Chapter I

Measure and Integration

This chapter is devoted to the basic notions of measurable spaces, measure, and integration. The coverage is limited to what probability theory requires as the entrance fee from its students. The presentation is in the form and style attuned to the modern treatments of probability theory and stochastic processes.

1 Measurable Spaces

Let E be a set. We use the usual notations for operations on subsets of E:

1.1 $$A \cup B, \quad A \cap B, \quad A \setminus B$$

denote, respectively, the union of A and B, the intersection of A and B, and the complement of B in A. In particular, $E \setminus B$ is called simply the complement of B and is also denoted by B^c. We write $A \subset B$ or $B \supset A$ to mean that A is a subset of B, that is, A is contained in B, or equivalently, B contains A. Note that $A = B$ if and only if $A \subset B$ and $A \supset B$. For an arbitrary collection $\{A_i : i \in I\}$ of subsets of E, we write

1.2 $$\bigcup_{i \in I} A_i, \qquad \bigcap_{i \in I} A_i$$

for the union and intersection, respectively, of all the sets A_i, $i \in I$.

The empty set is denoted by $\emptyset$. Sets A and B are said to be *disjoint* if $A \cap B = \emptyset$. A collection of sets is said to be *disjointed* if its every element is disjoint from every other. A countable disjointed collection of sets whose union is A is called a *partition* of A.

A collection $\mathcal{C}$ of subsets of E is said to be *closed under* intersections if $A \cap B$ belongs to $\mathcal{C}$ whenever A and B belong to $\mathcal{C}$. Of course, then, the

E. Çınlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, DOI 10.1007/978-0-387-87859-1_1,

intersection of every non-empty finite collection of sets in $\mathcal{C}$ is in $\mathcal{C}$. If the intersection of every countable collection of sets in $\mathcal{C}$ is in $\mathcal{C}$, then we say that $\mathcal{C}$ is closed under countable intersections. The notions of being closed under complements, unions, and countable unions, etc. are defined similarly.

Sigma-algebras

A non-empty collection $\mathcal{E}$ of subsets of E is called an *algebra* on E provided that it be closed under finite unions and complements. It is called a *σ-algebra* on E if it is closed under complements and countable unions, that is, if

1.3 a) $A \in \mathcal{E} \quad \Rightarrow \quad E \setminus A \in \mathcal{E}$,
b) $A_1, A_2, \ldots \in \mathcal{E} \quad \Rightarrow \quad \bigcup_n A_n \in \mathcal{E}$.

Since the intersection of a collection of sets is the complement of the union of the complements of those sets, a σ-algebra is also closed under countable intersections.

Every σ-algebra on E includes E and $\emptyset$ at least. Indeed, $\mathcal{E} = \{\emptyset, E\}$ is the simplest σ-algebra on E; it is called the *trivial σ-algebra*. The largest is the collection of all subsets of E, usually denoted by 2^E; it is called the *discrete σ-algebra* on E.

The intersection of an arbitrary (countable or uncountable) family of σ-algebras on E is again a σ-algebra on E. Given an arbitrary collection $\mathcal{C}$ of subsets of E, consider all the σ-algebras that contain $\mathcal{C}$ (there is at least one such σ-algebra, namely 2^E); take the intersection of all those σ-algebras; the result is the smallest σ-algebra that contains $\mathcal{C}$; it is called the σ-algebra *generated* by $\mathcal{C}$ and is denoted by $\sigma\mathcal{C}$.

If E is a topological space, then the σ-algebra generated by the collection of all open subsets of E is called the *Borel σ-algebra* on E; it is denoted by$\mathcal{B}_E$ or $\mathcal{B}(E)$; its elements are called *Borel sets*.

p-systems and d-systems

A collection $\mathcal{C}$ of subsets of E is called a p-system if it is closed under intersections; here, p is for product, the latter being an alternative term for intersection, and next, d is for Dynkin who introduced these systems into probability. A collection $\mathcal{D}$ of subsets of E is called a d-system on E if

1.4 a) $E \in \mathcal{D}$,
b) $A, B \in \mathcal{D}$ and $A \supset B \quad \Rightarrow A \setminus B \in \mathcal{D}$,
c) $(A_n) \subset \mathcal{D}$ and $A_n \nearrow A \quad \Rightarrow A \in \mathcal{D}$.

In the last line, we wrote $(A_n) \subset \mathcal{D}$ to mean that (A_n) is a sequence of elements of $\mathcal{D}$ and we wrote $A_n \nearrow A$ to mean that the sequence is increasing with limit A in the following sense:

1.5 $$A_1 \subset A_2 \subset \ldots, \quad \cup_n A_n = A.$$

It is obvious that a σ-algebra is both a p-system and a d-system, and the converse will be shown next. Thus, p-systems and d-systems are primitive structures whose superpositions yield σ-algebras.

1.6 PROPOSITION. *A collection of subsets of E is a σ-algebra if and only if it is both a p-system and a d-system on E.*

Proof. Necessity is obvious. To show the sufficiency, let $\mathcal{E}$ be a collection of subsets of E that is both a p-system and a d-system. First, $\mathcal{E}$ is closed under complements: $A \in \mathcal{E} \Rightarrow E \setminus A \in \mathcal{E}$, since $E \in \mathcal{E}$ and $A \subset E$ and $\mathcal{E}$ is a d-system. Second, it is closed under unions: $A, B \in \mathcal{E} \Rightarrow A \cup B \in \mathcal{E}$, because $A \cup B = (A^c \cap B^c)^c$ and $\mathcal{E}$ is closed under complements (as shown) and under intersections by the hypothesis that it is a p-system. Finally, this closure extends to countable unions: if $(A_n) \subset \mathcal{E}$, then $B_1 = A_1$ and $B_2 = A_1 \cup A_2$ and so on belong to $\mathcal{E}$ by the preceding step, and $B_n \nearrow \bigcup_n A_n$, which together imply that $\bigcup_n A_n \in \mathcal{E}$ since $\mathcal{E}$ is a d-system by hypothesis. □

The lemma next is in preparation for the main theorem of this section. Its proof is left as an exercise in checking the conditions 1.4 one by one.

1.7 LEMMA. *Let $\mathcal{D}$ be a d-system on E. Fix D in $\mathcal{D}$ and let*

$$\hat{\mathcal{D}} = \{A \in \mathcal{D} : \ A \cap D \in \mathcal{D}\}$$

Then, $\hat{\mathcal{D}}$ is again a d-system.

Monotone class theorem

This is a very useful tool for showing that certain collections are σ-algebras. We give it in the form found most useful in probability theory.

1.8 THEOREM. *If a d-system contains a p-system, then it contains also the σ-algebra generated by that p-system.*

Proof. Let $\mathcal{C}$ be a p-system. Let $\mathcal{D}$ be the smallest d-system on E that contains $\mathcal{C}$, that is, $\mathcal{D}$ is the intersection of all d-systems containing $\mathcal{C}$. The claim is that $\mathcal{D} \supset \sigma\mathcal{C}$. To show it, since $\sigma\mathcal{C}$ is the smallest σ-algebra containing $\mathcal{C}$, it is sufficient to show that $\mathcal{D}$ is a σ-algebra. In view of Proposition 1.6, it is thus enough to show that the d-system $\mathcal{D}$ is also a p-system.

To that end, fix B in $\mathcal{C}$ and let

$$\mathcal{D}_1 = \{A \in \mathcal{D} : \ A \cap B \in \mathcal{D}\}.$$

Since $\mathcal{C}$ is contained in $\mathcal{D}$, the set B is in $\mathcal{D}$; and Lemma 1.7 implies that $\mathcal{D}_1$ is a d-system. It also contains $\mathcal{C}$: if $A \in \mathcal{C}$ then $A \cap B \in \mathcal{C}$ since B is in $\mathcal{C}$ and $\mathcal{C}$ is a p-system. Hence, $\mathcal{D}_1$ must contain the smallest d-system containing $\mathcal{C}$, that is, $\mathcal{D}_1 \supset \mathcal{D}$. In other words, $A \cap B \in \mathcal{D}$ for every A in $\mathcal{D}$ and B in $\mathcal{C}$.

Consequently, for fixed A in $\mathcal{D}$, the collection

$$\mathcal{D}_2 = \{B \in \mathcal{D} : \ A \cap B \in \mathcal{D}\}$$

contains $\mathcal{C}$. By Lemma 1.7, $\mathcal{D}_2$ is a d-system. Thus, $\mathcal{D}_2$ must contain $\mathcal{D}$. In other words, $A \cap B \in \mathcal{D}$ whenever A and B are in $\mathcal{D}$, that is, $\mathcal{D}$ is a p-system. □

Measurable spaces

A *measurable space* is a pair $(E, \mathcal{E})$ where E is a set and $\mathcal{E}$ is a σ-algebra on E. Then, the elements of $\mathcal{E}$ are called *measurable sets*. When E is topological and $\mathcal{E} = \mathcal{B}_E$, the Borel σ-algebra on E, then measurable sets are also called *Borel sets*.

Products of measurable spaces

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. For $A \subset E$ and $B \subset F$, we write $A \times B$ for the set of all pairs (x, y) with x in A and y in B; it is called the *product* of A and B. If $A \in \mathcal{E}$ and $B \in \mathcal{F}$, then $A \times B$ is said to be a *measurable rectangle*. We let $\mathcal{E} \otimes \mathcal{F}$ denote the σ-algebra on $E \times F$ generated by the collection of all measurable rectangles; it is called the *product σ-algebra*. The measurable space $(E \times F, \mathcal{E} \otimes \mathcal{F})$ is called the product of $(E, \mathcal{E})$ and $(F, \mathcal{F})$, and the notation $(E, \mathcal{E}) \times (F, \mathcal{F})$ is used as well.

Exercises

1.9 *Partition generated σ-algebras.*

a) Let $\mathcal{C} = \{A, B, C\}$ be a partition of E. List the elements of $\sigma\mathcal{C}$.

b) Let $\mathcal{C}$ be a (countable) partition of E. Show that every element of $\sigma\mathcal{C}$ is a countable union of elements taken from $\mathcal{C}$. Hint: Let $\mathcal{E}$ be the collection of all sets that are countable unions of elements taken from $\mathcal{C}$. Show that $\mathcal{E}$ is a σ-algebra, and argue that $\mathcal{E} = \sigma\mathcal{C}$.

c) Let $E = \mathbb{R}$, the set of all real numbers. Let $\mathcal{C}$ be the collection of all singleton subsets of $\mathbb{R}$, that is, each element of $\mathcal{C}$ is a set that consists of exactly one point in $\mathbb{R}$. Show that every element of $\sigma\mathcal{C}$ is either a countable set or the complement of a countable set. Incidentally, $\sigma\mathcal{C}$ is much smaller than $\mathcal{B}(\mathbb{R})$; for instance, the interval $(0, 1)$ belongs to the latter but not to the former.

1.10 *Comparisons.* Let $\mathcal{C}$ and $\mathcal{D}$ be two collections of subsets of E. Show the following:

a) If $\mathcal{C} \subset \mathcal{D}$ then $\sigma\mathcal{C} \subset \sigma\mathcal{D}$
b) If $\mathcal{C} \subset \sigma\mathcal{D}$ then $\sigma\mathcal{C} \subset \sigma\mathcal{D}$
c) If $\mathcal{C} \subset \sigma\mathcal{D}$ and $\mathcal{D} \subset \sigma\mathcal{C}$, then $\sigma\mathcal{C} = \sigma\mathcal{D}$
d) If $\mathcal{C} \subset \mathcal{D} \subset \sigma\mathcal{C}$, then $\sigma\mathcal{C} = \sigma\mathcal{D}$

1.11 *Borel σ-algebra on* $\mathbb{R}$. Every open subset of $\mathbb{R} = (-\infty, +\infty)$, the real line, is a countable union of open intervals. Use this fact to show that $\mathcal{B}_{\mathbb{R}}$ is generated by the collection of all open intervals.

1.12 *Continuation.* Show that every interval of $\mathbb{R}$ is a Borel set. In particular, $(-\infty, x)$, $(-\infty, x]$, $(x, y]$, $[x, y]$ are all Borel sets. For each x, the singleton $\{x\}$ is a Borel set.

1.13 *Continuation.* Show that $\mathcal{B}_{\mathbb{R}}$ is also generated by any one of the following (and many others):

a) The collection of all intervals of the form $(-\infty, x]$.
b) The collection of all intervals of the form $(x, y]$.
c) The collection of all intervals of the form $[x, y]$.
d) The collection of all intervals of the form (x, ∞).

Moreover, in each case, x and y can be limited to be rationals.

1.14 *Lemma 1.7.* Prove.

1.15 *Trace spaces.* Let $(E, \mathcal{E})$ be a measurable space. Fix $D \subset E$ and let

$$\mathcal{D} = \mathcal{E} \cap D = \{A \cap D : \; A \in \mathcal{E}\}.$$

Show that $\mathcal{D}$ is a σ-algebra on D. It is called the *trace* of $\mathcal{E}$ on D, and $(D, \mathcal{D})$ is called the *trace* of $(E, \mathcal{E})$ on D.

1.16 *Single point extensions.* Let $(E, \mathcal{E})$ be a measurable space, and let Δ be an extra point, not in E. Let $\bar{E} = E \cup \{\Delta\}$. Show that

$$\bar{\mathcal{E}} = \mathcal{E} \cup \{A \cup \{\Delta\} : \; A \in \mathcal{E}\}$$

is a σ-algebra on $\bar{E}$; it is the σ-algebra on $\bar{E}$ generated by $\mathcal{E}$.

1.17 *Product spaces.* Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Show that the product σ-algebra $\mathcal{E} \otimes \mathcal{F}$ is also the σ-algebra generated by $\hat{\mathcal{E}} \cup \hat{\mathcal{F}}$, where

$$\hat{\mathcal{E}} = \{A \times F : \; A \in \mathcal{E}\}, \quad \hat{\mathcal{F}} = \{E \times B : \; B \in \mathcal{F}\}.$$

1.18 *Unions of σ-algebras.* Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be σ-algebras on the same set E. Their union is not a σ-algebra, except in some special cases. The σ-algebra generated by $\mathcal{E}_1 \cup \mathcal{E}_2$ is denoted by $\mathcal{E}_1 \vee \mathcal{E}_2$. More generally, if $\mathcal{E}_i$ is a σ-algebra on E for each i in some (countable or uncountable) index set I, then

$$\mathcal{E}_I = \bigvee_{i \in I} \mathcal{E}_i$$

denotes the σ-algebra generated by $\bigcup_{i \in I} \mathcal{E}_i$ (a similar notation for intersection is superfluous, since $\bigcap_{i \in I} \mathcal{E}_i$ is always a σ-algebra). Let $\mathcal{C}$ be the collection of all sets A having the form

$$A = \bigcap_{i \in J} A_i$$

for some finite subset J of I and sets A_i in $\mathcal{E}_i$, $i \in J$. Show that $\mathcal{C}$ contains all $\mathcal{E}_i$ and therefore $\bigcup_I \mathcal{E}_i$. Thus, $\mathcal{C}$ generates the σ-algebra $\mathcal{E}_I$. Show that $\mathcal{C}$ is a p-system.

2 Measurable Functions

Let E and F be sets. A *mapping* or *function* f from E into F is a rule that assigns an element $f(x)$ of F to each x in E, and then we write $f : E \mapsto F$ to indicate it. If $f(x)$ is an element of F for each x in E, we also write $f : x \mapsto f(x)$ to name the mapping involved; for example, $f : x \mapsto x^2 + 5$ is the function f from $\mathbb{R}$ into $\mathbb{R}_+$ satisfying $f(x) = x^2 + 5$. Given a mapping $f : E \mapsto F$ and a subset B of F, the *inverse image* of B under f is

2.1 $$f^{-1}B = \{x \in E : \ f(x) \in B\}.$$

We leave the proof of the next lemma as an exercise in ordinary logic.

2.2 Lemma. *Let f be a mapping from E into F. Then,*

$$f^{-1}\emptyset = \emptyset, \quad f^{-1}F = E, \quad f^{-1}(B \setminus C) = (f^{-1}B) \setminus (f^{-1}C),$$
$$f^{-1}\bigcup_i B_i = \bigcup_i f^{-1}B_i, \quad f^{-1}\bigcap_i B_i = \bigcap_i f^{-1}B_i$$

for all subsets B and C of F and arbitrary collections $\{B_i : \ i \in I\}$ of subsets of F.

Measurable functions

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A mapping $f : E \mapsto F$ is said to be *measurable* relative to $\mathcal{E}$ and $\mathcal{F}$ if $f^{-1}B \in \mathcal{E}$ for every B in $\mathcal{F}$. The following reduces the checks involved.

2.3 Proposition. *In order for $f : E \mapsto F$ to be measurable relative to $\mathcal{E}$ and $\mathcal{F}$, it is necessary and sufficient that, for some collection $\mathcal{F}_0$ that generates $\mathcal{F}$, we have $f^{-1}B \in \mathcal{E}$ for every B in $\mathcal{F}_0$.*

Proof. Necessity is trivial. To prove the sufficiency, let $\mathcal{F}_0$ be a collection of subsets of F such that $\sigma\mathcal{F}_0 = \mathcal{F}$, and suppose that $f^{-1}B \in \mathcal{E}$ for every B in $\mathcal{F}_0$. We need to show that

$$\mathcal{F}_1 = \{B \in \mathcal{F} : \ f^{-1}B \in \mathcal{E}\}$$

contains $\mathcal{F}$ and thus is equal to $\mathcal{F}$. Since $\mathcal{F}_1 \supset \mathcal{F}_0$ by assumption, once we show that $\mathcal{F}_1$ is a σ-algebra, we will have $\mathcal{F}_1 = \sigma\mathcal{F}_1 \supset \sigma\mathcal{F}_0 = \mathcal{F}$ as needed. But checking that $\mathcal{F}_1$ is a σ-algebra is straightforward using Lemma 2.2. □

Composition of functions

Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, and $(G, \mathcal{G})$ be measurable spaces. Let f be a mapping from E into F, and g a mapping from F into G. The *composition* of f and g is the mapping $g \circ f$ from E into G defined by

2.4 $$g \circ f(x) = g(f(x)), \quad x \in E.$$

The next proposition will be recalled by the phrase "measurable functions of measurable functions are measurable".

2.5 PROPOSITION. *If f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and g relative to $\mathcal{F}$ and $\mathcal{G}$, then $g \circ f$ is measurable relative to $\mathcal{E}$ and $\mathcal{G}$.*

Proof. Let f and g be measurable. For C in $\mathcal{G}$, observe that $(g \circ f)^{-1}C = f^{-1}(g^{-1}C)$. Now, $g^{-1}C \in \mathcal{F}$ by the measurability of g and, hence, $f^{-1}(g^{-1}C) \in \mathcal{E}$ by the measurability of f. So, $g \circ f$ is measurable. □

Numerical functions

Let $(E, \mathcal{E})$ be a measurable space. Recall that $\mathbb{R} = (-\infty, +\infty)$, $\bar{\mathbb{R}} = [-\infty, +\infty]$, $\mathbb{R}_+ = [0, +\infty)$, $\bar{\mathbb{R}}_+ = [0, +\infty]$. A *numerical function* on E is a mapping from E into $\bar{\mathbb{R}}$ or some subset of $\bar{\mathbb{R}}$. If all its values are in $\mathbb{R}$, it is said to be *real-valued.* If all its values are in $\bar{\mathbb{R}}_+$, it is said to be *positive.*

A numerical function on E is said to be $\mathcal{E}$-*measurable* if it is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\bar{\mathbb{R}})$, the latter denoting the Borel σ-algebra on $\bar{\mathbb{R}}$ as usual. If E is topological and $\mathcal{E} = \mathcal{B}(E)$, then $\mathcal{E}$-measurable functions are called *Borel functions.*

The following proposition is a corollary of Proposition 2.3 using the fact that $\mathcal{B}(\bar{\mathbb{R}})$ is generated by the collection of intervals $[-\infty, r]$ with r in $\mathbb{R}$. No proof seems needed.

2.6 PROPOSITION. *A mapping $f : E \mapsto \bar{\mathbb{R}}$ is $\mathcal{E}$-measurable if and only if, for every r in $\mathbb{R}$, $f^{-1}[-\infty, r] \in \mathcal{E}$.*

2.7 REMARKS. a) The proposition remains true if $[-\infty, r]$ is replaced by $[-\infty, r)$ or by $[r, \infty]$ or by $(r, \infty]$, because the intervals $[-\infty, r)$ with r in $\mathbb{R}$ generate $\mathcal{B}(\bar{\mathbb{R}})$ and similarly for the other two forms.

b) In the particular case $f : E \mapsto F$, where F is a countable subset of $\bar{\mathbb{R}}$, the mapping f is $\mathcal{E}$-measurable if and only if $f^{-1}\{a\} = \{x \in E : f(x) = a\}$ is in $\mathcal{E}$ for every a in F.

Positive and negative parts of a function

For a and b in $\bar{\mathbb{R}}$ we write $a \vee b$ for the maximum of a and b, and $a \wedge b$ for the minimum. The notation extends to numerical functions naturally: for instance, $f \vee g$ is the function whose value at x is $f(x) \vee g(x)$. Let $(E, \mathcal{E})$ be a measurable space. Let f be a numerical function on E. Then,

2.8 $$f^+ = f \vee 0, \qquad f^- = -(f \wedge 0)$$

are both positive functions and $f = f^+ - f^-$. The function f^+ is called the *positive part* of f, and f^- the *negative part.*

2.9 PROPOSITION. *The function f is $\mathcal{E}$-measurable if and only if both f^+ and f^- are.*

Proof is left as an exercise. The decomposition $f = f^+ - f^-$ enables us to obtain many results for arbitrary f from the corresponding results for positive functions.

Indicators and simple functions

Let $A \subset E$. Its indicator, denoted by 1_A, is the function defined by

2.10 $$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

We write simply 1 for 1_E. Obviously, 1_A is $\mathcal{E}$-measurable if and only if $A \in \mathcal{E}$.

A function f on E is said to be *simple* if it has the form

2.11 $$f = \sum_1^n a_i 1_{A_i}$$

for some n in $\mathbb{N}^* = \{1, 2, \ldots\}$, real numbers $a_1, \ldots, a_n$, and measurable sets $A_1, \ldots, A_n$ (belonging to the σ-algebra $\mathcal{E}$). It is clear that, then, there exist m in $\mathbb{N}^*$ and distinct real numbers $b_1, \ldots, b_m$ and a measurable partition $\{B_1, \ldots, B_m\}$ of E such that $f = \sum_1^m b_i 1_{B_i}$; this latter representation is called the *canonical form* of the simple function f.

It is immediate from Proposition 2.6 (or Remark 2.7b) applied to the canonical form that every simple function is $\mathcal{E}$-measurable. Conversely, if f is $\mathcal{E}$-measurable, takes only finitely many values, and all those values are real numbers, then f is a simple function. In particular, every constant is a simple function. Finally, if f and g are simple, then so are

2.12 $$f + g, \qquad f - g, \qquad fg, \qquad f/g, \qquad f \vee g, \qquad f \wedge g,$$

except that in the case of f/g one should make sure that g is nowhere zero.

Limits of sequences of functions

Let (f_n) be a sequence of numerical functions on E. The functions

2.13 $$\inf f_n, \qquad \sup f_n, \qquad \liminf f_n, \qquad \limsup f_n$$

are defined on E pointwise: for instance, the first is the function whose value at x is the infimum of the sequence of numbers $f_n(x)$. In general, limit inferior is dominated by the limit superior. If the two are equal, that is, if

2.14 $$\liminf f_n = \limsup f_n = f,$$

say, then the sequence (f_n) is said to have a pointwise limit f and we write $f = \lim f_n$ or $f_n \to f$ to express it.

If (f_n) is increasing, that is, if $f_1 \le f_2 \le \ldots$, then $\lim f_n$ exists and is equal to $\sup f_n$. We shall write $f_n \nearrow f$ to mean that (f_n) is increasing and has limit f. Similarly, $f_n \searrow f$ means that (f_n) is decreasing and has limit f.

The following shows that the class of measurable functions is closed under limits.

2.15 THEOREM. *Let (f_n) be a sequence of $\mathcal{E}$-measurable functions. Then, each one of the four functions in 2.13 is $\mathcal{E}$-measurable. Moreover, if it exists, $\lim f_n$ is $\mathcal{E}$-measurable.*

Proof. We start by showing that $f = \sup f_n$ is $\mathcal{E}$-measurable. For every x in E and r in $\mathbb{R}$, we note that $f(x) \le r$ if and only if $f_n(x) \le r$ for all n. Thus, for each r in $\mathbb{R}$,

$$f^{-1}[-\infty, r] = \{x : \ f(x) \le r\} = \bigcap_n \{x : \ f_n(x) \le r\} = \bigcap_n f_n^{-1}[-\infty, r].$$

The rightmost member belongs to $\mathcal{E}$: for each n, the set $f_n^{-1}[-\infty, r] \in \mathcal{E}$ by the $\mathcal{E}$-measurability of f_n, and $\mathcal{E}$ is closed under countable intersections. So, by Proposition 2.6, $f = \sup f_n$ is $\mathcal{E}$-measurable.

Measurability of $\inf f_n$ follows from the preceding step upon observing that $\inf f_n = -\sup(-f_n)$. It is now obvious that

$$\liminf f_n = \sup_m \inf_{n \ge m} f_n, \qquad \limsup f_n = \inf_m \sup_{n \ge m} f_n$$

are $\mathcal{E}$-measurable. If these two are equal, the common limit is the definition of $\lim f_n$, which is $\mathcal{E}$-measurable. □

Approximation of measurable functions

We start by approximating the identity function on $\bar{\mathbb{R}}_+$ by an increasing sequence of simple functions of a specific form (dyadic functions). We leave the proof of the next lemma as an exercise; drawing d_n for $n = 1, 2, 3$ should do.

2.16 LEMMA. *For each n in $\mathbb{N}^*$, let*

$$d_n(r) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)}(r) + n 1_{[n,\infty]}(r), \quad r \in \bar{\mathbb{R}}_+.$$

Then, each d_n is an increasing right-continuous simple function on $\bar{\mathbb{R}}_+$, and $d_n(r)$ increases to r for each r in $\bar{\mathbb{R}}_+$ as $n \to \infty$.

The following theorem is important: it reduces many a computation about measurable functions to a computation about simple functions followed by limit taking.

2.17 THEOREM. *A positive function on E is $\mathcal{E}$-measurable if and only if it is the limit of an increasing sequence of positive simple functions.*

Proof. Sufficiency is immediate from Theorem 2.15. To show the necessity part, let $f : E \mapsto \bar{\mathbb{R}}_+$ be $\mathcal{E}$-measurable. We are to show that there is a sequence (f_n) of positive simple functions increasing to f. To that end, let (d_n) be as in the preceding lemma and put $f_n = d_n \circ f$. Then, for each n, the function f_n is $\mathcal{E}$-measurable, since it is a measurable function of a measurable function. Also, it is positive and takes only finitely many values, because d_n is so. Thus, each f_n is positive and simple. Moreover, since $d_n(r)$ increases to r for each r in $\bar{\mathbb{R}}_+$ as $n \to \infty$, we have that $f_n(x) = d_n(f(x))$ increases to $f(x)$ for each x in E as $n \to \infty$.

Monotone classes of functions

Let $\mathcal{M}$ be a collection of numerical functions on E. We write $\mathcal{M}_+$ for the subcollection consisting of positive functions in $\mathcal{M}$, and $\mathcal{M}_b$ for the subcollection of bounded functions in $\mathcal{M}$.

The collection $\mathcal{M}$ is called a *monotone class* provided that it includes the constant function 1, and $\mathcal{M}_b$ is a linear space over $\mathbb{R}$, and $\mathcal{M}_+$ is closed under increasing limits; more explicitly, $\mathcal{M}$ is a monotone class if

2.18 a) $1 \in \mathcal{M}$,
b) $f, g \in \mathcal{M}_b$ and $a, b \in \mathbb{R} \;\Rightarrow\; af + bg \in \mathcal{M}$,
c) $(f_n) \subset \mathcal{M}_+,\ f_n \nearrow f \;\Rightarrow\; f \in \mathcal{M}$.

The next theorem is used often to show that a certain property holds for all $\mathcal{E}$-measurable functions. It is a version of Theorem 1.8, it is called the monotone class theorem for functions.

2.19 THEOREM. *Let $\mathcal{M}$ be a monotone class of functions on E. Suppose, for some p-system $\mathcal{C}$ generating $\mathcal{E}$, that $1_A \in \mathcal{M}$ for every A in $\mathcal{C}$. Then, $\mathcal{M}$ includes all positive $\mathcal{E}$-measurable functions and all bounded $\mathcal{E}$-measurable functions.*

Proof. We start by showing that $1_A \in \mathcal{M}$ for every A in $\mathcal{E}$. To this end, let

$$\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{M}\}.$$

Using the conditions 2.18, it is easy to check that $\mathcal{D}$ is a d-system. Since $\mathcal{D} \supset \mathcal{C}$ by assumption, and since $\mathcal{C}$ is a p-system that generates $\mathcal{E}$, we must have $\mathcal{D} \supset \mathcal{E}$ by the monotone class theorem 1.8. So, $1_A \in \mathcal{M}$ for every A in $\mathcal{E}$.

Therefore, in view of the property 2.18b, $\mathcal{M}$ includes all simple functions.

Let f be a positive $\mathcal{E}$-measurable function. By Theorem 2.17, there exists a sequence of positive simple functions f_n increasing to f. Since each f_n is in $\mathcal{M}_+$ by the preceding step, the property 2.18c implies that $f \in \mathcal{M}$.

Finally, let f be a bounded $\mathcal{E}$-measurable function. Then f^+ and f^- are in $\mathcal{M}$ by the preceding step and are bounded obviously. Thus, by 2.18b, we conclude that $f = f^+ - f^- \in \mathcal{M}$. □

Standard measurable spaces

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Let f be a bijection from E onto F, and let $\hat{f}$ denote its functional inverse ($\hat{f}(y) = x$ if and only if $f(x) = y$). Then, f is said to be an *isomorphism* of $(E, \mathcal{E})$ and $(F, \mathcal{F})$ if f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ and $\hat{f}$ is measurable relative to $\mathcal{F}$ and $\mathcal{E}$. The measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$ are said to be *isomorphic* if there exists an isomorphism between them.

A measurable space $(E, \mathcal{E})$ is said to be *standard* if it is isomorphic to $(F, \mathcal{B}_F)$ for some Borel subset F of $\mathbb{R}$.

The class of standard spaces is surprisingly large and includes almost all the spaces we shall encounter. Here are some examples: The spaces $\mathbb{R}$, $\mathbb{R}^d$, $\mathbb{R}^\infty$ together with their respective Borel σ-algebras are standard measurable spaces. If E is a complete separable metric space, then $(E, \mathcal{B}_E)$ is standard. If E is a Polish space, that is, if E is a topological space metrizable by a metric for which it is complete and separable, then $(E, \mathcal{B}_E)$ is standard. If E is a separable Banach space, or more particularly, a separable Hilbert space, then $(E, \mathcal{B}_E)$ is standard. Further examples will appear later.

Clearly, $[0, 1]$ and its Borel σ-algebra form a standard measurable space; so do $\{1, 2, \ldots, n\}$ and its discrete σ-algebra; so do $\mathbb{N} = \{0, 1, \ldots\}$ and its discrete σ-algebra. Every standard measurable space is isomorphic to one of these three (this is a deep result).

Notation

We shall use $\mathcal{E}$ both for the σ-algebra and for the collection of all the numerical functions that are measurable relative to it. Recall that, for an arbitrary collection $\mathcal{M}$ of numerical functions, we write $\mathcal{M}_+$ for the subcollection of positive functions in $\mathcal{M}$, and $\mathcal{M}_b$ for the subcollection of bounded ones in $\mathcal{M}$. Thus, for instance, $\mathcal{E}_+$ is the collection of all $\mathcal{E}$-measurable positive functions.

A related notation is $\mathcal{E}/\mathcal{F}$ which is used for the class of all functions $f : E \mapsto F$ that are measurable relative to $\mathcal{E}$ and $\mathcal{F}$. The notation $\mathcal{E}/\mathcal{F}$ is simplified to $\mathcal{E}$ when $F = \bar{\mathbb{R}}$ and $\mathcal{F} = \mathcal{B}(\mathbb{R})$.

Exercises and complements

2.20 *σ-algebra generated by a function.* Let E be a set and $(F, \mathcal{F})$ a measurable space. For $f : E \mapsto F$, define

$$f^{-1}\mathcal{F} = \{f^{-1}B : \ B \in \mathcal{F}\}$$

where $f^{-1}B$ is as defined in 2.1. Show that $f^{-1}\mathcal{F}$ is a σ-algebra on E. It is the smallest σ-algebra on E such that f is measurable relative to it and $\mathcal{F}$. It is called the σ-algebra generated by f. If $(E, \mathcal{E})$ is a measurable space, then f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ if and only if $f^{-1}\mathcal{F} \subset \mathcal{E}$; this is another way of stating the definition of measurability.

2.21 *Product spaces.* Let $(E, \mathcal{E})$, $(F, \mathcal{F})$, $(G, \mathcal{G})$ be measurable spaces. Let $f : E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and let $g : E \mapsto G$ be measurable relative to $\mathcal{E}$ and $\mathcal{G}$. Define $h : E \mapsto F \times G$ by

$$h(x) = (f(x), g(x)), \quad x \in E.$$

Show that h is measurable relative to $\mathcal{E}$ and $\mathcal{F} \otimes \mathcal{G}$.

2.22 *Sections.* Let $f : E \times F \mapsto G$ be measurable relative to $\mathcal{E} \otimes \mathcal{F}$ and $\mathcal{G}$. Show that, for fixed x_0 in E, the mapping $h : y \mapsto f(x_0, y)$ is measurable relative to $\mathcal{F}$ and $\mathcal{G}$. (Hint: Note that $h = f \circ g$ where $g : F \mapsto E \times F$ is defined by $g(y) = (x_0, y)$ and show that g is measurable relative to $\mathcal{F}$ and $\mathcal{E} \otimes \mathcal{F}$.) The mapping h is called the *section* of f at x_0.

2.23 *Proposition 2.9.* Prove.

2.24 *Discrete spaces.* Suppose that E is countable and $\mathcal{E} = 2^E$, the discrete σ-algebra on E. Then, $(E, \mathcal{E})$ is said to be *discrete*. Show that every function on E is $\mathcal{E}$-measurable.

2.25 Suppose that $\mathcal{E}$ is generated by a countable partition of E. Show that, then, a numerical function on E is $\mathcal{E}$-measurable if and only if it is constant over each member of that partition.

2.26 *Elementary functions.* A function f on E is said to be *elementary* if it has the form

$$f = \sum_1^\infty a_i 1_{A_i},$$

where $a_i \in \bar{\mathbb{R}}$ and $A_i \in \mathcal{E}$ for each i, the A_i being disjoint. Show that every such function is $\mathcal{E}$-measurable.

2.27 *Measurable functions.* Show that a positive function f on E is $\mathcal{E}$-measurable if and only if it has the form

$$f = \sum_1^\infty a_n 1_{A_n},$$

for some sequence $(a_n) \subset \bar{\mathbb{R}}_+$ and some sequence $(A_n) \subset \mathcal{E}$, disjointedness not required.

2.28 *Approximation by simple functions.* Show that a numerical function f on E is $\mathcal{E}$-measurable if and only if it is the limit of a sequence (f_n) of simple functions. Hint: For necessity, put $f_n = f_n^+ - f_n^-$, where $f_n^+ = d_n \circ f^+$ and $f_n^- = d_n \circ f^-$ with d_n as in Lemma 2.16.

2.29 *Arithmetic operations.* Let f and g be $\mathcal{E}$-measurable. Show that, then, each one of

$$f + g, \qquad f - g, \qquad f \cdot g, \qquad f/g$$

is $\mathcal{E}$-measurable provided that it be well-defined (the issue arises from the fact that $+\infty - \infty$, $(+\infty)(-\infty)$, $0/0$, ∞/∞ are undefined). Recall, however, that $0 \cdot \infty = 0$ is defined.

2.30 *Continuous functions.* Suppose that E is topological. Show that every continuous function $f : E \mapsto \bar{\mathbb{R}}$ is a Borel function. Hint: If f is continuous, then $f^{-1}B$ is open for every open subset of $\bar{\mathbb{R}}$.

2.31 *Step functions, right-continuous functions.* a) A function $f : \mathbb{R}_+ \mapsto \bar{\mathbb{R}}$ is said to be a right-continuous step function if there is a sequence (t_n) in $\mathbb{R}_+$ with $0 = t_0 < t_1 < \cdots$ and $\lim t_n = +\infty$ such that f is constant over each interval $[t_n, t_{n+1})$. Every such function is elementary and, thus, Borel measurable. b) Let $f : \mathbb{R}_+ \mapsto \bar{\mathbb{R}}$ be right-continuous, that is, $f(r_n) \to f(r)$ whenever (r_n) is a sequence decreasing to r. Show that f is Borel measurable. Hint: Note that $f = \lim f_n$, where $f_n = f \circ \bar{d}_n$ for n in $\mathbb{N}^*$ with

$$\bar{d}_n(r) = \sum_{k=1}^{\infty} \frac{k}{2^n} 1_{\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)}(r), \quad r \in \mathbb{R}_+.$$

Extend this to $f : \mathbb{R} \mapsto \bar{\mathbb{R}}$ by symmetry on $\mathbb{R} \setminus \mathbb{R}_+$. Similarly, every left-continuous function is Borel.

2.32 *Increasing functions.* Let $f : \mathbb{R} \mapsto \bar{\mathbb{R}}$ be increasing. Show that f is Borel measurable.

2.33 *Measurability of sets defined by functions.* We introduce the notational principle that $\{f \in B\}$, $\{f > r\}$, $\{f \leq g\}$, etc. stand for, respectively,

$$\{x \in E : \ f(x) \in B\}, \quad \{x \in E : \ f(x) > r\}, \quad \{x \in E : \ f(x) \leq g(x)\},$$

etc. For instance, $\{f \leq g\}$ is the set on which f is dominated by g.

Let f and g be $\mathcal{E}$-measurable functions on E. Show that the following sets are in $\mathcal{E}$:

$$\{f > g\}, \quad \{f < g\}, \quad \{f \neq g\}, \quad \{f = g\}, \quad \{f \geq g\}, \quad \{f \leq g\}.$$

Hint: $\{f > g\}$ is the set of all x for which $f(x) > r$ and $g(x) < r$ for some rational number r.

2.34 *Positive monotone classes.* This is a variant of the monotone class theorem 2.19: Let $\mathcal{M}_+$ be a collection of positive functions on E. Suppose that

a) $1 \in \mathcal{M}_+$
b) $f, g \in \mathcal{M}_+$ and $a, b \in \mathbb{R}$ and $af + bg \geq 0 \ \Rightarrow \ af + bg \in \mathcal{M}_+$
c) $(f_n) \subset \mathcal{M}_+, \ f_n \nearrow f \ \Rightarrow \ f \in \mathcal{M}_+$.

Suppose, for some p-system $\mathcal{C}$ generating $\mathcal{E}$ that $1_A \in \mathcal{M}_+$ for each A in $\mathcal{C}$. Then, $\mathcal{M}_+$ includes every positive $\mathcal{E}$-measurable function. Prove.

2.35 *Bounded monotone classes.* This is another variant of the monotone class theorem. Let $\mathcal{M}_b$ be a collection of bounded functions on E. Suppose that

a) $1 \in \mathcal{M}_b$,
b) $f, g \in \mathcal{M}_b$ and $a, b \in \mathbb{R} \Rightarrow af + bg \in \mathcal{M}_b$,
c) $(f_n) \subset \mathcal{M}_b, f_n \geq 0, f_n \nearrow f,$ and f is bounded $\Rightarrow f \in \mathcal{M}_b$.

Suppose, for some p-system $\mathcal{C}$ generating $\mathcal{E}$ that $1_A \in \mathcal{M}_b$ for each A in $\mathcal{C}$. Then, $\mathcal{M}_b$ includes every bounded $\mathcal{E}$-measurable function. Prove.

3 Measures

Let $(E, \mathcal{E})$ be a measurable space, that is, E is a set and $\mathcal{E}$ is a σ-algebra on E. A *measure* on $(E, \mathcal{E})$ is a mapping $\mu : \mathcal{E} \mapsto \bar{\mathbb{R}}_+$ such that

3.1 a) $\mu(\emptyset) = 0$,
b) $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ for every disjointed sequence (A_n) in $\mathcal{E}$.

The latter condition is called *countable additivity*. Note that $\mu(A)$ is always positive and can be $+\infty$; the number $\mu(A)$ is called the measure of A; we also write μA for it.

A *measure space* is a triplet $(E, \mathcal{E}, \mu)$, where $(E, \mathcal{E})$ is a measurable space and μ is a measure on it.

Examples

3.2 *Dirac measures.* Let $(E, \mathcal{E})$ be a measurable space, and let x be a fixed point of E. For each A in $\mathcal{E}$, put

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Then, δ_x is a measure on $(E, \mathcal{E})$. It is called the *Dirac measure* sitting at x.

3.3 *Counting measures.* Let $(E, \mathcal{E})$ be a measurable space. Let D be a fixed subset of E. For each A in $\mathcal{E}$, let $\nu(A)$ be the number of points in $A \cap D$. Then, ν is a measure on $(E, \mathcal{E})$. Such ν are called *counting measures*. Often, the set D is taken to be countable, in which case

$$\nu(A) = \sum_{x \in D} \delta_x(A), \quad A \in \mathcal{E}.$$

3.4 *Discrete measures.* Let $(E, \mathcal{E})$ be a measurable space. Let D be a countable subset of E. For each x in D, let $m(x)$ be a positive number. Define

$$\mu(A) = \sum_{x \in D} m(x)\, \delta_x(A), \quad A \in \mathcal{E}.$$

Then, μ is a measure on $(E, \mathcal{E})$. Such measures are said to be *discrete*. We may think of $m(x)$ as the mass attached to the point x, and then $\mu(A)$ is the mass on the set A. In particular, if $(E, \mathcal{E})$ is a discrete measurable space, then every measure μ on it has this form.

3.5 *Lebesgue measures.* A measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is called the Lebesgue measure on $\mathbb{R}$ if $\mu(A)$ is the length of A for every interval A. As with most measures, it is impossible to display $\mu(A)$ for every Borel set A, but one can do integration with it, which is the main thing measures are for. Similarly, the Lebesgue measure on $\mathbb{R}^2$ is the "area" measure, on $\mathbb{R}^3$ the "volume", etc. We shall write *Leb* for them. Also note the harmless vice of saying, for example, Lebesgue measure on $\mathbb{R}^2$ to mean Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

Some properties

3.6 Proposition. *Let μ be a measure on a measurable space $(E, \mathcal{E})$. Then, the following hold for all measurable sets A, B, and $A_1, A_2, \ldots$:*

Finite additivity: $A \cap B = \emptyset \;\Rightarrow\; \mu(A \cup B) = \mu(A) + \mu(B)$.

Monotonicity: $A \subset B \;\Rightarrow\; \mu(A) \le \mu(B)$.

Sequential continuity: $A_n \nearrow A \;\Rightarrow\; \mu(A_n) \nearrow \mu(A)$.

Boole's inequality: $\mu(\bigcup_n A_n) \le \sum_n \mu(A_n)$.

Proof. Finite additivity is a particular instance of countable additivity of μ: take $A_1 = A$, $A_2 = B$, $A_3 = A_4 = \ldots = \emptyset$ in 3.1b. Monotonicity follows from finite additivity and the positivity of μ: for $A \subset B$, we can write B as the union of disjoint sets A and $B \setminus A$, and hence

$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A),$$

since $\mu(B \setminus A) \ge 0$. Sequential continuity follows from countable additivity: Suppose that $A_n \nearrow A$. Then, $B_1 = A_1$, $B_2 = A_2 \backslash A_1$, $B_3 = A_3 \backslash A_2, \ldots$ are disjoint, their union is A, and the union of the first n is A_n. Thus, the sequence of numbers $\mu(A_n)$ increases and

$$\lim \mu(A_n) = \lim \mu(\cup_1^n B_i) = \lim \sum_1^n \mu(B_i) = \sum_1^\infty \mu(B_i) = \mu(A).$$

Finally, to show Boole's inequality, we start by observing that

$$\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \;\le\; \mu(A) + \mu(B)$$

for arbitrary A and B in $\mathcal{E}$. This extends to finite unions by induction:

$$\mu(\cup_1^n A_i) \;\le\; \sum_1^n \mu(A_i).$$

Taking limits on both sides completes the proof since the left side has limit $\mu(\cup_1^\infty A_i)$ by sequential continuity. □

Arithmetic of measures

Let $(E, \mathcal{E})$ be a measurable space. If μ is a measure on it and $c > 0$ is a constant, then $c\mu$ is again a measure on it. If μ and ν are measures on it, then so is $\mu + \nu$. If $\mu_1, \mu_2, \ldots$ are measures, then so is $\sum_n \mu_n$; this can be checked using the elementary fact that, if the numbers a_{mn} are positive,

$$\sum_m \sum_n a_{mn} = \sum_n \sum_m a_{mn}.$$

Finite, σ-finite, Σ-finite measures

Let μ be a measure on a measurable space $(E, \mathcal{E})$. It is said to be *finite* if $\mu(E) < \infty$; then $\mu(A) < \infty$ for all A in $\mathcal{E}$ by the monotonicity of μ. It is called a *probability measure* if $\mu(E) = 1$. It is said to be σ*-finite* if there exists a measurable partition (E_n) of E such that $\mu(E_n) < \infty$ for each n. Finally, it is said to be Σ-finite if there exists a sequence of finite measures μ_n such that $\mu = \sum_n \mu_n$. Every finite measure is obviously σ-finite. Every σ-finite measure is Σ-finite; see Exercise 3.13 for this point and for examples.

Specification of measures

Given a measure on $(E, \mathcal{E})$, its values over a p-system generating $\mathcal{E}$ determine its values over all of $\mathcal{E}$, generally. The following is the precise statement for finite measures. Its version for σ-finite measures is given in Exercise 3.18.

3.7 PROPOSITION. *Let $(E, \mathcal{E})$ be a measurable space. Let μ and ν be measures on it with $\mu(E) = \nu(E) < \infty$. If μ and ν agree on a p-system generating $\mathcal{E}$, then μ and ν are identical.*

Proof. Let $\mathcal{C}$ be a p-system generating $\mathcal{E}$. Suppose that $\mu(A) = \nu(A)$ for every A in $\mathcal{C}$, and $\mu(E) = \nu(E) < \infty$. We need to show that, then, $\mu(A) = \nu(A)$ for every A in $\mathcal{E}$, or equivalently, that

$$\mathcal{D} = \{A \in \mathcal{E} : \ \mu(A) = \nu(A)\}$$

contains $\mathcal{E}$. Since $\mathcal{D} \supset \mathcal{C}$ by assumption, it is enough to show that $\mathcal{D}$ is a d-system, for, then, the monotone class theorem 1.8 yields the desired conclusion that $\mathcal{D} \supset \mathcal{E}$. So, we check the conditions for $\mathcal{D}$ to be a d-system. First, $E \in \mathcal{D}$ by the assumption that $\mu(E) = \nu(E)$. If $A, B \in \mathcal{D}$, and $A \supset B$, then $A \setminus B \in \mathcal{D}$, because

$$\mu(A \setminus B) = \mu(A) - \mu(B) = \nu(A) - \nu(B) = \nu(A \setminus B),$$

where we used the finiteness of μ to solve $\mu(A) = \mu(B) + \mu(A \setminus B)$ for $\mu(A \setminus B)$ and similarly for $\nu(A \setminus B)$. Finally, suppose that $(A_n) \subset \mathcal{D}$ and $A_n \nearrow A$; then, $\mu(A_n) = \nu(A_n)$ for every n, the left side increases to $\mu(A)$ by the sequential continuity of μ, and the right side to $\nu(A)$ by the same for ν; hence, $\mu(A) = \nu(A)$ and $A \in \mathcal{D}$. □

3.8 COROLLARY. *Let μ and ν be probability measures on $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$. Then, $\mu = \nu$ if and only if $\mu[-\infty, r] = \nu[-\infty, r]$ for every r in $\mathbb{R}$.*

Proof is immediate from the preceding proposition: $\mu(\bar{\mathbb{R}}) = \nu(\bar{\mathbb{R}}) = 1$ since μ and ν are probability measures, and the intervals $[-\infty, r]$ with r in $\mathbb{R}$ form a p-system generating the Borel σ-algebra on $\bar{\mathbb{R}}$.

Atoms, purely atomic measures, diffuse measures

Let $(E, \mathcal{E})$ be a measurable space. Suppose that the singleton $\{x\}$ belongs to $\mathcal{E}$ for every x in E; this is true for all standard measurable spaces. Let μ be a measure on $(E, \mathcal{E})$. A point x is said to be an *atom* of μ if $\mu\{x\} > 0$. The measure μ is said to be *diffuse* if it has no atoms. It is said to be *purely atomic* if the set D of its atoms is countable and $\mu(E \setminus D) = 0$. For example, Lebesgue measures are diffuse, a Dirac measure is purely atomic with one atom, discrete measures are purely atomic.

The following proposition applies to Σ-finite (and therefore, to finite and σ-finite) measures. We leave the proof as an exercise; see 3.15.

3.9 PROPOSITION. *Let μ be a Σ-finite measure on $(E, \mathcal{E})$. Then,*

$$\mu = \lambda + \nu,$$

where λ is a diffuse measure and ν is purely atomic.

Completeness, negligible sets

Let $(E, \mathcal{E}, \mu)$ be a measure space. A measurable set B is said to be *negligible* if $\mu(B) = 0$. An arbitrary subset of E is said to be *negligible* if it is contained in a measurable negligible set. The measure space is said to be *complete* if every negligible set is measurable. If it is not complete, the following shows how to enlarge $\mathcal{E}$ to include all negligible sets and to extend μ onto the enlarged $\mathcal{E}$. We leave the proof to Exercise 3.16. The measure space $(E, \bar{\mathcal{E}}, \bar{\mu})$ described is called the *completion* of $(E, \mathcal{E}, \mu)$. When $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$ and $\mu = Leb$, the elements of $\bar{\mathcal{E}}$ are called the *Lebesgue measurable* sets.

3.10 PROPOSITION. *Let $\mathcal{N}$ be the collection of all negligible subsets of E. Let $\bar{\mathcal{E}}$ be the σ-algebra generated by $\mathcal{E} \cup \mathcal{N}$. Then,*

a) every B in $\bar{\mathcal{E}}$ has the form $B = A \cup N$, where $A \in \mathcal{E}$ and $N \in \mathcal{N}$,

b) the formula $\bar{\mu}(A \cup N) = \mu(A)$ defines a unique measure $\bar{\mu}$ on $\bar{\mathcal{E}}$, we have $\bar{\mu}(A) = \mu(A)$ for $A \in \mathcal{E}$, and the measure space $(E, \bar{\mathcal{E}}, \bar{\mu})$ is complete.

Almost everywhere

If a proposition holds for all but a negligible set of x in E, then we say that it holds for almost every x, or almost everywhere. If the measure μ used to define negligibility needs to be indicated, we say μ-almost every x or μ-almost everywhere. If E is replaced by a measurable set A, we say almost everywhere on A. For example, given numerical functions f and g on E, and a measurable set A, saying that $f = g$ almost everywhere on A is equivalent to saying that $\{x \in A : f(x) \neq g(x)\}$ is negligible, which is then equivalent to saying that there exists a measurable set M with $\mu(M) = 0$ such that $f(x) = g(x)$ for every x in $A \setminus M$.

Exercises and complements

3.11 *Restrictions and traces.* Let $(E, \mathcal{E})$ be a measurable space, and μ a measure on it. Let $D \in \mathcal{E}$.

a) Define $\nu(A) = \mu(A \cap D)$, $A \in \mathcal{E}$. Show that ν is a measure on $(E, \mathcal{E})$; it is called the trace of μ on D.

b) Let $\mathcal{D}$ be the trace of $\mathcal{E}$ on D (see 1.15). Define $\nu(A) = \mu(A)$ for A in $\mathcal{D}$. Show that ν is a measure on $(D, \mathcal{D})$; it is called the restriction of μ to D.

3.12 *Extensions.* Let $(E, \mathcal{E})$ be a measurable space, let $D \in \mathcal{E}$, and let $(D, \mathcal{D})$ be the trace of $(E, \mathcal{E})$ on D. Let μ be a measure on $(D, \mathcal{D})$ and define ν by

$$\nu(A) = \mu(A \cap D), \quad A \in \mathcal{E}.$$

Show that ν is a measure on $(E, \mathcal{E})$. This device allows us to regard a "measure on D" as a "measure on E".

3.13 *σ-and Σ-finiteness*

a) Let $(E, \mathcal{E})$ be a measurable space. Let μ be a σ-finite measure on it. Then, μ is Σ-finite. Show. Hint: Let (E_n) be a measurable partition of E such that $\mu(E_n) < \infty$ for each n; define μ_n to be the trace of μ on E_n as in Exercise 3.11a; show that $\mu = \sum_n \mu_n$.

b) Show that the Lebesgue measure on $\mathbb{R}$ is σ-finite.

c) Let μ be the discrete measure of Example 3.4 with $(E, \mathcal{E})$ discrete. Show that it is σ-finite if and only if $m(x) < \infty$ for every x in D. Show that it is always Σ-finite.

d) Let $E = [0, 1]$ and $\mathcal{E} = \mathcal{B}(E)$. For A in $\mathcal{E}$, define $\mu(A)$ to be 0 if $\text{Leb}\, A = 0$ and $+\infty$ if $\text{Leb}\, A > 0$. Show that μ is not σ-finite but is Σ-finite.

e) Let $(E, \mathcal{E})$ be as in (d) here. Define $\mu(A)$ to be the counting measure on it (see Example 3.3 and take $D = E$). Show that μ is neither σ-finite nor Σ-finite.

3.14 *Atoms.* Show that a finite measure has at most countably many atoms. Show that the same is true for Σ-finite measures. Hint: If $\mu(E) < \infty$ then the number of atoms with $\mu\{x\} > \frac{1}{n}$ is at most $n\mu(E)$.

3.15 *Proof of Proposition 3.9.* Let D be the set of all atoms of the given Σ-finite measure μ. Then, D is countable by the preceding exercise and, thus, measurable by the measurability of singletons. Define

$$\lambda(A) = \mu(A \setminus D), \qquad \nu(A) = \mu(A \cap D), \qquad A \in \mathcal{E}.$$

Show that λ is a diffuse measure, ν purely atomic, and $\mu = \lambda + \nu$. Note that ν has the form in Example 3.4 with $m(x) = \mu\{x\}$ for each atom x.

3.16 *Proof of Proposition 3.10.* Let $\mathcal{F}$ be the collection of all sets having the form $A \cup N$ with A in $\mathcal{E}$ and N in $\mathcal{N}$. Show that $\mathcal{F}$ is a σ-algebra on E. Argue that $\mathcal{F} = \bar{\mathcal{E}}$, thus proving part (a). To show (b), we need to show that, if $A \cup N = A' \cup N'$ with A and A' in $\mathcal{E}$ and N and N' in $\mathcal{N}$, then $\mu(A) = \mu(A')$. To this end pick M in $\mathcal{E}$ such that $\mu(M) = 0$ and $M \supset N$, and pick M' similarly for N'. Show that $A \subset A' \cup M'$ and $A' \subset A \cup M$. Use this, monotonicity of μ, Boole's inequality, etc. several times to show that $\mu(A) = \mu(A')$.

3.17 *Measurability on completions.* Let $(E, \mathcal{E}, \mu)$ be a measure space, and $(E, \bar{\mathcal{E}}, \bar{\mu})$ its completion. Let f be a numerical function on E. Show that f is $\bar{\mathcal{E}}$-measurable if and only if there exists an $\mathcal{E}$-measurable function g such that $f = g$ almost everywhere. Hint: For sufficiency, choose M in $\mathcal{E}$ such that $\mu(M) = 0$ and $f = g$ outside M, and note that $\{f \le r\} = A \cup N$ where $A = \{g \le r\} \setminus M$ and $N \subset M$. For necessity, assuming f is positive $\bar{\mathcal{E}}$-measurable, write $f = \sum_1^\infty a_n 1_{A_n}$ with $A_n \in \bar{\mathcal{E}}$ for each n (see Exercise 2.27) and choosing B_n in $\mathcal{E}$ such that $A_n = B_n \cup N_n$ for some negligible N_n, define $g = \sum_1^\infty a_n 1_{B_n}$, and show that $\{f \ne g\} \subset \bigcup_n N_n = N$, which is negligible.

3.18 *Equality of measures.* This is to extend Proposition 3.7 to σ-finite measures. Let μ and ν be such measures on $(E, \mathcal{E})$. Suppose that they agree on a p-system $\mathcal{C}$ that generates $\mathcal{E}$. Suppose further that $\mathcal{C}$ contains a partition (E_n) of E such that $\mu(E_n) = \nu(E_n) < \infty$ for each n. Then, $\mu = \nu$. Prove this.

3.19 *Existence of probability measures.* Let E be a set, $\mathcal{D}$ an algebra on it, and put $\mathcal{E} = \sigma\mathcal{D}$. Suppose that $\lambda : \mathcal{D} \mapsto [0, 1]$ is such that $\lambda(E) = 1$ and $\lambda(A \cup B) = \lambda(A) + \lambda(B)$ whenever A and B are disjoint sets in $\mathcal{D}$. Is it possible to extend λ to a probability measure on $\mathcal{E}$? In other words, does there exist a measure μ on $(E, \mathcal{E})$ such that $\mu(A) = \lambda(A)$ for every A in $\mathcal{D}$? If such a measure exists, then it is unique by Proposition 3.7, since $\mathcal{D}$ is a p-system that generates $\mathcal{E}$.
The answer is provided by *Caratheodory's extension theorem*, a classical result. Such a probability measure μ exists provided that λ be countably additive on $\mathcal{D}$, that is, if (A_n) is a disjointed sequence in $\mathcal{D}$ with $A = \bigcup_n A_n \in \mathcal{D}$, then we must have $\lambda(A) = \sum_n \lambda(A_n)$, or equivalently, if $(A_n) \subset \mathcal{D}$ and $A_n \searrow \emptyset$ then we must have $\lambda(A_n) \searrow 0$.

4 Integration

Let $(E, \mathcal{E}, \mu)$ be a measure space. Recall that $\mathcal{E}$ stands also for the collection of all $\mathcal{E}$-measurable functions on E and that $\mathcal{E}_+$ is the sub-collection consisting of positive $\mathcal{E}$-measurable functions. Our aim is to define the "integral of f with respect to μ" for all reasonable functions f in $\mathcal{E}$. We shall denote it by any of the following:

4.1 $$\mu f = \mu(f) = \int_E \mu(dx) f(x) = \int_E f \, d\mu.$$

As the notation μf suggests, integration is a kind of multiplication; this will become clear when we show that the following hold for all a, b in $\mathbb{R}_+$ and f, g, f_n in $\mathcal{E}_+$:

4.2 a) *Positivity*: $\mu f \geq 0$; $\mu f = 0$ if $f = 0$.
b) *Linearity*: $\mu(af + bg) = a\,\mu f + b\,\mu g$.
c) *Monotone convergence theorem*: If $f_n \nearrow f$, then $\mu f_n \nearrow \mu f$.

We start with the definition of the integral and proceed to proving the properties 4.2 and their extensions. At the end, we shall also show that 4.2 characterizes integration.

4.3 Definition. *a) Let f be simple and positive. If its canonical form is $f = \sum_1^n a_i 1_{A_i}$, then we define*

$$\mu f = \sum_1^n a_i\, \mu(A_i).$$

b) Let $f \in \mathcal{E}_+$. Put $f_n = d_n \circ f$, where the d_n are as in Lemma 2.16. Then each f_n is simple and positive, and the sequence (f_n) increases to f as shown in the proof of 2.17. The integral μf_n is defined for each n by the preceding step, and the sequence of numbers μf_n is increasing (see Remark 4.4d below). We define

$$\mu f = \lim \mu f_n.$$

c) Let $f \in \mathcal{E}$. Then, $f^+ = f \vee 0$ and $f^- = -(f \wedge 0)$ belong to $\mathcal{E}_+$, and their integrals $\mu(f^+)$ and $\mu(f^-)$ are defined by the preceding step. Noting that $f = f^+ - f^-$, we define

$$\mu f = \mu(f^+) - \mu(f^-)$$

provided that at least one term on the right side be finite. Otherwise, if $\mu(f^+) = \mu(f^-) = +\infty$, then μf is undefined.

4.4 Remarks. Let f, g, etc. be simple and positive.

a) The formula for μf remains the same even when $f = \sum a_i 1_{A_i}$ is not the canonical representation of f. This is easy to check using the finite additivity of μ.

b) If a and b are in $\mathbb{R}_+$, then $af + bg$ is simple and positive, and the linearity property holds:

$$\mu(af + bg) = a\,\mu f + b\,\mu g.$$

This can be checked using the preceding remark.

c) If $f \leq g$ then $\mu f \leq \mu g$. This follows from the linearity property above applied to the simple positive functions f and $g - f$:

$$\mu f \leq \mu f + \mu(g - f) = \mu(f + g - f) = \mu g.$$

d) In step (b) of the definition, we have $f_1 \leq f_2 \leq \ldots$. The preceding remark on monotonicity shows that $\mu f_1 \leq \mu f_2 \leq \ldots$. Thus, $\lim \mu f_n$ exists as claimed (it can be $+\infty$).

Examples

a) *Discrete measures.* Fix x_0 in E and consider the Dirac measure δ_{x_0} sitting at x_0. Going through the steps of the definition of the integral, we see that $\delta_{x_0} f = f(x_0)$ for every f in $\mathcal{E}$. This extends to discrete measures: if $\mu = \sum_{x \in D} m(x)\delta_x$ for some countable set D and positive masses $m(x)$, then

$$\mu f = \sum_{x \in D} m(x)\ f(x)$$

for every f in $\mathcal{E}_+$. A similar result holds for purely atomic measures as well.

b) *Discrete spaces.* Suppose that $(E, \mathcal{E})$ is discrete, that is, E is countable and $\mathcal{E} = 2^E$. Then, every numerical function on E is $\mathcal{E}$-measurable, and every measure μ has the form in the preceding example with $D = E$ and $m(x) = \mu\{x\}$. Thus, for every positive function f on E,

$$\mu f = \sum_{x \in E} \mu\{x\} f(x).$$

In this case, and especially when E is finite, every function can be thought as a vector, and similarly for every measure. Further, we think of functions as column vectors and of measures as row vectors. Then, the integral μf is seen to be the product of the row vector μ and the column vector f. So, the notation is well-chosen in this case and extends to arbitrary spaces in a most suggestive manner.

c) *Lebesgue integrals.* Suppose that E is a Borel subset of $\mathbb{R}^d$ for some $d \geq 1$ and suppose that $\mathcal{E} = \mathcal{B}(E)$, the Borel subsets of E. Suppose that μ is the restriction of the Lebesgue measure on $\mathbb{R}^d$ to $(E, \mathcal{E})$. For f in $\mathcal{E}$, we employ the following notations for the integral μf:

$$\mu f = \text{Leb}_E\ f = \int_E \text{Leb}(dx)\ f(x) = \int_E dx\ f(x),$$

the last using dx for $\text{Leb}(dx)$ in keeping with tradition. This integral is called the Lebesgue integral of f on E.

If the Riemann integral of f exists, then so does the Lebesgue integral, and the two integrals are equal. The converse is false; the Lebesgue integral exists for a larger class of functions than does the Riemann integral. For example, if $E = [0, 1]$, and f is the indicator of the set of all rational numbers in E, then the Lebesgue integral of f is well-defined by 4.3a to be zero, but the Riemann integral does not exist because the discontinuity set of f in E is E itself and $\text{Leb}\, E = 1 \neq 0$ (recall that a Borel function is Riemann integrable over an interval $[a, b]$ if and only if its points of discontinuity in $[a, b]$ form a set of Lebesgue measure 0).

Integrability

A function f in $\mathcal{E}$ is said to be *integrable* if μf exists and is a real number. Thus, f in $\mathcal{E}$ is integrable if and only if $\mu f^+ < \infty$ and $\mu f^- < \infty$, or equivalently, if and only if the integral of $|f| = f^+ + f^-$ is a finite number. We leave it as an exercise to show that every integrable function is real-valued almost everywhere.

Integral over a set

Let $f \in \mathcal{E}$ and let A be a measurable set. Then, $f1_A \in \mathcal{E}$, and the *integral of f over A* is defined to be the integral of $f1_A$. The following notations are used for it:

4.5 $$\mu(f1_A) = \int_A \mu(dx)f(x) = \int_A f\,d\mu.$$

The following shows that, for each f in $\mathcal{E}_+$, the set function $A \mapsto \mu(f1_A)$ is finitely additive. This property extends to countable additivity as a corollary to the monotone convergence theorem 4.8 below.

4.6 LEMMA. *Let $f \in \mathcal{E}_+$. Let A and B be disjoint sets in $\mathcal{E}$ with union C. Then*

$$\mu(f1_A) + \mu(f1_B) = \mu(f1_C).$$

Proof. If f is simple, this is immediate from the linearity property of Remark 4.4b. For arbitrary f in $\mathcal{E}_+$, putting $f_n = d_n \circ f$ as in Definition 4.3b, we get

$$\mu(f_n1_A) + \mu(f_n1_B) = \mu(f_n1_C)$$

since the f_n are simple. Observing that $f_n1_D = d_n \circ (f1_D)$ for $D = A, B, C$ and taking limits as $n \to \infty$ we get the desired result through Definition 4.3b. □

Positivity and monotonicity

4.7 PROPOSITION. *If $f \in \mathcal{E}_+$, then $\mu f \geq 0$. If f and g are in $\mathcal{E}_+$ and $f \leq g$, then $\mu f \leq \mu g$.*

Proof. Positivity of μf for f positive is immediate from Definition 4.3. To show monotonicity, let $f_n = d_n \circ f$ and $g_n = d_n \circ g$ as in step 4.3b. Since each d_n is an increasing function (see Lemma 2.16), $f \leq g$ implies that $f_n \leq g_n$ for each n which in turn implies that $\mu f_n \leq \mu g_n$ for each n by Remark 4.4c. Letting $n \to \infty$, we see from Definition 4.3b that $\mu f \leq \mu g$. □

Monotone Convergence Theorem

This is the main theorem of integration. It is the key tool for interchanging the order of taking limits and integrals. It states that the mapping $f \mapsto \mu f$ from $\mathcal{E}_+$ into $\bar{\mathbb{R}}_+$ is continuous under increasing limits. As such, it is an extension of the sequential continuity of measures.

4.8 THEOREM. *Let (f_n) be an increasing sequence in $\mathcal{E}_+$. Then,*

$$\mu(\lim f_n) = \lim \mu f_n.$$

Proof. Let $f = \lim f_n$; it is well-defined since (f_n) is increasing. Clearly, $f \in \mathcal{E}_+$, and μf is well-defined. Since (f_n) is increasing, the integrals μf_n form an increasing sequence of numbers by the monotonicity property shown by Proposition 4.7. Hence, $\lim \mu f_n$ exists. We want to show that the limit is μf. Since $f \geq f_n$ for each n, we have $\mu f \geq \mu f_n$ by the monotonicity property. It follows that $\mu f \geq \lim \mu f_n$. The following steps show that the reverse inequality holds as well.

a) Fix b in $\mathbb{R}_+$ and B in $\mathcal{E}$. Suppose that $f(x) > b$ for every x in the set B. Since the sets $\{f_n > b\}$ are increasing to $\{f > b\}$, the sets $B_n = B \cap \{f_n > b\}$ are increasing to B, and

4.9 $$\lim \mu(B_n) = \mu(B)$$

by the sequential continuity of μ. On the other hand,

$$f_n 1_B \geq f_n 1_{B_n} \geq b 1_{B_n},$$

which yields via monotonicity that

$$\mu(f_n 1_B) \geq \mu(b 1_{B_n}) = b\mu(B_n).$$

Taking note of 4.9 we conclude that

4.10 $$\lim \mu(f_n 1_B) \geq b\mu(B).$$

This remains true if $f(x) \geq b$ for all x in B: If $b = 0$ then this is trivially true. If $b > 0$ then choose a sequence (b_m) strictly increasing to b; then, 4.10 holds with b replaced by b_m; and letting $m \to \infty$ we obtain 4.10 again.

b) Let g be a positive simple function such that $f \geq g$. If $g = \sum_1^m b_i 1_{B_i}$ is its canonical representation, then $f(x) \geq b_i$ for every x in B_i, and 4.10 yields

$$\lim_n \mu(f_n 1_{B_i}) \geq b_i \mu(B_i), \qquad i = 1, \ldots, m.$$

Hence, by the finite additivity of $A \mapsto \mu(f_n 1_A)$ shown in Lemma 4.6,

4.11 $$\lim_n \mu f_n = \lim_n \sum_{i=1}^m \mu(f_n 1_{B_i}) = \sum_{i=1}^m \lim_n \mu(f_n 1_{B_i}) \geq \sum_{i=1}^m b_i \mu(B_i) = \mu g.$$

c) Recall that $\mu f = \lim \mu(d_k \circ f)$ by Definition 4.3b. For each k, the function $d_k \circ f$ is simple and $f \geq d_k \circ f$. Hence, taking $g = d_k \circ f$ in 4.11, we have

$$\lim_n \mu f_n \geq \mu(d_k \circ f)$$

for all k. Letting $k \to \infty$ we obtain the desired inequality that $\lim \mu f_n \geq \mu f$. □

Linearity of integration

4.12 Proposition. *For f and g in $\mathcal{E}_+$ and a and b in $\mathbb{R}_+$,*

$$\mu(af + bg) = a\,\mu f + b\,\mu g.$$

The same is true for integrable f and g in $\mathcal{E}$ and arbitrary a and b in $\mathbb{R}$.

Proof. Suppose that f, g, a, b are all positive. If f and g are simple, the linearity can be checked directly as remarked in 4.4b. If not, choose (f_n) and (g_n) to be sequences of simple positive functions increasing to f and g respectively. Then,

$$\mu(af_n + bg_n) = a\,\mu f_n + b\,\mu g_n,$$

and the monotone convergence theorem applied to both sides completes the proof. The remaining statements follow from Definition 4.3c and the linearity for positive functions after putting $f = f^+ - f^-$ and $g = g^+ - g^-$. □

Insensitivity of the integral

We show next that the integral of a function remains unchanged if the values of the function are changed over a negligible set.

4.13 Proposition. *If A in $\mathcal{E}$ is negligible, then $\mu(f1_A) = 0$ for every f in $\mathcal{E}$. If f and g are in $\mathcal{E}_+$ and $f = g$ almost everywhere, then $\mu f = \mu g$. If $f \in \mathcal{E}_+$ and $\mu f = 0$, then $f = 0$ almost everywhere.*

Proof. a) Let A be measurable and negligible. If $f \in \mathcal{E}_+$ and simple, then $\mu(f1_A) = 0$ by Definition 4.3a. This extends to the non-simple case by the monotone convergence theorem using a sequence of simple f_n increasing to f: then $\mu(f_n 1_A) = 0$ for all n and $\mu(f1_A)$ is the limit of the left side. For f in $\mathcal{E}$ arbitrary, we have $\mu(f^+1_A) = 0$ and $\mu(f^-1_A) = 0$ and hence $\mu(f1_A) = 0$ since $(f1_A)^+ = f^+1_A$ and $(f1_A)^- = f^-1_A$.

b) If f and g are in $\mathcal{E}_+$ and $f = g$ almost everywhere, then $A = \{f \neq g\}$ is measurable and negligible, and the integrals of f and g on A both vanish. Thus, with $B = A^c$, we have $\mu f = \mu(f1_B)$ and $\mu g = \mu(g1_B)$, which imply $\mu f = \mu g$ since $f(x) = g(x)$ for all x in B.

c) Let $f \in \mathcal{E}_+$ and $\mu f = 0$. We need to show that the set $N = \{f > 0\}$ has measure 0. Take a sequence of numbers $\varepsilon_k > 0$ decreasing to 0, let $N_k = \{f > \varepsilon_k\}$, and observe that $N_k \nearrow N$, which implies that $\mu(N_k) \nearrow \mu(N)$ by the sequential continuity of μ. Thus, it is enough to show that $\mu(N_k) = 0$ for every k. This is easy to show: $f \geq \varepsilon_k 1_{N_k}$ implies that $\mu f \geq \varepsilon_k \mu(N_k)$, and since $\mu f = 0$ and $\varepsilon_k > 0$, we must have $\mu(N_k) = 0$. □

Fatou's lemma

We return to the properties of the integral under limits. Next is a useful consequence of the monotone convergence theorem.

4.14 LEMMA. *Let* $(f_n) \subset \mathcal{E}_+$. *Then* $\mu(\liminf f_n) \le \liminf \mu f_n$.

Proof. Define $g_m = \inf_{n \ge m} f_n$ and recall that $\liminf f_n$ is the limit of the increasing sequence (g_m) in $\mathcal{E}_+$. Hence, by the monotone convergence theorem,

$$\mu(\liminf f_n) = \lim \mu g_m.$$

On the other hand, $g_m \le f_n$ for all $n \ge m$, which implies that $\mu g_m \le \mu f_n$ for all $n \ge m$ by the monotonicity of integration, which in turn means that $\mu g_m \le \inf_{n \ge m} \mu f_n$. Hence, as desired,

$$\lim \mu g_m \le \liminf \mu f_n. \qquad \square$$

4.15 COROLLARY. *Let* $(f_n) \subset \mathcal{E}$. *If there is an integrable function* g *such that* $f_n \ge g$ *for every* n, *then*

$$\mu(\liminf f_n) \le \liminf \mu f_n.$$

If there is an integrable function g *such that* $f_n \le g$ *for every* n, *then*

$$\mu(\limsup f_n) \ge \limsup \mu f_n.$$

Proof. Let g be integrable. Then, the complement of the measurable set $A = \{g \in \mathbb{R}\}$ is negligible (see Exercise 4.24 for this). Hence, $f_n 1_A = f_n$ almost everywhere, $g1_A = g$ almost everywhere, and $g1_A$ is real-valued. The first statement follows from Fatou's Lemma applied to the well-defined sequence $(f_n 1_A - g1_A)$ in $\mathcal{E}_+$ together with the linearity and insensitivity of integration. The second statement follows again from Fatou's lemma, now applied to the well-defined sequence $(g1_A - f_n 1_A)$ in $\mathcal{E}_+$ together with the linearity and insensitivity, and the observation that $\limsup r_n = -\liminf(-r_n)$ for every sequence (r_n) in $\bar{\mathbb{R}}$. $\square$

Dominated convergence theorem

This is the second important tool for interchanging the order of taking limits and integrals. A function f is said to be *dominated* by the function g if $|f| \le g$; note that $g \ge 0$ necessarily. A sequence (f_n) is said to be dominated by g if $|f_n| \le g$ for every n. If so, and if g can be taken to be a finite constant, then (f_n) is said to be *bounded*.

4.16 THEOREM. *Let* $(f_n) \subset \mathcal{E}$. *Suppose that* (f_n) *is dominated by some integrable function* g. *If* $\lim f_n$ *exists, then it is integrable and*

$$\mu(\lim f_n) = \lim \mu f_n.$$

Proof. By assumption, $-g \le f_n \le g$ for every n, and both g and $-g$ are integrable. Thus, both statements of the last corollary apply:

4.17 $$\mu(\liminf f_n) \le \liminf \mu f_n \le \limsup \mu f_n \le \mu(\limsup f_n).$$

If $\lim f_n$ exists, then $\liminf f_n = \limsup f_n = \lim f_n$, and $\lim f_n$ is integrable since it is dominated by g. Hence, the extreme members of 4.17 are finite and equal, and all inequality signs are in fact equalities. □

If (f_n) is bounded, say by the constant b, and if the measure μ is finite, then we can take $g = b$ in the preceding theorem. The resulting corollary is called the *bounded convergence theorem*:

4.18 THEOREM. *Let $(f_n) \subset \mathcal{E}$. Suppose that (f_n) is bounded and μ is finite. If $\lim f_n$ exists, then it is a bounded integrable function and*

$$\mu(\lim f_n) = \lim \mu f_n.$$

Almost everywhere versions

The insensitivity of integration to changes over negligible sets enables us to re-state all the results above by allowing the conditions to fail over negligible sets. We start by extending the definition of integration somewhat.

4.19 CONVENTION. *Let f be a numerical function on E. Suppose that there exists an $\mathcal{E}$-measurable function g such that $f(x) = g(x)$ for almost every x in E. Then, we define the integral μf of f to be the number μg provided that μg is defined. Otherwise, if μg does not exist, μf does not exist either.*

The definition here is without ambiguities: if h is another measurable function such that $f = h$ almost everywhere, then $g = h$ almost everywhere; if μg exists, then so does μh and $\mu g = \mu h$ by the insensitivity property; if μg does not exist, then neither does μh.

In fact, the convention here is one of notation making, almost. Let $g \in \mathcal{E}$ and $f = g$ almost everywhere. Let $(E, \bar{\mathcal{E}}, \bar{\mu})$ be the completion of $(E, \mathcal{E}, \mu)$. Then, $f \in \bar{\mathcal{E}}$ (see Exercise 3.17 for this), and the integral $\bar{\mu} f$ makes sense by Definition 4.3 applied on the measurable space $(E, \bar{\mathcal{E}}, \bar{\mu})$. Since $\mathcal{E} \subset \bar{\mathcal{E}}$, the function g is $\bar{\mathcal{E}}$-measurable as well, and $\bar{\mu} g$ makes sense and it is clear that $\bar{\mu} g = \mu g$. Since f and g are $\bar{\mathcal{E}}$-measurable and $f = g$ $\bar{\mu}$-almost everywhere, $\bar{\mu} f = \bar{\mu} g$ by insensitivity. So, the convention above amounts to writing μf instead of $\bar{\mu} f$.

With this convention in place, we now re-state the monotone convergence theorem in full generality.

4.20 THEOREM. *Let (f_n) be a sequence of numerical functions on E. Suppose that, for each n, there is g_n in $\mathcal{E}$ such that $f_n = g_n$ almost everywhere. Further, suppose for each n that $f_n \ge 0$ almost everywhere and $f_n \le f_{n+1}$ almost everywhere. Then, $\lim f_n$ exists almost everywhere, is positive almost everywhere, and $\mu(\lim f_n) = \lim \mu f_n$.*

We discuss this fully to indicate its meaning and the issues involved. Let $\mathcal{N}$ denote the collection of all measurable negligible sets, that is, every N in $\mathcal{N}$ belongs to $\mathcal{E}$ and $\mu(N) = 0$. Now fix n. To say that $f_n = g_n$ almost everywhere is to say that there is N_n in $\mathcal{N}$ such that $f_n = g_n$ outside N_n (that is, $f_n(x) = g_n(x)$ whenever $x \notin N_n$). Similarly, $f_n \geq 0$ almost everywhere means that there is M_n in $\mathcal{N}$ such that $f_n \geq 0$ outside M_n. And, since $f_n \leq f_{n+1}$ almost everywhere, there is L_n in $\mathcal{N}$ such that $f_n \leq f_{n+1}$ outside L_n. These are the conditions. The claim of the theorem is as follows. First, there is an $\mathcal{E}$-measurable function f, and a set N in $\mathcal{N}$ such that $\lim f_n(x)$ exists and is equal to $f(x)$ for every x outside N. Also, there is M in $\mathcal{N}$ such that $f \geq 0$ outside M. Finally, $\mu f = \lim \mu f_n$, where the μf_n are defined by convention 4.19 to be the numbers μg_n.

Proof. Let

$$N = \bigcup_{n=1}^{\infty} (L_n \cup M_n \cup N_n).$$

Then, $N \in \mathcal{E}$ and $\mu(N) = 0$ by Boole's inequality, that is, $N \in \mathcal{N}$.

For x outside N, we have

$$0 \leq f_1(x) = g_1(x) \leq f_2(x) = g_2(x) \leq \ldots,$$

and hence $\lim f_n(x)$ exists and is equal to $\lim g_n(x)$. Define

$$f(x) = \begin{cases} \lim f_n(x) & \text{if } x \notin N \\ 0 & \text{if } x \in N \end{cases}$$

Clearly, f is the limit of the increasing sequence $(g_n 1_{E \setminus N})$ in $\mathcal{E}_+$. So, f is in $\mathcal{E}_+$ and we may take $M = \emptyset$. There remains to show that $\mu f = \lim \mu g_n$. Now in fact

$$\mu f = \mu(\lim g_n 1_{E\setminus N}) = \lim \mu(g_n 1_{E\setminus N}) = \lim \mu g_n,$$

where we used the monotone convergence theorem to justify the second equality, and the insensitivity to justify the third. □

The reader is invited to formulate the "almost everywhere version" of the dominated convergence theorem and to prove it carefully once. We shall use such versions without further ado whenever the need drives us.

Characterization of the integral

Definition 4.3 defines the integral μf for every f in $\mathcal{E}_+$. Thus, in effect, integration extends the domain of μ from the measurable sets (identified with their indicator functions) to the space $\mathcal{E}_+$ of all positive measurable functions (and beyond), and hence we may regard μ as the mapping $f \mapsto \mu f$ from $\mathcal{E}_+$ into $\bar{\mathbb{R}}_+$. The mapping $\mu : \mathcal{E}_+ \mapsto \bar{\mathbb{R}}_+$ is necessarily positive, linear, and continuous under increasing limits; these were promised in 4.2 and proved as Proposition 4.7, Proposition 4.12, and Theorem 4.8. We end this section with the following very useful converse.

4.21 THEOREM. *Let $(E, \mathcal{E})$ be a measurable space. Let L be a mapping from $\mathcal{E}_+$ into $\bar{\mathbb{R}}_+$. Then there exists a unique measure μ on $(E, \mathcal{E})$ such that $L(f) = \mu f$ for every f in $\mathcal{E}_+$ if and only if*

4.22
a) $f = 0 \Rightarrow L(f) = 0$.
b) $f, g \in \mathcal{E}_+$ and $a, b \in \mathbb{R}_+ \Rightarrow L(af + bg) = aL(f) + bL(g)$.
c) $(f_n) \subset \mathcal{E}_+$ and $f_n \nearrow f \Rightarrow L(f_n) \nearrow L(f)$.

Proof. Necessity of the conditions is immediate from the properties of the integral: (a) follows from the definition of μf, (b) from linearity, and (c) from the monotone convergence theorem.

To show the sufficiency, suppose that L has the properties (a)-(c). Define

4.23 $$\mu(A) = L(1_A), \qquad A \in \mathcal{E}.$$

We show that μ is a measure. First, $\mu(\emptyset) = L(1_\emptyset) = L(0) = 0$. Second, if $A_1, A_2, \ldots$ are disjoint sets in $\mathcal{E}$ with union A, then the indicator of $\bigcup_1^n A_i$ is $\sum_1^n 1_{A_i}$, the latter is increasing to 1_A, and hence,

$$\mu(A) = L(1_A) = \lim_n L(\sum_1^n 1_{A_i}) = \lim_n \sum_1^n L(1_{A_i}) = \lim_n \sum_1^n \mu(A_i) = \sum_1^\infty \mu(A_i),$$

where we used the conditions (c) and (b) to justify the second and third equality signs.

So, μ is a measure on $(E, \mathcal{E})$. It is unique by the necessity of 4.23. Now, $L(f) = \mu f$ for simple f in $\mathcal{E}_+$ by the linearity property (b) of L and the linearity of integration. This in turn implies that, for every f in $\mathcal{E}_+$, choosing simple $f_n \nearrow f$,

$$L(f) = \lim L(f_n) = \lim \mu f_n = \mu f$$

by condition (c) and the monotone convergence theorem. □

Exercises and complements

4.24 *Integrability.* If $f \in \mathcal{E}_+$ and $\mu f < \infty$, then f is real-valued almost everywhere. Show this. More generally, if f is integrable then it is real-valued almost everywhere.

4.25 *Test for vanishing.* Let $f \in \mathcal{E}_+$. Then $\mu f = 0$ if and only if $f = 0$ almost everywhere. Prove.

4.26 *Alternative form of the monotone convergence theorem.* If $f_1, f_2, \ldots$ are in $\mathcal{E}_+$ then

$$\mu \sum_1^\infty f_n = \sum_1^\infty \mu f_n.$$

4.27 *Sums of measures.* Recall that if $\mu_1, \mu_2, \ldots$ are measures on $(E, \mathcal{E})$, so is $\mu = \sum \mu_n$. Show that, for every f in $\mathcal{E}_+$,

$$\mu f = \sum_n \mu_n f.$$

4.28 *Absolute values.* Assuming that μf exists, show that $|\mu f| \leq \mu |f|$.

4.29 *Mean value theorem.* If $\mu(A) > 0$ and $a \leq f(x) \leq b$ for every x in A, then show that

$$a \leq \frac{1}{\mu(A)} \int_A f d\mu \ \leq \ b.$$

4.30 *Generalization of the monotone convergence theorem.* If $f_n \geq g$ for all n for some integrable function g, and if (f_n) increases to f, then μf exists and is equal to $\lim \mu f_n$. If $f_n \leq g$ for all n for some integrable function g and if (f_n) decreases to f, then μf exists and is equal to $\lim \mu f_n$.

4.31 *On dominated convergence.* In the dominated convergence theorem, the condition that (f_n) be dominated by an integrable g is necessary. Suppose that $E = (0, 1)$, $\mathcal{E} = \mathcal{B}_E$, $\mu = Leb$. Take, for $n = 1, 2, \ldots$,

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Then, $f_n(x) \to 0$ for every x in E, the integral $\mu f_n = 1$ for every n, but $0 = \mu(\lim f_n) \neq \lim \mu f_n = 1$.

4.32 *Test for σ-finiteness.* A measure μ on $(E, \mathcal{E})$ is σ-finite if and only if there exists a strictly positive function f in $\mathcal{E}$ such that $\mu f < \infty$. Prove this. Hint for the sufficiency part: Let $E_n = \{f > \frac{1}{n}\}$ and note that $E_n \nearrow E$ whereas $\frac{1}{n}\mu(E_n) \leq \mu(f 1_{E_n}) \leq \mu f < \infty$.

5 Transforms and Indefinite Integrals

This section is about measures defined from other measures via various means and the relationships among integrals with respect to them.

Image measures

Let $(F, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces. Let ν be a measure on $(F, \mathcal{F})$ and let $h : F \mapsto E$ be measurable relative to $\mathcal{F}$ and $\mathcal{E}$. We define a mapping $\nu \circ h^{-1}$ from the σ-algebra $\mathcal{E}$ into $\bar{\mathbb{R}}_+$ by

5.1 $$\nu \circ h^{-1}(B) = \nu(h^{-1}B), \qquad B \in \mathcal{E},$$

which is well-defined since $h^{-1}B \in \mathcal{F}$ by the measurability of h. It is easy to check that $\nu \circ h^{-1}$ is a measure on $(E, \mathcal{E})$; it is called the *image* of ν under h. Other notations current are $h \circ \nu$, $h(\nu)$, $\nu \circ h$, ν_h.

If ν is finite, then so is its image. If ν is Σ-finite, again, so is its image. But, the image of a σ-finite measure generally fails to be σ-finite (but is Σ-finite).

The following relates integrals with respect to $\nu \circ h^{-1}$ to integrals with respect to ν.

5.2 THEOREM. *For every f in $\mathcal{E}_+$ we have $(\nu \circ h^{-1})f = \nu(f \circ h)$.*

Proof. Define $L : \mathcal{E}_+ \mapsto \bar{\mathbb{R}}_+$ by setting $L(f) = \nu(f \circ h)$. It can be checked that L satisfies the conditions of the integral characterization theorem 4.21. Thus, $L(f) = \mu f$ for some unique measure μ on $(E, \mathcal{E})$. That μ is precisely the measure $\nu \circ h^{-1}$, because

$$\mu(B) = L(1_B) = \nu(1_B \circ h) = \nu(h^{-1}B), \qquad B \in \mathcal{E}. \qquad \square$$

The limitation to positive $\mathcal{E}$-measurable functions can be removed: for arbitrary f in $\mathcal{E}$ the same formula holds provided that the integral on one side be well-defined (and then both sides are well-defined).

The preceding theorem is a generalization of the change of variable formula from calculus. In more explicit notation, with $\mu = \nu \circ h^{-1}$, the theorem is that

5.3 $$\int_F \nu(dx) f(h(x)) = \int_E \mu(dy) f(y),$$

that is, if $h(x)$ is replaced with y then $\nu(dx)$ must be replaced with $\mu(dy)$. In calculus, it is often the case that $E = F = \mathbb{R}^d$ for some fixed dimension d, and μ and ν are expressed in terms of the Lebesgue measure on $\mathbb{R}^d$ and the Jacobian of the transformation h. In probability theory, often, the measure ν is defined implicitly through the formula 5.3 by stating the transformation h and the corresponding image measure μ. We take up still another use next.

Images of the Lebesgue measure

Forming image measures is a convenient method of creating new measures from the old, and if the old measure ν is convenient enough as an integrator, then 5.3 provides a useful formula for the integrals with respect to the new measure μ. In fact, the class of measures that can be represented as images of the Lebesgue measure on $\mathbb{R}_+$ is very large. The following is the precise statement; combined with the preceding theorem it reduces integrals over abstract spaces to integrals on $\mathbb{R}_+$ with respect to the Lebesgue measure.

5.4 THEOREM. *Let $(E, \mathcal{E})$ be a standard measurable space. Let μ be a Σ-finite measure on $(E, \mathcal{E})$ and put $b = \mu(E)$, possibly $+\infty$. Then, there exists a mapping h from $[0, b)$ into E, measurable relative to $\mathcal{B}_{[0,b)}$ and $\mathcal{E}$, such that*

$$\mu = \lambda \circ h^{-1},$$

where λ is the Lebesgue measure on $[0, b)$.

Proof will be sketched in Exercises 5.15 and 5.16 in a constructive fashion.

Indefinite integrals

Let $(E, \mathcal{E}, \mu)$ be a measure space. Let p be a positive $\mathcal{E}$-measurable function. Define

5.5 $$\nu(A) = \mu(p1_A) = \int_A \mu(dx)p(x), \qquad A \in \mathcal{E}.$$

It follows from the monotone convergence theorem (alternative form) that ν is a measure on $(E, \mathcal{E})$. It is called the *indefinite integral* of p with respect to μ.

5.6 PROPOSITION. *For every f in $\mathcal{E}_+$, we have $\nu f = \mu(pf)$.*

Proof. Let $L(f) = \mu(pf)$ and check that L satisfies the conditions of Theorem 4.21. Thus, there exists a unique measure $\hat{\mu}$ on $(E, \mathcal{E})$ such that $L(f) = \hat{\mu} f$ for every f in $\mathcal{E}_+$. We have $\hat{\mu} = \nu$, since

$$\hat{\mu}(A) = L(1_A) = \mu(p1_A) = \nu(A), \qquad A \in \mathcal{E}. \qquad \square$$

The formula 5.5 is another convenient tool for creating new measures from the old. Written in more explicit notation, the preceding proposition becomes

5.7 $$\int_E \nu(dx)\, f(x) = \int_E \mu(dx)\, p(x)\, f(x) \qquad f \in \mathcal{E}_+,$$

which can be expressed informally by writing

5.8 $$\nu(dx) = \mu(dx)\, p(x), \qquad x \in E,$$

once it is understood that μ and ν are measures on $(E, \mathcal{E})$ and that p is positive $\mathcal{E}$-measurable.

Heuristically, we may think of $\mu(dx)$ as the amount of mass put by μ on an "infinitesimal neighborhood" dx of the point x, and similarly of $\nu(dx)$. Then, 5.8 takes on the meaning that $p(x)$ is the mass density, at x, of the measure ν with respect to μ. For this reason, the function p is called the density function of ν relative to μ, and the following notations are used for it:

5.9 $$p = \frac{d\nu}{d\mu}; \qquad p(x) = \frac{\nu(dx)}{\mu(dx)}, \qquad x \in E.$$

The expressions 5.5-5.9 are equivalent ways of saying the same thing: ν is the indefinite integral of p with respect to μ, or p is the density of ν relative to μ.

Radon-Nikodym theorem

Let μ and ν be measures on a measurable space $(E, \mathcal{E})$. Then, ν is said to be *absolutely continuous* with respect to μ if, for every set A in $\mathcal{E}$,

5.10 $$\mu(A) = 0 \;\Rightarrow\; \nu(A) = 0.$$

If ν is the indefinite integral of some positive $\mathcal{E}$-measurable function with respect to μ, then it is evident from 5.5 that ν is absolutely continuous with respect to μ. The following, called the Radon-Nikodym theorem, shows that the converse is true as well, at least when μ is σ-finite. We list it here without proof. We shall give two proofs of it later.

5.11 THEOREM. *Suppose that μ is σ-finite, and ν is absolutely continuous with respect to μ. Then, there exists a positive $\mathcal{E}$-measurable function p such that*

5.12 $$\int_E \nu(dx)\, f(x) = \int_E \mu(dx)\, p(x)\, f(x), \qquad f \in \mathcal{E}_+.$$

Moreover, p is unique up to equivalence: if 5.12 holds for another $\hat{p}$ in $\mathcal{E}_+$, then $\hat{p}(x) = p(x)$ for μ-almost every x in E.

The function p in question can be denoted by $d\nu/d\mu$ in view of the equivalence of 5.5-5.9 and 5.12; and the function p is also called the *Radon-Nikodym derivative* of ν with respect to μ. See Exercises 5.17-5.20 for some remarks.

A matter of style

When an explicit expression is desired for a measure μ, there are several choices. One can go with the definition and give a formula for $\mu(A)$. Equivalently, and usually with greater ease and clarity, one can display a formula for the integral μf for arbitrary f in $\mathcal{E}_+$. In those cases where μ has a density with respect to some well-known measure like the Lebesgue measure, it is better to give the formula for μf or, to be more brief, to give a formula like $\mu(dx) = \lambda(dx)\, p(x)$ by using the form 5.8, with λ denoting the Lebesgue measure. All things considered, if a uniform style is desired, it is best to display an expression for μf. We shall do either that or use the form 5.8 when the form of p is important.

Exercises and complements

5.13 *Time changes.* Let c be an increasing right-continuous function from $\mathbb{R}_+$ into $\bar{\mathbb{R}}_+$. Define

$$a(u) = \inf\{t \in \mathbb{R}_+ :\ c(t) > u\}, \qquad u \in \mathbb{R}_+,$$

with the usual convention that $\inf \emptyset = \infty$.

a) Show that the function $a:\ \mathbb{R}_+ \mapsto \bar{\mathbb{R}}_+$ is increasing and right-continuous, and that

$$c(t) = \inf\{u \in \mathbb{R}_+ :\ a(u) > t\}, \qquad t \in \mathbb{R}_+.$$

Thus, a and c are right-continuous "functional inverses" of each other. See Figure 1 below.

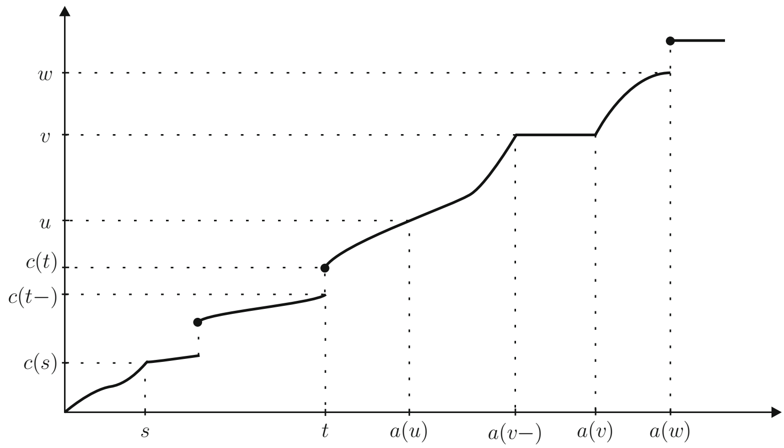

Figure 1: Both c and a are increasing right-continuous. They are functional inverses of each other.

b) Suppose $c(t) < \infty$. Show that $a(c(t)) \geq t$, with equality if and only if $c(t+\varepsilon) > c(t)$ for every $\varepsilon > 0$.

Imagine a clock whose mechanism is so rigged that it points to the number $c(t)$ when the actual time is t. Then, when the clock points to the number u, the actual time is $a(u)$. Hence the term "time change" for the operations involved.

5.14 *Distribution functions and measures on* $\mathbb{R}_+$. Let μ be a measure on $\mathbb{R}_+$ (with its Borel σ-algebra) such that $c(t) = \mu[0,t]$ is finite for every t in $\mathbb{R}_+$. The limit $b = c(\infty) = \lim_{t\to\infty} c(t)$ is allowed to be $+\infty$.

a) Show that c is increasing and right-continuous. It is called the *cumulative distribution function* associated with μ.

b) Define $a(u)$ as in 5.13 for $u \in [0,b)$, let λ denote the Lebesgue measure on $[0,b)$. Show that

$$\mu = \lambda \circ a^{-1} \ .$$

This demonstrates Theorem 5.4 in the case of measures like the present μ. Incidentally, we have also shown that to every increasing right-continuous function c from $\mathbb{R}_+$ into $\mathbb{R}_+$ there corresponds a unique measure μ on $\mathbb{R}_+$ whose cumulative distribution function is c.

5.15 *Representation of measures: Finite case.* Let μ be a finite measure on a standard measurable space $(E, \mathcal{E})$. We aim to prove Theorem 5.4 in this case assuming that $(E, \mathcal{E})$ is isomorphic to $(D, \mathcal{B}_D)$ where $D = [0,1]$. The remaining cases where E is finite or countably infinite are nearly trivial.

The idea is simple: First, use the isomorphism to carry μ from E into a measure $\hat{\mu}$ on D. Second, follow the steps of 5.14 to write $\hat{\mu} = \lambda \circ a^{-1}$ where λ is the Lebesgue measure on $B = [0, b)$ with $b = \hat{\mu}(D) = \mu(E)$. Finally, use the inverse of the isomorphism to carry $\hat{\mu}$ back onto E. Here are the details.
Let $f : E \mapsto D$ be the isomorphism involved. Let $g : D \mapsto E$ be the functional inverse of f, that is, $g(t) = x$ if and only if $f(x) = t$. Define $\hat{\mu} = \mu \circ f^{-1}$; then $\hat{\mu}$ is a measure on D with total mass $\hat{\mu}(D) = \mu(E) = b$. Put $B = [0, b)$, $\mathcal{B} = \mathcal{B}_B$ and λ the Lebesgue measure on B.
Define $c(t) = \hat{\mu}[0, t]$ for t in D. Define $a(u)$ by 5.13 for u in B. Note that $a : B \mapsto D$ is measurable and that $\hat{\mu} = \lambda \circ a^{-1}$. Define $h(u) = g \circ a(u)$ for u in B. Observe that $\lambda \circ h^{-1} = \lambda \circ a^{-1} \circ g^{-1} = \mu$ as needed.

5.16 *Continuation: Σ-finite case.* Let $(E, \mathcal{E})$ be isomorphic to $(D, \mathcal{B}_D)$ where $D = [0, 1]$. Let μ be Σ-finite on $(E, \mathcal{E})$, say $\mu = \sum \mu_n$ with each μ_n finite. Since the case of finite μ is already covered, we assume that $b = \mu(E) = +\infty$. Let $B = [0, b) = \mathbb{R}_+$, $\mathcal{B} = \mathcal{B}(\mathbb{R}_+)$, and λ the Lebesgue measure on $\mathbb{R}_+$. Let $f : E \mapsto D$ and $g : D \mapsto E$ as before.
Let $D_n = 2n + D = [2n, 2n + 1]$, $n = 0, 1, 2, \ldots$; note that $D_0, D_1, \ldots$ are disjoint. Define $f_n : E \mapsto D_n$ by setting $f_n(x) = 2n + f(x)$ and let $g_n : D_n \mapsto E$ be the functional inverse of f_n, that is, $g_n(t) = g(t - 2n)$. Now, $\hat{\mu}_n = \mu_n \circ f_n^{-1}$ is a measure on D_n. Define

$$\hat{\mu}(C) = \sum_0^\infty \hat{\mu}_n(C \cap D_n), \qquad C \in \mathcal{B}(\mathbb{R}_+).$$

This defines a measure $\hat{\mu}$ on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $c(t) = \hat{\mu}[0, t] < \infty$ for every t in $\mathbb{R}_+$, as in Exercise 5.14. Let a be defined as in 5.13, and observe that $\hat{\mu} = \lambda \circ a^{-1}$. Also observe that, by the way a is defined, $a(u)$ belongs to the set $\bigcup_n D_n$ for each u. Finally, put

$$h(u) = g_n \circ a(u) \qquad \text{if } a(u) \in D_n,$$

and show that $\mu = \lambda \circ h^{-1}$ as claimed.

5.17 *Absolute continuity for atomic measures.* Let ν be a Σ-finite purely atomic measure on some measurable space $(E, \mathcal{E})$ such that the singletons $\{x\}$ belong to $\mathcal{E}$ for each x in E. Let D be the collection of all atoms, and recall that D is countable. Let $\mu(A)$ be the number of points in $A \cap D$. Then, ν is absolutely continuous with respect to μ. Find the density $p = d\nu/d\mu$.

5.18 *Radon-Nikodym derivatives.* Let μ be a measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $c(t) = \mu[0, t]$ is finite for every t in $\mathbb{R}_+$. If μ is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}_+$, then the cumulative distribution function c is differentiable at λ-almost every t in $\mathbb{R}_+$ and

$$p(t) = \frac{\mu(dt)}{\lambda(dt)} = \frac{d}{dt} c(t) \qquad \text{for } \lambda\text{−almost every } t.$$

5.19 *Radon-Nikodym and σ-finiteness.* The condition that μ be σ-finite cannot be removed in general. Let ν be the Lebesgue measure on $E = [0,1]$, $\mathcal{E} = \mathcal{B}(E)$, and let $\mu(A)$ be 0 or $+\infty$ according as $\nu(A)$ is 0 or strictly positive, $A \in \mathcal{E}$. Then, μ is Σ-finite, and ν is absolutely continuous with respect to μ. Show that the conclusion of Theorem 5.11 fails in this case.

5.20 *On Σ-finiteness.* Let μ be a Σ-finite measure on an arbitrary measurable space $(E, \mathcal{E})$, say with the decomposition $\mu = \sum \mu_n$, where $\mu_n(E) < \infty$ for each n. Define $\nu(A) = \sum_n \mu_n(A)/2^n \mu_n(E)$, $A \in \mathcal{E}$. Show that ν is a finite measure, and μ is absolutely continuous with respect to ν. Thus, there exists $p \in \mathcal{E}_+$ such that

$$\mu(dx) = \nu(dx)\, p(x)\,, \qquad x \in E.$$

If μ is σ-finite, show that p is real-valued ν-almost everywhere. Show that, in the converse direction, if μ is absolutely continuous with respect to a finite measure ν, then μ is Σ-finite.

5.21 *Singularity.* Let μ and ν be measures on some measurable space $(E, \mathcal{E})$. Then, ν is said to be *singular* with respect to μ if there exists a set D in $\mathcal{E}$ such that

$$\mu(D) = 0 \qquad \text{and} \qquad \nu(E \setminus D) = 0.$$

The notion is the opposite of absolute continuity. Show that, if ν is purely atomic and μ is diffuse then ν is singular with respect to μ. This does not exhaust the possibilities, however, as the famous example next illustrates.

5.22 *Cantor set, Cantor measure.* Start with the interval $E = [0,1]$. Delete the set $D_{0,1} = (\frac{1}{3}, \frac{2}{3})$ which forms the middle third of E; this leaves two closed intervals. Delete the middle thirds of those, that is, delete $D_{1,1} = (\frac{1}{9}, \frac{2}{9})$ and $D_{1,2} = (\frac{7}{9}, \frac{8}{9})$; there remain four closed intervals. Delete the middle thirds of those four intervals, and continue in this fashion. At the end, the deleted intervals form the open set

$$D = \bigcup_{i=0}^{\infty} \bigcup_{j=1}^{2^i} D_{i,j},$$

and the set of points that remain is

$$C = E \setminus D.$$

The closed set C is called the *Cantor set.*

Next we construct a continuous function $c: E \mapsto [0,1]$ that remains constant over each interval $D_{i,j}$ and increases (only) on C. Define $c(t) = \frac{1}{2}$ for t in $D_{0,1}$; let $c(t) = \frac{1}{4}$ for t in $D_{1,1}$ and $c(t) = \frac{3}{4}$ for t in $D_{1,2}$; let $c(t) = \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$ according as t is in $D_{2,1}$, $D_{2,2}$, $D_{2,3}$, $D_{2,4}$; and so on. This defines a uniformly continuous increasing function from D into $[0,1]$. Since D is dense in E, we may extend c onto E by continuity. The resulting function $c: E \mapsto [0,1]$ is called the Cantor function.

a) Show that $\text{Leb}(D) = 1$, $\text{Leb}(C) = 0$.

b) Let ν be the measure on E corresponding to the (cumulative distribution) function c, that is, $\nu = \lambda \circ a^{-1}$ where λ is the Lebesgue measure on $[0, 1)$ and $a : [0, 1) \mapsto E$ is the inverse of c as in 5.13. We call ν the Cantor measure. Show that $\nu(C) = 1$ and $\nu(D) = 0$. Conclude that ν is a diffuse measure on E and that ν is singular with respect to the Lebesgue measure on E.

c) Show that the range of a is $C \setminus C_0$ where C_0 consists of the point 1 and the countable collection of points that are the left-end-points of the intervals $D_{i,j}$. Thus, a is a one-to-one mapping from $[0, 1)$ onto $C \setminus C_0$, and it follows that $C \setminus C_0$ has the power of the continuum. Thus, the Cantor set has the power of the continuum, even though its Lebesgue measure is 0.

d) The Cantor set is everywhere dense in itself, that is, for every t in C there exists $(t_n) \subset C \setminus \{t\}$ such that $t = \lim t_n$. Incidentally, a closed set that is everywhere dense in itself is said to be *perfect.*

5.23 *Lebesgue-Stieltjes integrals.* Let c be an increasing right-continuous function from $\mathbb{R}_+$ into $\mathbb{R}_+$. Let μ be the measure on $\mathbb{R}_+$ that has c as its cumulative distribution function (see Exercise 5.14). For each positive Borel function f on $\mathbb{R}_+$, define

$$\int_{\mathbb{R}_+} f(t)\, dc(t) = \int_{\mathbb{R}_+} \mu(dt)\, f(t).$$

The left side is called the Lebesgue-Stieltjes integral of f with respect to c. Note that, with the notation of 5.14,

$$\int_{\mathbb{R}_+} \mu(dt) f(t) = \int_0^b du f(a(u)).$$

Replacing f by $f 1_A$ one obtains the same integral over the interval A. Extensions to arbitrary Borel functions f on $\mathbb{R}_+$ are as usual for μf, namely, by using the decomposition $f = f^+ - f^-$. Extension from the space $\mathbb{R}_+$ onto $\mathbb{R}$ is obvious. Finally, extensions to functions c that can be decomposed as $c = c_1 - c_2$ with both c_1 and c_2 increasing and right-continuous (see the next exercise) can be done by setting

$$\int_{\mathbb{R}} f(t)\, dc(t) = \int_{\mathbb{R}} f(t)\, dc_1(t) - \int_{\mathbb{R}} f(t)\, dc_2(t)$$

for those f for which the integrals on the right make sense and are not both $+\infty$ or both $-\infty$.

5.24 *Functions of bounded variation.* Let f be a function from $\mathbb{R}_+$ into $\mathbb{R}$. Think of $f(t)$ as the position, at time t, of an insect moving on the line $\mathbb{R}$. We are interested in the total amount of traveling done during a finite interval $(s, t]$. Here is the precise version.

A *subdivision* of $[s,t]$ is a finite collection $\mathcal{A}$ of disjoint intervals of the form $(\ ,\ ]$ whose union is $(s,t]$. We define

$$V_f(s,t) = \sup_{\mathcal{A}} \sum_{(u,v]\in\mathcal{A}} |f(v) - f(u)|$$

where the supremum is over all subdivisions $\mathcal{A}$ of $[s,t]$. The number $V_f(s,t)$ is called the *total variation* of f on $(s,t]$. The function f is said to be of *bounded variation* on $[s,t]$ if $V_f(s,t) < \infty$.

Show the following:

a) If f is increasing on $[s,t]$, then $V_f(s,t) = f(t) - f(s)$.

b) If f is differentiable and its derivative is bounded by b on $[s,t]$, then $V_f(s,t) \le (t-s)\cdot b$.

c) $V_f(s,t) + V_f(t,u) = V_f(s,u)$ for $s < t < u$.

d) $V_{f+g}(s,t) \le V_f(s,t) + V_g(s,t)$. Thus, if f and g are of bounded variation on $[s,t]$, then so are $f+g$ and $f-g$.

e) The function f is of bounded variation on $[s,t]$ if and only if $f = g - h$ for some real-valued positive functions g and h that are increasing on $[s,t]$.

Hint: To show the necessity, define $g(r)$ and $h(r)$ for r in $(s,t]$ by

$$2g(r) = V_f(s,r) + f(r) + f(s), \qquad 2h(r) = V_f(s,r) - f(r) + f(s)$$

and show that g and h are increasing and $f = g - h$.

The class of functions f for which Lebesgue-Stieltjes integrals $\int g\,df$ are defined is the class of f that are of bounded variation over bounded intervals.

6 Kernels and Product Spaces

Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces. Let K be a mapping from $E\times\mathcal{F}$ into $\bar{\mathbb{R}}_+$. Then, K is called a *transition kernel* from $(E,\mathcal{E})$ into $(F,\mathcal{F})$ if

6.1 a) the mapping $x \mapsto K(x,B)$ is $\mathcal{E}$-measurable for every set B in $\mathcal{F}$, and

b) the mapping $B \mapsto K(x,B)$ is a measure on $(F,\mathcal{F})$ for every x in E.

For example, if ν is a finite measure on $(F,\mathcal{F})$, and k is a positive function on $E\times F$ that is measurable with respect to the product σ-algebra $\mathcal{E}\otimes\mathcal{F}$, then it will be seen shortly that

6.2 $$K(x,B) = \int_B \nu(dy)\,k(x,y)\,, \qquad x\in E,\ B\in\mathcal{F}\,,$$

defines a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. In the further special case where $E = \{1, \ldots, m\}$ and $F = \{1, \ldots, n\}$ with their discrete σ-algebras, the transition kernel K is specified by the numbers $K(x, \{y\})$ and can be regarded as an m by n matrix of positive numbers. This special case will inform the choice of notations like Kf and μK below (recall that functions are thought as generalizations of column vectors and measures as generalizations of row vectors).

Measure-kernel-function

6.3 THEOREM. *Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Then,*

$$Kf(x) = \int_F K(x, dy)\, f(y)\,, \qquad x \in E,$$

defines a function Kf that is in $\mathcal{E}_+$ for every function f in $\mathcal{F}_+$;

$$\mu K(B) = \int_E \mu(dx)\, K(x, B)\,, \qquad B \in \mathcal{F},$$

defines a measure μK on $(F, \mathcal{F})$ for each measure μ on $(E, \mathcal{E})$; and

$$(\mu K)f = \mu(Kf) = \int_E \mu(dx) \int_F K(x, dy)\, f(y)$$

for every measure μ on $(E, \mathcal{E})$ and function f in $\mathcal{F}_+$.

Proof. a) Let $f \in \mathcal{F}_+$. Then Kf is a well-defined positive function on E, since the number $Kf(x)$ is the integral of f with respect to the measure $B \mapsto K(x, B)$. We show that Kf is $\mathcal{E}$-measurable in two steps: First, if f is simple, say $f = \sum_1^n b_i 1_{B_i}$, then $Kf(x) = \sum b_i K(x, B_i)$, which shows that Kf is $\mathcal{E}$-measurable since it is a linear combination of the $\mathcal{E}$-measurable functions $x \mapsto K(x, B_i)$, $i = 1, \ldots n$. Second, if f in $\mathcal{F}_+$ is not simple, we choose simple f_n in $\mathcal{F}_+$ increasing to f; then $Kf(x) = \lim_n Kf_n(x)$ for each x by the monotone convergence theorem for the measure $B \mapsto K(x, B)$; and, hence Kf is $\mathcal{E}$-measurable since it is the limit of $\mathcal{E}$-measurable functions Kf_n.

b) We prove the remaining two claims together. Fix a measure μ on $(E, \mathcal{E})$. Define $L : \mathcal{F}_+ \mapsto \bar{\mathbb{R}}_+$ by setting

$$L(f) = \mu(Kf).$$

If $f = 0$ then $L(f) = 0$. If f and g are in $\mathcal{F}_+$, and a and b in $\mathbb{R}_+$, then

$$\begin{aligned} L(af + bg) = \mu(K(af + bg)) &= \mu(aKf + bKg) \\ &= a\mu(Kf) + b\mu(Kg) = aL(f) + bL(g), \end{aligned}$$

where the second equality is justified by the linearity of the integration with respect to the measure $B \mapsto K(x, B)$ for each x, and the third equality by the linearity of the integration with respect to μ. Finally, if $(f_n) \subset \mathcal{F}_+$ and $f_n \nearrow f$, then $Kf_n(x) \nearrow Kf(x)$ by the monotone convergence theorem for $B \mapsto K(x, B)$, and

$$L(f_n) = \mu(Kf_n) \nearrow \mu(Kf) = L(f)$$

by the monotone convergence theorem for μ. Hence, by Theorem 4.21, there exists a measure ν on $(F, \mathcal{F})$ such that $L(f) = \nu f$ for every f in $\mathcal{F}_+$. Taking $f = 1_B$, we see that $\nu(B) = \mu K(B)$ for every set B in $\mathcal{F}$, that is, $\nu = \mu K$. So, μK is a measure on $(F, \mathcal{F})$, and $(\mu K)f = \nu f = L(f) = \mu(Kf)$ as claimed.

6.4 Remark. To specify a kernel K from $(E, \mathcal{E})$ into $(F, \mathcal{F})$ it is more than enough to specify Kf for every f in $\mathcal{F}_+$. Conversely, as an extension of Theorem 4.21, it is easy to see that a mapping $f \mapsto Kf$ from $\mathcal{F}_+$ into $\mathcal{E}_+$ specifies a transition kernel K if and only if

a) $K0 = 0$,
b) $K(af + bg) = aKf + bKg$ for f and g in $\mathcal{F}_+$ and a and b in $\mathbb{R}_+$,
c) $Kf_n \nearrow Kf$ for every sequence (f_n) in $\mathcal{F}_+$ increasing to f.

Obviously, then, $K(x, B) = K1_B(x)$.

Products of kernels, Markov kernels

Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$ and let L be a transition kernel from $(F, \mathcal{F})$ into $(G, \mathcal{G})$. Then, their *product* is the transition kernel KL from $(E, \mathcal{E})$ into $(G, \mathcal{G})$ defined by

6.5 $$(KL)f = K(Lf), \qquad f \in \mathcal{G}_+.$$

Remark 6.4 above can be used to show that KL is indeed a kernel. Obviously,

$$KL(x, B) = \int_F K(x, dy)\, L(y, B), \qquad x \in E,\ B \in \mathcal{G}.$$

A transition kernel from $(E, \mathcal{E})$ into $(E, \mathcal{E})$ is called simply a transition kernel on $(E, \mathcal{E})$. Such a kernel K is called a *Markov* kernel on $(E, \mathcal{E})$ if $K(x, E) = 1$ for every x, and a *sub-Markov* kernel if $K(x, E) \le 1$ for every x.

If K is a transition kernel on $(E, \mathcal{E})$, its *powers* are the kernels on $(E, \mathcal{E})$ defined recursively by

6.6 $$K^0 = I, \quad K^1 = K, \quad K^2 = KK, \quad K^3 = KK^2, \ldots,$$

where I is the identity kernel on $(E, \mathcal{E})$:

6.7 $$I(x, A) = \delta_x(A) = 1_A(x), \qquad x \in E,\ A \in \mathcal{E}.$$

Note that $If = f$, $\mu I = \mu$, $\mu If = \mu f$, $IK = KI = K$ always. If K is Markov, so is K^n for every integer $n \ge 0$.

Kernels finite and bounded

Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. In analogy with measures, K is said to be *finite* if $K(x, F) < \infty$ for each x, and *σ-finite* if $B \mapsto K(x, B)$ is σ-finite for each x. It is said to be *bounded* if $x \mapsto K(x, F)$ is bounded, and *σ-bounded* if there exists a measurable partition (F_n) of F such that $x \mapsto K(x, F_n)$ is bounded for each n. It is said to be *Σ-finite* if $K = \sum_1^\infty K_n$ for some sequence of finite kernels K_n, and *Σ-bounded* if the K_n can be chosen to be bounded. In the very special case where $K(x, F) = 1$ for all x, the kernel is said to be a *transition probability* kernel. Markov kernels are transition probability kernels. Some connections between these notions are put in exercises.

Functions on product spaces

We start by re-stating the content of Exercise 2.22: sections of a measurable function are measurable.

6.8 PROPOSITION. *Let $f \in \mathcal{E} \otimes \mathcal{F}$. Then, $x \mapsto f(x, y)$ is in $\mathcal{E}$ for each y in F, and $y \mapsto f(x, y)$ is in $\mathcal{F}$ for each x in E.*

Unfortunately, the converse is not true: it is possible that the conclusions hold, and yet f is not $\mathcal{E} \otimes \mathcal{F}$-measurable. One needs something stronger than measurability in at least one of the variables to conclude that f is in $\mathcal{E} \otimes \mathcal{F}$. See Exercise 6.28 for such an example.

The following is a generalization of the operation $f \mapsto Kf$ of Theorem 6.3 to functions f defined on the product space.

6.9 PROPOSITION. *Let K be a Σ-finite kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Then, for every positive function f in $\mathcal{E} \otimes \mathcal{F}$,*

6.10 $$Tf(x) = \int_F K(x, dy)\, f(x, y), \qquad x \in E,$$

defines a function Tf in $\mathcal{E}_+$. Moreover, the transformation $T : (\mathcal{E} \otimes \mathcal{F})_+ \mapsto \mathcal{E}_+$ is linear and continuous under increasing limits, that is,

a) $T(af + bg) = aTf + bTg$ for positive f and g in $\mathcal{E} \otimes \mathcal{F}$, and a and b in $\mathbb{R}_+$,

b) $Tf_n \nearrow Tf$ for every positive sequence $(f_n) \subset \mathcal{E} \otimes \mathcal{F}$ with $f_n \nearrow f$.

Proof. Let f be a positive function in $\mathcal{E} \otimes \mathcal{F}$. Then, for each x in E, the section $f_x : \; y \mapsto f(x, y)$ is a function in $\mathcal{F}_+$ by Proposition 6.8, and $Tf(x)$ is the integral of f_x with respect to the measure $K_x : \; B \mapsto K(x, B)$. Thus, $Tf(x)$ is a well-defined positive number for each x, and the linearity property (a) is immediate from the linearity of integration with respect to K_x for all x, and the continuity property (b) follows from the monotone convergence theorem for the measures K_x. There remains to show that Tf is $\mathcal{E}$-measurable.

We show this by a monotone class argument assuming that K is bounded. Boundedness of K implies that Tf is well-defined by 6.10 and is bounded for each bounded f in $\mathcal{E} \otimes \mathcal{F}$, and it is checked easily that

$$\mathcal{M} = \{f \in \mathcal{E} \otimes \mathcal{F} : \quad f \text{ is positive or bounded, } Tf \in \mathcal{E}\}$$

is a monotone class. Moreover, $\mathcal{M}$ includes the indicator of every measurable rectangle $A \times B$, since

$$T1_{A\times B}(x) = \int_F K(x, dy) 1_A(x) 1_B(y) = 1_A(x) K(x, B)$$

and the right side defines an $\mathcal{E}$-measurable function. Since the measurable rectangles generate the σ-algebra $\mathcal{E} \otimes \mathcal{F}$, it follows from the monotone class theorem 2.19 that $\mathcal{M}$ includes all positive (or bounded) f in $\mathcal{E} \otimes \mathcal{F}$ assuming that K is bounded. See Exercise 6.29 for extending the proof to Σ-finite K. □

Measures on the product space

The following is the general method for constructing measures on the product space $(E \times F, \mathcal{E} \otimes \mathcal{F})$.

6.11 THEOREM. *Let μ be a measure on $(E, \mathcal{E})$. Let K be a Σ-finite transition kernel from $(E, \mathcal{E})$ to $(F, \mathcal{F})$. Then,*

6.12 $$\pi f = \int_E \mu(dx) \int_F K(x, dy) f(x, y), \qquad f \in (\mathcal{E} \otimes \mathcal{F})_+$$

defines a measure π on the product space $(E \times F, \mathcal{E} \otimes \mathcal{F})$. Moreover, if μ is σ-finite and K is σ-bounded, then π is σ-finite and is the unique measure on that product space satisfying

6.13 $$\pi(A \times B) = \int_A \mu(dx) K(x, B), \qquad A \in \mathcal{E},\ B \in \mathcal{F}.$$

Proof. In the notation of the last proposition, the right side of 6.12 is $\mu(Tf)$, the integral of Tf with respect to μ. To see that it defines a measure, we use Theorem 4.21. Define $L(f) = \mu(Tf)$ for f in $\mathcal{E} \otimes \mathcal{F}$ positive. Then, $L(0) = 0$ obviously, L is linear since T is linear and integration is linear, and L is continuous under increasing limits by the same property for T and the monotone convergence theorem for μ. Hence, there is a unique measure, call it π, such that $L(f)$ is the integral of f with respect to π for every positive f in $\mathcal{E} \otimes \mathcal{F}$. This proves the first claim.

To prove the second, start by observing that π satisfies 6.13. Supposing that μ is σ-finite and K is σ-bounded, there remains to show that π is σ-finite and is the only measure satisfying 6.13. To that end, let $\hat{\pi}$ be another measure satisfying 6.13. Since μ is σ-finite, there is a measurable partition (E_m) of E such that $\mu(E_m) < \infty$ for each m. Since K is σ-bounded, there is a measurable partition (F_n) of F such that $x \mapsto K(x, F_n)$ is bounded for each n. Note that the measurable rectangles $E_m \times F_n$ form a partition of $E \times F$ and that, by the formula 6.13 for π and $\hat{\pi}$,

$$\pi(E_m \times F_n) = \hat{\pi}(E_m \times F_n) < \infty$$

for each m and n. Thus, the measures π and $\hat{\pi}$ are σ-finite, they agree on the p-system of measurable rectangles generating $\mathcal{E} \otimes \mathcal{F}$, and that p-system contains a partition of $E \times F$ over which π and $\hat{\pi}$ are finite. It follows from Exercise 3.18 that $\pi = \hat{\pi}$. □

Product measures and Fubini

In the preceding theorem, if the kernel K has the special form $K(x, B) = \nu(B)$ for some Σ-finite measure ν on $(F, \mathcal{F})$, then the measure π is called the *product* of μ and ν and is denoted by $\mu \times \nu$. The following theorem, generally referred to as Fubini's, is concerned with integration with respect to $\pi = \mu \times \nu$. Its main point is the formula 6.15: under reasonable conditions, in repeated integration, one can change the order of integration with impunity.

6.14 Theorem. *Let μ and ν be Σ-finite measures on $(E, \mathcal{E})$ and $(F, \mathcal{F})$, respectively.*

a) There exists a unique Σ-finite measure π on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ such that, for every positive f in $\mathcal{E} \otimes \mathcal{F}$,

6.15 $$\pi f = \int_E \mu(dx) \int_F \nu(dy)\, f(x, y) = \int_F \nu(dy) \int_E \mu(dx)\, f(x, y).$$

b) If $f \in \mathcal{E} \otimes \mathcal{F}$ and is π-integrable, then $y \mapsto f(x, y)$ is ν-integrable for μ-almost every x, and $x \mapsto f(x, y)$ is μ-integrable for ν-almost every y, and 6.15 holds again.

6.16 Remark. a) Since we have more than one measure, for notions like integrability and negligibility, one needs to point out the measure associated. So, π-integrable means "integrable with respect to the measure π".

b) It is clear from 6.15 that

6.17 $$\pi(A \times B) = \mu(A)\nu(B), \qquad A \in \mathcal{E}, B \in \mathcal{F},$$

and for this reason we call π the product of μ and ν and we use the notation $\pi = \mu \times \nu$.

c) If both μ and ν are σ-finite, then Theorem 6.11 applies with $K(x, B) = \nu(B)$ and implies that π is the only measure satisfying 6.17. Otherwise, it is possible that there are measures $\hat{\pi}$ satisfying $\hat{\pi}(A \times B) = \mu(A)\nu(B)$ for all A in $\mathcal{E}$ and B in $\mathcal{F}$ but with $\hat{\pi} f$ differing from πf for some positive f in $\mathcal{E} \otimes \mathcal{F}$.

Proof. a) Let πf be defined by the first integral in 6.15. Taking $K(x, B) = \nu(B)$ in Theorem 6.11 shows that this defines a measure π on the product space. Since $\mu = \sum \mu_i$ and $\nu = \sum \nu_j$ for some finite measures μ_i and ν_j, we have

$$\pi f = \sum_i \sum_j \int_E \mu_i(dx) \int_F \nu_j(dy)\, f(x, y) = \sum_{i,j} (\mu_i \times \nu_j) f$$

by Exercise 4.27 and the monotone convergence theorem. Thus, $\pi = \sum_{i,j} \mu_i \times \nu_j$ and, arranging the pairs (i, j) into a sequence, we see that $\pi = \sum \pi_n$ for some sequence of finite measures π_n.

b) To prove the equality of the integrals in 6.15, we start by observing that the second integral is in fact an integral over $F \times E$: defining $\hat{f}: F \times E \mapsto \bar{\mathbb{R}}_+$ by $\hat{f}(y, x) = f(x, y)$, the second integral is

$$\begin{aligned}\hat{\pi} \hat{f} &= \int_F \nu(dy) \int_E \mu(dx)\, \hat{f}(y, x) = \sum_j \sum_i \int_F \nu_j(dy) \int_E \mu_i(dx)\, \hat{f}(y, x) \\ &= \sum_{i,j} (\nu_j \times \mu_i)\, \hat{f}.\end{aligned}$$

Hence, to prove that $\pi f = \hat{\pi} \hat{f}$, it is sufficient to show that $(\mu_i \times \nu_j) f = (\nu_j \times \mu_i) \hat{f}$ for each pair of i and j. Fixing i and j, this amounts to showing that

$$\pi f = (\mu \times \nu) f = (\nu \times \mu) \hat{f} = \hat{\pi} \hat{f}$$

under the assumption that μ and ν are both finite.

c) Assume μ and ν finite. Let $h: E \times F \mapsto F \times E$ be the transposition mapping $(x, y) \mapsto (y, x)$. It is obviously measurable relative to $\mathcal{E} \otimes \mathcal{F}$ and $\mathcal{F} \otimes \mathcal{E}$. For sets A in $\mathcal{E}$ and B in $\mathcal{F}$,

$$\pi \circ h^{-1}(B \times A) = \pi(A \times B) = \mu(A)\nu(B) = \hat{\pi}(B \times A),$$

which implies via Proposition 3.7 that $\hat{\pi} = \pi \circ h^{-1}$. Hence, $\hat{\pi} \hat{f} = (\pi \circ h^{-1}) \hat{f} = \pi(\hat{f} \circ h) = \pi f$ since $\hat{f} \circ h(x, y) = \hat{f}(y, x) = f(x, y)$.

d) Let f be π-integrable. Then 6.15 holds for f^+ and f^- separately, and $\pi f = \pi f^+ - \pi f^-$ with both terms finite. Hence, 6.15 holds for f. As to the integrability of sections, we observe that the integrability of f implies that $x \mapsto \int_F \nu(dy)\, f(x, y)$ is real-valued for μ-almost every x, which in turn is equivalent to saying that $y \mapsto f(x, y)$ is ν-integrable for μ-almost every x. By symmetry, the finiteness for the second integral implies that $x \mapsto f(x, y)$ is μ-integrable for ν-almost every y.

Finite products

The concepts and results above extend easily to products of finitely many spaces. Let $(E_1, \mathcal{E}_1)$, ..., $(E_n, \mathcal{E}_n)$ be measurable spaces. Their *product* is denoted by any of the following three:

6.18 $$\bigotimes_{i=1}^{n}(E_i, \mathcal{E}_i) = (\bigtimes_{i=1}^{n} E_i, \bigotimes_{i=1}^{n} \mathcal{E}_i) = (E_1 \times \cdots \times E_n, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n),$$

where $E_1 \times \cdots \times E_n$ is the set of all n-tuples $(x_1, \ldots, x_n)$ with x_i in E_i for $i = 1, \ldots, n$, and $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$ is the σ-algebra generated by the *measurable rectangles* $A_1 \times \cdots \times A_n$ with A_i in $\mathcal{E}_i$, $i = 1, \ldots n$.

Let $\mu_1, \ldots, \mu_n$ be Σ-finite measures on $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectively. Then, their *product* $\pi = \mu_1 \times \cdots \times \mu_n$ is the measure defined on the measurable product space by analogy with Theorem 6.14: for positive functions f in $\bigotimes \mathcal{E}_i$,

6.19 $$\pi f = \int_{E_1} \mu_1(dx_1) \int_{E_2} \mu_2(dx_2) \cdots \int_{E_n} \mu_n(dx_n)\, f(x_1, \ldots, x_n).$$

It is usual to denote the resulting measure space

6.20 $$\bigotimes_{i=1}^{n}(E_i, \mathcal{E}_i, \mu_i).$$

Fubini's theorem is generalized to this space and shows that, if f is positive or π-integrable, the integrals on the right side of 6.19 can be performed in any order desired.

More general measures can be defined on the product space 6.18 with the help of kernels. We illustrate the technique for $n = 3$: Let μ_1 be a measure on $(E_1, \mathcal{E}_1)$, let K_2 be a transition kernel from $(E_1, \mathcal{E}_1)$ into $(E_2, \mathcal{E}_2)$, and let K_3 be a transition kernel from $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ into $(E_3, \mathcal{E}_3)$. Consider the formula

6.21 $$\pi f = \int_{E_1} \mu_1(dx_1) \int_{E_2} K_2(x_1, dx_2) \int_{E_3} K_3((x_1, x_2), dx_3)\, f(x_1, x_2, x_3)$$

for positive f in $\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3$. Assuming that K_2 and K_3 are Σ-finite, repeated applications of Theorem 6.11 show that this defines a measure π on $(E_1 \times E_2 \times E_3, \mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3)$.

In situations like this, we shall omit as many parentheses as we can and use a notation analogous to 5.8. For instance, instead of 6.21, we write

6.22 $$\pi(dx_1, dx_2, dx_3) = \mu_1(dx_1)\, K_2(x_1, dx_2)\, K_3(x_1, x_2, dx_3).$$

The notation

6.23 $$\pi = \mu_1 \times K_2 \times K_3$$

is also used for the same thing and is in accord with the notation for product measures.

Infinite products

Let T be an arbitrary set, countable or uncountable. It will play the role of an index set; we think of it as the time set. For each t in T, let $(E_t, \mathcal{E}_t)$ be a measurable space. Let x_t be a point in E_t for each t in T. Then we write $(x_t)_{t\in T}$ for the resulting collection and think of it as a function on T; this is especially appropriate when $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ for all t, because, then, $x = (x_t)_{t\in T}$ can be regarded as the mapping $t \mapsto x_t$ from T into E. The set F of all such functions $x = (x_t)_{t\in T}$ is called the *product space* defined by $\{E_t : t \in T\}$; and the notation $\times_{t\in T} E_t$ is used for F.

A *rectangle* in F is a subset of the form

6.24 $$\underset{t\in T}{\times} A_t = \{x \in F : x_t \in A_t \text{ for each } t \text{ in } T\}$$

where A_t differs from E_t for only a finite number of t. It is said to be measurable if $A_t \in \mathcal{E}_t$ for every t (for which A_t differs from E_t). The σ-algebra on F generated by the collection of all measurable rectangles is called the product σ-algebra and is denoted by $\bigotimes_{t\in T} \mathcal{E}_t$. The resulting measurable space is denoted variously by

6.25 $$\bigotimes_{t\in T}(E_t, \mathcal{E}_t) = (\underset{t\in T}{\times} E_t, \bigotimes_{t\in T} \mathcal{E}_t).$$

In the special case where $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ for all t, the following notations are also in use for the same:

6.26 $$(E, \mathcal{E})^T = (E^T, \mathcal{E}^T)$$

Although this is the logical point to describe the construction of measures on the product space, we shall delay it until the end of Chapter IV, at which point the steps involved should look intuitive. For the present, we list the following proposition which allows an arbitrary collection of measurable functions to be thought as one measurable function. It is a many-dimensional generalization of the result in Exercise 2.21.

6.27 PROPOSITION. *Let $(\Omega, \mathcal{H})$ be a measurable space. Let $(F, \mathcal{F}) = \otimes_{t\in T}(E_t, \mathcal{E}_t)$. For each t in T, let f_t be a mapping from Ω into E_t. For each ω in Ω, define $f(\omega)$ to be the point $(f_t(\omega))_{t\in T}$ in F. Then, the mapping $f : \Omega \mapsto F$ is measurable relative to $\mathcal{H}$ and $\mathcal{F}$ if and only if f_t is measurable relative to $\mathcal{H}$ and $\mathcal{E}_t$ for every t in T.*

Proof. Suppose that f is measurable relative to $\mathcal{H}$ and $\mathcal{F}$. Then, $\{f \in B\} \in \mathcal{H}$ for every B in $\mathcal{F}$. In particular, taking B to be the rectangle in 6.24 with $A_t = E_t$ for all t except $t = s$ for some fixed s, we see that $\{f \in B\} = \{f_s \in A_s\} \in \mathcal{H}$ for A_s in $\mathcal{E}_s$. Thus, f_s is measurable relative to $\mathcal{H}$ and $\mathcal{E}_s$ for every s fixed.

Suppose that each f_t is measurable relative to $\mathcal{H}$ and $\mathcal{E}_t$. If B is a measurable rectangle in F, then $\{f \in B\}$ is the intersection of finitely many sets of the form $\{f_t \in A_t\}$ with A_t in $\mathcal{E}_t$, and hence, $\{f \in B\} \in \mathcal{H}$. Since measurable rectangles generate the product σ-algebra $\mathcal{F}$, this implies via Proposition 2.3 that f is measurable relative to $\mathcal{H}$ and $\mathcal{F}$. □

Exercises

6.28 *Measurability in the product space.* Suppose that $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, and let $(F, \mathcal{F})$ be arbitrary. Let $f : E \times F \mapsto \bar{\mathbb{R}}$ be such that $y \mapsto f(x, y)$ is $\mathcal{F}$-measurable for each x in E and that $x \mapsto f(x, y)$ is right-continuous (or left-continuous) for each y in F. Show that, then, f is in $\mathcal{E} \otimes \mathcal{F}$.

6.29 *Image measures and kernels.* Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. Let $h : E \mapsto F$ be measurable relative to $\mathcal{E}$ and $\mathcal{F}$. Define

$$K(x, B) = 1_B \circ h(x), \qquad x \in E, B \in \mathcal{F}.$$

Show that K is a transition probability kernel. Show that, in the measure-kernel-function notation of Theorem 6.3,

$$Kf = f \circ h, \quad \mu K = \mu \circ h^{-1}, \quad \mu K f = \mu(f \circ h).$$

6.30 *Transition densities.* Let ν be a σ-finite measure on $(F, \mathcal{F})$, and let k be a positive function in $\mathcal{E} \otimes \mathcal{F}$. Define K by 6.2, that is, in differential notation,

$$K(x, dy) = \nu(dy)\, k(x, y).$$

Show that K is a transition kernel. Then, k is called the transition density function of K with respect to ν.

6.31 *Finite spaces.* Let $E = \{1, \dots, m\}$, $F = \{1, \dots, n\}$, $G = \{1, \dots, p\}$ with their discrete σ-algebras. Functions on such spaces can be regarded as column vectors, measures as row vectors, and kernels as matrices. Show that, with these interpretations, the notations Kf, μK, $\mu K f$, KL used in Theorem 6.3 and Definition 6.5 are in accord with the usual notations used in linear algebra.

6.32 *Finite and bounded kernels.* Let K be a finite transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Define

$$h(x) = \begin{cases} K(x, F) & \text{if } K(x, F) > 0, \\ 1 & \text{if } K(x, F) = 0, \end{cases}$$

and define H by solving

$$K(x, B) = h(x)\, H(x, B).$$

Show that $h \in \mathcal{E}_+$ and that H is a bounded kernel.

6.33 *Proof of Proposition 6.9.* Complete the proof. Hint: Use the preceding exercise to extend the proof from the bounded kernels to finite ones, and finally extend it to Σ-finite kernels.

6.34 *Fubini and Σ-finiteness.* In general, in order for 6.15 to hold, it is necessary that μ and ν be Σ-finite. For instance, let $E = F = [0, 1]$ with their Borel σ-algebras, and let μ be the Lebesgue measure on E, and ν the counting measure on F (that is, $\nu(A)$ is the number of points in A). Then, for $f(x, y) = 1$ if $x = y$ and 0 otherwise, the first integral in 6.15 is equal to 1, but the second is equal to 0.

Complements

6.35 *Product and Borel σ-algebras.* For each t in some index set T, let E_t be a topological space and let $\mathcal{E}_t = \mathcal{B}(E_t)$, the Borel σ-algebra on E_t. Let $(F, \mathcal{F}) = \bigotimes_T (E_t, \mathcal{E}_t)$ be the product measurable space. The product space F can be given the product topology, and let $\mathcal{B}(F)$ be the Borel σ-algebra corresponding to that topology on F.

In general, $\mathcal{B}(F) \supset \mathcal{F}$. If T is countable and if every $\mathcal{E}_t$ has a countable open base, then $\mathcal{F} = \mathcal{B}(F)$. In particular, $\mathbb{R}^n$ and $\mathbb{R}^\infty = \mathbb{R}^\mathbb{N}$ are topological spaces and their Borel σ-algebras coincide with the appropriate product σ-algebras; more precisely

$$(\mathcal{B}_\mathbb{R})^T = \mathcal{B}(\mathbb{R}^T)$$

for $T = \{1, 2, \ldots, n\}$ for every integer $n \geq 1$ and also for $T = \mathbb{N}^* = \{1, 2, \ldots\}$. This equality fails when T is uncountable, $\mathcal{B}(\mathbb{R}^T)$ being the larger then.

6.36 *Standard measurable spaces.* Let $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2), \ldots$ be standard measurable spaces, and let $(F, \mathcal{F})$ be their product. Then, $(F, \mathcal{F})$ is also standard.