\mathcal{T}$-measurable random variable, then X = constant a.s.

In accord with notations for certain sets introduced in the second paragraph of Section 2.2, if T is a map from a set Ω to a set S, the set $T^{-1}A$, $A \subset S$, will be denoted by $\{T \in A\}$; and if $T_\alpha : \Omega \to S_\alpha$, $\alpha \in I$, then $\bigcap_{\alpha\in I} T_\alpha^{-1} A_\alpha$, $A_\alpha \subset S_\alpha$, is denoted by $\{T_\alpha \in A_\alpha, \alpha \in I\}$; in particular, if $X_1, \ldots, X_k$ are random variables, then $\bigcap_{j=1}^k \{X_j \in B_j\} = \{X_1 \in B_1, \ldots, X_k \in B_k\}$. When a probability measure P is concerned, $P(\{\cdots\})$ will be abbreviated to $P(\cdots)$.

Given a family $\{X_\alpha\}$ of r.v.'s, the smallest σ-algebra relative to which every X_α is measurable is denoted by $\sigma(X'_\alpha s)$; in particular, $\sigma(X_1, \ldots, X_k)$ is the smallest σ-algebra relative to which $X_1, \ldots, X_k$ are measurable.

Exercise 7.5.2 If $X_1, \ldots, X_k$ are r.v.'s, let $X = (X_1, \ldots, X_k)$ be the map from Ω to $\mathbb{R}^k$ defined by $X(\omega) = (X_1(\omega), \ldots, X_k(\omega))$ for $\omega \in \Omega$. Show that $\sigma(X_1, \ldots, X_k) = \{X^{-1}B : B \in \mathcal{B}^k\}$.

We shall call a map $X : \Omega \to \mathbb{R}^k$, $k \geq 2$ a **random vector** if $X^{-1}B \in \Sigma$ for all $B \in \mathcal{B}^k$; in other words, X is a random vector if X is $\Sigma|\mathcal{B}^k$-measurable. Put $X = (X_1, \ldots, X_k)$, where $X_1, \ldots, X_k$ are the component functions of X. Since $\{X^{-1}B : B \in \mathcal{B}^k\} \supset \bigcup_{j=1}^k \{X_j^{-1}B_j : B_j \in \mathcal{B}\}$, we conclude that if X is a random vector, then $X_1, \ldots, X_k$ are r.v.'s; on the other hand, if $X_1, \ldots, X_k$ are r.v.'s, then X is a random vector, by Exercise 7.5.2. Thus, $X = (X_1, \ldots, X_k)$ is a random vector if and only if $X_1, \ldots, X_k$ are r.v.'s.

A family $\{X_\alpha\}$ of r.v.'s is said to be **independent** if $\{\sigma(X_\alpha)\}$ is independent; then we also say that $X_\alpha' s$ are independent.

Exercise 7.5.3 Suppose that $\{X_\alpha\}$ is an independent family of r.v.'s and that $\{g_\alpha\}$ is a family of Borel functions on $\mathbb{R}$. Show that $\{g_\alpha \circ X_\alpha\}$ is an independent family of r.v.'s.

Lemma 7.5.1 *If $X_1, \ldots, X_n$, $n \geq 2$, are independent r.v.'s then for integer j, $1 \leq j < n$, $\sigma(X_1, \ldots, X_j)$ and $\sigma(X_{j+1}, \ldots, X_n)$ are independent.*

Proof Put $\widehat{X} = (X_1, \ldots, X_j)$ and $\widehat{Y} = (X_{j+1}, \ldots, X_n)$. In view of Exercise 7.5.2, we need to show that

$$P(\widehat{X} \in B, \widehat{Y} \in C) = P(\widehat{X} \in B) \cdot P(\widehat{Y} \in C) \tag{7.16}$$

for all $B \in \mathcal{B}^j$ and $C \in \mathcal{B}^{n-j}$. Consider $B_l \in \mathcal{B}$, $l = 1, \ldots, n$; we have

$$P(\widehat{X} \in B_1 \times \cdots \times B_j, \widehat{Y} \in B_{j+1} \times \cdots \times B_n) = P(X_1 \in B_1, \ldots, X_n \in B_n)$$
$$= \prod_{l=1}^n P(X_l \in B_l) = P(\widehat{X} \in B_1 \times \cdots \times B_j) \cdot P(\widehat{Y} \in B_{j+1} \times \cdots \times B_n),$$

hence, (7.16) holds for $B = B_1 \times \cdots \times B_j$ and $C = B_{j+1} \times \cdots \times B_n$. Fix $B_{j+1}, \ldots, B_n$ and let $\mathcal{N} = \{B \in \mathcal{B}^j : P(\widehat{X} \in B, \widehat{Y} \in B_{j+1} \times \cdots \times B_n) = P(\widehat{X} \in B)P(\widehat{Y} \in B_{j+1} \times \cdots \times B_n)\}$. Evidently, $\mathcal{N}$ is a λ-system containing the family $\mathcal{P}$ of all sets of the form $B_1 \times \cdots \times B_j$, where $B_1, \ldots, B_j$ are in $\mathcal{B}$. Now $\mathcal{P}$ is a π-system and $\sigma(\mathcal{P}) = \mathcal{B}^j$, therefore $\mathcal{B}^j \supset \mathcal{N} \supset \sigma(\mathcal{P}) = \mathcal{B}^j$. Thus $\mathcal{N} = \mathcal{B}^j$. This means that

$$P(\widehat{X} \in B, \widehat{Y} \in B_{j+1} \times \cdots \times B_n) = P(\widehat{X} \in B) \cdot P(\widehat{Y} \in B_{j+1} \times \cdots \times B_n)$$

for $B \in B^j$ and $B_{j+1}, \ldots, B_n$ in $\mathcal{B}$. Next fix $B \in \mathcal{B}^j$ and let

$$\mathcal{N}' = \{C \in \mathcal{B}^{n-j} : P(\widehat{X} \in B, \widehat{Y} \in C) = P(\widehat{X} \in B) \cdot P(\widehat{Y} \in C)\}.$$

Argue as in the immediately preceding part of the proof, we infer that $\mathcal{N}' = \mathcal{B}^{n-j}$ and finish the proof. ■

Lemma 7.5.2 *Suppose that $X_1, \ldots, X_n$, $n \geq 2$, are independent r.v.'s, and let $1 \leq j < n$ be an integer. Then $g_1 \circ (X_1, \ldots, X_j)$ and $g_2 \circ (X_{j+1}, \ldots, X_n)$ are independent if g_1 and g_2 are Borel functions on $\mathbb{R}^j$ and $\mathbb{R}^{n-j}$ respectively.*

Proof Let B and C be Borel sets of $\mathbb{R}$. Since $\{g_1 \circ (X_1, \ldots, X_j) \in B\} = \{(X_1, \ldots, X_j) \in g_1^{-1}B\}$ and $\{g_2 \circ (X_{j+1}, \ldots, X_n) \in C\} = \{(X_{j+1}, \ldots, X_n) \in g_2^{-1}C\}$, and since $g_1^{-1}B$

and $g_2^{-1}C$ are in $\mathcal{B}^j$ and B^{n-j} respectively, we know from Exercise 7.5.2 that $\{g_1 \circ (X_1, \ldots, X_j) \in B\}$ and $\{g_2 \circ (X_{j+1}, \ldots, X_n) \in C\}$ are in $\sigma(X_1, \ldots, X_j)$ and $\sigma(X_{j+1}, \ldots, X_n)$ respectively. It then follows from Lemma 7.5.1 that

$$\begin{aligned} &P(g_1 \circ (X_1, \ldots, X_j) \in B, g_2 \circ (X_{j+1}, \ldots, X_n)) \\ &= P(g_1 \circ (X_1, \ldots, X_j) \in B) \cdot P(g_2 \circ (X_{j+1}, \ldots, X_n) \in C). \end{aligned} \qquad \blacksquare$$

Theorem 7.5.2 *If X and Y are independent integrable r.v.'s, then XY is integrable and $E(XY) = E(X) \cdot E(Y)$.*

Proof By Exercise 7.5.3, X^{ε_1} and X^{ε_2} are independent, where each of the symbols ε_1 and ε_2 is either + or −. We may therefore assume that both X and Y are nonnegative. Observe then that if S_1 and S_2 are simple functions measurable w.r.t. $\sigma(X)$ and $\sigma(Y)$ respectively, then $E(S_1S_2) = E(S_1) \cdot E(S_2)$. Now, choose increasing sequences $\{S_n^{(1)}\}$ and $\{S_n^{(2)}\}$ of simple functions such that each $S_n^{(1)}$ is $\sigma(X)$-measurable and each $S_n^{(2)}$ is $\sigma(Y)$-measurable; and furthermore $S_n^{(1)} \nearrow X$ and $S_n^{(2)} \nearrow Y$ pointwise. Using the monotone convergence theorem, we have

$$\begin{aligned} E(X) \cdot E(Y) &= \left[\lim_{n\to\infty} E(S_n^{(1)})\right]\left[\lim_{n\to\infty} E(S_n^{(2)})\right] = \lim_{n\to\infty} [E(S_n^{(1)}) \cdot E(S_n^{(2)})] \\ &= \lim_{n\to\infty} E(S_n^{(1)}S_n^{(2)}) = E(XY). \end{aligned} \qquad \blacksquare$$

Corollary 7.5.1 *If $X_1, \ldots, X_n$, $n \geq 2$ are independent integrable r.v.'s, then $X_1 \cdots X_n$ is integrable and $E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$.*

Proof When $n = 2$, this is Theorem 7.5.2. Suppose now that $n \geq 3$; then $X_1 \cdots X_{n-1}$ and X_n are independent, by Lemma 7.5.2, and the corollary follows by induction on n. $\blacksquare$

Corollary 7.5.2 *If $X_1, \ldots, X_n$ are independent and integrable, then*

$$\operatorname{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \operatorname{Var}(X_j).$$

Proof

$$\begin{aligned} \operatorname{Var}\left(\sum_{j=1}^n X_j\right) &= E\left(\left[\sum_{j=1}^n \{X_j - E(X_j)\}\right]^2\right) \\ &= E\left(\sum_{j=1}^n \{X_j - E(X_j)\}^2 + \sum_{j\neq k}\{X_j - E(X_j)\}\{X_k - E(X_k)\}\right) \\ &= \sum_{j=1}^n \operatorname{Var}(X_j) + \sum_{j\neq k} E(\{X_j - E(X_j)\}\{X_k - E(X_k)\} \\ &= \sum_{j=1}^n \operatorname{Var}(X_j), \end{aligned}$$

because $X_j - E(X_j)$ and $X_k - E(X_k)$ are independent, by Exercise 7.5.3 if $j \neq k$, and hence $E(\{X_j - E(X_j)\}\{X_k - E(X_k)\}) = E(\{X_j - E(X_j)\}) \cdot E(\{X_k - E(X_k)\}) = 0$, by Theorem 7.5.2. ■

A probability measure μ on $\mathcal{B}$ is called a **probability distribution** and the distribution $X_\# P$ of a r.v. X is called the **probability distribution** of X (recall that $X_\# P(B) = P(X \in B)$ for $B \in \mathcal{B}$). A family of r.v.'s is said to be **identically distributed** if random variables of the family have identical probability distribution. For $p > 0$, $E(|X|^p)$ is called the p-th **absolute moment** of the r.v. X; while if $m \in \mathbb{N}$, $E(X^m)$ is referred to as the m-th **moment** of X.

Example 7.5.1 A r.v. X is said to be normally distributed with mean m and variance σ^2 if for $B \in \mathcal{B}$,

$$X_\# P(B) = P(X \in B) = \frac{1}{\sqrt{2\pi}\sigma} \int_B \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} dx,$$

where as usual we write $\exp\{\beta\}$ for e^β if the expression for β is complicated. If X is normally distributed with mean m and variance σ^2, then

$$\begin{aligned}
E(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} x dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{-t^2}(\sqrt{2}\sigma t + m)\sqrt{2}\sigma\, dt \\
&= \frac{m}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-t^2} dt = m; \\
\mathrm{Var}(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} (x-m)^2 dx \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-t^2} t^2 dt = \sigma^2.
\end{aligned}$$

Thus X actually has m as its expectation and σ^2 its variance. The probability distribution μ, defined by

$$\mu(B) = \frac{1}{\sqrt{2\pi}\sigma} \int_B \exp\left\{\frac{-(x-m)^2}{2\sigma^2}\right\} dx, \quad B \in \mathcal{B},$$

is called the **normal distribution with mean m and variance σ^2** and is denoted by $N(m, \sigma^2)$. The distribution $N(0, 1)$ is called the **standard normal distribution**.

Example 7.5.2 Consider the Bernoulli sequence space $(\Omega, \sigma(\mathcal{Q}), P)$ of Example 3.4.6. Recall that $\Omega = \{\omega = (\omega_k) : \omega_k \in \{0, 1\}, k \in \mathbb{N}\}$; $\mathcal{Q}$ is the smallest algebra on Ω that contains all sets of the form $E(\varepsilon_1, \ldots, \varepsilon_n) = \{\omega = (\omega_k) : \omega_1 = \varepsilon_1, \ldots, \omega_n = \varepsilon_n\}$, $n \in \mathbb{N}$ and $\varepsilon_j \in \{0, 1\}, j = 1, \ldots, n$, and P is the unique probability measure on $\sigma(\mathcal{Q})$ such that $P(E(\varepsilon_1, \ldots, \varepsilon_n)) = 2^{-n}$. If for $j \in \mathbb{N}$ and $\varepsilon \in \{0, 1\}$ let $E^j_\varepsilon = \{\omega = (\omega_k) : \omega_j = \varepsilon\}$, then we know from Exercise 1.3.2 that

$$P\left(E^{j_1}_{\varepsilon_1} \cap \cdots \cap E^{j_k}_{\varepsilon_k}\right) = \prod_{l=1}^{k} P(E^{j_l}_{\varepsilon_l}) = 2^{-k} \tag{7.17}$$

if $1 \leq j_1 < \cdots < j_k$ is any finite sequence in $\mathbb{N}$. Now for $j \in \mathbb{N}$, define a r.v. X_j by $X_j(\omega) = \omega_j$, then $\sigma(X_j) = \{\emptyset, E_0^j, E_1^j, \Omega\}$; and therefore we infer from (7.17) that $\{\sigma(X_j)\}$ is independent and consequently the r.v.'s $X_1, \ldots, X_j, \ldots$ are independent. Clearly, the probability distribution of each X_j is the measure μ on $\mathcal{B}$ such that $\mu(\{0\}) = \mu(\{1\}) = \frac{1}{2}$. Hence the sequence $\{X_j\}$ is identically distributed; furthermore $E(X_j) = \frac{1}{2}$, $\text{Var}(X_j) = \frac{1}{8}$, and m-th moment of X_j is $\frac{1}{2}$ for $m \in \mathbb{N}$.

Return now to the general discussion and consider an independent and identically distributed sequence $\{X_j\}$ of r.v.'s. Such a sequence is usually referred to as an **i.i.d.** sequence. Suppose that the common probability distribution of $X_j's$ is μ, then for any Borel function g on $\mathbb{R}$ such that $\int_{\mathbb{R}} g d\mu$ exists, we know from (4.1) that $\int_{\mathbb{R}} g d\mu = \int_{\Omega} g \circ X_j dP$; in particular, the m-th moment is the same for all $X_j's$ if it exists for one of them. Thus $E(X_j^2) = E(X_1^2)$ for all j. Assume now that $E(X_1^2) < \infty$ and let $S_n = \sum_{j=1}^n X_j$, $n \in \mathbb{N}$. Then, $E(S_n) = nE(X_1)$ or $E(\frac{S_n}{n}) = E(X_1)$, and hence from the Chebyshev inequality (6.3), we have

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) \leq \varepsilon^{-2}\,\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n\varepsilon^2}\,\text{Var}(X_1)$$

for any given $\varepsilon > 0$. This is stated as a theorem.

Theorem 7.5.3 (Weak law of large numbers) *Suppose that $\{X_j\}$ is an i.i.d. sequence of r.v.'s with finite second moment, then*

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2}\,\text{Var}(X_1) \tag{7.18}$$

for any given $\varepsilon > 0$, where $S_n = \sum_{j=1}^n X_j$.

A sequence $\{Y_j\}$ of r.v.'s is said to **converge in probability** to a r.v. Y if $\lim_{j\to\infty} P(|Y_j - Y| \geq \varepsilon) = 0$ for every $\varepsilon > 0$; the notation $Y_j \to Y[P]$ is used to mean that $\{Y_j\}$ converges to Y in probability. Apparently, convergence of Y_j to Y a.s. or in L^p-norm as $j \to \infty$ implies that $Y_j \to Y[P]$, hence convergence in probability is weaker than convergence a.s. and convergence in L^p-norm. Since Theorem 7.5.3 implies that $\frac{S_n}{n} \to E(X_1)[P]$, it is usually referred to as the **weak law** of large numbers.

Theorem 7.5.4 (Strong law of large numbers) *Suppose that $\{X_j\}$ is an independent sequence of r.v.'s such that $E(X_j) = 0$ and $E(X_j^4) \leq C < \infty$ for $j \in \mathbb{N}$. Let $S_n = \sum_{j=1}^n X_j$, then $\frac{S_n}{n} \to 0$ a.s. as $n \to \infty$.*

Proof Observe that (cf. Exercise 7.5.3):

(i) $E(X_iX_j^3) = E(X_i)E(X_j^3) = 0$ if $i \neq j$;

(ii) $E(X_iX_j^2X_k) = 0$ if i, j, k are different from one another; and

(iii) $E(X_iX_jX_kX_l) = 0$ if i, j, k, l are different from one another;
and note that

(iv) $\{E(X_j^2)\}^2 \le E(X_j^4) \le C$ for all j by Jensen's inequality (6.4).

Now since $E(S_n^4) = \sum_{i,j,k,l} E(X_iX_jX_kX_l)$, we conclude from (i), (ii), and (iii) that

$$\begin{aligned} E(S_n^4) &= \sum_{j=1}^{n} E(X_j^4) + \binom{4}{2} \sum_{1\le i<j\le n} E(X_i^2X_j^2) \\ &\le nC + 6 \sum_{1\le i<j\le n} E(X_i^2)E(X_j^2); \end{aligned}$$

but $E(X_i^2)E(X_j^2) \le \frac{1}{2}\{E(X_i^2)^2 + E(X_j^2)^2\} \le \frac{1}{2}\{E(X_i^4) + E(X_j^4)\}$, by (iv), for each pair $i < j$, and consequently

$$E(S_n^4) \le nC + 6\frac{n(n-1)}{2}C \le 3Cn^2,$$

or

$$E\left(\left(\frac{S_n}{n}\right)^4\right) \le \frac{3C}{n^2}.$$

The last inequality implies that $E(\sum_{n=1}^{\infty}(\frac{S_n}{n})^4) \le 3C\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$, and hence $\sum_{n=1}^{\infty}(\frac{S_n}{n})^4 < \infty$ a.s. Then, $\lim_{n\to\infty}\frac{S_n}{n} = 0$ a.s. follows. ■

Corollary 7.5.3 *Let $\{X_j\}$ be an independent sequence of r.v.'s with bounded fourth moment such that $E(X_j) = E(X_1)$ for all $j \in \mathbb{N}$; then $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} X_j = E(X_1)$ a.s.*

Proof Put $Y_j = X_j - E(X_j)$; then $E(Y_j) = 0$ for all j and $\{E(Y_j^4)\}$ is bounded. We then apply Theorem 7.5.4 to conclude the proof. ■

Now apply Corollary 7.5.3 to the sequence $\{X_j\}$ of Example 7.5.2; we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{2} \quad \text{a.s.}$$

i.e. the event $\{\omega \in \Omega : \lim_{n\to\infty}\frac{S_n(\omega)}{n} = \frac{1}{2}\}$ occurs with probability one, where $S_n = \sum_{j=1}^{n} X_j$; in other words, if we interpret $\{X_j\}$ as a sequence of tossing of a fair coin, the relative frequency with which heads appears in the first n tosses approaches $\frac{1}{2}$ as $n \to \infty$ almost certainly. This is what we proclaim in the last paragraph of Section 1.3.

As we know in Example 4.3.2, the Bernoulli sequence space $(\Omega, \sigma(\mathcal{Q}), P)$ and $([0,1], \mathcal{B}|[0,1], \lambda)$ are measure-theoretically the same space; it is therefore worthwhile considering the counterpart of the sequence $\{X_j\}$ of Example 7.5.2 in the space $([0,1], \mathcal{B}|[0,1], \lambda)$. For $x \in [0,1]$, let $0.x_1 \ldots x_k \ldots$ be the binary expansion of x with the convention that in case where two expansions are possible, the expansion with infinitely many $1's$ is chosen, and for $j \in \mathbb{N}$, define a r.v. Z_j by $Z_j(x) = x_j$. From the discussion in Example 3.4.6, one verifies readily from the independence of the sequence $\{X_j\}$ of Example 7.5.2 that $\{Z_j\}$ is independent, $E(Z_j) = E(Z_1) = \frac{1}{2}$, and $E(Z_j^4) = \frac{1}{2}$. Then,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Z_j = \frac{1}{2} \quad \text{a.s.}$$

i.e.

$$\lim_{n\to\infty} \frac{1}{n}\{\text{number of } 1's \text{ in } x_1, \ldots, x_n\} = \frac{1}{2} \tag{7.19}$$

for almost every x of $[0, 1]$. We call a number x in $[0, 1]$ **a normal number** if (7.19) holds. Then (7.19) can be stated as follows.

Theorem 7.5.5 (Borel) *Almost all numbers in* $[0, 1]$ *are normal.*

We now come to introduce the Fourier integral for probability distributions. The Fourier integral φ of a probability distribution μ is a function on $\mathbb{R}$, defined by

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} d\mu(x), \quad t \in \mathbb{R}. \tag{7.20}$$

We call attention to inconsistency in the definition of Fourier integral for functions and for probability distributions; should consistency of definition be preferred, the function φ, defined by (7.20), would be called the Fourier inverse integral of μ. It is readily seen that $\varphi(0) = 1$, $|\varphi(t)| \leq 1$, and φ is uniformly continuous on $\mathbb{R}$. In probability theory, φ is called the **characteristic function** of μ; and if a r.v. X has μ as its probability distribution, φ is also referred to as the **characteristic function of** X. Note that if φ is the characteristic function of X, then,

$$\varphi(t) = E(e^{itX}), \quad t \in \mathbb{R}.$$

Exercise 7.5.4 Let φ be the characteristic function of the r.v. X, and suppose that $E(|X|) < \infty$. Show that $\varphi \in C^1(\mathbb{R})$ and $\varphi'(t) = E(iXe^{itX})$.

Exercise 7.5.5 Show that the characteristic function φ of $N(0, 1)$ is given by $\varphi(t) = e^{-\frac{t^2}{2}}$.

Exercise 7.5.6 Suppose that φ is the characteristic function of a probability distribution μ. Show that for $u > 0$,

$$\mu\left(\left(-\infty, \frac{-2}{u}\right] \cup \left[\frac{2}{u}, \infty\right)\right) \leq \frac{1}{u}\int_{-u}^{u}(1 - \varphi(t))dt.$$

(Hint: $\frac{1}{u}\int_{-u}^{u}(1 - \varphi(t))dt = 2\int_{-\infty}^{\infty}(1 - \frac{\sin ux}{ux})d\mu(x) \geq 2\int_{|x|\geq\frac{2}{\mu}}(1 - \frac{1}{|ux|})d\mu(x)$.)

Theorem 7.5.6 *Suppose that X_1 and X_2 are independent random variables with characteristic functions φ_1 and φ_2 respectively, and let φ be the characteristic function of $X_1 + X_2$, then $\varphi = \varphi_1\varphi_2$.*

Proof Since e^{itX_1} and e^{itX_2} are independent, we have

$$\begin{aligned}\varphi(t) &= E(e^{it(X_1+X_2)}) = E(e^{itX_1} \cdot e^{itX_2}) \\ &= E(e^{itX_1}) \cdot E(e^{itX_2}) = \varphi_1(t)\varphi_2(t).\end{aligned}$$

■

Theorem 7.5.7 (Inversion formula) *Let μ be a probability distribution with characteristic function φ; then for $a < b$ in $\mathbb{R}$,*

$$\mu((a,b]) = \lim_{L\to\infty} \frac{1}{2\pi} \int_{-L}^{L} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt$$

if $\mu(\{a\}) = \mu(\{b\}) = 0$.

Proof Put $S(L) = \int_0^L \frac{\sin t}{t}dt$, $L > 0$. Then $\int_0^L \frac{\sin\theta t}{t}dt = \operatorname{sgn}\theta S(L|\theta|)$ and $\lim_{L\to\infty} S(L) = \frac{\pi}{2}$. Now consider the integral

$$\begin{aligned}\mathcal{I}(L) &= \frac{1}{2\pi}\int_{-L}^{L}\left[\frac{e^{-ita} - e^{-itb}}{it}\right]\varphi(t)dt \\ &= \frac{1}{2\pi}\int_{-L}^{L}\left(\int_{-\infty}^{\infty}\frac{e^{it(x-a)} - e^{it(x-b)}}{it}d\mu(x)\right)dt.\end{aligned}$$

From the elementary inequality $|e^{i\theta} - 1| \leq |\theta|$, we have

$$\left|\frac{e^{it(x-a)} - e^{it(x-b)}}{it}\right| = \frac{1}{|t|}\left|e^{it(b-a)} - 1\right| \leq b - a$$

for any $x \in \mathbb{R}$. We may therefore apply the Fubini theorem to the integral defining $\mathcal{I}(L)$:

$$\begin{aligned}\mathcal{I}(L) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-L}^{L}\frac{e^{it(x-a)} - e^{it(x-b)}}{it}dt\right)d\mu(x) \\ &= \int_{-\infty}^{\infty}\left(\int_{0}^{L}\left[\frac{\sin t(x-a)}{\pi t} - \frac{\sin t(x-b)}{\pi t}\right]dt\right)d\mu(x) \\ &= \int_{-\infty}^{\infty}\left[\frac{\operatorname{sgn}(x-a)}{\pi}S(L|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi}S(L|x-b|)\right]d\mu(x).\end{aligned}$$

Let us denote the integrand of this last integral by $\theta_L(a, b; x)$ and put $\theta_{ab}(x) = \lim_{L\to\infty} \theta_L(a, b; x)$; then,

$$\theta_{ab}(x) = \begin{cases} 0 & \text{if } x < a \text{ or } x > b; \\ \dfrac{1}{2} & \text{if } x = a \text{ or } x = b; \\ 1 & \text{if } a < x < b. \end{cases}$$

From the second mean-value theorem, $\{\int_0^\alpha \frac{\sin t}{t} dt\}_{\alpha>0}$ is bounded and therefore $|\theta_L(a, b; x)| \le M < \infty$ for all $L > 0$ and $x \in \mathbb{R}$. Hence by LDCT, we conclude that

$$\lim_{L\to\infty} \mathcal{I}(L) = \int_{-\infty}^{\infty} \theta_{ab}(x) d\mu(x)$$

$$= \frac{1}{2}\mu(\{a\}) + \mu((a, b)) + \frac{1}{2}\mu(\{b\}) = \mu((a, b]). \qquad \blacksquare$$

Exercise 7.5.7 Let μ be the probability measure on $\mathcal{B}$ concentrated at 0. Find the characteristic function of μ and use Theorem 7.5.7 to show that

$$\lim_{L\to\infty} \int_0^L \frac{\sin at}{t} dt = \int_0^\infty \frac{\sin at}{t} dt = \frac{\pi}{2}$$

for all $a > 0$.

Corollary 7.5.4 *If the probability distributions μ and ν have the same characteristic function, then $\mu = \nu$.*

Proof Let $\Pi = \{(a, b] : \mu(\{a\}) = \mu(\{b\}) = \nu(\{a\}) = \nu(\{b\}) = 0\} \cup \{\emptyset\}$, and $\mathcal{N} = \{B \in \mathcal{B} : \mu(B) = \nu(B)\}$. Theorem 7.5.7 implies that $\mathcal{N} \supset \Pi$. But Π is a π-system, $\mathcal{N}$ is a λ-system, and $\sigma(\Pi) = \mathcal{B}$, hence it follows from the $(\pi$-$\lambda)$ theorem that $\mathcal{N} = \mathcal{B}$. $\blacksquare$

Corollary 7.5.4 means that the characteristic function of a probability distribution μ uniquely determines μ and is therefore named the characteristic function of μ.

We are ready to state and prove the central limit theorem in probability theory. Suppose that $\{X_j\}$ is an i.i.d. sequence of random variables such that $E(X_j) = 0, \mathrm{Var}(X_j) = E(X_j^2) = 1$, and $E(|X_j|^3) < \infty$. For $n \in \mathbb{N}$, put $Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j$.

Theorem 7.5.8 (Central limit theorem) *The characteristic function of Y_n converges to the characteristic function of $N(0, 1)$ uniformly on any given finite interval.*

Proof Denote by φ the common characteristic function of X_j's and by μ the common distribution of X_j's. Using the fundamental theorem of calculus repeatedly, we have

$$\begin{aligned} e^{itx} &= 1 + i\int_0^{tx} e^{i\theta}\,d\theta = 1 + i\int_0^{tx}\left(1 + i\int_0^{\theta} e^{is}ds\right)d\theta \\ &= 1 + itx - \int_0^{tx}\left(\int_0^{\theta} e^{is}ds\right)d\theta \\ &= 1 + itx - \int_0^{tx}\left(\int_0^{\theta}\left(1 + i\int_0^{s} e^{i\tau}d\tau\right)ds\right)d\theta \\ &= 1 + itx - \frac{1}{2}t^2x^2 - i\int_0^{tx}\left(\int_0^{\theta}\left(\int_0^{s} e^{i\tau}d\tau\right)ds\right)d\theta \\ &= 1 + itx - \frac{1}{2}t^2x^2 + h(tx), \end{aligned}$$

where $|h(tx)| \le \frac{1}{6}|tx|^3$; consequently,

$$\varphi(t) = \int_{\mathbb{R}} e^{itx}d\mu(x) = 1 - \frac{1}{2}t^2 + \int_{\mathbb{R}} h(tx)d\mu(x) = 1 - \frac{1}{2}t^2 + H(t). \tag{7.21}$$

Note that $\int_{\mathbb{R}} itxd\mu(x) = itE(X_j) = 0$ and $\int_{\mathbb{R}} t^2x^2d\mu(x) = t^2E(X_j^2) = t^2$ have been used in deriving (7.21), and that

$$|H(t)| \le \frac{1}{6}E(|X_j|^3)|t|^3 \equiv C|t|^3. \tag{7.22}$$

Now let I be a finite interval in $\mathbb{R}$; then for some $b > 0$, $|t| \le b$ for $t \in I$, and hence there is $n_0 \in \mathbb{N}$, such that

$$\left(1 - \frac{1}{2}\frac{t^2}{n}\right) \ge \frac{1}{2}, \quad t \in I \tag{7.23}$$

if $n \ge n_0$. Denote now by φ_n the characteristic function of Y_n. We know from Theorem 7.5.6 that

$$\begin{aligned} \varphi_n(t) &= E\left(\exp\left\{\frac{it}{\sqrt{n}}\sum_{j=1}^{n} X_j\right\}\right) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n \\ &= \left[1 - \frac{1}{2}\frac{t^2}{n} + H\left(\frac{t}{\sqrt{n}}\right)\right]^n \\ &= \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n \left\{1 + \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^{-1} H\left(\frac{t}{\sqrt{n}}\right)\right\}^n \\ &= \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n (1 + G(t,n))^n, \end{aligned}$$

of which for $n \geq n_0$ and $t \in I$, we have from (7.22) and (7.23),

$$\left|G(t,n)\right| \leq 2Cb^3 n^{-\frac{3}{2}}.$$

Observe now from the mean-value theorem in differential calculus that, for $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\begin{aligned}\left(1+\frac{\alpha}{n}\right)^n &= \exp\left\{\ln\left(1+\frac{\alpha}{n}\right)^n\right\}\\ &= \exp\left\{n[\ln(n+\alpha)-\ln n]\right\} = \exp\left\{\frac{n\alpha}{n+\alpha_n}\right\},\end{aligned}$$

where $n+\alpha_n$ is between n and $n+\alpha$, and consequently

$$\lim_{n\to\infty}\left(1+\frac{\alpha}{n}\right)^n = e^{\alpha}$$

uniformly for $|\alpha| \leq B$ if $B > 0$ is fixed. As a consequence,

$$\lim_{n\to\infty}\left(1-\frac{1}{2}\frac{t^2}{n}\right)^n = e^{-\frac{t^2}{2}}$$

uniformly for $t \in I$; and since $\left|nG(t,n)\right| \leq 2Cb^3 n^{-\frac{1}{2}}$ for $n \geq n_0$ and $t \in I$, for any given $\varepsilon > 0$ there is $n_1 \geq n_0$ in $\mathbb{N}$ such that if $n \geq n_1$ and $t \in I$, then,

$$\left|\left(1+\frac{nG(t,n)}{n}\right)^n - e^{nG(t,n)}\right| < \frac{\varepsilon}{2}. \tag{7.24}$$

We may choose n_1 sufficiently large so that, if $n \geq n_1$ and $t \in I$, then $\left|nG(t,n)\right|$ will be small enough so that $\left|nG(t,n)\right| < \frac{\varepsilon}{4}$, and

$$1-\left|nG(t,n)\right| \leq e^{nG(t,n)} \leq 1+2\left|nG(t,n)\right|. \tag{7.25}$$

Finally, using (7.24) and (7.25), we have for $n \geq n_1$ and $t \in I$,

$$\begin{aligned}(1+G(t,n))^n - 1 &> e^{nG(t,n)} - \frac{\varepsilon}{2} - 1 \geq 1 - \left|nG(t,n)\right| - \frac{\varepsilon}{2} - 1 > -\varepsilon;\\ (1+G(t,n))^n - 1 &< e^{nG(t,n)} + \frac{\varepsilon}{2} - 1 \leq 1 + 2\left|nG(t,n)\right| + \frac{\varepsilon}{2} - 1 \leq \varepsilon.\end{aligned}$$

Thus, $\left|(1+G(t,n))^n - 1\right| < \varepsilon$ if $n \geq n_1$ and $t \in I$ i.e. $\lim_{n\to\infty}(1+G(t,n))^n = 1$ uniformly for $t \in I$. Summing up, we have shown that $\varphi_n(t)$ converges to $e^{-\frac{t^2}{2}}$ uniformly for $t \in I$. But $e^{-\frac{t^2}{2}}$ is the characteristic function of $N(0,1)$ (cf. Exercise 7.5.5). ■

The following exercise illustrates the relevance of the central limit theorem.

Exercise 7.5.8 Let Y_n be as in Theorem 7.5.8 and μ_n the probability distribution of Y_n; and let ν be $N(0,1)$. Furthermore, put $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \nu((-\infty, x])$ for $x \in \mathbb{R}$.

(i) Given that $\varepsilon > 0$. Show that there is $a > 0$ such that

$$\nu(\{|x| \geq a\}) < \varepsilon;$$
$$\mu_n(\{|x| \geq a\}) < \varepsilon,\ n = 1, 2, 3, \ldots,$$

(Hint: cf. Exercise 7.5.6 and central limit theorem.)

(ii) Show that if f is a bounded continuous function on $\mathbb{R}$, then

$$\lim_{n\to\infty} \int_{\mathbb{R}} f d\mu_n = \int_{\mathbb{R}} f d\nu.$$

(Hint: use (i) and Theorem 7.5.7.)

(iii) For $-\infty < \alpha < \beta < \infty$, define a continuous function $f_{\alpha,\beta}$ as follows:

$$f_{\alpha,\beta}(t) = \begin{cases} 1, & t \leq \alpha; \\ 0, & t \geq \beta; \\ \frac{\beta - t}{\beta - \alpha}, & \alpha < t < \beta. \end{cases}$$

Now let $-\infty < u < x < y < \infty$. By applying (ii) for $f = f_{x,y}$ and $f_{u,x}$ in this order, show that

$$\limsup_{n\to\infty} F_n(x) \leq F(y); \quad F(x-) \leq \liminf_{n\to\infty} F_n(x),$$

and then conclude that $\lim_{n\to\infty} F_n(x) = F(x)$ for $x \in \mathbb{R}$.

(iv) Show that for any finite interval I in $\mathbb{R}$,

$$\lim_{n\to\infty} \mu_n(I) = \frac{1}{\sqrt{2\pi}} \int_I e^{-\frac{t^2}{2}} dt.$$