I.3 The Lebesgue integral

We have just seen that $C[a, b]$ has two quite reasonable metrics on it. In Section I.5 we will see that it is a complete metric space in the metric

$$d_1(f, g) = \sup_{x \in [a, b]} |f(x) - g(x)|$$

In the other metric we considered, $d_2(f, g) = \|f - g\|_1$ with $\|h\|_1 = \int_a^b |h(x)|\, dx$, $C[a, b]$ is *not complete*. To see this for $C[0, 1]$, let f_n be given as in Figure I.3. It is not hard to see that f_n is Cauchy in $\|\cdot\|_1$, but it does not converge to any function in $C[a, b]$; rather, in an *intuitive sense*, it "converges" to the characteristic function of $[\frac{3}{4}, \frac{1}{4}]$ (which is, of course, not in $C[0, 1]$!).

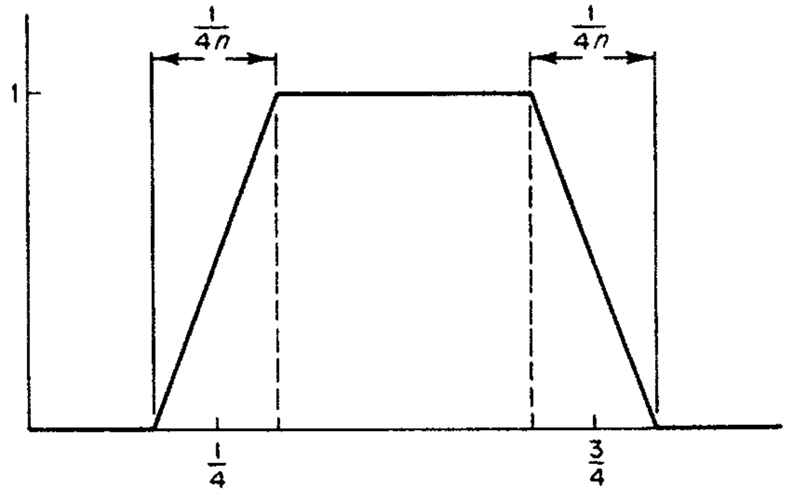

FIGURE I.3 The graph of f_n.

We can always complete $C[a, b]$ in $\|\cdot\|_1$ realizing elements of the completion as equivalence classes of Cauchy sequences of continuous functions; this realization is not noteworthy for its transparency. The example above suggests we might also be able to realize elements of the completion as functions. If we do realize them as functions, we should be able to define the integral $\int_a^b |f(x)|\, dx$ (merely as $d_2(f, 0)$!) for any f in the completion.

The simplest way to realize elements of the completion as functions is to turn the above analysis around: one introduces an extended notion of integral on a bigger space than $C[a, b]$; call it $L^1[a, b]$. We will prove L^1 is complete, so by general arguments the closure of C in L^1 is complete (and it turns out $\bar{C} = L^1$).

Now, how can one extend the notion of Riemann integral? The usual definition of the Riemann integral is based on dividing the *domain* of f into finer and finer pieces. For "nasty" functions, this method does not work and

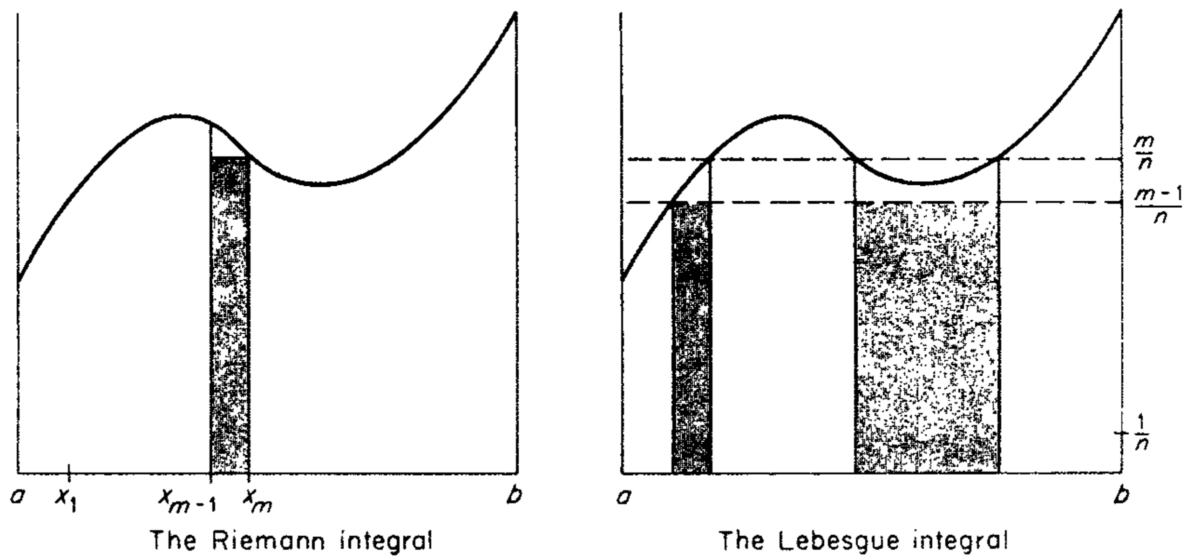

FIGURE I.4

so a different method is needed—the simplest modification is to divide the range into finer and finer pieces (Figure I.4). This method depends more on the function and so has the possibility of working for more types of functions. We are thus interested in sets $f^{-1}[a, b]$ and their size. We suppose we have a size function μ on sets which generalizes $\mu([a, b]) = b - a$. We will shortly return to this size function and see that not all sets have a "size." We will then restrict the types of f by demanding that $f^{-1}[a, b]$ have a "size." Looking at Figure I.4, we define for $f \geq 0$

$$\sum_n(f) = \sum_{m=0}^{\infty} \frac{m}{n} \mu\left(f^{-1}\left[\left[\frac{m}{n}, \frac{m+1}{n}\right)\right]\right)$$

Then $\sum_{2n}(f) \geq \sum_n(f)$ so that $\lim_{n\to\infty} \sum_{2^n}(f) = \sup_n (\sum_{2^n}(f))$ exists (it may be ∞). This limit is defined to be $\int f\,dx$. We remark that for technical purposes (that is, proving theorems!) one makes a different definition which can be shown to agree with this definition only after a lot of work. The definition as $\lim \sum_{2^n}(f)$ is however the best to keep in mind when thinking intuitively.

Thus, we have transferred the problem to one of defining an extended notion of size. We must first decide what sets are to have a size. Why not all sets? There is a classical example (see also Problem 13) which shows that not all sets in $\mathbb{R}^3$ can have a size if we want that size to be invariant under rotations and translations (and not to be trivial, such as assigning zero to all sets): it is possible to break up a unit ball into a finite number of wild pieces, move the pieces around by rotation and translation and reassemble the pieces to get two balls of radius one (Banach–Tarski paradox). Thus, all sets cannot have a size, and so some family $\mathscr{B}$ of sets will be the "measurable sets." What properties do we want $\mathscr{B}$ to have? We would like both $f^{-1}[[0, a)]$ and $f^{-1}[[a, \infty)]$ to be measurable ($f \geq 0$) so we would like $\mathscr{B}$ to have the property: $A \in \mathscr{B}$ implies $\mathbb{R}\backslash A \in \mathscr{B}$. Also, when f is continuous, we want $f^{-1}[(a, b)]$ to be in $\mathscr{B}$, so $\mathscr{B}$ should contain the open sets. Finally, we want to have

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

if the A_n are mutually disjoint (to meet our intuitive notion of size) so we would like $\bigcup_{n=1}^{\infty} A_n \in \mathscr{B}$ if each A_n is in $\mathscr{B}$.

Definition The **Borel sets** of $\mathbb{R}$ is the smallest family of subsets of $\mathbb{R}$ with the following properties:

(i) The family is closed under complements.
(ii) The family is closed under countable unions.
(iii) The family contains each open interval.

To see that such a *smallest* family exists we note that if $\{\mathscr{B}_\alpha\}_{\alpha \in A}$ is a collection of families obeying (i), (ii), and (iii), then so does $\bigcap_{\alpha \in A} \mathscr{B}_\alpha$. Thus the intersection of all families obeying (i)–(iii) is the smallest such family.

Now we define the Lebesgue measures of sets in $\mathscr{B}$, the Borel sets in $\mathbb{R}$.

Definition Let $\mathscr{I}$ be the family of all countable unions of disjoint open intervals (which is just the family of open sets) and let

$$\mu\left(\bigcup_{i=1}^{\infty}(a_i, b_i)\right) \equiv \sum_{i=1}^{\infty}(b_i - a_i)$$

(which may be infinite). For any $B \in \mathscr{B}$, define

$$\mu(B) = \inf_{\substack{I \in \mathscr{I} \\ B \subset I}} \mu(I)$$

This notion of size has four crucial properties:

Theorem I.8

(a) $\mu(\emptyset) = 0$

(b) If $\{A_n\}_{n=1}^{\infty} \subset \mathscr{B}$ and the A_n are mutually disjoint ($A_n \cap A_m = \emptyset$, all $m \neq n$), then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

(c) $\mu(B) = \inf\{\mu(I) \mid B \subset I, I \text{ is open}\}$

(d) $\mu(B) = \sup\{\mu(C) \mid C \subset B, C \text{ is compact}\}$

The infinite sum in (b) contains only positive terms, so it either converges to a finite number or diverges to infinity, in which case we set it equal to ∞. (c) and (d) say that any Borel set can be approximated "from the outside" by open sets and from the inside by compact sets. We remind the reader that *on the real line* a set is compact if and only if it is closed and bounded.

We have thus extended the usual notion of size of intervals and we define the family of functions we will consider in the obvious way:

Definition A function f is called a **Borel function** if and only if $f^{-1}[(a, b)]$ is a Borel set for all a, b.

It is often convenient to allow our functions to take the values $\pm\infty$ on small sets in which case we require $f^{-1}[\{\pm\infty\}]$ to be Borel.

Proposition f is a Borel if and only if, for all $B \in \mathscr{B}$, $f^{-1}[B] \in \mathscr{B}$ (see Problem 14).

This last proposition implies that the composition of two Borel functions is Borel. Many books deal with a slightly larger class of functions than the Borel class. They first define a set M to be measurable if one can write $M \cup A_1 = B \cup A_2$ where B is Borel and $A_i \subset B_i$ with B_i Borel and $\mu(B_i) = 0$ (thus they add and subtract "unimportant" sets from Borel sets). A measurable function is then defined as a function, f, for which $f^{-1}[(a, b)]$ is always measurable. It is no longer true that $f \circ g$ is measurable if f and g are, and many technical problems arise. *In any event, we deal only with Borel sets and functions and use the words Borel and measurable interchangeably.*

Borel functions are closed under many operations:

Proposition (a) If f, g are Borel, then so are $f + g$, fg, $\max\{f, g\}$ and $\min\{f, g\}$. If f is Borel and $\lambda \in \mathbb{R}$, λf is Borel.

(b) If each f_n is Borel, $n = 1, 2, \ldots$, and $f_n(p) \to f(p)$ for all p, then f is Borel.

Since $|f| = \max\{f, -f\}$, $|f|$ is measurable if f is.

As we sketched above, given $f \geq 0$, one can define $\int f\,dx$ (which may be ∞). If $\int |f|\,dx < \infty$, we write $f \in \mathscr{L}^1$ and define $\int f\,dx = \int f_+\,dx - \int f_-\,dx$ where $f_+ = \max\{f, 0\}$; $f_- = \max\{-f, 0\}$. $\mathscr{L}^1(a, b)$ is the set of functions on (a, b) which are in $\mathscr{L}^1$ if we extend them to the whole real line by defining them to be zero outside of (a, b). If $f \in \mathscr{L}^1(a, b)$, we write $\int f\,dx = \int_a^b f\,dx$. We then have:

Theorem I.9 Let f and g be measurable functions. Then

(a) If $f, g \in \mathscr{L}^1(a, b)$, so are $f + g$ and λf, for all $\lambda \in \mathbb{R}$.
(b) If $|g| \leq f$ and $f \in \mathscr{L}^1$, then $g \in \mathscr{L}^1$.
(c) $\int (f + g)\,dx = \int f\,dx + \int g\,dx$ if f and g are in $\mathscr{L}^1$.
(d) $|\int f\,dx| \leq \int |f|\,dx$ if f is in $\mathscr{L}^1$.
(e) If $f \leq g$, then $\int f\,dx \leq \int g\,dx$, if f and g are in $\mathscr{L}^1$.
(f) If f is bounded and measurable on $-\infty < a < b < \infty$, then $f \in \mathscr{L}^1$ and

$$\left|\int_a^b f\,dx\right| \leq |b - a| \left(\sup_{a \leq x \leq b} |f(x)| \right)$$

This theorem shows that $\int$ has all the nice properties of the Riemann integral even though it is defined for a larger class of functions.

The properties that make the space L^1 (which we will shortly define) complete are the following absolutely essential convergence theorems:

Theorem I.10 (monotone convergence theorem) Let $f_n \geq 0$ be measurable. Suppose $f_n(p) \to f(p)$ for each p and that $f_{n+1}(p) \geq f_n(p)$ all p and n (in which case we write $f_n \nearrow f$). If $\int f_n(p) \ dp < C$ for all n, then $f \in \mathscr{L}^1$ and $\int |f(p) - f_n(p)| \, dp \to 0$ as $n \to \infty$.

Theorem I.11 (dominated convergence theorem) Let $f_n(p) \to f(p)$ for each p and suppose $|f_n(p)| \leq G(p)$ for all n and some $G \in \mathscr{L}^1$. Then $f \in \mathscr{L}^1$ and $\int |f(p) - f_n(p)| \ dp \to 0$ as $n \to \infty$.

In the latter case, we say G dominates the pointwise convergence. That a dominating function exists is crucial. For example, let $f_n(x) = (1/n)\chi_{[-n,n]}(x)$. Then $f_n(x) \to 0$ for each x, but $\int |f_n| \ dx = 2$ so $\int |f_n(x)| \ dx$ does not go to zero. In this case, it is not hard to see that $\sup_n |f_n(x)| = G(x)$ is not in $\mathscr{L}^1$.

We are almost ready to define $\mathscr{L}^1$ as a metric space by letting $\rho(f, g) = \int |f - g| \ dx$. We cannot quite do this because $\int |f - g| \ dx = 0$ does not imply $f \equiv g$ (for example, f and g might differ at a single point). Thus, we first define the notion of almost everywhere (a.e.):

Definition We say a condition $C(x)$ holds almost everywhere (a.e.) if $\{x \mid C(x) \text{ is false}\}$ is a subset of a set of measure zero.

Definition We say two functions $f, g \in \mathscr{L}^1$ are equivalent if $f(x) = g(x)$ a.e. (this is the same as saying $\int |f - g| \ dx = 0$).

Definition The set of equivalence classes in $\mathscr{L}^1$ is denoted by as L^1. L^1 with the norm $\|f\|_1 = \int |f| \ dx$ is a normed linear space.

Thus an element of L^1 is an equivalence class of functions equal a.e. In particular when $f \in L^1$, the symbol $f(x)$ for a particular x does not make sense. Nevertheless we continue to write "$f(x)$" but only in situations where statements are independent of a choice from the equivalence class. Thus, for example, $f_n(x) \to f(x)$ *for almost all* x is independent of the representatives chosen for f and f_n. By this replacement of pointwise convergence with pointwise convergence almost everywhere, the two convergence theorems carry over from $\mathscr{L}^1$ to L^1.

Having cautioned the reader that $f(x)$ is "technically meaningless" for $f \in L^1$, we remark that in *certain special cases* it is meaningful. Suppose $f \in L^1$

has a representative $\tilde{f}$ (that is, $\tilde{f}$ is a function; f an equivalence class of functions) which is continuous. Then no other representative of f is continuous, so it is natural to write $f(x)$ for $\tilde{f}(x)$.

The critical fact about L^1 is:

2.7 THEOREM (Completeness of L^p-spaces)

Let $1 \le p \le \infty$ and let f^i, for $i = 1, 2, 3, \dots$, be a **Cauchy sequence** *in $L^p(\Omega)$, i.e., $\|f^i - f^j\|_p \to 0$ as $i, j \to \infty$. (This means that for each $\varepsilon > 0$ there is an N such that $\|f^i - f^j\|_p < \varepsilon$ when $i > N$ and $j > N$.) Then there exists a unique function $f \in L^p(\Omega)$ such that $\|f^i - f\|_p \to 0$ as $i \to \infty$. We denote this latter fact by*

$$f^i \to f \quad \text{as} \quad i \to \infty,$$

and we say that f^i **converges strongly to** f **in** $L^p(\Omega)$.

Moreover, there exists a subsequence $f^{i_1}, f^{i_2}, \dots$ (with $i_1 < i_2 < \cdots$, of course) and a nonnegative function F in $L^p(\Omega)$ such that

(i) *Domination :* $|f^{i_k}(x)| \le F(x)$ *for all k and μ-almost every x.* (1)

(ii) *Pointwise convergence:* $\lim_{k\to\infty} f^{i_k}(x) = f(x)$ *for μ-almost every x.* (2)

REMARK. 'Convergence' and 'strong convergence' are used interchangeably. The phrase norm convergence is also used.

PROOF. The first, and most important remark, concerns a strategy that is frequently very useful. Namely, it suffices to show the strong convergence for *some* subsequence. To prove this sufficiency, let f^{i_k} be a subsequence that converges strongly to f in $L^p(\Omega)$ as $k \to \infty$. Since, by the triangle inequality,

$$\|f^i - f\|_p \le \|f^i - f^{i_k}\|_p + \|f^{i_k} - f\|_p,$$

we see that for any $\varepsilon > 0$ we can make the last term on the right side less than $\varepsilon/2$ by choosing k large. The first term on the right can be made smaller than $\varepsilon/2$ by choosing i and k large enough, since f^i is a Cauchy sequence. Thus, $\|f^i - f\|_p < \varepsilon$ for i large enough and we can conclude convergence for the *whole sequence,* i.e., $f^i \to f$. This also proves, incidentally, that the limit—if it exists—is unique.

To obtain such a convergent subsequence pick a number i_1 such that $\|f^{i_1} - f^n\|_p \le 1/2$ for all $n \ge i_1$. That this is possible is precisely the definition of a Cauchy sequence. Now choose i_2 such that $\|f^{i_2} - f^n\|_p < 1/4$ for all $n \ge i_2$ and so on. Thus we have obtained a subsequence of the integers, i_k, with the property that $\|f^{i_k} - f^{i_{k+1}}\|_p \le 2^{-k}$ for $k = 1, 2, \dots$. Consider the monotone sequence of positive functions

$$F_l(x) := |f^{i_1}(x)| + \sum_{k=1}^{l} |f^{i_k}(x) - f^{i_{k+1}}(x)|. \tag{3}$$

By the triangle inequality

$$\|F_l\|_p \leq \|f^{i_1}\|_p + \sum_{k=1}^{l} 2^{-k} \leq \|f^{i_1}\|_p + 1.$$

Thus, by the monotone convergence theorem, F_l converges pointwise μ-a.e. to a positive function F which is in $L^p(\Omega)$ and hence is finite almost everywhere. The sequence

$$f^{i_{k+1}}(x) = f^{i_1}(x) + (f^{i_2}(x) - f^{i_1}(x)) + \cdots + (f^{i_{k+1}}(x) - f^{i_k}(x)) \qquad (4)$$

thus converges absolutely for almost every x, and hence it also converges for the same x's to some number $f(x)$. Since $|f^{i_k}(x)| \leq F(x)$ and $F \in L^p(\Omega)$, we know by dominated convergence that f is in $L^p(\Omega)$. Again by dominated convergence $\|f^{i_k} - f\|_p \to 0$ as $k \to \infty$ since $|f^{i_k}(x) - f(x)| \leq F(x) + |f(x)| \in L^p(\Omega)$. Thus, the subsequence f^{i_k} converges strongly in $L^p(\Omega)$ to f. ■

As a final result which brings us full circle to our original motivation:

Proposition $C[a, b]$ is dense (in $\|\cdot\|_1$) in $L^1[a, b]$, i.e. L^1 is the completion of C.

Proof See Problem 18.

We defined $L^1[a, b]$ as a space of real-valued functions. It is often convenient to deal with complex-valued functions, f, whose real and imaginary parts are in $L^1[a, b]$. When no confusion arises, we will denote this space, with the norm

$$\|f\|_1 = \int_a^b |f| \, dx$$

also by $L^1[a, b]$. The integral of a complex-valued function is defined by

$$\int f \, dx = \int \text{Re}(f) \, dx + i \int \text{Im}(f) \, dx$$

I.4 Abstract measure theory

One of the most important tools which one combines with abstract functional analysis in the study of various concrete models is "general" measure theory, that is, the theory of the last section extended to a more abstract setting.

The simplest way to generalize the Lebesgue integral is to work with functions on the real line and with Borel sets but to generalize the underlying measure; we consider this special case of abstract measure theory first. Recall that the Lebesgue integral was constructed as follows. We started with a notion of size for intervals, $\mu([a, b]) = b - a$, and extended this in a unique way to a notion of size for arbitrary Borel sets. Armed with this notion of size for Borel sets, the integral of Borel functions was obtained by measuring sets of the form $f^{-1}([a, b])$. We found the vector space $L^1([0, 1], dx)$ constructed in the last section is just the completion of $C[0, 1]$ with the metric $d_2(f, g) = \int_0^1 |f(x) - g(x)| \, dx$, where we needed only the Riemann integral to define d_2.

Now suppose an arbitrary monotone function $\alpha(x)$ is given (that is, $x > y$ implies $\alpha(x) \geq \alpha(y)$). It is not hard to see that the limit from the right, $\lim_{\varepsilon \to 0} \alpha(x + |\varepsilon|)$ and the limit from the left, $\lim_{\varepsilon \to 0} \alpha(x - |\varepsilon|)$ exist; we write them as $\alpha(x + 0)$ and $\alpha(x - 0)$ respectively. Since (a, b) does not include the points a and b, it is natural to define $\mu_\alpha((a, b)) = \alpha(b - 0) - \alpha(a + 0)$. From this notion of size for intervals, one can construct a measure μ_α on Borel sets of $\mathbb{R}$, that is, a map $\mu_\alpha \colon \mathscr{B} \to [0, \infty]$ with $\mu_\alpha(\bigcup B_i) = \sum_{i=1}^\infty \mu_\alpha(B_i)$ if $B_i \cap B_j = \emptyset$ and $\mu_\alpha(\emptyset) = 0$. By construction, this measure has the regularity property

$$\begin{aligned} \mu_\alpha(B) &= \sup\{\mu(C) \mid C \subset B, \quad C \text{ compact}\} \\ &= \inf\{\mu(O) \mid B \subset O, \quad O \text{ open}\} \end{aligned}$$

Also, $\mu(C) < \infty$ for any compact set C. A measure with these two regularity properties is called a **Borel measure**. In particular, $\mu_\alpha([a, b]) = \alpha(b + 0) - \alpha(a - 0)$. One can then construct an integral $f \to \int f\, d\mu_\alpha$ (we will also write $\int f\, d\alpha$) which has properties (a)–(e) of Theorem I.9; it is called a Lebesgue–Stieltjes integral. $L^1([a, b], d\alpha)$ and $L^1(\mathbb{R}, d\alpha)$ can be formed as before. These spaces of equivalence classes of functions are complete in the metric $\rho(f, g) = \int |f - g|\, d\alpha$, and analogues of the monotone and dominated convergence theorems hold. The continuous functions $C[a, b]$ form a dense subspace of $L^1([a, b], d\alpha)$; put differently, $L^1([a, b], d\alpha)$ is the completion of $C[a, b]$ with the metric $\rho_\alpha(f, g) = \int_a^b |f - g|\, d\alpha$ where we need only use the Riemann–Stieltjes integral to define ρ_α (see Problem 11).

Let us consider three examples which illustrate the variety of Lebesgue–Stieltjes measures.

Example 1 Suppose α is continuously differentiable. Then $\mu_\alpha(a, b) = \int_a^b (d\alpha/dx)\, dx$ where dx is Lebesgue measure, so it is to be expected (and is indeed true!) that

$$\int f\, d\alpha = \int f\left(\frac{d\alpha}{dx}\right) dx$$

Thus, these measures can essentially be described in terms of Lebesgue measure.

Example 2 Suppose that $\alpha(x)$ is the characteristic function of $[0, \infty)$. Then $\mu_\alpha(a, b) = 1$ if $0 \in (a, b)$ and is 0 if $0 \notin (a, b)$. The measure one gets out is very easy to describe: $\mu_\alpha(B) = 1$ if $0 \in B$, and $\mu_\alpha(B) = 0$ if $0 \notin B$. The reader is invited to construct explicitly the integral and convince himself that

$$\int f\, d\alpha = f(0)$$

This measure $d\alpha$ is known as the Dirac measure (since it is just like a δ function). Let us consider $L^1(\mathbb{R}, d\alpha)$ in this case. In $\mathscr{L}^1$ we have $\rho(f, g) = |f(0) - g(0)|$ so $\rho(f, g) = 0$ if and only if $f(0) = g(0)$. As a result, we see that the equivalence classes in L^1 are completely described by the value $f(0)$ so that $L^1(\mathbb{R}, d\alpha)$ is just a one-dimensional vector space! Notice how different this is from the case of $L^1(\mathbb{R}, dx)$ where the value of a "function" at a single point is not defined (since elements of L^1 are equivalence classes).

Example 3 Our last example makes use of a fairly pathological function, $\alpha(x)$, which we first construct. Let S be the subset of $[0, 1]$

$$S = (\tfrac{1}{3}, \tfrac{2}{3}) \cup (\tfrac{1}{9}, \tfrac{2}{9}) \cup (\tfrac{7}{9}, \tfrac{8}{9}) \cup (\tfrac{1}{27}, \tfrac{2}{27}) \cup \cdots$$

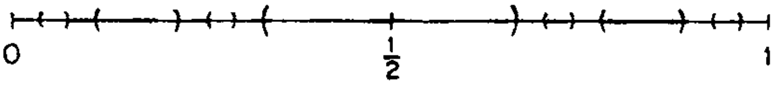

FIGURE I.5 The Cantor set.

that is, remove the middle third of what is not in S at each stage and add it to S, see Figure I.5. The Lebesgue measure of S is $\frac{1}{3} + 2(\frac{1}{9}) + 4(\frac{1}{27}) + \cdots = 1$. Let $C = [0, 1]\backslash S$. It has Lebesgue measure 0. C, which is known as the Cantor set, is easy to describe if we write each $x \in [0, 1]$ in its base three decimal expansion. Then $x \in C$ if and only if this base 3 expansion has no 1's. Thus C is an uncountable set of measure 0. To see this, map C in a one-one way *onto* $[0, 1]$ by changing 2's into 1's and viewing the end result as a base 2 number. Now construct $\alpha(x)$ as follows: set $\alpha(x) = \frac{1}{2}$ on $(\frac{1}{3}, \frac{2}{3})$; $\alpha(x) = \frac{1}{4}$ on $(\frac{1}{9}, \frac{2}{9})$; $\alpha(x) = \frac{3}{4}$ on $(\frac{7}{9}, \frac{8}{9})$, etc.; see Figure I.6. Extend α to $[0, 1]$ by making it con-

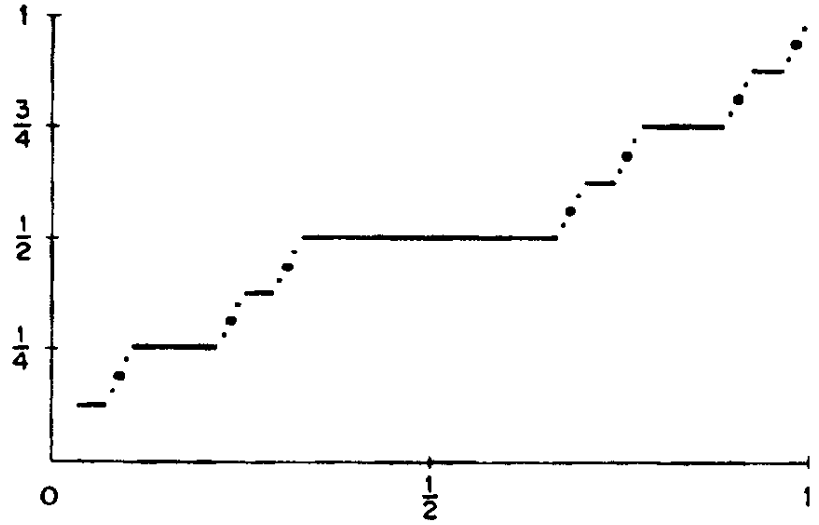

FIGURE I.6 The Cantor function.

tinuous. Then α is a nonconstant continuous function with the strange property that $\alpha'(x)$ exists a.e. (with respect to Lebesgue measure) and is zero a.e. Now, we can form the measure μ_α. Since α is continuous, $\mu_\alpha(\{p\}) = 0$ for any set $\{p\}$ with only one point. Nevertheless, μ_α is concentrated on the set C in the sense that $\mu_\alpha([0, 1]\backslash C) = \mu_\alpha(S) = 0$. On the other hand, the Lebesgue measure of C is zero. Thus μ_α and Lebesgue measure "live" on completely different sets.

In a sense we now make precise, these three examples are models of the most general Lebesgue–Stieltjes measures. Suppose μ is a Borel measure on $\mathbb{R}$.

First, let $P = \{x \mid \mu(\{x\}) \neq 0\}$, that is, P is the set of **pure points** of μ. Since μ is Borel [$\mu(C) < \infty$ for any compact set], P is a countable set. Define

$$\mu_{\rm pp}(X) = \sum_{x \in P \cap X} \mu(\{x\}) = \mu(P \cap X)$$

Then $\mu_{\rm pp}$ is a measure and $\mu_{\rm cont} = \mu - \mu_{\rm pp}$ is positive. $\mu_{\rm cont}$ has the property $\mu_{\rm cont}(\{p\}) = 0$ for all p, that is, it has no pure points and $\mu_{\rm pp}$ has only pure points in the sense that $\mu_{\rm pp}(X) = \sum_{x \in X} \mu_{\rm pp}(\{x\})$.

Definition A Borel measure μ on $\mathbb{R}$ is called **continuous** if it has no pure points. μ is called a **pure point measure** if $\mu(X) = \sum_{x \in X} \mu(x)$ for any Borel set X.

Thus, we have seen:

Theorem I.13 Any Borel measure can be decomposed uniquely into a sum $\mu = \mu_{\rm pp} + \mu_{\rm cont}$ where $\mu_{\rm cont}$ is continuous and $\mu_{\rm pp}$ is a pure point measure.

We have thus generalized Example 2 by allowing sums of Dirac measures. Is there any generalization of Examples 1 and 3?

Definition We say that μ is **absolutely coutinuous with respect to (w.r.t.) Lebesgue measure** if there is a function, f, locally L^1 (that is, $\int_a^b |f(x)|\, dx < \infty$ for any finite interval (a, b)) so that

$$\int g\, d\mu = \int gf\, dx$$

for any Borel function g in $L^1(\mathbb{R}, d\mu)$. We then write $d\mu = f\, dx$.

This definition generalizes Example 1; we will eventually make a different (but equivalent!) definition of absolute continuity.

Definition We say μ is **singular relative to Lebesgue measure** if and only if $\mu(S) = 0$ for some set S where $\mathbb{R} \backslash S$ has Lebesque measure 0.

The fundamental result is:

Theorem I.14 (Lebesgue decomposition theorem) Let μ be a Borel measure. Then $\mu = \mu_{\rm ac} + \mu_{\rm sing}$ in a unique way with $\mu_{\rm ac}$ absolutely continuous w.r.t. Lebesgue measure and with $\mu_{\rm sing}$ singular relative to Lebesgue measure.

Thus Theorems I.13 and I.14 tell us that any measure μ on $\mathbb{R}$ has a canonical decomposition $\mu = \mu_{pp} + \mu_{ac} + \mu_{sing}$ where μ_{pp} is pure point, μ_{ac} absolutely continuous with respect to Lebesgue measure, and μ_{sing} is *continuous and singular* relative to Lebesgue measure. This decomposition will recur in a quantum-mechanical context where any state will be a sum of bound states, scattering states, and states with no physical interpretation (one of our hardest jobs will be to show that this last type of state does not occur; that is, that certain measures have $\mu_{sing} = 0$; (see Chapter XIII).)

This completes our study of measures on $\mathbb{R}$. The next level of generalization involves measures on sets with some underlying topological structure; we will return to study this case of intermediate generality in Section IV.4. The most general setting lets us deal with an arbitrary set. We first need an abstraction of Borel sets:

Definition A nonempty family $\mathscr{R}$ of subsets of a set M is called a **σ-ring** if and only if

(a) $A_i \in \mathscr{R}$, $i = 1, 2, \ldots$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathscr{R}$.
(b) If $A, B \in \mathscr{R}$, then $A \backslash B \in \mathscr{R}$.

If $M \in \mathscr{R}$, we say that $\mathscr{R}$ is a **σ-field.**

The definition of measure is obvious(!):

Definition A **measure** on a set M with σ-ring $\mathscr{R}$ is a map $\mu\colon \mathscr{R} \to [0, \infty]$ with the properties:

(a) $\mu(\emptyset) = 0$

(b) $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$, if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

We shall often speak of the measure space $\langle M, \mu \rangle$ without explicitly mentioning $\mathscr{R}$, but the σ-ring is a crucial element of the definition. Occasionally, we will write $\langle M, \mathscr{R}, \mu \rangle$. For certain pathologically "big" spaces, one wants to use the notion of σ-ring rather than σ-field, but to keep things simple, we will consider measures on σ-fields and will suppose the whole space isn't too big in the sense:

Definition A measure μ on a σ-field $\mathscr{F}$ is called **σ-finite** if and only if $M = \bigcup_{i=1}^{\infty} A_i$ with each $\mu(A_i) < \infty$.

We will suppose all our underlying measures are σ-finite.

Definition Let M, N be sets with σ-fields $\mathscr{R}$ and $\mathscr{F}$. A map $T\colon M \to N$ is called **measurable** (w.r.t. $\mathscr{R}$ and $\mathscr{F}$) if and only if $\forall A \in \mathscr{F}$, $T^{-1}[A] \in \mathscr{R}$. A map $f\colon M \to \mathbb{R}$ is called measurable if it is measurable w.r.t. $\mathscr{R}$ and the Borel sets of $\mathbb{R}$.

Given a measure μ on a measure space M, we can define $\int f\,d\mu$ for any positive real-valued measurable function on M and we can form $\mathscr{L}^1(M, d\mu)$, the set of integrable functions and $L^1(M, d\mu)$, the equivalence classes of functions in $\mathscr{L}^1$ equal a.e.[μ]. As in the case $\langle M, d\mu\rangle = \langle \mathbb{R}, dx\rangle$, the following crucial theorems hold:

Theorem I.15 (monotone convergence theorem) If $f_n \in \mathscr{L}^1(M, d\mu)$, $0 \le f_1(x) \le f_2(x) \le \cdots$ and $f(x) = \lim_{n\to\infty} f_n(x)$, then $f \in \mathscr{L}^1$ if and only if $\lim_{n\to\infty} \|f_n\|_1 < \infty$ and in that case $\lim_{n\to\infty} \|f - f_n\|_1 = 0$ and $\lim_{n\to\infty} \|f_n\|_1 = \|f\|_1$.

Theorem I.16 (dominated convergence theorem) If $f_n \in L^1(M, d\mu)$, $\lim_{n\to\infty} f_n(x) = f(x)$ a.e.[μ], and if there is a $G \in L^1$ with $|f_n(x)| \le G(x)$ a.e.[μ], for all n, then $f \in L^1$ and $\lim_{n\to\infty} \|f - f_n\|_1 = 0$.

Theorem I.17 (Fatou's lemma) If $f_n \in \mathscr{L}^1$, each $f_n(x) \ge 0$ and if $\underline{\lim}\,\|f_n\|_1 < \infty$, then $f(x) = \underline{\lim}\, f_n(x)$ is in $\mathscr{L}^1$ and $\|f\|_1 \le \underline{\lim}\,\|f_n\|_1$.

Note In Fatou's lemma nothing is said about $\lim_{n\to\infty} \|f - f_n\|_1$.

Theorem I.18 (Riesz–Fisher theorem) $L^1(M, d\mu)$ is complete.

One also has the idea of mutually singular:

Definition Let μ, ν be two measures on a space M with σ-field $\mathscr{R}$. We say that μ and ν are mutually singular if there is a set $A \in \mathscr{R}$ with $\mu(A) = 0$, $\nu(M\backslash A) = 0$.

It is useful to take a weaker looking definition of absolute continuity which is essentially the opposite of singular:

Definition We say ν is absolutely continuous w.r.t. μ if and only if $\mu(A) = 0$ implies $\nu(A) = 0$.

That this definition is the same as the previous one is a consequence of:

Theorem I.19 (Radon–Nikodym theorem) ν is absolutely continuous w.r.t. μ if and only if there is a measurable function f so that

$$\nu(A) = \int f(x)\chi_A(x)\, d\mu(x)$$

for any measurable set A. f is uniquely determined a.e. (w.r.t. μ).

Finally the Lebesgue decomposition theorem has an abstract form:

Theorem I.20 (Lebesgue decomposition theorem) Let μ, ν be two measures on a measure space $\langle M, \mathscr{R}\rangle$. Then ν can be written uniquely as $\nu = \nu_{\text{ac}} + \nu_{\text{sing}}$ where μ and ν_{sing} are mutually singular and ν_{ac} is absolutely continuous w.r.t. μ.

There is one final subject in measure theory which we must consider and that involves changing the order of integration in a multiple integral. We first must consider what functions can be multiply integrated:

Definition Let $\langle M, \mathscr{R}\rangle$, $\langle N, \mathscr{F}\rangle$ be two sets with associated σ-fields. Then the σ-field, $\mathscr{R} \otimes \mathscr{F}$ of subsets of $M \times N$ is defined to be the smallest σ-field containing $\{R \times F \mid R \in \mathscr{R}, F \in \mathscr{F}\}$.

Notice that if $f\colon M \times N \to \mathbb{R}$ is measurable (w.r.t. $\mathscr{R} \otimes \mathscr{F}$), then for any $m \in M$, the function $n \mapsto f(m, n)$ is measurable (w.r.t. $\mathscr{F}$). If ν is a measure on N such that $\int f(m, n)\, d\nu(n)$ exists for all m, then one can show that $m \mapsto \int f(m, n)\, d\nu(n)$ is measurable (w.r.t. $\mathscr{R}$). There is a direct analogue of the fact that absolute convergent sums can be rearranged at will:

Theorem I.21 (Fubini's theorem) Let f be a measurable function on $M \times N$. Let μ be a measure on M, ν a measure on N. Then

$$\int_M \left(\int_N |f(m, n)|\, d\nu(n) \right) d\mu(m) < \infty$$

if and only if

$$\int_N \left(\int_M |f(m, n)|\, d\mu(m) \right) d\nu(n) < \infty$$

and if one (and thus both) of these integrals is finite, then

$$\int_N \left(\int_M f(m, n)\, d\mu(m) \right) d\nu(n) = \int_M \left(\int_N f(m, n)\, d\nu(n) \right) d\mu(m)$$

In Problem 25, the reader will see that the finiteness of the integral of the absolute value is critical.

Fubini's theorem can be put into perspective by the notion of product measure:

Theorem I.22 Let μ be a σ-finite measure on $\langle M, \mathscr{R} \rangle$ and ν a σ-finite measure on $\langle N, \mathscr{F} \rangle$. Then, there is a unique measure $\mu \otimes \nu$ on $\langle M \times N, \mathscr{R} \otimes \mathscr{F} \rangle$ obeying

$$(\mu \otimes \nu)(R \times F) = \mu(R)\nu(F)$$

(where $0 \cdot \infty = 0$). If f is a measurable function on $M \times N$, then

$$\int_M \left(\int_N |f(m, n)| \, d\nu(n) \right) d\mu(m) < \infty$$

if and only if

$$\int_{M \times N} |f| \, d(\mu \otimes \nu) < \infty$$

and in that case

$$\int_{M \times N} f \, d(\mu \otimes \nu) = \int_M \left(\int_N f \, d\nu \right) d\mu$$

One can describe the measure $\mu \otimes \nu$ quite explicitly. If $M \in \mathscr{R} \times \mathscr{F}$ and $M \subset \bigcup_{i=1}^{\infty} R_i \times F_i$ we have $(\mu \otimes \nu)(M) \leq \sum_{i=1}^{\infty} \mu(R_i)\nu(F_i)$. In fact, for any $M \in \mathscr{R} \times \mathscr{F}$,

$$(\mu \otimes \nu)(M) = \inf \left\{ \sum_{i=1}^{\infty} \mu(R_i)\nu(F_i) \,\middle|\, M \subset \bigcup_{i=1}^{\infty} R_i \times F_i \right\}$$

In particular, we can approximate M with a countable union of rectangles making an arbitrarily small error.