Appendix A: Measure Theory

Complete treatments of the results stated in this appendix may be found in any measure theory book; see for example Parthasarathy [281], Royden [321] or Kingman and Taylor [195]. A similar summary of measure theory without proofs may be found in Walters [375, Chap. 0]. Some of this appendix will use terminology from Appendix B.

A.1 Measure Spaces

Let X be a set, which will usually be infinite, and denote by $\mathbb{P}(X)$ the collection of all subsets of X.

Definition A.1. *A set $\mathscr{S} \subseteq \mathbb{P}(X)$ is called a* semi-algebra *if*

(1) $\varnothing \in \mathscr{S}$,
(2) $A, B \in \mathscr{S}$ *implies that* $A \cap B \in \mathscr{S}$, *and*
(3) *if* $A \in \mathscr{S}$ *then the complement* $X \smallsetminus A$ *is a finite union of pairwise disjoint elements in* $\mathscr{S}$*;*

if in addition

(4) $A \in \mathscr{S}$ *implies that* $X \smallsetminus A \in \mathscr{S}$,

then it is called an algebra. *If* $\mathscr{S}$ *satisfies the additional property*

(5) $A_1, A_2, \dots \in \mathscr{S}$ *implies that* $\bigcup_{n=1}^{\infty} A_n \in \mathscr{S}$,

then $\mathscr{S}$ *is called a* σ*-algebra. For any collection of sets* $\mathscr{A}$*, write* $\sigma(\mathscr{A})$ *for the smallest* σ*-algebra containing* $\mathscr{A}$ *(this is possible since the intersection of* σ*-algebras is a* σ*-algebra).*

Example A.2. The collection of intervals in $[0, 1]$ forms a semi-algebra.

Definition A.3. *A collection* $\mathscr{M} \subseteq \mathbb{P}(X)$ *is called a* monotone class *if*

$$A_1 \subseteq A_2 \subseteq \cdots \text{ and } A_n \in \mathscr{M} \text{ for all } n \geqslant 1 \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{M}$$

and

$$B_1 \supseteq B_2 \supseteq \cdots \text{ and } B_n \in \mathscr{M} \text{ for all } n \geqslant 1 \implies \bigcap_{n=1}^{\infty} B_n \in \mathscr{M}.$$

The intersection of two monotone classes is a monotone class, so there is a well-defined smallest monotone class $\mathscr{M}(\mathscr{A})$ containing any given collection of sets $\mathscr{A}$. This gives an alternative characterization of the σ-algebra generated by an algebra.

Theorem A.4. *Let* $\mathscr{A}$ *be an algebra. Then the smallest monotone class containing* $\mathscr{A}$ *is* $\sigma(\mathscr{A})$.

A function $\mu : \mathscr{S} \to \mathbb{R}_{\geqslant 0} \cup \{\infty\}$ is *finitely additive* if $\mu(\varnothing) = 0$ and*

$$\mu(A \cup B) = \mu(A) + \mu(B) \tag{A.1}$$

for any disjoint elements A and B of $\mathscr{S}$ with $A \sqcup B \in \mathscr{S}$, and is *countably additive* if

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

if $\{A_n\}$ is a collection of disjoint elements of $\mathscr{S}$ with $\bigsqcup_{n=1}^{\infty} A_n \in \mathscr{S}$.

The main structure of interest in ergodic theory is that of a *probability space* or *finite measure space*.

Definition A.5. *A triple* $(X, \mathscr{B}, \mu)$ *is called a finite measure space if* $\mathscr{B}$ *is a* σ-*algebra and* μ *is a countably additive measure defined on* $\mathscr{B}$ *with* $\mu(X) < \infty$. *A triple* $(X, \mathscr{B}, \mu)$ *is called a* σ-*finite measure space if* X *is a countable union of elements of* $\mathscr{B}$ *of finite measure. If* $\mu(X) = 1$ *then a finite measure space is called a probability space.*

A probability measure μ is said to be *concentrated* on a measurable set A if $\mu(A) = 1$.

Theorem A.6. *If* $\mu : \mathscr{S} \to \mathbb{R}_{\geqslant 0}$ *is a countably additive measure defined on a semi-algebra, then there is a unique countably additive measure defined on* $\sigma(\mathscr{S})$ *which extends* μ.

* The conventions concerning ∞ in this setting are that $\infty + c = \infty$ for any c in $\mathbb{R}_{\geqslant 0} \cup \{\infty\}$, $c \cdot \infty = \infty$ for any $c > 0$, and $0 \cdot \infty = 0$.

Theorem A.7. *Let $\mathscr{A} \subseteq \mathscr{B}$ be an algebra in a probability space $(X, \mathscr{B}, \mu)$. Then the collection of sets B with the property that for any $\varepsilon > 0$ there is an $A \in \mathscr{A}$ with $\mu(A \triangle B) < \varepsilon$ is a σ-algebra.*

As discussed in Section 2.1, the basic objects of ergodic theory are measure-preserving maps (see Definition 2.1). The next result gives a convenient way to check whether a transformation is measure-preserving.

Theorem A.8. *Let $(X, \mathscr{B}_X, \mu)$ and $(Y, \mathscr{B}_Y, \nu)$ be probability spaces, and let $\mathscr{S}$ be a semi-algebra which generates $\mathscr{B}_Y$. A measurable map $\phi : X \to Y$ is* measure-preserving *if and only if*

$$\mu(\phi^{-1}B) = \nu(B)$$

for all $B \in \mathscr{S}$.

Proof. Let

$$\mathscr{S}' = \{B \in \mathscr{B}_Y \mid \phi^{-1}(B) \in \mathscr{B}_X, \mu(\phi^{-1}B) = \nu(B)\}.$$

Then $\mathscr{S} \subseteq \mathscr{S}'$, and (since each member of the algebra generated by $\mathscr{S}$ is a finite disjoint union of elements of $\mathscr{S}$) the algebra generated by $\mathscr{S}$ lies in $\mathscr{S}'$. It is clear that $\mathscr{S}'$ is a monotone class, so Theorem A.4 shows that $\mathscr{S}' = \mathscr{B}_Y$ as required. □

The next result is an important lemma from probability; what it means is that if the sum of the probabilities of a sequence of events is finite, then the probability that infinitely many of them occur is zero.

Theorem A.9 (Borel–Cantelli[102]). *Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $(A_n)_{n \geqslant 1}$ be a sequence of measurable sets with $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. Then*

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right)\right) = 0.$$

If the sequence of sets are pairwise independent, that is if

$$\mu(A_i \cap A_j) = \mu(A_i)\mu(A_j)$$

for all $i \neq j$, then $\sum_{n=1}^{\infty} \mu(A_n) = \infty$ implies that

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \mu\left(\bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right)\right) = 1.$$

The elements of a σ-algebra are typically very complex, and it is often enough to approximate sets by a convenient smaller collection of sets.

Theorem A.10. *If $(X, \mathscr{B}, \mu)$ is a probability space and $\mathscr{A}$ is an algebra which generates $\mathscr{B}$ (that is, with $\sigma(\mathscr{A}) = \mathscr{B}$), then for any $B \in \mathscr{B}$ and $\varepsilon > 0$ there is an $A \in \mathscr{A}$ with $\mu(A \triangle B) < \varepsilon$.*

A measure space is called *complete* if any subset of a null set is measurable.

If X is a topological space, then there is a distinguished collection of sets to start with, namely the open sets. The σ-algebra generated by the open sets is called the *Borel σ-algebra*. If the space is second countable, then the *support* of a measure is the largest closed set with the property that every open neighborhood of every point in the set has positive measure; equivalently the support of a measure is the complement of the largest open set of zero measure.

If X is a metric space, then any Borel probability measure μ on X (that is, any probability measure defined on the Borel σ-algebra $\mathscr{B}$ of X) is *regular*$^{(103)}$: for any Borel set $B \subseteq X$ and $\varepsilon > 0$ there is an open set O and a closed set C with $C \subseteq B \subseteq O$ and $\mu(O \smallsetminus C) < \varepsilon$.

A.2 Product Spaces

Let $I \subseteq \mathbb{Z}$ and assume that for each $i \in I$ a probability space $\mathsf{X}_i = (X_i, \mathscr{B}_i, \mu_i)$ is given. Then the product space $X = \prod_{i \in I} X_i$ may be given the structure of a probability space $(X, \mathscr{B}, \mu)$ as follows. Any set of the form

$$\prod_{i \in I, i < \min(F)} X_i \times \prod_{i \in F} A_i \times \prod_{i \in I, i > \max(F)} X_i,$$

or equivalently of the form

$$\{x = (x_i)_{i \in I} \in X \mid x_i \in A_i \text{ for } i \in F\},$$

for some finite set $F \subseteq I$, is called a *measurable rectangle*. The collection of all measurable rectangles forms a semi-algebra $\mathscr{S}$, and the product σ-algebra is $\mathscr{B} = \sigma(\mathscr{S})$. The product measure μ is obtained by defining the measure of the measurable rectangle above to be $\prod_{i \in F} \mu_i(A_i)$ and then extending to $\mathscr{B}$.

The main extension result in this setting is the Kolmogorov consistency theorem, which allows measures on infinite product spaces to be built up from measures on finite product spaces.

Theorem A.11. *Let $X = \prod_{i \in I} X_i$ with $I \subseteq \mathbb{Z}$ and each X_i a probability space. Suppose that for every finite subset $F \subseteq I$ there is a probability measure μ_F defined on $X_F = \prod_{i \in F} X_i$, and that these measures are consistent in the sense that if $E \subseteq F$ then the projection map*

$$\left(\prod_{i \in F} X_i, \mu_F \right) \longrightarrow \left(\prod_{i \in E} X_i, \mu_E \right)$$

is measure-preserving. Then there is a unique probability measure μ on the probability space $\prod_{i \in I} X_i$ with the property that for any $F \subseteq I$ the projection map

$$\left(\prod_{i\in I} X_i, \mu\right) \longrightarrow \left(\prod_{i\in F} X_i, \mu_F\right)$$

is measure-preserving.

In the construction of an infinite product $\prod_{i\in I}\mu_i$ of probability measures above, the finite products $\mu_F = \prod_{i\in F}\mu_i$ satisfy the compatibility conditions needed in Theorem A.11.

In many situations each $X_i = (X_i, \mathsf{d}_i)$ is a fixed compact metric space with $0 < \operatorname{diam}(X_i) < \infty$. In this case the product space $X = \prod_{n\in\mathbb{Z}} X_n$ is also a compact metric space with respect to the metric

$$\mathsf{d}(x,y) = \sum_{n\in\mathbb{Z}} \frac{d_n(x_n, y_n)}{2^n \operatorname{diam}(X_n)},$$

and the Borel σ-algebra of X coincides with the product σ-algebra defined above.

A.3 Measurable Functions

Let $(X, \mathscr{B}, \mu)$ be a probability space. Natural classes of measurable functions on X are built up from simpler functions, just as the σ-algebra $\mathscr{B}$ may be built up from simpler collections of sets.

A function $f : X \to \mathbb{R}$ is called *simple* if

$$f(x) = \sum_{j=1}^{m} c_j \chi_{A_j}(x)$$

for constants $c_j \in \mathbb{R}$ and disjoint sets $A_j \in \mathscr{B}$. The integral of f is then defined to be

$$\int f \,\mathrm{d}\mu = \sum_{j=1}^{m} c_j \mu(A_j).$$

A function $g : X \to \mathbb{R}$ is called *measurable* if $g^{-1}(A) \in \mathscr{B}$ for any (Borel) measurable set $A \subseteq \mathbb{R}$. The basic approximation result states that for any measurable function $g : X \to \mathbb{R}_{\geqslant 0}$ there is a pointwise increasing sequence of simple functions $(f_n)_{n\geqslant 1}$ with $f_n(x) \nearrow g(x)$ for each $x \in X$. This allows us to define

$$\int g \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu,$$

which is guaranteed to exist since

$$f_n(x) \leqslant f_{n+1}(x)$$

for all $n \geqslant 1$ and $x \in X$ (in contrast to the usual terminology from calculus, we include the possibility that the integral and the limit are infinite). It may be shown that this is well-defined (independent of the choice of the sequence of simple functions).

A measurable function $g : X \to \mathbb{R}_{\geqslant 0}$ is *integrable* if $\int g \,\mathrm{d}\mu < \infty$. In general, a measurable function $g : X \to \mathbb{R}$ has a unique decomposition into $g = g^+ - g^-$ with $g^+(x) = \max\{g(x), 0\}$; both g^+ and g^- are measurable. The function g is said to be integrable if both g^+ and g^- are integrable, and the integral is defined by $\int g \,\mathrm{d}\mu = \int g^+ \,\mathrm{d}\mu - \int g^- \,\mathrm{d}\mu$. If f is integrable and g is measurable with $|g| \leqslant f$, then g is integrable. The integral of an integrable function f over a measurable set A is defined by

$$\int_A f \,\mathrm{d}\mu = \int f \chi_A \,\mathrm{d}\mu.$$

For $1 \leqslant p < \infty$, the space $\mathscr{L}^p_\mu$ (or $\mathscr{L}^p(X)$, $\mathscr{L}^p(X, \mu)$ and so on) comprises the measurable functions $f : X \to \mathbb{R}$ with $\int |f|^p \,\mathrm{d}\mu < \infty$. Define an equivalence relation on $\mathscr{L}^p_\mu$ by $f \sim g$ if $\int |f - g|^p \,\mathrm{d}\mu = 0$ and write $L^p_\mu = \mathscr{L}^p_\mu / \sim$ for the space of equivalence classes. Elements of L^p_μ will be described as functions rather than equivalence classes, but it is important to remember that this is an abuse of notation (for example, in the construction of conditional measures on page 139). In particular the value of an element of L^p_μ at a specific point does not make sense, unless that point itself has positive μ-measure. The function $\|\cdot\|_p$ defined by

$$\|f\|_p = \left(\int |f|^p \,\mathrm{d}\mu \right)^{1/p}$$

is a norm (see Appendix B), and under this norm L^p is a Banach space.

The case $p = \infty$ is distinguished: the *essential supremum* is the generalization to measurable functions of the supremum of a continuous function, and is defined by

$$\|f\|_\infty = \inf \big\{ \alpha \mid \mu(\{x \in X \mid f(x) > \alpha\}) = 0 \big\} .$$

The space $\mathscr{L}^\infty_\mu$ is then defined to be the space of measurable functions f with $\|f\|_\infty < \infty$, and once again L^∞_μ is defined to be $\mathscr{L}^\infty_\mu / \sim$. The norm $\|\cdot\|_\infty$ makes L^∞_μ into a Banach space. For $1 \leqslant p < q \leqslant \infty$ we have $L^p \supseteq L^q$ for any finite measure space, with strict inclusion except in some degenerate cases.

In practice we will more often use $\mathscr{L}^\infty$, which denotes the bounded functions.

An important consequences of the Borel–Cantelli lemma is that norm convergence in L^p forces pointwise convergence along a subsequence.

Corollary A.12. *If (f_n) is a sequence convergent in L^p_μ $(1 \leqslant p \leqslant \infty)$ to f, then there is a subsequence (f_{n_k}) converging pointwise almost everywhere to f.*

Proof. Choose the sequence (n_k) so that

$$\|f_{n_k} - f\|_p^p < \frac{1}{k^{2+p}}$$

for all $k \geqslant 1$. Then

$$\mu\left(\left\{x \in X \,\middle|\, |f_{n_k}(x) - f(x)| > \frac{1}{k}\right\}\right) < \frac{1}{k^2}.$$

It follows by Theorem A.9 that for almost every x, $|f_{n_k}(x) - f(x)| > \frac{1}{k}$ for only finitely many k, so $f_{n_k}(x) \to f(x)$ for almost every x. □

Finally we turn to integration of functions of several variables; a measure space $(X, \mathscr{B}, \mu)$ is called σ-finite if there is a sequence $A_1, A_2, \dots$ of measurable sets with $\mu(A_n) < \infty$ for all $n \geqslant 1$ and with $X = \bigcup_{n \geqslant 1} A_n$.

Theorem A.13 (Fubini–Tonelli[104]). *Let f be a non-negative integrable function on the product of two σ-finite measure spaces $(X, \mathscr{B}, \mu)$ and $(Y, \mathscr{C}, \nu)$. Then, for almost every $x \in X$ and $y \in Y$, the functions*

$$h(x) = \int_Y f(x, y)\,\mathrm{d}\nu, \quad g(y) = \int_X f(x, y)\,\mathrm{d}\mu$$

are integrable, and

$$\int_{X \times Y} f\,\mathrm{d}(\mu \times \nu) = \int_X h\,\mathrm{d}\mu = \int_Y g\,\mathrm{d}\nu. \tag{A.2}$$

This may also be written in a more familiar form as

$$\begin{aligned}\int_{X \times Y} f(x, y)\,\mathrm{d}(\mu \times \nu)(x, y) &= \int_X \left(\int_Y f(x, y)\,\mathrm{d}\nu(y)\right)\mathrm{d}\mu(x) \\ &= \int_Y \left(\int_X f(x, y)\,\mathrm{d}\mu(x)\right)\mathrm{d}\nu(y).\end{aligned}$$

We note that integration makes sense for functions taking values in some other spaces as well, and this will be discussed further in Section B.7.

A.4 Radon–Nikodym Derivatives

One of the fundamental ideas in measure theory concerns the properties of a probability measure viewed from the perspective of a given measure. Fix a σ-finite measure space $(X, \mathscr{B}, \mu)$ and some measure ν defined on $\mathscr{B}$.

- The measure ν is *absolutely continuous* with respect to μ, written $\nu \ll \mu$, if $\mu(A) = 0 \implies \nu(A) = 0$ for any $A \in \mathscr{B}$.
- If $\nu \ll \mu$ and $\mu \ll \nu$ then μ and ν are said to be *equivalent*.
- The measures μ and ν are mutually *singular*, written $\mu \perp \nu$, if there exist disjoint sets A and B in $\mathscr{B}$ with $A \cup B = X$ and with $\mu(A) = \nu(B) = 0$.

These notions are related by two important theorems.

Theorem A.14 (Lebesgue decomposition). *Given σ-finite measures μ and ν on $(X, \mathscr{B})$, there are measures ν_0 and ν_1 with the properties that*

(1) $\nu = \nu_0 + \nu_1$;
(2) $\nu_0 \ll \mu$; *and*
(3) $\nu_1 \perp \mu$.

The measures ν_0 and ν_1 are uniquely determined by these properties.

Theorem A.15 (Radon–Nikodym derivative[105]). *If $\nu \ll \mu$ then there is a measurable function $f \geqslant 0$ on X with the property that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu$$

for any set $A \in \mathscr{B}$.

By analogy with the fundamental theorem of calculus (Theorem A.25), the function f is written $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and is called the *Radon–Nikodym derivative* of ν with respect to μ. Notice that for any two measures μ_1, μ_2 we can form a new measure $\mu_1 + \mu_2$ simply by defining $(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A)$ for any measurable set A. Then $\mu_i \ll \mu_1 + \mu_2$, so there is a Radon–Nikodym derivative of μ_i with respect to $\mu_1 + \mu_2$ for $i = 1, 2$.

A.5 Convergence Theorems

The most important distinction between integration on L^p spaces as defined above and Riemann integration on bounded Riemann-integrable functions is that the L^p functions are closed under several natural limiting operations, allowing for the following important convergence theorems.

Theorem A.16 (Monotone Convergence Theorem). *If $f_1 \leqslant f_2 \leqslant \cdots$ is a pointwise increasing sequence of integrable functions on the probability space $(X, \mathscr{B}, \mu)$, then $f = \lim_{n\to\infty} f_n$ satisfies*

$$\int f \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

In particular, if $\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu < \infty$, then f is finite almost everywhere.

Theorem A.17 (Fatou's Lemma). *Let $(f_n)_{n\geqslant 1}$ be a sequence of measurable real-valued functions on a probability space, all bounded below by some integrable function. If $\liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu < \infty$ then $\liminf_{n\to\infty} f_n$ is integrable, and*

$$\int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

Theorem A.18 (Dominated Convergence Theorem). *If $h : X \to \mathbb{R}$ is an integrable function and $(f_n)_{n \geqslant 1}$ is a sequence of measurable real-valued functions which are dominated by h in the sense that $|f_n| \leqslant h$ for all $n \geqslant 1$, and $\lim_{n\to\infty} f_n = f$ exists almost everywhere, then f is integrable and*

$$\int f \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

A.6 Well-behaved Measure Spaces

It is convenient to slightly extend the notion of a Borel probability space as follows (cf. Definition 5.13).

Definition A.19. *Let X be a dense Borel subset of a compact metric space $\overline{X}$, with a probability measure μ defined on the restriction of the Borel σ-algebra $\mathscr{B}$ to X. The resulting probability space $(X, \mathscr{B}, \mu)$ is a* Borel probability space*.

For our purposes, this is the most convenient notion of a measure space that is on the one hand sufficiently general for the applications needed, while on the other has enough structure to permit explicit and convenient proofs.

A circle of results called Lusin's theorem [237] (or Luzin's theorem) show that measurable functions are continuous off a small set. These results are true in almost any context where continuity makes sense, but we state a form of the result here in the setting needed.

Theorem A.20 (Lusin). *Let $(X, \mathscr{B}, \mu)$ be a Borel probability space and let $f : X \to \mathbb{R}$ be a measurable function. Then, for any $\varepsilon > 0$, there is a continuous function $g : X \to \mathbb{R}$ with the property that $\mu(\{x \in X \mid f(x) \neq g(x)\}) < \varepsilon$.*

As mentioned in the endnote to Definition 5.13, there is a slightly different formulation of the standard setting for ergodic theory, in terms of Lebesgue spaces.

Definition A.21. *A probability space is a* Lebesgue space *if it is isomorphic as a measure space to*

$$\left([0, s] \sqcup A, \mathscr{B}, m_{[0,s]} + \sum_{a\in A} p_a \delta_a\right)$$

for some countable set A of atoms *and numbers $s, p_a > 0$ with*

* Commonly the σ-algebra $\mathscr{B}$ is enlarged to its *completion* $\mathscr{B}_\mu$, which is the smallest σ-algebra containing both $\mathscr{B}$ and all subsets of null sets with respect to μ. It is also standard to allow any probability space that is isomorphic to $(X, \mathscr{B}_\mu, \mu)$ in Definition A.19 as a measure space to be called a Lebesgue space.

$$s + \sum_{a \in A} p_a = 1,$$

where $\mathscr{B}$ comprises unions of Lebesgue measurable sets in $[0, s]$ and arbitrary subsets of A, $m_{[0,s]}$ is the Lebesgue measure on $[0, s]$, and δ_a is the Dirac measure defined by $\delta_a(B) = \chi_B(a)$.

The next result shows, *inter alia*, that this notion agrees with that used in Definition A.19 (a proof of this may be found in the book of Parthasarathy [281, Chap. V]) up to completion of the measure space (a measure space is complete if all subsets of a null set are measurable and null). We will not use this result here.

Theorem A.22. *A probability space is a Lebesgue space in the sense of Definition* A.21 *if and only if it is isomorphic to $(X, \mathscr{B}_\mu, \mu)$ for some probability measure μ on the completion $\mathscr{B}_\mu$ of the Borel σ-algebra $\mathscr{B}$ of a complete separable metric space X.*

The function spaces from Section A.3 are particularly well-behaved for Lebesgue spaces.

Theorem A.23 (Riesz–Fischer[(106)]). *Let $(X, \mathscr{B}, \mu)$ be a Lebesgue space. For any p, $1 \leqslant p < \infty$, the space L^p_μ is a separable Banach space with respect to the $\|\cdot\|_p$-norm. In particular, L^2_μ is a separable Hilbert space.*

A.7 Lebesgue Density Theorem

The space $\mathbb{R}$ together with the usual metric and Lebesgue measure $m_{\mathbb{R}}$ is a particularly important and well-behaved special case, and here it is possible to say that a set of positive measure is thick in a precise sense.

Theorem A.24 (Lebesgue[(107)]). *If $A \subseteq \mathbb{R}$ is a measurable set, then*

$$\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} m_{\mathbb{R}} \left(A \cap (a - \varepsilon, a + \varepsilon) \right) = 1$$

for $m_{\mathbb{R}}$-almost every $a \in A$.

A point a with this property is said to be a *Lebesgue density point* or a *point with Lebesgue density* 1. An equivalent and more familiar formulation of the result is a form of the fundamental theorem of calculus.

Theorem A.25. *If $f : \mathbb{R} \to \mathbb{R}$ is an integrable function then*

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_s^{s+\varepsilon} f(t) \, \mathrm{d}t = f(s)$$

for $m_{\mathbb{R}}$-almost every $s \in [0, \infty)$.

The equivalence of Theorem A.24 and A.25 may be seen by approximating an integrable function with simple functions.

A.8 Substitution Rule

Let $O \subseteq \mathbb{R}^n$ be an open set, and let $\phi : O \to \mathbb{R}^n$ be a C^1-map with Jacobian $J_\phi = |\det \mathrm{D}\,\phi|$. Then for any measurable function $f \geqslant 0$ (or for any integrable function f) defined on $\phi(O) \subseteq \mathbb{R}^n$ we have[(108)].

$$\int_O f(\phi(\mathbf{x}))J_\phi(\mathbf{x})\,\mathrm{d}m_{\mathbb{R}^n}(\mathbf{x}) = \int_{\phi(O)} f(\mathbf{y})\,\mathrm{d}m_{\mathbb{R}^n}(\mathbf{y}). \tag{A.3}$$

We recall the definition of the push-forward of a measure. Let $(X, \mathscr{B}_X)$ and $(Y, \mathscr{B}_Y)$ be two spaces equipped with σ-algebras. Let μ be a measure on X defined on $\mathscr{B}_X$, and let $\phi : X \to Y$ be measurable. Then the push-forward $\phi_*\mu$ is the measure on $(Y, \mathscr{B}_Y)$ defined by $(\phi_*\mu)(B) = \mu(\phi^{-1}(B))$ for all $B \in \mathscr{B}_Y$.

The substitution rule allows us to calculate the push-forward of the Lebesgue measure under smooth maps as follows.

Lemma A.26. *Let $O \subseteq \mathbb{R}^n$ be open, let $\phi : O \to \mathbb{R}^n$ be a smooth injective map with non-vanishing Jacobian $J_\phi = |\det \mathrm{D}\,\phi|$. Then the push-forward $\phi_* m_O$ of the Lebesgue measure $m_O = m_{\mathbb{R}^n}|_O$ restricted to O is absolutely continuous with respect to $m_{\mathbb{R}^n}$ and is given by*

$$\mathrm{d}\phi_* m_O = J_\phi^{-1} \circ \phi^{-1}\,\mathrm{d}m_{\phi(O)}.$$

Moreover, if we consider a measure $\mathrm{d}\mu = F\,\mathrm{d}m_O$ absolutely continuous with respect to m_O, then similarly

$$\mathrm{d}\phi_*\mu = F \circ \phi^{-1} J_\phi^{-1} \circ \phi^{-1}\,\mathrm{d}m_{\phi(O)}.$$

Proof. Recall that under the assumptions of the lemma, ϕ^{-1} is smooth and $J_{\phi^{-1}} = J_\phi^{-1} \circ \phi^{-1}$. Therefore, by equation (A.3) and the definition of the push-forward,

$$\begin{aligned}
\int_{\phi(O)} f(x)J_\phi^{-1}\left(\phi^{-1}(x)\right)\,\mathrm{d}m_{\mathbb{R}^n}(x) &= \int_{\phi(O)} f\left(\phi(\phi^{-1}(x))\right) J_{\phi^{-1}}(x)\,\mathrm{d}m_{\mathbb{R}^n}(x) \\
&= \int_O f(\phi(y))\,\mathrm{d}m_{\mathbb{R}^n}(y) \\
&= \int_{\phi(O)} f(x)\,\mathrm{d}\phi_* m_O(x)
\end{aligned}$$

for any characteristic function $f = \chi_B$ of a measurable set $B \subseteq \phi(O)$. This implies the first claim. Moreover, for any measurable functions $f \geqslant 0, F \geqslant 0$ defined on $\phi(O), O$ respectively,

$$\int_{\phi(O)} f(x)F(\phi^{-1}(x))J_\phi^{-1}(\phi^{-1}(x))\,\mathrm{d}m_{\mathbb{R}^n}(x) = \int_O f(\phi(y))F(y)\,\mathrm{d}m_{\mathbb{R}^n},$$

which implies the second claim. $\square$

Notes to Appendix A

(102)(Page 397) This result was stated by Borel [40, p. 252] for independent events as part of his study of normal numbers, but as pointed out by Barone and Novikoff [18] there are some problems with the proofs. Cantelli [46] noticed that half of the theorem holds without independence; this had also been noted by Hausdorff [142] in a special case. Erdős and Rényi [84] showed that the result holds under the much weaker assumption of pairwise independence.

(103)(Page 398) This is shown, for example, in Parthasarathy [281, Th. 1.2]: defining a Borel set A to be regular if, for any $\varepsilon > 0$, there is an open set O_ε and a closed set C_ε with $C_\varepsilon \subseteq A \subseteq O_\varepsilon$ and $\mu(O_\varepsilon \smallsetminus C_\varepsilon) < \varepsilon$, it may be shown that the collection of all regular sets forms a σ-algebra and contains the closed sets.

(104)(Page 401) A form of this theorem goes back to Cauchy for continuous functions on the reals, and this was extended by Lebesgue [220] to bounded measurable functions. Fubini [97] extended this to integrable functions, showing that if $f : [a,b] \times [c,d] \to \mathbb{R}$ is integrable then $y \mapsto f(x,y)$ is integrable for almost every x, and proving equation (A.2). Tonelli [364] gave the formulation here, for non-negative functions on products of σ-finite spaces. Complete proofs may be found in Royden [321] or Lieb and Loss [229, Th. 1.12]. While the result is robust and of central importance, some hypotheses are needed: if the function is not integrable or the spaces are not σ-finite, the integrals may have different values. A detailed treatment of the minimal hypotheses needed for a theorem of Fubini type, along with counterexamples and applications, is given by Fremlin [96, Sect. 252].

(105)(Page 402) This result is due to Radon [298] when μ is Lebesgue measure on $\mathbb{R}^n$, and to Nikodym [273] in the general case.

(106)(Page 404) This result emerged in several notes of Riesz and two notes of Fischer [91], [92], with a full treatment of the result that $L^2(\mathbb{R})$ is complete appearing in a paper of Riesz [312].

(107)(Page 404) This is due to Lebesgue [220], and a convenient source for the proof is the monograph of Oxtoby [277]. Notice that Theorem A.24 expresses how constrained measurable sets are: it is impossible, for example, to find a measurable subset A of $[0,1]$ with the property that $m_{\mathbb{R}}(A \cap [a,b]) = \frac{1}{2}(b-a)$ for all $b > a$. While a measurable subset of measure $\frac{1}{2}$ may have an intricate structure, it cannot occupy only half of the space on all possible scales.

(108)(Page 405) The usual hypotheses are that the map ϕ is injective and the Jacobian non-vanishing; these may be relaxed considerably, and the theorem holds in very general settings both measurable (see Hewitt and Stromberg [152]) and smooth (see Spivak [351]).