Chapter 7
L^p-Spaces and the Radon–Nikodym Theorem

In this chapter, we study the spaces of functions whose pth power is integrable. In Section 7.2, we first derive some of the important inequalities (Hölder, Minkowski, Jensen) and then in Section 7.3 investigate the case $p = 2$ in more detail. Apart from the inequalities, the important results for probability theory are Lebesgue's decomposition theorem and the Radon–Nikodym theorem in Section 7.4. At first reading, some readers might wish to skip some of the more analytic parts of this chapter.

7.1 Definitions

We always assume that $(\Omega, \mathcal{A}, \mu)$ is a σ-finite measure space. In Definition 4.16, for measurable $f : \Omega \to \overline{\mathbb{R}}$, we defined

$$\|f\|_p := \left(\int |f|^p \, d\mu\right)^{1/p} \quad \text{for } p \in [1, \infty)$$

and

$$\|f\|_\infty := \inf\bigl\{K \ge 0 : \mu\bigl(|f| > K\bigr) = 0\bigr\}.$$

Further, we defined the spaces of functions where these expressions are finite:

$$\mathcal{L}^p(\Omega, \mathcal{A}, \mu) = \mathcal{L}^p(\mathcal{A}, \mu) = \mathcal{L}^p(\mu) = \{f : \Omega \to \overline{\mathbb{R}} \text{ measurable and } \|f\|_p < \infty\}.$$

We saw that $\|\cdot\|_1$ is a seminorm on $\mathcal{L}^1(\mu)$. Here our first goal is to change $\|\cdot\|_p$ into a proper norm for all $p \in [1, \infty]$. Apart from the fact that we still have to show the triangle inequality, to this end, we have to change the space a little bit since we only have

$$\|f - g\|_p = 0 \quad \iff \quad f = g \quad \mu\text{-a.e.}$$

For a proper norm (that is, not only a seminorm), the left-hand side has to imply equality (not only a.e.) of f and g. Hence we now consider f and g as equivalent if

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_7,

$f = g$ almost everywhere. Thus let

$$\mathcal{N} = \{f \text{ is measurable and } f = 0\ \mu\text{-a.e.}\}.$$

For any $p \in [1, \infty]$, $\mathcal{N}$ is a subvector space of $\mathcal{L}^p(\mu)$. Thus formally we can build the factor space. This is the standard procedure in order to change a seminorm into a proper norm.

Definition 7.1 (Factor space) For any $p \in [1, \infty]$, define

$$L^p(\Omega, \mathcal{A}, \mu) := \mathcal{L}^p(\Omega, \mathcal{A}, \mu)/\mathcal{N} = \big\{\bar{f} := f + \mathcal{N} : f \in \mathcal{L}^p(\mu)\big\}.$$

For $\bar{f} \in L^p(\mu)$, define $\|\bar{f}\|_p = \|f\|_p$ for any $f \in \bar{f}$. Also let $\int \bar{f}\, d\mu = \int f\, d\mu$ if this expression is defined for f.

Note that $\|\bar{f}\|_p$ and $\int \bar{f}\, d\mu$ do not depend on the choice of the representative $f \in \bar{f}$. Recall from Theorem 4.19 that $\int \bar{f}\, d\mu$ is well-defined if $f \in \mathcal{L}^p(\mu)$ and if μ is finite but it need not be if μ is infinite.

We first investigate convergence with respect to $\|\cdot\|_p$. To this end, we extend the corresponding theorem (Theorem 6.25) on convergence with respect to $\|\cdot\|_1$.

Definition 7.2 Let $p \in [1, \infty]$ and $f, f_1, f_2, \ldots \in \mathcal{L}^p(\mu)$. If $\|f_n - f\|_p \overset{n\to\infty}{\longrightarrow} 0$, then we say that $(f_n)_{n\in\mathbb{N}}$ converges to f in $L^p(\mu)$ and we write $f_n \overset{L^p}{\longrightarrow} f$.

Theorem 7.3 *Let* $p \in [1, \infty]$ *and* $f_1, f_2, \ldots \in \mathcal{L}^p(\mu)$. *Then the following statements are equivalent*:

(i) *There is an* $f \in \mathcal{L}^p(\mu)$ *with* $f_n \overset{L^p}{\longrightarrow} f$.
(ii) $(f_n)_{n\in\mathbb{N}}$ *is a Cauchy sequence in* $\mathcal{L}^p(\mu)$.

If $p < \infty$, *then*, *in addition*, (i) *and* (ii) *are equivalent to*:

(iii) $(|f_n|^p)_{n\in\mathbb{N}}$ *is uniformly integrable and there exists a measurable* f *with* $f_n \overset{\text{meas}}{\longrightarrow} f$.

The limits in (i) *and* (ii) *coincide.*

Proof For $p = \infty$, the equivalence of (i) and (ii) is a simple consequence of the triangle inequality.

Now let $p \in [1, \infty)$. The proof is similar to the proof of Theorem 6.25.

"(i) $\Longrightarrow$ (ii)" Note that $|x + y|^p \le 2^p(|x|^p + |y|^p)$ for all $x, y \in \mathbb{R}$. Hence

$$\|f_m - f_n\|_p^p \le 2^p\big(\|f_m - f\|_p^p + \|f_n - f\|_p^p\big) \overset{n\to\infty}{\longrightarrow} 0 \quad \text{for } m, n \to \infty.$$

"(ii) $\Longrightarrow$ (iii)" This works as in the proof of Theorem 6.25.

"(iii) $\Longrightarrow$ (i)" Since $|f_n|^p \overset{n\to\infty}{\longrightarrow} |f|^p$ in measure, by Theorem 6.25, we have $|f|^p \in \mathcal{L}^1(\mu)$ and hence $f \in \mathcal{L}^p(\mu)$. For $n \in \mathbb{N}$, define $g_n = |f_n - f|^p$. Then

$g_n \overset{n\to\infty}{\longrightarrow} 0$ in measure, and $(g_n)_{n\in\mathbb{N}}$ is uniformly integrable since $g_n \le 2^p(|f_n|^p + |f|^p)$. Hence we get (by Theorem 6.25) $\|f_n - f\|_p^p = \|g_n\|_1 \overset{n\to\infty}{\longrightarrow} 0$. □

Exercise 7.1.1 Let $(X_i)_{i\in\mathbb{N}}$ be independent, square integrable random variables with $\mathbf{E}[X_i] = 0$ for all $i \in \mathbb{N}$.

(i) Show that $\sum_{i=1}^{\infty} \mathbf{Var}[X_i] < \infty$ implies that there exists a real random variable X with $\sum_{i=1}^{n} X_i \overset{n\to\infty}{\longrightarrow} X$ almost surely.
(ii) Does the converse implication hold in (i)?

Exercise 7.1.2 Let $f : \Omega \to \mathbb{R}$ be measurable. Show that the following hold.

(i) If $\int |f|^p \, d\mu < \infty$ for some $p \in (0, \infty)$, then $\|f\|_p \overset{p\to\infty}{\longrightarrow} \|f\|_\infty$.
(ii) The integrability condition in (i) cannot be waived.

Exercise 7.1.3 Let $p \in (1, \infty)$, $f \in \mathcal{L}^p(\lambda)$, where λ is the Lebesgue measure on $\mathbb{R}$. Let $T : \mathbb{R} \to \mathbb{R}$, $x \mapsto x + 1$. Show that

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \overset{n\to\infty}{\longrightarrow} 0 \quad \text{in } L^p(\lambda).$$

7.2 Inequalities and the Fischer–Riesz Theorem

We present one of the most important inequalities of probability theory, Jensen's inequality for convex functions, and indicate how to derive from it Hölder's inequality and Minkowski's inequality. They in turn yield the triangle inequality for $\|\cdot\|_p$ and help in determining the dual space of $L^p(\mu)$. However, for the formal proofs of the latter inequalities, we will follow a different route.

Before stating Jensen's inequality, we give a primer on the basics of convexity of sets and functions.

Definition 7.4 A subset G of a vector space (or of an affine linear space) is called *convex* if, for any two points $x, y \in G$ and any $\lambda \in [0, 1]$, we have $\lambda x + (1-\lambda)y \in G$.

Example 7.5

(i) The convex subsets of $\mathbb{R}$ are the intervals.
(ii) A linear subspace of a vector space is convex.
(iii) The set of all probability measures on a measurable space is a convex set. ◊

Definition 7.6 Let G be a convex set. A map $\varphi : G \to \mathbb{R}$ is called *convex* if for any two points $x, y \in G$ and any $\lambda \in [0, 1]$, we have

$$\varphi\big(\lambda x + (1-\lambda)y\big) \le \lambda\varphi(x) + (1-\lambda)\varphi(y).$$

φ is called *concave* if $(-\varphi)$ is convex.

Let $I \subset \mathbb{R}$ be an interval. Let $\varphi : I \to \mathbb{R}$ be continuous and in the interior I° twice continuously differentiable with second derivative φ''. Then φ is convex if and only if $\varphi''(x) \geq 0$ for all $x \in I^\circ$. To put it differently, the first derivative φ' of a convex function is a monotone increasing function. In the next theorem, we will see that this is still true even if φ is not twice continuously differentiable when we pass to the right-sided derivative $D^+\varphi$ (or to the left-sided derivative), which we show always exists.

Theorem 7.7 *Let $I \subset \mathbb{R}$ be an interval with interior I° and let $\varphi : I \to \mathbb{R}$ be a convex map. Then*:

(i) *φ is continuous on I° and hence measurable with respect to $\mathcal{B}(I)$.*

(ii) *For $x \in I^\circ$, define the function of difference quotients*

$$g_x(y) := \frac{\varphi(y) - \varphi(x)}{y - x} \quad \textit{for } y \in I \setminus \{x\}.$$

Then g_x is monotone increasing and there exist the left-sided and right-sided derivatives

$$D^-\varphi(x) := \lim_{y \uparrow x} g_x(y) = \sup\big\{g_x(y) : y < x\big\}$$

and

$$D^+\varphi(x) := \lim_{y \downarrow x} g_x(y) = \inf\big\{g_x(y) : y > x\big\}.$$

(iii) *For $x \in I^\circ$, we have $D^-\varphi(x) \leq D^+\varphi(x)$ and*

$$\varphi(x) + (y - x)t \leq \varphi(y) \quad \textit{for any } y \in I \quad \iff \quad t \in \big[D^-\varphi(x), D^+\varphi(x)\big].$$

Hence $D^-\varphi(x)$ and $D^+\varphi(x)$ are the minimal and maximal slopes of a tangent at x.

(iv) *The maps $x \mapsto D^-\varphi(x)$ and $x \mapsto D^+\varphi(x)$ are monotone increasing. $x \mapsto D^-\varphi(x)$ is left continuous and $x \mapsto D^+\varphi(x)$ is right continuous. We have $D^-\varphi(x) = D^+\varphi(x)$ at all points of continuity of $D^-\varphi$ and $D^+\varphi$.*

(v) *φ is differentiable at x if and only if $D^-\varphi(x) = D^+\varphi(x)$. In this case, the derivative is $\varphi'(x) = D^+\varphi(x)$.*

(vi) *φ is almost everywhere differentiable and $\varphi(b) - \varphi(a) = \int_a^b D^+\varphi(x)\,dx$ for $a, b \in I^\circ$.*

Proof (i) Let $x \in I^\circ$. Assume that $\liminf_{n\to\infty} \varphi(x - 1/n) \leq \varphi(x) - \varepsilon$ for some $\varepsilon > 0$. Since φ is convex, we have

$$\varphi(y) \geq \varphi(x) + n(y - x)\big(\varphi(x) - \varphi(x - 1/n)\big) \quad \text{for all } y > x \text{ and } n \in \mathbb{N}.$$

Combining this with the assumption, we get $\varphi(y) = \infty$ for all $y > x$. Hence the assumption was false. A similar argument for the right-hand side yields continuity of φ at x.

(ii) Monotonicity is implied by convexity. The other claims are evident.

(iii) By monotonicity of g_x, we have $D^-\varphi(x) \le D^+\varphi(x)$. By construction, $\varphi(x) + (y-x)t \le \varphi(y)$ for all $y < x$ if and only if $t \ge D^-\varphi(x)$. The inequality holds for all $y > x$ if and only if $t \le D^+\varphi(x)$.

(iv) For $\varepsilon > 0$, by the convexity, the map $x \mapsto g_x(x+\varepsilon)$ is monotone increasing and is continuous by (i). Being an infimum of monotone increasing and continuous functions the map $x \mapsto D^+\varphi(x)$ is monotone increasing and right continuous. The statement for $D^-\varphi$ follows similarly. As $x \mapsto g_x(y)$ is monotone, we get $D^+\varphi(x') \ge D^-\varphi(x') \ge D^+\varphi(x)$ for $x' > x$. If $D^+\varphi$ is continuous at x, then $D^-\varphi(x) = D^+\varphi(x)$.

(v) This is obvious since $D^-\varphi$ and $D^+\varphi$ are the limits of the sequences of slopes of the left-sided and right-sided secant lines, respectively.

(vi) For $\varepsilon > 0$, let $A_\varepsilon = \{x \in I : D^+\varphi(x) \ge \varepsilon + \lim_{y\uparrow x} D^+\varphi(y)\}$ be the set of points of discontinuity of size at least ε. For any two points $a, b \in I$ with $a < b$, we have $\#(A_\varepsilon \cap (a,b)) \le \varepsilon^{-1}(D^+\varphi(b) - D^+\varphi(a))$; hence $A_\varepsilon \cap (a,b)$ is a finite set. Thus A_ε is countable. Hence also $A = \bigcup_{n=1}^\infty A_{1/n}$ is countable and thus a null set. By (iv) and (v), φ is differentiable in $I^\circ \setminus A$ with derivative $D^+\varphi$. □

If I is an interval, then a map $g : I \to \mathbb{R}$ is called *affine linear* if there are numbers $a, b \in \mathbb{R}$ such that $g(x) = ax + b$ for all $x \in I$. If $\varphi : I \to \mathbb{R}$ is a map, then we write

$$L(\varphi) := \{g : I \to \mathbb{R} \text{ is affine linear and } g \le \varphi\}.$$

As a shorthand, we write $\sup L(\varphi)$ for the map $x \mapsto \sup\{f(x) : f \in L(\varphi)\}$.

Corollary 7.8 *Let $I \subset \mathbb{R}$ be an open interval and let $\varphi : I \to \mathbb{R}$ be a map. Then the following are equivalent.*

(i) *φ is convex.*
(ii) *For any $x_0 \in I$, there exists a $g \in L(\varphi)$ with $g(x_0) = \varphi(x_0)$.*
(iii) *$L(\varphi)$ is nonempty and $\varphi = \sup L(\varphi)$.*
(iv) *There is a sequence $(g_n)_{n\in\mathbb{N}}$ in $L(\varphi)$ with $\varphi = \lim_{n\to\infty} \max\{g_1, \ldots, g_n\}$.*

Proof "(ii) $\Longrightarrow$ (iii) $\Longleftrightarrow$ (iv)" This is obvious.

"(iii) $\Longrightarrow$ (i)" The supremum of convex functions is convex and any affine linear map is convex. Hence $\sup L(\varphi)$ is convex if $L(\varphi) \ne \emptyset$.

"(i) $\Longrightarrow$ (ii)" By Theorem 7.7(iii), for any $x_0 \in I$, the map

$$x \mapsto \varphi(x_0) + (x - x_0)\, D^+\varphi(x_0)$$

is in $L(\varphi)$. □

Theorem 7.9 (Jensen's inequality) *Let $I \subset \mathbb{R}$ be an interval and let X be an I-valued random variable with $\mathbf{E}[|X|] < \infty$. If φ is convex, then $\mathbf{E}[\varphi(X)^-] < \infty$ and*

$$\mathbf{E}\big[\varphi(X)\big] \ge \varphi\big(\mathbf{E}[X]\big).$$

Proof As $L(\varphi) \neq \emptyset$ by Corollary 7.8(iii), we can choose numbers $a, b \in \mathbb{R}$ such that $ax + b \leq \varphi(x)$ for all $x \in I$. Hence

$$\mathbf{E}\big[\varphi(X)^-\big] \leq \mathbf{E}\big[(aX+b)^-\big] \leq |b| + |a| \cdot \mathbf{E}\big[|X|\big] < \infty.$$

We distinguish the cases where $\mathbf{E}[X]$ is in the interior I° or at the boundary ∂I.

Case 1. If $\mathbf{E}[X] \in I^\circ$, then let $t^+ := D^+\varphi(\mathbf{E}[X])$ be the maximal slope of a tangent of φ at $\mathbf{E}[X]$. Then $\varphi(x) \geq t^+ \cdot (x - \mathbf{E}[X]) + \varphi(\mathbf{E}[X])$ for all $x \in I$; hence

$$\mathbf{E}\big[\varphi(X)\big] \geq t^+\mathbf{E}\big[X - \mathbf{E}[X]\big] + \mathbf{E}\big[\varphi\big(\mathbf{E}[X]\big)\big] = \varphi\big(\mathbf{E}[X]\big).$$

Case 2. If $\mathbf{E}[X] \in \partial I$, then $X = \mathbf{E}[X]$ a.s.; hence $\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi(\mathbf{E}[X])] = \varphi(\mathbf{E}[X])$. □

Jensen's inequality can be extended to $\mathbb{R}^n$. To this end, we need a representation of convex functions of many variables as a supremum of affine linear functions. Recall that a function $g : \mathbb{R}^n \to \mathbb{R}$ is called affine linear if there is an $a \in \mathbb{R}^n$ and a $b \in \mathbb{R}$ such that $g(x) = \langle a, x\rangle + b$ for all x. Here $\langle\,\cdot\,,\,\cdot\,\rangle$ denotes the usual inner product on $\mathbb{R}^n$.

Theorem 7.10 *Let $G \subset \mathbb{R}^n$ be open and convex and let $\varphi : G \to \mathbb{R}$ be a map. Then Corollary 7.8 holds with I replaced by G. If φ is convex, then φ is continuous and hence measurable. If φ is twice continuously differentiable, then φ is convex if and only if the Hessian matrix is positive semidefinite.*

Proof As we need these statements only in the proof of the multidimensional Jensen inequality, which will not play a central role in the following, we only give references for the proofs. In Rockafellar's book [145], continuity follows from Theorem 10.1, and the statements of Corollary 7.8 follow from Theorem 12.1 and Theorem 18.8. The claim about the Hessian matrix can be found in Theorem 4.5. □

Theorem 7.11 (Jensen's inequality in $\mathbb{R}^n$) *Let $G \subset \mathbb{R}^n$ be a convex set and let $X_1, \ldots, X_n$ be integrable real random variables with $\mathbf{P}[(X_1, \ldots, X_n) \in G] = 1$. Further, let $\varphi : G \to \mathbb{R}$ be convex. Then $\mathbf{E}[\varphi(X_1, \ldots, X_n)^-] < \infty$ and*

$$\mathbf{E}\big[\varphi(X_1, \ldots, X_n)\big] \geq \varphi\big(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n]\big).$$

Proof First consider the case where G is open. Here, the argument is similar to the proof of Theorem 7.9. Let $g \in L(\varphi)$ with

$$g\big(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n]\big) = \varphi\big(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n]\big).$$

As $g \leq \varphi$ is linear, we get

$$\mathbf{E}\big[\varphi(X_1, \ldots, X_n)\big] \geq \mathbf{E}\big[g(X_1, \ldots, X_n)\big] = g\big(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n]\big).$$

Integrability of $\varphi(X_1,\dots,X_n)^-$ can be derived in a similar way to the one-dimensional case.

Now consider the general case where G is not necessarily open. Here the problem that arises when $(\mathbf{E}[X_1],\dots,\mathbf{E}[X_n]) \in \partial G$ is a bit more tricky than in the one-dimensional case since ∂G can have flat pieces that in turn, however, are convex. Hence one cannot infer that $(X_1,\dots,X_n)$ equals its expectation almost surely. We only sketch the argument. First infer that $(X_1,\dots,X_n)$ is almost surely in one of those flat pieces. This piece is necessarily of dimension smaller than n. Now restrict φ to that flat piece and inductively reduce its dimension until reaching a point, the case that has already been treated above. Details can be found, e.g., in [37, Theorem 10.2.6]. □

Example 7.12 Let X be a real random variable with $\mathbf{E}[X^2] < \infty$, $I = \mathbb{R}$ and $\varphi(x) = x^2$. By Jensen's inequality, we get

$$\mathbf{Var}[X] = \mathbf{E}\big[X^2\big] - \big(\mathbf{E}[X]\big)^2 \geq 0. \qquad \Diamond$$

Example 7.13 Let $G = [0,\infty) \times [0,\infty)$, $\alpha \in (0,1)$ and $\varphi(x,y) = x^\alpha y^{1-\alpha}$. Then φ is concave (exercise!); hence, for nonnegative random variables X and Y with finite expectation (by Theorem 7.11),

$$\mathbf{E}\big[X^\alpha Y^{1-\alpha}\big] \leq \big(\mathbf{E}[X]\big)^\alpha \big(\mathbf{E}[Y]\big)^{1-\alpha}. \qquad \Diamond$$

Example 7.14 Let G, X and Y be as in Example 7.13. Let $p \in (1,\infty)$. Then $\psi(x,y) = (x^{1/p} + y^{1/p})^p$ is concave. Hence (by Theorem 7.11)

$$\big(\mathbf{E}[X]^{1/p} + \mathbf{E}[Y]^{1/p}\big)^p \geq \mathbf{E}\big[\big(X^{1/p} + Y^{1/p}\big)^p\big]. \qquad \Diamond$$

Before we present Hölder's inequality and Minkowski's inequality, we need a preparatory lemma.

Lemma 7.15 (Young's inequality) *For $p,q \in (1,\infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$ and for $x,y \in [0,\infty)$,*

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}. \tag{7.1}$$

Proof Fix $y \in [0,\infty)$ and define $f(x) := \frac{x^p}{p} + \frac{y^q}{q} - xy$ for $x \in [0,\infty)$. f is twice continuously differentiable in $(0,\infty)$ with derivatives $f'(x) = x^{p-1} - y$ and $f''(x) = (p-1)x^{p-2}$. In particular, f is strictly convex and hence assumes its (unique) minimum at $x_0 = y^{1/(p-1)}$. By assumption, $q = \frac{p}{p-1}$; hence $x_0^p = y^q$ and thus

$$f(x_0) = \left(\frac{1}{p} + \frac{1}{q}\right) y^q - y^{1/(p-1)} y = 0. \qquad \square$$

Theorem 7.16 (Hölder's inequality) *Let* $p, q \in [1, \infty]$ *with* $\frac{1}{p} + \frac{1}{q} = 1$ *and* $f \in \mathcal{L}^p(\mu)$, $g \in \mathcal{L}^q(\mu)$. *Then* $(fg) \in \mathcal{L}^1(\mu)$ *and*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Proof The cases $p = 1$ and $p = \infty$ are trivial. Hence, let $p \in (1, \infty)$. Let $f \in \mathcal{L}^p(\mu)$ and $g \in \mathcal{L}^q(\mu)$ be nontrivial. By passing to $f/\|f\|_p$ and $g/\|g\|_q$, we may assume that $\|f\|_p = \|g\|_q = 1$. By Lemma 7.15, we have

$$\begin{aligned}\|fg\|_1 = \int |f| \cdot |g| \, d\mu &\leq \frac{1}{p} \int |f|^p \, d\mu + \frac{1}{q} \int |g|^q \, d\mu \\ &= \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p \cdot \|g\|_q.\end{aligned}$$

□

Theorem 7.17 (Minkowski's inequality) *For* $p \in [1, \infty]$ *and* $f, g \in \mathcal{L}^p(\mu)$,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{7.2}$$

Proof The case $p = \infty$ is trivial. Hence, let $p \in [1, \infty)$. The left-hand side in (7.2) does not decrease if we replace f and g by $|f|$ and $|g|$. Hence we may assume $f \geq 0$ and $g \geq 0$ and (to avoid trivialities) $\|f + g\|_p > 0$.

Now $(f + g)^p \leq 2^p(f^p \vee g^p) \leq 2^p(f^p + g^p)$; hence $f + g \in \mathcal{L}^p(\mu)$. Apply Hölder's inequality to $f \cdot (f + g)^{p-1}$ and to $g \cdot (f + g)^{p-1}$ to get

$$\begin{aligned}\|f + g\|_p^p = \int (f + g)^p \, d\mu &= \int f(f + g)^{p-1} \, d\mu + \int g(f + g)^{p-1} \, d\mu \\ &\leq \|f\|_p \cdot \left\|(f + g)^{p-1}\right\|_q + \|g\|_p \cdot \left\|(f + g)^{p-1}\right\|_q \\ &= \big(\|f\|_p + \|g\|_p\big) \cdot \|f + g\|_p^{p-1}.\end{aligned}$$

Note that in the last step, we used the fact that $p - p/q = 1$. Dividing both sides by $\|f + g\|_p^{p-1}$ yields (7.2). □

In Theorem 7.17, we verified the triangle inequality and hence that $\|\cdot\|_p$ is a norm. Theorem 7.3 says that this norm is complete (i.e., every Cauchy sequence converges). A complete normed vector space is called a *Banach space*. Summing up, we have shown the following theorem.

Theorem 7.18 (Fischer–Riesz) $(L^p(\mu), \|\cdot\|_p)$ *is a Banach space for every* $p \in [1, \infty]$.

Exercise 7.2.1 Show Hölder's inequality by applying Jensen's inequality to the function of Example 7.13.

Exercise 7.2.2 Show Minkowski's inequality by applying Jensen's inequality to the function of Example 7.14.

Exercise 7.2.3 Let X be a real random variable and let $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Show that X is in $\mathcal{L}^p(\mathbf{P})$ if and only if there exists a $C < \infty$ such that $|\mathbf{E}[XY]| \le C\|Y\|_q$ for any bounded random variable Y.

7.3 Hilbert Spaces

In this section, we study the case $p = 2$ in more detail. The main goal is the representation theorem for continuous linear functionals on Hilbert spaces due to Riesz and Fréchet. This theorem is a cornerstone for a functional analytic proof of the Radon–Nikodym theorem in Section 7.4.

Definition 7.19 Let V be a real vector space. A map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called an *inner product* if:

(i) (Linearity) $\langle x, \alpha y + z \rangle = \alpha \langle x, y \rangle + \langle x, z \rangle$ for all $x, y, z \in V$ and $\alpha \in \mathbb{R}$.
(ii) (Symmetry) $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$.
(iii) (Positive definiteness) $\langle x, x \rangle > 0$ for all $x \in V \setminus \{0\}$.

If only (i) and (ii) hold and $\langle x, x \rangle \ge 0$ for all x, then $\langle \cdot, \cdot \rangle$ is called a positive semidefinite symmetric bilinear form, or a *semi-inner product*.

If $\langle \cdot, \cdot \rangle$ is an inner product, then $(V, \langle \cdot, \cdot \rangle)$ is called a (real) *Hilbert space* if the norm defined by $\|x\| := \langle x, x \rangle^{1/2}$ is complete; that is, if $(V, \|\cdot\|)$ is a Banach space.

Definition 7.20 For $f, g \in \mathcal{L}^2(\mu)$, define

$$\langle f, g \rangle := \int f g \, d\mu.$$

For $\bar{f}, \bar{g} \in L^2(\mu)$, define $\langle \bar{f}, \bar{g} \rangle := \langle f, g \rangle$, where $f \in \bar{f}$ and $g \in \bar{g}$.

Note that this definition is independent of the particular choices of the representatives of f and g.

Theorem 7.21 *$\langle \cdot, \cdot \rangle$ is an inner product on $L^2(\mu)$ and a semi-inner product on $\mathcal{L}^2(\mu)$. In addition, $\|f\|_2 = \langle f, f \rangle^{1/2}$.*

Proof This is left as an exercise. □

As a corollary to Theorem 7.18, we get the following.

Corollary 7.22 *$(L^2(\mu), \langle \cdot, \cdot \rangle)$ is a real Hilbert space.*

Lemma 7.23 *If $\langle \cdot, \cdot \rangle$ is a semi-inner product on the real vector space V, then $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is continuous (with respect to the product topology of the topology on V that is generated by the pseudo-metric $d(x, y) = \langle x - y, x - y \rangle^{1/2}$).*

Proof This is obvious. □

Definition 7.24 (Orthogonal complement) Let V be a real vector space with inner product $\langle \cdot, \cdot \rangle$. If $W \subset V$, then the orthogonal complement of W is the following linear subspace of V:

$$W^{\perp} := \big\{ v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W \big\}.$$

Theorem 7.25 (Orthogonal decomposition) *Let $(V, \langle \cdot, \cdot \rangle)$ be a Hilbert space and let $W \subset V$ be a closed linear subspace. For any $x \in V$, there is a unique representation $x = y + z$ where $y \in W$ and $z \in W^{\perp}$.*

Proof Let $x \in V$ and $c := \inf\{\|x - w\| : w \in W\}$. Further, let $(w_n)_{n\in\mathbb{N}}$ be a sequence in W with $\|x - w_n\| \overset{n\to\infty}{\longrightarrow} c$. The parallelogram law yields

$$\|w_m - w_n\|^2 = 2\|w_m - x\|^2 + 2\|w_n - x\|^2 - 4\left\|\frac{1}{2}(w_m + w_n) - x\right\|^2.$$

As W is linear, we have $(w_m + w_n)/2 \in W$; hence $\|\frac{1}{2}(w_m + w_n) - x\| \ge c$. Thus $(w_n)_{n\in\mathbb{N}}$ is a Cauchy sequence: $\|w_m - w_n\| \longrightarrow 0$ if $m, n \to \infty$.

Since V is complete and W is closed, W is also complete; hence there is a $y \in W$ with $w_n \overset{n\to\infty}{\longrightarrow} y$. Now let $z := x - y$. Then $\|z\| = \lim_{n\to\infty} \|w_n - x\| = c$ by continuity of the norm (Lemma 7.23).

Consider an arbitrary $w \in W \setminus \{0\}$. We define $\varrho := -\langle z, w \rangle / \|w\|^2$ and get $y + \varrho w \in W$; hence

$$c^2 \le \big\|x - (y + \varrho w)\big\|^2 = \|z\|^2 + \varrho^2\|w\|^2 + 2\varrho\langle z, w \rangle = c^2 - \varrho^2\|w\|^2.$$

Concluding, we have $\langle z, w \rangle = 0$ for all $w \in W$ and thus $z \in W^{\perp}$.

Uniqueness of the decomposition is easy: If $x = y' + z'$ is an orthogonal decomposition, then $y - y' \in W$ and $z - z' \in W^{\perp}$ as well as $y - y' + z - z' = 0$; hence

$$\begin{aligned} 0 &= \big\|y - y' + z - z'\big\|^2 \\ &= \big\|y - y'\big\|^2 + \big\|z - z'\big\|^2 + 2\big\langle y - y', z - z'\big\rangle \\ &= \big\|y - y'\big\|^2 + \big\|z - z'\big\|^2, \end{aligned}$$

whence $y = y'$ and $z = z'$. □

Theorem 7.26 (Riesz–Fréchet representation theorem) *Let $(V, \langle\cdot,\cdot\rangle)$ be a Hilbert space and let $F : V \to \mathbb{R}$ be a map. Then the following are equivalent.*

(i) *F is continuous and linear.*
(ii) *There is an $f \in V$ with $F(x) = \langle x, f\rangle$ for all $x \in V$.*

The element $f \in V$ in (ii) *is uniquely determined.*

Proof "(ii)$\Longrightarrow$ (i)" For any $f \in V$, by definition of the inner product, the map $x \mapsto \langle x, f\rangle$ is linear. By Lemma 7.23, this map is also continuous.

"(i)$\Longrightarrow$ (ii)" If $F \equiv 0$, then choose $f = 0$. Now assume F is not identically zero. As F is continuous, the kernel $W := F^{-1}(\{0\})$ is a closed (proper) linear subspace of V. Let $v \in V \setminus W$ and let $v = y + z$ for $y \in W$ and $z \in W^{\perp}$ be the orthogonal decomposition of v. Then $z \neq 0$ and $F(z) = F(v) - F(y) = F(v) \neq 0$. Hence we can define $u := z/F(z) \in W^{\perp}$. Clearly, $F(u) = 1$ and for any $x \in V$, we have $F(x - F(x)u) = F(x) - F(x)F(u) = 0$; hence $x - F(x)u \in W$ and thus $\langle x - F(x)u, u\rangle = 0$. Consequently, $F(x) = \langle x, u\rangle/\|u\|^2$. Now define $f := u/\|u\|^2$. Then $F(x) = \langle x, f\rangle$ for all $x \in V$.

"Uniqueness" Let $\langle x, f\rangle = \langle x, g\rangle$ for all $x \in V$. Letting $x = f - g$, we get $0 = \langle f - g, f - g\rangle$; hence $f = g$. □

In the following section, we will need the representation theorem for the space $\mathcal{L}^2(\mu)$, which, unlike $L^2(\mu)$, is not a Hilbert space. However, with a little bit of *abstract nonsense*, one can apply the preceding theorem to $\mathcal{L}^2(\mu)$. Recall that $\mathcal{N} = \{f \in \mathcal{L}^2(\mu) : \langle f, f\rangle = 0\}$ is the subspace of functions that equal zero almost everywhere. Let $L^2(\mu) = \mathcal{L}^2(\mu)/\mathcal{N}$ be the factor space. This is a special case of the situation where $(V, \langle\cdot,\cdot\rangle)$ is a linear space with complete semi-inner product. In this case, $\mathcal{N} := \{v \in V : \langle v, v\rangle = 0\}$ and $V_0 = V/\mathcal{N} := \{f + \mathcal{N} : f \in V\}$. Denote $\langle v + \mathcal{N}, w + \mathcal{N}\rangle_0 := \langle v, w\rangle$ to obtain a Hilbert space $(V_0, \langle\cdot,\cdot\rangle_0)$.

Corollary 7.27 *Let $(V, \langle\cdot,\cdot\rangle)$ be a linear vector space with complete semi-inner product. The map $F : V \to \mathbb{R}$ is continuous and linear if and only if there is an $f \in V$ with $F(x) = \langle x, f\rangle$ for all $x \in V$.*

Proof One implication is trivial. Hence, let F be continuous and linear. Then $F(0) = 0$ since F is linear. Note that $F(v) = F(0) = 0$ for all $v \in \mathcal{N}$ since F is continuous. Indeed, v lies in *every* open neighborhood of 0; hence F assumes at v the same value as at 0. Thus F induces a continuous linear map $F_0 : V_0 \to \mathbb{R}$ by $F_0(x + \mathcal{N}) = F(x)$. By Theorem 7.26, there is an $f + \mathcal{N} \in V_0$ with $F_0(x + \mathcal{N}) = \langle x + \mathcal{N}, f + \mathcal{N}\rangle_0$ for all $x + \mathcal{N} \in V_0$. However, $F(x) = \langle x, f\rangle$ for all $x \in V$ by the definition of F_0 and $\langle\cdot,\cdot\rangle_0$. □

Corollary 7.28 *The map $F : \mathcal{L}^2(\mu) \to \mathbb{R}$ is continuous and linear if and only if there is an $f \in \mathcal{L}^2(\mu)$ with $F(g) = \int gf \, d\mu$ for all $g \in \mathcal{L}^2(\mu)$.*

Proof The space $\mathcal{L}^2(\mu)$ fulfills the conditions of Corollary 7.27. □

Exercise 7.3.1 (Fourier series) For $n \in \mathbb{N}_0$, define $S_n, C_n : [0,1] \to [0,1]$ by $S_n(x) = \sqrt{2}\sin(2\pi nx)$, $C_n(x) = \sqrt{2}\cos(2\pi nx)$. For two square summable sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}_0}$, let $h_{a,b} := b_0 + \sum_{n=1}^{\infty}(a_n S_n + b_n C_n)$. Further, let W be the vector space of such $h_{a,b}$.

Show the following:

(i) The functions $C_0, S_n, C_n, n \in \mathbb{N}$ form an orthogonal system in $L^2([0,1],\lambda)$.
(ii) The series defining $h_{a,b}$ converges in $L^2([0,1],\lambda)$.
(iii) W is a closed linear subspace of $L^2([0,1],\lambda)$.
(iv) $W = L^2([0,1],\lambda)$. More precisely, for any $f \in \mathcal{L}^2([0,1],\lambda)$, there exist uniquely defined square summable sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}_0}$ such that $f = h_{a,b}$. Furthermore, $\|f\|_2^2 = b_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)$.

Hint: Show (iv) first for step functions (see Exercise 4.2.6).

7.4 Lebesgue's Decomposition Theorem

In this section, we employ the properties of Hilbert spaces that we derived in the last section in order to decompose a measure into a singular part and a part that is absolutely continuous, both with respect to a second given measure. Furthermore, we show that the absolutely continuous part has a density. Let μ and ν be measures on $(\Omega, \mathcal{A})$. By Definition 4.13, a measurable function $f : \Omega \to [0,\infty)$ is called a *density* of ν with respect to μ if

$$\nu(A) := \int f \mathbb{1}_A \, d\mu \quad \text{for all } A \in \mathcal{A}. \tag{7.3}$$

On the other hand, for any measurable $f : \Omega \to [0,\infty)$, Eq. (7.3) defines a measure ν on $(\Omega, \mathcal{A})$. In this case, we also write

$$\nu = f\mu \quad \text{and} \quad f = \frac{d\nu}{d\mu}. \tag{7.4}$$

For example, the normal distribution $\nu = \mathcal{N}_{0,1}$ has the density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ with respect to the Lebesgue measure $\mu = \lambda$ on $\mathbb{R}$.

If $g : \Omega \to [0,\infty]$ is measurable, then (by Theorem 4.15)

$$\int g \, d\nu = \int g f \, d\mu. \tag{7.5}$$

Hence $g \in \mathcal{L}^1(\nu)$ if and only if $gf \in \mathcal{L}^1(\mu)$, and in this case (7.5) holds.

If $\nu = f\mu$, then $\nu(A) = 0$ for all $A \in \mathcal{A}$ with $\mu(A) = 0$. The situation is quite the opposite for, e.g., the Poisson distribution $\mu = \mathrm{Poi}_\varrho$ with parameter $\varrho > 0$ and $\nu = \mathcal{N}_{0,1}$. Here $\mathbb{N}_0 \subset \mathbb{R}$ is a ν-null set with $\mu(\mathbb{R} \setminus \mathbb{N}_0) = 0$. We say that ν is *singular* to μ.

The main goal of this chapter is to show that an arbitrary σ-finite measure ν on a measurable space $(\Omega, \mathcal{A})$ can be decomposed into a part that is singular to the σ-finite measure μ and a part that has a density with respect to μ (Lebesgue's decomposition theorem, Theorem 7.33).

Theorem 7.29 (Uniqueness of the density) *Let ν be σ-finite. If f_1 and f_2 are densities of ν with respect to μ, then $f_1 = f_2$ μ-almost everywhere. In particular, the density $\frac{d\nu}{d\mu}$ is unique up to equality μ-almost everywhere.*

Proof Let $E_n \uparrow \Omega$ with $\nu(E_n) < \infty$, $n \in \mathbb{N}$. Let $A_n = E_n \cap \{f_1 > f_2\}$ for $n \in \mathbb{N}$. Then $\nu(A_n) < \infty$; hence

$$0 = \nu(A_n) - \nu(A_n) = \int_{A_n} (f_1 - f_2)\, d\mu.$$

By Theorem 4.8(i), $f_2 \mathbb{1}_{A_n} = f_1 \mathbb{1}_{A_n}$ μ-a.e. As $f_1 > f_2$ on A_n, we infer $\mu(A_n) = 0$ and

$$\mu\big(\{f_1 > f_2\}\big) = \mu\Bigg(\bigcup_{n\in\mathbb{N}} A_n\Bigg) = 0.$$

Similarly, we get $\mu(\{f_1 < f_2\}) = 0$; hence $f_1 = f_2$ μ-a.e. □

Definition 7.30 Let μ and ν be two measures on $(\Omega, \mathcal{A})$.

(i) ν is called *absolutely continuous* with respect to μ (symbolically $\nu \ll \mu$) if

$$\nu(A) = 0 \quad \text{for all } A \in \mathcal{A} \text{ with } \mu(A) = 0. \tag{7.6}$$

The measures μ and ν are called *equivalent* (symbolically $\mu \approx \nu$) if $\nu \ll \mu$ and $\mu \ll \nu$.

(ii) μ is called *singular* to ν (symbolically $\mu \perp \nu$) if there exists an $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $\nu(\Omega \setminus A) = 0$.

Remark 7.31 Clearly, $\mu \perp \nu \iff \nu \perp \mu$. ◊

Example 7.32

(i) Let μ be a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with density f with respect to the Lebesgue measure λ. Then $\mu(A) = \int_A f\, d\lambda = 0$ for every $A \in \mathcal{A}$ with $\lambda(A) = 0$; hence $\mu \ll \lambda$. If λ-almost everywhere $f > 0$, then $\mu(A) = \int_A f\, d\lambda > 0$ if $\lambda(A) > 0$; hence $\mu \approx \lambda$. If $\lambda(\{f = 0\}) > 0$, then (since $\mu(\{f = 0\}) = 0$) $\lambda \not\ll \mu$.

(ii) Consider the Bernoulli distributions Ber_p and Ber_q for $p, q \in [0, 1]$. If $p \in (0, 1)$, then $\mathrm{Ber}_q \ll \mathrm{Ber}_p$. If $p \in \{0, 1\}$, then $\mathrm{Ber}_q \ll \mathrm{Ber}_p$ if and only if $p = q$, and $\mathrm{Ber}_q \perp \mathrm{Ber}_p$ if and only if $q = 1 - p$.

(iii) Consider the Poisson distributions Poi_α and Poi_β for $\alpha, \beta \ge 0$. We have $\mathrm{Poi}_\alpha \ll \mathrm{Poi}_\beta$ if and only if $\beta > 0$ or $\alpha = 0$.

(iv) Consider the infinite product measures (see Theorem 1.64) $(\mathrm{Ber}_p)^{\otimes\mathbb{N}}$ and $(\mathrm{Ber}_q)^{\otimes\mathbb{N}}$ on $\Omega = \{0, 1\}^{\mathbb{N}}$. Then $(\mathrm{Ber}_p)^{\otimes\mathbb{N}} \perp (\mathrm{Ber}_q)^{\otimes\mathbb{N}}$ if $p \neq q$. Indeed, for $n \in \mathbb{N}$, let $X_n((\omega_1, \omega_2, \ldots)) = \omega_n$ be the projection of Ω to the nth coordinate. Then under $(\mathrm{Ber}_r)^{\otimes\mathbb{N}}$ the family $(X_n)_{n\in\mathbb{N}}$ is independent and Bernoulli-distributed with parameter r (see Example 2.18). By the strong law of large numbers, for any $r \in \{p, q\}$, there exists a measurable set $A_r \subset \Omega$ with $(\mathrm{Ber}_r)^{\otimes \mathbb{N}}(\Omega \setminus A_r) = 0$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = r \quad \text{for all } \omega \in A_r.$$

In particular, $A_p \cap A_q = \emptyset$ if $p \neq q$ and thus $(\mathrm{Ber}_p)^{\otimes \mathbb{N}} \perp (\mathrm{Ber}_q)^{\otimes \mathbb{N}}$. ◊

Theorem 7.33 (Lebesgue's decomposition theorem) *Let μ and ν be σ-finite measures on $(\Omega, \mathcal{A})$. Then ν can be uniquely decomposed into an absolutely continuous part ν_a and a singular part ν_s (with respect to μ):*

$$\nu = \nu_a + \nu_s, \quad \text{where } \nu_a \ll \mu \text{ and } \nu_s \perp \mu.$$

ν_a has a density with respect to μ, and $\frac{d\nu_a}{d\mu}$ is $\mathcal{A}$-measurable and finite μ-a.e.

Corollary 7.34 (Radon–Nikodym theorem) *Let μ and ν be σ-finite measures on $(\Omega, \mathcal{A})$. Then*

$$\nu \text{ has a density w.r.t. } \mu \quad \Longleftrightarrow \quad \nu \ll \mu.$$

In this case, $\frac{d\nu}{d\mu}$ is $\mathcal{A}$-measurable and finite μ-a.e. $\frac{d\nu}{d\mu}$ is called the Radon–Nikodym derivative of ν with respect to μ.

Proof One direction is trivial. Hence, let $\nu \ll \mu$. By Theorem 7.33, we get that $\nu = \nu_a$ has a density with respect to μ. □

Proof of Theorem 7.33 The idea goes back to von Neumann. We follow the exposition in [37].

By the usual exhaustion arguments, we can restrict ourselves to the case where μ and ν are finite. By Theorem 4.19, the canonical inclusion $i : \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \hookrightarrow \mathcal{L}^1(\Omega, \mathcal{A}, \mu + \nu)$ is continuous. Since $\nu \leq \mu + \nu$, the linear functional $\mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \to \mathbb{R}$, $h \mapsto \int h\, d\nu$ is continuous. By the Riesz–Fréchet theorem (here Corollary 7.28), there exists a $g \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu)$ such that

$$\int h\, d\nu = \int hg\, d(\mu + \nu) \quad \text{for all } h \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu) \tag{7.7}$$

or equivalently

$$\int f(1 - g)\, d(\mu + \nu) = \int f\, d\mu \quad \text{for all } f \in \mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu). \tag{7.8}$$

If in (7.7) we choose $h = \mathbb{1}_{\{g<0\}}$, then we get that $(\mu + \nu)$-almost everywhere $g \geq 0$. Similarly, with $f = \mathbb{1}_{\{g>1\}}$ in (7.8), we obtain that $(\mu + \nu)$-almost everywhere $g \leq 1$. Hence $0 \leq g \leq 1$.

Now let $f \geq 0$ be measurable and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative functions in $\mathcal{L}^2(\Omega, \mathcal{A}, \mu + \nu)$ with $f_n \uparrow f$. By the monotone convergence theorem (applied to the measure $(1-g)(\mu+\nu)$; that is, the measure with density $(1-g)$ with respect to $\mu + \nu$), we obtain that (7.8) holds for all measurable $f \geq 0$. Similarly, we get (7.7) for all measurable $h \geq 0$.

Let $E := g^{-1}(\{1\})$. If we let $f = \mathbb{1}_E$ in (7.8), then we get $\mu(E) = 0$. Define the measures ν_a and ν_s for $A \in \mathcal{A}$ by

$$\nu_a(A) := \nu(A \setminus E) \quad \text{and} \quad \nu_s(A) := \nu(A \cap E).$$

Clearly, $\nu = \nu_a + \nu_s$ and $\nu_s(\Omega \setminus E) = 0$; hence $\nu_s \perp \mu$. If now $A \cap E = \emptyset$ and $\mu(A) = 0$, then $\int \mathbb{1}_A \, d\mu = 0$. Hence, by (7.8), also $\int_A (1-g)\, d(\mu + \nu) = 0$. On the other hand, we have $1 - g > 0$ on A; hence $\mu(A) + \nu(A) = 0$ and thus $\nu_a(A) = \nu(A) = 0$. If, more generally, B is measurable with $\mu(B) = 0$, then $\mu(B \setminus E) = 0$; hence, as shown above, $\nu_a(B) = \nu_a(B \setminus E) = 0$. Consequently, $\nu_a \ll \mu$ and $\nu = \nu_a + \nu_s$ is the decomposition we wanted to construct.

In order to obtain the density of ν_a with respect to μ, we define $f := \frac{g}{1-g}\mathbb{1}_{\Omega\setminus E}$. For any $A \in \mathcal{A}$, by (7.8) and (7.7) with $h = \mathbb{1}_{A\setminus E}$,

$$\int_A f \, d\mu = \int_{A\cap E^c} g \, d(\mu + \nu) = \nu(A \setminus E) = \nu_a(A).$$

Hence $f = \frac{d\nu_a}{d\mu}$. □

Exercise 7.4.1 For every $x \in (0, 1]$, let $x = (0, x_1x_2x_3\ldots) := \sum_{n=1}^{\infty} x_n 2^{-n}$ be the dyadic expansion (with $\limsup_{n\to\infty} x_n = 1$ for definiteness). Define a map $F : (0, 1] \to (0, 1]$ by

$$F(x) = (0, x_1x_1x_2x_2x_3x_3\ldots) = \sum_{n=1}^{\infty} 3x_n 4^{-n}.$$

Let U be a random variable that is uniformly distributed on $(0, 1]$ and denote by $\mu := \mathbf{P}_{U\circ F^{-1}}$ the distribution of $F(U)$.

Show that the probability measure μ has a continuous distribution function and that μ is singular to the Lebesgue measure $\lambda\big|_{(0,1]}$.

Exercise 7.4.2 Let $n \in \mathbb{N}$ and $p, q \in [0, 1]$. For which values of p and q do we have $b_{n,p} \ll b_{n,q}$? Compute the Radon–Nikodym derivative $\frac{db_{n,p}}{db_{n,q}}$.

7.5 Supplement: Signed Measures

In this section, we show the decomposition theorems for signed measures (Hahn, Jordan) and deliver an alternative proof for Lebesgue's decomposition theorem. We owe some of the proofs to [89].

Definition 7.35 Let μ and ν be two measures on $(\Omega, \mathcal{A})$. ν is called *totally continuous* with respect to μ if, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $A \in \mathcal{A}$

$$\mu(A) < \delta \quad \text{implies} \quad \nu(A) < \varepsilon. \tag{7.9}$$

Remark 7.36 The definition of total continuity is similar to that of uniform integrability (see Theorem 6.24(iii)), at least for finite μ. We will come back to this connection in the framework of the martingale convergence theorem that will provide an alternative proof of the Radon–Nikodym theorem (Corollary 7.34). ◊

Theorem 7.37 *Let μ and ν be measures on $(\Omega, \mathcal{A})$. If ν is totally continuous with respect to μ, then $\nu \ll \mu$. If $\nu(\Omega) < \infty$, then the converse also holds.*

Proof "$\Longrightarrow$ " Let ν be totally continuous with respect to μ. Let $A \in \mathcal{A}$ with $\mu(A) = 0$. For all $\varepsilon > 0$, by assumption, $\nu(A) < \varepsilon$; hence $\nu(A) = 0$ and thus $\nu \ll \mu$.

"$\Longleftarrow$ " Let ν be finite but not totally continuous with respect to μ. Then there exist an $\varepsilon > 0$ and sets $A_n \in \mathcal{A}$ with $\mu(A_n) < 2^{-n}$ but $\nu(A_n) \geq \varepsilon$ for all $n \in \mathbb{N}$. Define

$$A := \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Then

$$\mu(A) = \lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mu(A_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} 2^{-k} = 0.$$

Since ν is finite and upper semicontinuous (Theorem 1.36), we have

$$\nu(A) = \lim_{n\to\infty} \nu\left(\bigcup_{k=n}^{\infty} A_k\right) \geq \inf_{n\in\mathbb{N}} \nu(A_n) \geq \varepsilon > 0.$$

Thus $\nu \not\ll \mu$. □

Example 7.38 In the converse implication of the theorem, the assumption of finiteness is essential. For example, let $\mu = \mathcal{N}_{0,1}$ be the standard normal distribution on $\mathbb{R}$ and let ν be the Lebesgue measure on $\mathbb{R}$. Then ν has the density $f(x) = \sqrt{2\pi}\, e^{x^2/2}$ with respect to μ. In particular, we have $\nu \ll \mu$. On the other hand, $\mu([n, \infty)) \stackrel{n\to\infty}{\longrightarrow} 0$ and $\nu([n, \infty)) = \infty$ for any $n \in \mathbb{N}$. Hence ν is not totally continuous with respect to μ. ◊

Example 7.39 Let $(\Omega, \mathcal{A})$ be a measurable space and let μ and ν be finite measures on $(\Omega, \mathcal{A})$. Denote by $\mathcal{Z}$ the set of finite partitions of Ω into pairwise disjoint measurable sets. That is, $Z \in \mathcal{Z}$ is a finite subset of $\mathcal{A}$ such that the sets $C \in Z$ are pairwise disjoint and $\bigcup_{C\in Z} C = \Omega$ for all Z. For $Z \in \mathcal{Z}$, define a function $f_Z : \Omega \to \mathbb{R}$

by

$$f_Z(\omega) = \sum_{C \in Z : \mu(C) > 0} \frac{\nu(C)}{\mu(C)} \mathbb{1}_C(\omega).$$

We show that the following three statements are equivalent.

(i) The family $(f_Z : Z \in \mathcal{Z})$ is uniformly integrable in $\mathcal{L}^1(\mu)$ and $\int f_Z \, d\mu = \nu(\Omega)$ for any $Z \in \mathcal{Z}$.
(ii) $\nu \ll \mu$.
(iii) ν is totally continuous with respect to μ.

The equivalence of (ii) and (iii) was established in the preceding theorem. If (ii) holds, then, for all $Z \in \mathcal{Z}$,

$$\int f_Z \, d\mu = \sum_{C \in Z : \mu(C) > 0} \nu(C) = \nu(\Omega)$$

since $\nu(C) = 0$ for those C that do not appear in the sum. Now fix $\varepsilon > 0$. Since (ii) implies (iii), there is a $\delta' > 0$ such that $\nu(A) < \varepsilon/2$ for all $A \in \mathcal{A}$ with $\mu(A) \le \delta'$. Let $K := \nu(\Omega)/\delta'$ and $\delta < \varepsilon/(2K)$. Then

$$\mu\Bigg(\bigcup_{C \in Z : K\mu(C) \le \nu(C)} C \Bigg) = \sum_{C \in Z : K\mu(C) \le \nu(C)} \mu(C) \le \frac{1}{K} \nu(\Omega) = \delta';$$

hence

$$\sum_{C \in Z : K\mu(C) \le \nu(C)} \nu(C) = \nu\Bigg(\bigcup_{C \in Z : K\mu(C) \le \nu(C)} C \Bigg) < \frac{\varepsilon}{2}.$$

We conclude that for all $A \in \mathcal{A}$ with $\mu(A) < \delta$,

$$\begin{aligned}
\int_A f_Z \, d\mu &= \sum_{C \in Z : \mu(C) > 0} \mu(A \cap C) \frac{\nu(C)}{\mu(C)} \\
&= \sum_{0 < K\mu(C) \le \nu(C)} \mu(A \cap C) \frac{\nu(C)}{\mu(C)} + \sum_{K\mu(C) > \nu(C)} \mu(A \cap C) \frac{\nu(C)}{\mu(C)} \\
&\le \frac{\varepsilon}{2} + \sum_{K\mu(C) > \nu(C)} K\mu(A \cap C) \le \frac{\varepsilon}{2} + K\mu(A) < \varepsilon.
\end{aligned}$$

Hence $(f_Z, Z \in \mathcal{Z})$ is uniformly integrable by Theorem 6.24(iii).

Now assume (i). If $\mu = 0$, then $\int f \, d\mu = 0$ for all f; hence $\nu(\Omega) = 0$ and thus $\nu \ll \mu$. Hence, let $\mu \ne 0$. Let $A \in \mathcal{A}$ with $\mu(A) = 0$. Then $Z = \{A, A^c\} \in \mathcal{Z}$ and $f_Z = \mathbb{1}_{A^c} \nu(A^c)/\mu(A^c)$. By assumption, $\nu(\Omega) = \int f_Z \, d\mu = \nu(A^c)$; hence $\nu(A) = 0$ and thus $\nu \ll \mu$. ◊

Definition 7.40 (Signed measure) A set function $\varphi : \mathcal{A} \to \mathbb{R}$ is called a *signed measure* on $(\Omega, \mathcal{A})$ if it is σ-additive; that is, if for any sequence of pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$,

$$\varphi\left(\biguplus_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \varphi(A_n). \tag{7.10}$$

The set of all signed measures will be denoted by $\mathcal{M}^{\pm} = \mathcal{M}^{\pm}(\Omega, \mathcal{A})$.

Remark 7.41

(i) If φ is a signed measure, then in (7.10) we automatically have absolute convergence. Indeed, the value of the left-hand side does not change if we change the order of the sets $A_1, A_2, \ldots$. In order for this to hold for the right-hand side, by Weierstraß's theorem on rearrangements of series, the series has to converge absolutely. In particular, for any sequence $(A_n)_{n\in\mathbb{N}}$ of pairwise disjoint sets, we have $\lim_{n\to\infty} \sum_{k=n}^{\infty} |\varphi(A_k)| = 0$.
(ii) If $\varphi \in \mathcal{M}^{\pm}$, then $\varphi(\emptyset) = 0$ since $\mathbb{R} \ni \nu(\emptyset) = \sum_{n\in\mathbb{N}} \nu(\emptyset)$.
(iii) In general, $\varphi \in \mathcal{M}^{\pm}$ is not σ-subadditive. ◊

Example 7.42 If μ^+, μ^- are finite measures, then $\varphi := \mu^+ - \mu^- \in \mathcal{M}^{\pm}$. We will see that every signed measure has such a representation. ◊

Theorem 7.43 (Hahn's decomposition theorem) *Let φ be a signed measure. Then there is a set $\Omega^+ \in \mathcal{A}$ with $\varphi(A) \geq 0$ for all $A \in \mathcal{A}$, $A \subset \Omega^+$ and $\varphi(A) \leq 0$ for all $A \in \mathcal{A}$, $A \subset \Omega^- := \Omega \setminus \Omega^+$. Such a decomposition $\Omega = \Omega^- \uplus \Omega^+$ is called a Hahn decomposition of Ω (with respect to φ).*

Proof Let $\alpha := \sup\{\varphi(A) : A \in \mathcal{A}\}$. We have to show that φ attains the maximum α; that is, there exists an $\Omega^+ \in \mathcal{A}$ with $\varphi(\Omega^+) = \alpha$. If this is the case, then $\alpha \in \mathbb{R}$ and for $A \subset \Omega^+$, $A \in \mathcal{A}$, we would have

$$\alpha \geq \varphi\big(\Omega^+ \setminus A\big) = \varphi\big(\Omega^+\big) - \varphi(A) = \alpha - \varphi(A);$$

hence $\varphi(A) \geq 0$. For $A \subset \Omega^-$, $A \in \mathcal{A}$, we would have $\varphi(A) \leq 0$ since

$$\alpha \geq \varphi\big(\Omega^+ \cup A\big) = \varphi\big(\Omega^+\big) + \varphi(A) = \alpha + \varphi(A).$$

We now construct Ω^+ with $\varphi(\Omega^+) = \alpha$. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ with $\alpha = \lim_{n\to\infty} \varphi(A_n)$. Let $A := \bigcup_{n=1}^{\infty} A_n$. As each A_n could still contain "portions with negative mass", we cannot simply choose $\Omega^+ = A$. Rather, we have to peel off the negative portions layer by layer.

Define $A_n^0 := A_n$, $A_n^1 := A \setminus A_n$, and let

$$\mathcal{P}_n := \left\{ \bigcap_{i=1}^{n} A_i^{s(i)} : s \in \{0,1\}^n \right\}$$

be the partition of A that is generated by $A_1, \ldots, A_n$. Clearly, for any $B, C \in \mathcal{P}_n$, either $B = C$ or $B \cap C = \emptyset$ holds. In addition, we have

$$A_n = \biguplus_{\substack{B \in \mathcal{P}_n \\ B \subset A_n}} B.$$

Define

$$\mathcal{P}_n^- := \{B \in \mathcal{P}_n : \varphi(B) < 0\}, \qquad \mathcal{P}_n^+ := \mathcal{P}_n \setminus \mathcal{P}_n^-$$

and

$$C_n := \bigcup_{B \in \mathcal{P}_n^+} B.$$

Due to the finite additivity of φ, we have

$$\varphi(A_n) = \sum_{\substack{B \in \mathcal{P}_n \\ B \subset A_n}} \varphi(B) \le \sum_{\substack{B \in \mathcal{P}_n^+ \\ B \subset A_n}} \varphi(B) \le \sum_{B \in \mathcal{P}_n^+} \varphi(B) = \varphi(C_n).$$

For $m \le n$, let $E_m^n = C_m \cup \ldots \cup C_n$. Hence, for $m < n$, we have $E_m^n \setminus E_m^{n-1} \subset C_n$ and thus

$$E_m^n \setminus E_m^{n-1} = \biguplus_{\substack{B \in \mathcal{P}_n^+ \\ B \subset E_m^n \setminus E_m^{n-1}}} B.$$

In particular, this implies $\varphi(E_m^n \setminus E_m^{n-1}) \ge 0$. For $E_m := \bigcup_{n \ge m} C_n$, we also have $E_m^n \uparrow E_m$ $(n \to \infty)$ and

$$\begin{aligned}\varphi(A_m) \le \varphi(C_m) = \varphi\big(E_m^m\big) &\le \varphi\big(E_m^m\big) + \sum_{n=m+1}^{\infty} \varphi\big(E_m^n \setminus E_m^{n-1}\big) \\ &= \varphi\left(E_m^m \cup \bigcup_{n=m+1}^{\infty} \big(E_m^n \setminus E_m^{n-1}\big)\right) = \varphi\left(\bigcup_{n=m}^{\infty} E_m^n\right) = \varphi(E_m).\end{aligned}$$

Now define $\Omega^+ = \bigcap_{m=1}^{\infty} E_m$; hence $E_m \downarrow \Omega^+$. Then

$$\begin{aligned}\varphi(E_m) &= \varphi\Big(\Omega^+ \uplus \biguplus_{n \ge m} (E_n \setminus E_{n+1})\Big) \\ &= \varphi\big(\Omega^+\big) + \sum_{n=m}^{\infty} \varphi(E_n \setminus E_{n+1}) \overset{m \to \infty}{\longrightarrow} \varphi\big(\Omega^+\big).\end{aligned}$$

In the last step, we used Remark 7.41(i). Summing up, we have

$$\alpha = \lim_{m \to \infty} \varphi(A_m) \le \lim_{m \to \infty} \varphi(E_m) = \varphi\big(\Omega^+\big).$$

However, by definition, $\alpha \ge \varphi(\Omega^+)$; hence $\alpha = \varphi(\Omega^+)$. This finishes the proof. $\square$

Corollary 7.44 (Jordan's decomposition theorem) *Assume* $\varphi \in \mathcal{M}^{\pm}(\Omega, \mathcal{A})$ *is a signed measure. Then there exist uniquely determined finite measures* φ^+, φ^- *with* $\varphi = \varphi^+ - \varphi^-$ *and* $\varphi^+ \perp \varphi^-$.

Proof Let $\Omega = \Omega^+ \uplus \Omega^-$ be a Hahn decomposition. Define $\varphi^+(A) := \varphi(A \cap \Omega^+)$ and $\varphi^-(A) := -\varphi(A \cap \Omega^-)$.

The uniqueness of the decomposition is trivial. □

Corollary 7.45 *Let* $\varphi \in \mathcal{M}^{\pm}(\Omega, \mathcal{A})$ *and let* $\varphi = \varphi^+ - \varphi^-$ *be the Jordan decomposition of* φ. *Let* $\Omega = \Omega^+ \uplus \Omega^-$ *be a Hahn decomposition of* Ω. *Then*

$$\begin{aligned}\|\varphi\|_{TV} &:= \sup\big\{\varphi(A) - \varphi(\Omega \setminus A) : A \in \mathcal{A}\big\} \\ &= \varphi\big(\Omega^+\big) - \varphi\big(\Omega^-\big) \\ &= \varphi^+(\Omega) + \varphi^-(\Omega)\end{aligned}$$

defines a norm on $\mathcal{M}^{\pm}(\Omega, \mathcal{A})$, *the so-called total variation norm.*

Proof We only have to show the triangle inequality. Let $\varphi_1, \varphi_2 \in \mathcal{M}^{\pm}$. Let $\Omega = \Omega^+ \uplus \Omega^-$ be a Hahn decomposition with respect to $\varphi := \varphi_1 + \varphi_2$ and let $\Omega = \Omega_i^+ \uplus \Omega_i^-$ be a Hahn decomposition with respect to φ_i, $i = 1, 2$. Then

$$\begin{aligned}\|\varphi_1 + \varphi_2\|_{TV} &= \varphi_1\big(\Omega^+\big) - \varphi_1\big(\Omega^-\big) + \varphi_2\big(\Omega^+\big) - \varphi_2\big(\Omega^-\big) \\ &\le \varphi_1\big(\Omega_1^+\big) - \varphi_1\big(\Omega_1^-\big) + \varphi_2\big(\Omega_2^+\big) - \varphi_2\big(\Omega_2^-\big) \\ &= \|\varphi_1\|_{TV} + \|\varphi_2\|_{TV}.\end{aligned}$$ □

With a lemma, we prepare for an alternative proof of Lebesgue's decomposition theorem (Theorem 7.33).

Lemma 7.46 *Let* μ, ν *be finite measures on* $(\Omega, \mathcal{A})$ *that are not mutually singular; in short,* $\mu \not\perp \nu$. *Then there is an* $A \in \mathcal{A}$ *with* $\mu(A) > 0$ *and an* $\varepsilon > 0$ *with*

$$\varepsilon\mu(E) \le \nu(E) \quad \text{for all } E \in \mathcal{A} \text{ with } E \subset A.$$

Proof For $n \in \mathbb{N}$, let $\Omega = \Omega_n^+ \uplus \Omega_n^-$ be a Hahn decomposition for $(\nu - \frac{1}{n}\mu) \in \mathcal{M}^{\pm}$. Define $M := \bigcap_{n\in\mathbb{N}} \Omega_n^-$. Clearly, $(\nu - \frac{1}{n}\mu)(M) \le 0$; hence $\nu(M) \le \frac{1}{n}\mu(M)$ for all $n \in \mathbb{N}$ and thus $\nu(M) = 0$. Since $\mu \not\perp \nu$, we get $\mu(\Omega \setminus M) = \mu(\bigcup_{n\in\mathbb{N}} \Omega_n^+) > 0$. Thus $\mu(\Omega_{n_0}^+) > 0$ for some $n_0 \in \mathbb{N}$. Define $A := \Omega_{n_0}^+$ and $\varepsilon := \frac{1}{n_0}$. Then $\mu(A) > 0$ and $(\nu - \varepsilon\mu)(E) \ge 0$ for all $E \subset A$, $E \in \mathcal{A}$. □

Alternative proof of Theorem 7.33 We show only the existence of a decomposition. By choosing a suitable sequence $\Omega_n \uparrow \Omega$, we can assume that ν is finite. Consider the set of functions

$$\mathcal{G} := \left\{g : \Omega \to [0, \infty] : g \text{ is measurable and } \int_A g\, d\mu \le \nu(A) \text{ for all } A \in \mathcal{A}\right\},$$

and define

$$\gamma := \sup\left\{ \int g \, d\mu : g \in \mathcal{G} \right\}.$$

Our aim is to find a maximal element f in $\mathcal{G}$ (i.e., an f for which $\int f \, d\mu = \gamma$). This f will be the density of ν_a.

Clearly, $0 \in \mathcal{G}$; hence $\mathcal{G} \neq \emptyset$. Furthermore,

$$f, g \in \mathcal{G} \quad \text{implies} \quad f \vee g \in \mathcal{G}. \tag{7.11}$$

Indeed, letting $E := \{f \geq g\}$, for all $A \in \mathcal{A}$, we have

$$\int_A (f \vee g) \, d\mu = \int_{A \cap E} f \, d\mu + \int_{A \setminus E} g \, d\mu \leq \nu(A \cap E) + \nu(A \setminus E) = \nu(A).$$

Choose a sequence $(g_n)_{n \in \mathbb{N}}$ in $\mathcal{G}$ such that $\int g_n \, d\mu \overset{n \to \infty}{\longrightarrow} \gamma$, and define the function $f_n = g_1 \vee \ldots \vee g_n$. Now (7.11) implies $f_n \in \mathcal{G}$. Letting $f := \sup\{f_n : n \in \mathbb{N}\}$, the monotone convergence theorem yields

$$\int_A f \, d\mu = \sup_{n \in \mathbb{N}} \int_A f_n \, d\mu \leq \nu(A) \quad \text{for all } A \in \mathcal{A}$$

(that is, $f \in \mathcal{G}$), and

$$\int f \, d\mu = \sup_{n \in \mathbb{N}} \int f_n \, d\mu \geq \sup_{n \in \mathbb{N}} \int g_n \, d\mu = \gamma.$$

Hence $\int f \, d\mu = \gamma \leq \nu(\Omega)$. Now define, for any $A \in \mathcal{A}$,

$$\nu_a(A) := \int_A f \, d\mu \quad \text{and} \quad \nu_s(A) := \nu(A) - \nu_a(A).$$

By construction, $\nu_a \ll \mu$ is a finite measure with density f with respect to μ. Since $f \in \mathcal{G}$, we have $\nu_s(A) = \nu(A) - \int_A f \, d\mu \geq 0$ for all $A \in \mathcal{A}$, and thus also ν_s is a finite measure. It remains to show $\nu_s \perp \mu$.

At this point we use Lemma 7.46. Assume that we had $\nu_s \not\perp \mu$. Then there would be an $\varepsilon > 0$ and an $A \in \mathcal{A}$ with $\mu(A) > 0$ such that $\varepsilon \mu(E) \leq \nu_s(E)$ for all $E \subset A$, $E \in \mathcal{A}$. Then, for $B \in \mathcal{A}$, we would have

$$\begin{aligned} \int_B (f + \varepsilon \mathbb{1}_A) \, d\mu &= \int_B f \, d\mu + \varepsilon \mu(A \cap B) \\ &\leq \nu_a(B) + \nu_s(A \cap B) \leq \nu_a(B) + \nu_s(B) = \nu(B). \end{aligned}$$

In other words, $(f + \varepsilon \mathbb{1}_A) \in \mathcal{G}$ and thus $\int (f + \varepsilon \mathbb{1}_A) \, d\mu = \gamma + \varepsilon \mu(A) > \gamma$, contradicting the definition of γ. Hence in fact $\nu_s \perp \mu$. □

Exercise 7.5.1 Let μ be a σ-finite measure on $(\Omega, \mathcal{A})$ and let φ be a signed measure on $(\Omega, \mathcal{A})$. Show that, analogously to the Radon–Nikodym theorem, the following two statements are equivalent:

(i) $\varphi(A) = 0$ for all $A \in \mathcal{A}$ with $\mu(A) = 0$.
(ii) There is an $f \in \mathcal{L}^1(\mu)$ with $\varphi = f\mu$; hence $\int_A f\, d\mu = \varphi(A)$ for all $A \in \mathcal{A}$.

Exercise 7.5.2 Let μ, ν, α be finite measures on $(\Omega, \mathcal{A})$ with $\nu \ll \mu \ll \alpha$.

(i) Show the chain rule for the Radon–Nikodym derivative:

$$\frac{d\nu}{d\alpha} = \frac{d\nu}{d\mu}\frac{d\mu}{d\alpha} \quad \alpha\text{-a.e.}$$

(ii) Show that $f := \frac{d\nu}{d(\mu+\nu)}$ exists and that $\frac{d\nu}{d\mu} = \frac{f}{1-f}$ holds μ-a.e.

7.6 Supplement: Dual Spaces

By the Riesz–Fréchet theorem (Theorem 7.26), every continuous linear functional $F : L^2(\mu) \to \mathbb{R}$ has a representation $F(g) = \langle f, g\rangle$ for some $f \in L^2(\mu)$. On the other hand, for any $f \in L^2(\mu)$, the map $L^2(\mu) \to \mathbb{R}$, $g \mapsto \langle f, g\rangle$ is continuous and linear. Hence $L^2(\mu)$ is canonically isomorphic to its topological dual space $(L^2(\mu))'$. This dual space is defined as follows.

Definition 7.47 (Dual space) Let $(V, \|\cdot\|)$ be a Banach space. The *dual space* V' of V is defined by

$$V' := \{F : V \to \mathbb{R} \text{ is continuous and linear}\}.$$

For $F \in V'$, we define $\|F\|' := \sup\{|F(f)| : \|f\| = 1\}$.

Remark 7.48 As F is continuous, for any $\delta > 0$, there exists an $\varepsilon > 0$ such that $|F(f)| < \delta$ for all $f \in V$ with $\|f\| < \varepsilon$. Hence $\|F\|' \le \delta/\varepsilon < \infty$. ◊

We are interested in the case $V = L^p(\mu)$ for $p \in [1, \infty]$ and write $\|F\|'_p$ for the norm of $F \in V'$. In the particular case $V = L^2(\mu)$, by the Cauchy–Schwarz inequality, we have $\|F\|'_2 = \|f\|_2$. This can be generalized:

Lemma 7.49 *Let* $p, q \in [1, \infty]$ *with* $\frac{1}{p} + \frac{1}{q} = 1$. *The canonical map*

$$\kappa : L^q(\mu) \to \big(L^p(\mu)\big)'$$
$$\kappa(f)(g) = \int fg\, d\mu \quad \text{for } f \in L^q(\mu), g \in L^p(\mu)$$

is an isometry; *that is*, $\|\kappa(f)\|'_p = \|f\|_q$.

Proof We show equality by showing the two inequalities separately.

"$\le$" This follows from Hölder's inequality.

"$\ge$" For any admissible pair p, q and all $f \in \mathcal{L}^q(\mu)$, $g \in \mathcal{L}^p(\mu)$, by the definition of the operator norm, $\|\kappa(f)\|'_p \|g\|_p \ge |\int fg\, d\mu|$. Define the sign function $\operatorname{sign}(x) = \mathbb{1}_{(0,\infty)}(x) - \mathbb{1}_{(-\infty,0)}(x)$. Replacing g by $\tilde{g} := |g| \operatorname{sign}(f)$ (note that $\|\tilde{g}\|_p = \|g\|_p$), we obtain

$$\big\|\kappa(f)\big\|'_p \|g\|_p \ge \left| \int f \tilde{g}\, d\mu \right| = \|fg\|_1. \tag{7.12}$$

First consider the case $q = 1$ and $f \in \mathcal{L}^1(\mu)$. Applying (7.12) with $g \equiv 1 \in \mathcal{L}^\infty(\mu)$ yields $\|\kappa(f)\|'_\infty \ge \|f\|_1$.

Now let $q \in (1, \infty)$. Let $g := |f|^{q-1}$. Since $\frac{q-1}{q} = \frac{1}{p}$, we have

$$\big\|\kappa(f)\big\|'_p \cdot \|g\|_p \ge \|fg\|_1 = \big\| |f|^q \big\|_1 = \|f\|_q^q = \|f\|_q \cdot \|f\|_q^{q-1} = \|f\|_q \cdot \|g\|_p.$$

Finally, let $q = \infty$. Without loss of generality, assume $\|f\|_\infty \in (0, \infty)$. Let $\varepsilon > 0$. Then there exists an $A_\varepsilon \in \mathcal{A}$ with $0 < \mu(A_\varepsilon) < \infty$ such that

$$A_\varepsilon \subset \big\{ |f| > (1 - \varepsilon)\|f\|_\infty \big\}.$$

If we let $g = \frac{1}{\mu(A_\varepsilon)} \mathbb{1}_{A_\varepsilon}$, then $\|g\|_1 = 1$ and $\|\kappa(f)\|'_1 \ge \|fg\|_1 \ge (1 - \varepsilon)\|f\|_\infty$. □

Theorem 7.50 *Let $p \in [1, \infty)$ and assume $\frac{1}{p} + \frac{1}{q} = 1$. Then $L^q(\mu)$ is isomorphic to its dual space $(L^p(\mu))'$ by virtue of the isometry κ.*

Proof The proof makes use of the Radon–Nikodym theorem (Corollary 7.34). However, here we only sketch the proof since we do not want to go into the details of signed measures and signed contents. A signed content ν is an additive set function that is the difference $\nu = \nu^+ - \nu^-$ of two finite contents. This definition is parallel to that of a signed measure that is the difference of two finite measures.

As κ is an isometry, κ in particular is injective. Hence we only have to show that κ is surjective. Let $F \in (L^p(\mu))'$. Then $\nu(A) = F(\mathbb{1}_A)$ is a signed content on $\mathcal{A}$ and we have

$$\big|\nu(A)\big| \le \|F\|'_p \big(\mu(A)\big)^{1/p}.$$

Since μ is $\emptyset$-continuous, ν is also $\emptyset$-continuous and is thus a signed measure on $\mathcal{A}$. We even have $\nu \ll \mu$. By the Radon–Nikodym theorem (Corollary 7.34) (applied to the measures ν^- and ν^+; see Exercise 7.5.1), ν admits a density with respect to μ; that is, a measurable function f with $\nu = f\mu$.

Let

$$\mathbb{E}_f := \big\{ g : g \text{ is a simple function with } \mu(g \ne 0) < \infty \big\}$$

and let

$$\mathbb{E}_f^+ := \{g \in \mathbb{E}_f : g \ge 0\}.$$

Then, for $g \in \mathbb{E}_f$,

$$F(g) = \int g f \, d\mu. \tag{7.13}$$

In order to show that (7.13) holds for all $g \in \mathcal{L}^p(\mu)$, we first show $f \in \mathcal{L}^q(\mu)$. To this end, we distinguish two cases.

Case 1: $p = 1$ For every $\alpha > 0$,

$$\begin{aligned}\mu\big(\{|f| > \alpha\}\big) &\le \frac{1}{\alpha}\nu\big(\{|f| > \alpha\}\big)\\ &= \frac{1}{\alpha}F(\mathbb{1}_{\{|f|>\alpha\}}) \le \frac{1}{\alpha}\|F\|_1' \cdot \|\mathbb{1}_{\{|f|>\alpha\}}\|_1 = \frac{1}{\alpha}\|F\|_1' \cdot \mu\big(\{|f| > \alpha\}\big).\end{aligned}$$

This implies $\mu(\{|f| > \alpha\}) = 0$ if $\alpha > \|F\|_1'$; hence $\|f\|_\infty \le \|F\|_1' < \infty$.

Case 2: $p \in (1, \infty)$ By Theorem 1.96, there are $g_1, g_2, \ldots \in \mathbb{E}_f^+$ such that $g_n \uparrow |f|$ μ-a.e. Define $h_n = \operatorname{sign}(f)(g_n)^{q-1} \in \mathbb{E}_f$; hence

$$\begin{aligned}\|g_n\|_q^q &\le \int h_n f \, d\mu = F(h_n)\\ &\le \|F\|_p' \cdot \|h_n\|_p = \|F\|_p' \cdot \big(\|g_n\|_q\big)^{q-1}.\end{aligned}$$

Thus we have $\|g_n\|_q \le \|F\|_p'$. Monotone convergence (Theorem 4.20) now yields $\|f\|_q \le \|F\|_p' < \infty$; hence $f \in \mathcal{L}^q(\mu)$.

Concluding, the map $\widetilde{F} : g \mapsto \int g f \, d\mu$ is in $(L^p(\mu))'$, and $\widetilde{F}(g) = F(g)$ for every $g \in \mathbb{E}_f$. Since $\widetilde{F}$ is continuous and $\mathbb{E}_f \subset L^p(\mu)$ is dense, we get $\widetilde{F} = F$. □

Remark 7.51 For $p = \infty$, the statement of Theorem 7.50 is false in general. (For finite $\mathcal{A}$, the claim is trivially true even for $p = \infty$.) For example, let $\Omega = \mathbb{N}$, $\mathcal{A} = 2^\Omega$ and let μ be the counting measure. Thus we consider sequence spaces $\ell^p = L^p(\mathbb{N}, 2^{\mathbb{N}}, \mu)$. For the subspace $\ell^K \subset \ell^\infty$ of convergent sequences, $F : \ell^K \to \mathbb{R}$, $(a_n)_{n\in\mathbb{N}} \mapsto \lim_{n\to\infty} a_n$ is a continuous linear functional. By the Hahn–Banach theorem of functional analysis (see, e.g., [87] or [173]), F can be extended to a continuous linear functional on ℓ^∞. However, clearly there is no sequence $(b_n)_{n\in\mathbb{N}} \in \ell^1$ with $F((a_n)_{n\in\mathbb{N}}) = \sum_{m=1}^\infty a_m b_m$. ◊

Exercise 7.6.1 Show that $\mathbb{E}_f \subset L^p(\mu)$ is dense if $p \in [1, \infty)$.