Graduate Texts in Mathematics 143

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

Graduate Texts in Mathematics

continued after index

J.L. Doob

Measure Theory

Springer

J.L. Doob
101 West Windsor Road
Urbana, IL 61801
USA

Mathematics Subject Classification (1991): 28-01, 60A10, 60Bxx, 60Gxx

Library of Congress Cataloging-in-Publication Data
Doob, J. L.
Measure theory / J. L. Doob
p. cm. — (Graduate texts in mathematics ; 143)
Includes bibliographical references and index.
ISBN 0-387-94055-3 (New York). — ISBN 3-540-94055-3 (Berlin)
1. Measure theory. 2. Probabilities. I. Title. II. Series.
QA325.D66 1993
515'.42—dc20 93-17497

ISBN 0-387-94055-3 Springer-Verlag New York Berlin Heidelberg
ISBN 0-387-94055-3 Springer-Verlag Berlin Heidelberg New York SPIN 10636463

Introduction

This book was planned originally not as a work to be published, but as an excuse to buy a computer, incidentally to give me a chance to organize my own ideas on what measure theory every would-be analyst should learn, and to detail my approach to the subject. When it turned out that Springer-Verlag thought that the point of view in the book had general interest and offered to publish it, I was forced to try to write more clearly and search for errors. The search was productive.

Readers will observe the stress on the following points.

The application of pseudometric spaces. Pseudometric, rather than metric spaces, are applied to obviate the artificial replacement of functions by equivalence classes, a replacement that makes the use of "almost everywhere" either improper or artificial. The words "function" and "the set on which a function has values at least ε" can be taken literally in this book. Pseudometric space properties are applied in many contexts. For example, outer measures are used to pseudometrize classes of sets and the extension of a finite measure from an algebra to a σ algebra is thereby reduced to finding the closure of a subset of a pseudometric space.

Probability concepts are introduced in their appropriate place, not consigned to a ghetto. Mathematical probability is an important part of measure theory, and every student of measure theory should be acquainted with the fundamental concepts and function relations specific to this part. Moreover, probability offers a wide range of measure theoretic examples and applications both in and outside pure mathematics. At an elementary level, probability-inspired examples free students from the delusions that product measures are the only important multidimensional measures and that continuous distributions are the only important distributions. At a more sophisticated level, it is absurd that analysts should be familiar with mutual orthogonality but not with mutual independence of functions, that they should be familiar with theorems on convergence of series of orthogonal functions but not on convergence of martingales.

Convergence of sequences of measures is treated both in the general Vitali-Hahn-Saks setting and in the mathematical setting of Borel measures on the metric spaces of classical analysis: the compact metric spaces and the locally compact separable metric spaces. The general discussion is applied in detail to finite Lebesgue-Stieltjes measures on the line, in particular to probability measures.

Contents

0
Conventions and Notation

1. Notation: Euclidean space

$\mathbf{R}^N$ denotes Euclidean N-space; $\mathbf{R} = \mathbf{R}^1$; $\mathbf{R}^+$ is the half line $[0,\infty)$; $\bar{\mathbf{R}}^+$ is the extended half-line $[0,+\infty]$; $\bar{\mathbf{R}}$ is the extended line $[-\infty,+\infty]$. The extended half-lines and lines can be metrized by giving them the metric of their images under the transformation $s' = \arctan s$.

2. Operations involving $\pm\infty$

$$
\begin{aligned}
a(\pm\infty) &= \pm\infty && \text{if } a > 0, \\
&= 0 && \text{if } a = 0, \\
&= \mp\infty && \text{if } a < 0.
\end{aligned}
$$

If a is finite, $a \pm \infty = \pm\infty$; if $a = +\infty$, $a + (+\infty) = +\infty$; if $a = -\infty$, $a + (-\infty) = -\infty$.

3. Inequalities and inclusions

"Positive" means "≥ 0"; "strictly positive" means "> 0." The symbols $\subset$ and $\supset$ allow equality. "Monotone" allows equality unless modified by "strictly." Thus the identically 0 function on $\mathbf{R}$ is both monotone increasing and decreasing, but is not strictly monotone in either direction.

4. A space and its subsets

If S is a space, the class of all its subsets is denoted by 2^S. The complement of a subset A of a space is denoted by $\tilde{A}$. If A and B are subsets of S, $\tilde{A} \cap B$ is sometimes denoted by $B - A$. The *indicator function* of a subset A of S, defined on S as 1 on A and 0 on $\tilde{A}$, is denoted by $\mathbf{1}_A$. In particular, the identically 1 function $\mathbf{1}_S$ will be denoted by $\mathbf{1}$ and the identically 0 function $\mathbf{1}_\emptyset$ by $\mathbf{0}$.

5. Notation: generation of classes of sets

If $\mathbf{A}$ is a class of subsets of a space, the classes $\mathbf{A}_\sigma$, $\mathbf{A}_\delta$, and $\tilde{\mathbf{A}}$ are, respectively, the classes of countable unions, countable intersections, and complements of the sets in $\mathbf{A}$.

6. Product sets

If $S_1,\ldots,S_n$ are sets, $S_1\times\cdots\times S_n$ is the *product set*

$$\{(s_1,\ldots,s_n): s_i \in S_i, (i \le n)\}.$$

If $\mathbf{A}_i$ is a class of subsets of S_i, $\mathbf{A}_1\times\cdots\times\mathbf{A}_n$ is the class

$$\{A_1\times\cdots\times A_n: A_i \in \mathbf{A}_i\ (i \le n)\}$$

of product sets. The corresponding definitions are made for infinite (not necessarily countable) products.

7. Dot notation for an index set

"$B_\bullet$" is shorthand for $\{B_i, i \in I\}$, where I is a specified not necessarily countable index set. Unless the subscript range is otherwise described, "a finite sequence $B_\bullet$" means the sequence $B_1,\ldots,B_n$, for some strictly positive integer n, and "a sequence $B_\bullet$" means the infinite sequence $B_1,B_2,\ldots$. The notation $\Sigma B_\bullet$ means the sum over the values of the subscript, and corresponding dot notation will be applied to (not necessarily countable) set unions and intersections. If $a_\bullet$ is a sequence, the notation $\lim a_\bullet$ means $\lim_{n\to\infty} a_n$, and corresponding dot notation will be applied to inferior and superior limits. When dots appear more than once in an expression, the missing symbol is to be the same in each place. Thus if $A_\bullet$ and $B_\bullet$ are sequences of sets, $\cup(A_\bullet\cap B_\bullet)$ is the union of intersections $A_n\cap B_n$.

8. Notation: sets defined by conditions on functions

If f is a function from a space S into a space S' and if A' is a subset of S', the set notation $\{s \in S: f(s) \in A'\}$ will sometimes be abbreviated to $\{f \in A'\}$. Here f may represent a set of functions. Thus if $g_1,\ldots,g_n$ are functions from S into S' and if B' is a subset of S'^n, the notation $\{s \in S: [g_1(s),\ldots,g_n(s)] \in B'\}$ may be abbreviated to $\{(g_1,\ldots,g_n) \in A'\}$.

9. Notation: open and closed sets

The classes of open and closed subsets of a topological space will be denoted, respectively, by $\mathbb{G}$ and $\mathbb{F}$.

10. Limit of a function at a point

The limit of a function at a point depends somewhat on the nationality and background of the writer. In this book, the limit does not involve the value of the function at the point. Thus the function $1_{\{0\}}$, defined on $\mathbf{R}$ as 0 except at the origin, where the function is defined as 1, has limit 0 at the origin in this book even though the function does not have a Bourbaki limit at the origin.

11. Metric spaces

Recall that a *metric space* is a space coupled with a *metric*. A metric for a space S is a *distance function* d, a function from $S \times S$ into $\mathbf{R}^+$ satisfying the following conditions.

(a) Symmetry: $d(s,t) = d(t,s)$.
(b) Identity: $d(s,t) = 0$ if and only if $s = t$.
(c) Triangle inequality: $d(s,u) \leq d(s,t) + d(t,u)$.

A *ball* in S is an open set $\{s: d(s,s_0) < r\}$; s_0 is the *center*, r is the *radius*.
It is a useful fact that if d is a metric for S and if c is a strictly positive constant, the function $d \wedge c$ is also a metric for S and determines the same topology as d. That is, the class of open sets is the same for $d \wedge c$ as for d. If d is a function from $S \times S$ into $\bar{\mathbf{R}}^+$ and satisfies (a), (b), and (c), the function $d \wedge c$ is a finite valued function satisfying these conditions and can therefore serve as a metric.

12. Standard metric space theorems

The following standard metric space theorems will be used. Proofs are sketched to facilitate checking by the reader that they are valid for the pseudometric spaces to be defined in Section 13.

(a) *A metric space (S,d) can be completed, that is, can be augmented by addition of new points to be complete.* To prove this theorem, let S' be the class of Cauchy sequences of points of S. The space S' is partitioned into equivalence classes, putting two Cauchy sequences $s_\bullet$ and $t_\bullet$ into the same equivalence class if and only if $\lim d(s_\bullet,t_\bullet) = 0$. If s' and t' are equivalence classes, define $d'(s',t') =$

$\lim d(s_\bullet,t_\bullet) = 0$. If s' and t' are equivalence classes, define $d'(s',t') = \lim d(s_\bullet,t_\bullet)$ for $s_\bullet$ in s' and $t_\bullet$ in t'. This limit exists, does not depend on the choice of Cauchy sequences in their equivalence classes, and (S',d') is a complete metric space. Define a function f from S into S' by $f(s) = s,s,s,\ldots$. This map preserves distance, and if S is identified with its image in S', S' is the desired completion of S.

(b) *A uniformly continuous function g from a dense subset of a metric space S into a complete metric space S' has a unique uniformly continuous extension to S.* To prove this theorem, observe that if s is not already in the domain of g, and if $s_\bullet$ is a sequence in the domain of g, with limit s, the uniform continuity of g implies that $\lim g(s_\bullet)$ exists and does not depend on the choice of $s_\bullet$. The value $g(s)$ is defined as this limit, and as so extended g is uniformly continuous on S. The uniqueness assertion is trivial.

(c) *If a complete metric space S is a countable union of closed sets, at least one summand has an inner point.* To prove this theorem, let $\cup S_\bullet$ be the union of a sequence of closed nowhere dense subsets of S. There is a closed ball B_1 of radius ≤ 1 in the open set $\tilde{S}_1$. Similarly there is a closed ball B_2 of radius $\leq 1/2$ in $B_1 \cap \tilde{S}_2$, and so on. The intersection of these closed balls is a point of S in no summand. Hence the union cannot be S, that is, if S is the union of a sequence of closed sets, at least one is not nowhere dense, and therefore has an inner point.

(d) *If $f_\bullet$ is a sequence of bounded continuous functions from a complete metric space S into* **R**, *and if* $\sup |f_\bullet(s)| < +\infty$ *for each point s of S, then there is a number* γ *for which the set* $\{s \in S\colon \sup |f_\bullet(s)| \leq \gamma\}$ *has an inner point.* This theorem follows at once from (c) because for each value of γ the set in question is closed, and as γ increases through the positive integers the set tends to S.

(e) A sequence $f_\bullet$ of functions from a metric space (S,d) into a metric space (S',d') is said to *converge uniformly at a point* s_0 of S, if there is convergence at s_0, and if to every strictly positive ε there corresponds a strictly positive δ and an integer k, with the property that $d'\big(f_m(s),f_n(s)\big) < \varepsilon$ whenever $n \geq k,\ m \geq k$, and $d(s,s_0) < \delta$. An equivalent condition is that there is a point s' of S' with the property that whenever $t_\bullet$ is a sequence in S, with limit s_0, then $\lim f_\bullet(t_\bullet) = s'$. *If $f_\bullet$ is a convergent sequence of continuous functions from S into S', the limit function f is continuous at every point of uniform convergence of the sequence.* In fact, if s_0 is a point of uniform convergence, if ε, δ, k are as just described, and if δ is decreased, if necessary, to make $d'\big(f_k(s),f_k(s_0)\big) < \varepsilon$ whenever $d(s,s_0) < \delta$, then

$$(12.1)\quad d'\big(f(s),f(s_0)\big) < d'\big(f(s),f_k(s)\big)+d'\big(f_k(s),f_k(s_0)\big)+d'\big(f_k(s_0),f(s_0)\big) < 3\varepsilon$$

whenever $d(s,s_0) < \delta$. Hence f is continuous at s_0, as asserted.

(f) *If a sequence $f_\bullet$ of continuous functions from a complete metric space (S,d) into a metric space (S',d') is convergent, there must be at least one point of uniform convergence.* (Since this assertion can be applied to the restrictions of the functions to an arbitrary closed ball in S, the set of points of uniform continuity of the sequence, and therefore the set of continuity points of the limit function, is actually dense in S.) This assertion is reduced to (c) as follows. For each pair of strictly positive integers m, k, the set

(12.2) $$\bigcap_{n>m}\{s: d'(f_n(s), f_m(s)) \leq 1/k\}$$

is a closed subset of S. When k is fixed and m increases, the union of these closed sets is S. It follows that there is a closed ball B_k in one of these sets of radius at most $1/k$. If this argument is carried through with S replaced successively by $B_1, B_2, \ldots,$the argument yields a monotone decreasing sequence $B_\bullet$ of balls whose intersection is a point of uniform convergence of the sequence $f_\bullet$.

13. Pseudometric spaces

A *pseudometric space* is a space coupled with a *pseudometric*. A pseudometric for a space S is a *pseudometric distance function d,* a function from $S\times S$ into $\mathbf{R}^+$ that satisfies 11(a) and 11(c), but 11(b) is weakened to

(11b') $$d(s,s) = 0.$$

There are two approaches to a pseudometric space (S,d). The most common approach is to define a space S^* of equivalence classes of subsets of S, putting two points s and t of S in the same equivalence class if and only if $d(s,t) = 0$. If s^* and t^* are equivalence classes define $d^*(s^*,t^*)$ as $d(s,t)$, for s in s^* and t in t^*. This definition does not depend on the choice of s and t in their equivalence classes, and d^* is a distance function making S^* a metric space.

A second approach, used in this book, is to stay with the pseudometric space, making the same definitions as formulated for metric spaces: open and closed sets, separable spaces, complete spaces, and so on. Note that if a sequence of points of a pseudometric space is convergent to a point, the sequence is also convergent to every point at zero distance from that point, and that therefore if a point is in an open (or closed) set of a pseudometric space every point at zero distance from it is also in that set. The theorems and proofs of the theorems in Section 12 remain valid for pseudometric spaces. It may seem that in fact there is not much difference between handling S and S^* except that S^* is simpler, but in fact in many measure theoretic contexts, the pseudometric space is less clumsy.

I
Operations on Sets

In this chapter, certain relations between and operations on subsets of an abstract space are described. When numbered relations are paired, as in (1.1), each relation of the pair yields the other relation when the sets involved are replaced by their complements. Proofs of easily verifiable assertions are omitted.

1. Unions and intersections

If $A_\bullet$ and $B_\bullet$ are collections of subsets of a space S,

(1.1) $$(\cup A_\bullet)^{\sim} = \cap \tilde{A}_\bullet, \qquad (\cap A_\bullet)^{\sim} = \cup \tilde{A}_\bullet,$$

(1.2) $$(\cup_s A_s) \cap (\cup_t B_t) = \cup_{s,t} (A_s \cap B_t),$$

$$(\cap_s A_s) \cup (\cap_t B_t) = \cap_{s,t} (A_s \cup B_t).$$

Obviously $1_{A \cap B} = 1_A 1_B$, $1_{\tilde{A}} = 1 - 1_A = 1 + 1_A \pmod 2$, and

(1.3) $$1_{A \cup B} = 1_A + 1_B - 1_A 1_B.$$

2. The symmetric difference operator Δ

In this section A, B, C, and D are subsets of a space S. The symmetric difference $A \Delta B$ is defined by

(2.1) $$A \Delta B = (A - B) \cup (B - A)$$

or, equivalently,

(2.2) $$1_{A \Delta B} = 1_A + 1_B \pmod 2.$$

The latter form provides easy proofs of some of the relations listed below. Obviously

(2.3) $A\Delta\emptyset = A,\ A\Delta S = \tilde{A}, A\Delta A = \emptyset, (A\Delta B)^{\sim} = \tilde{A}\Delta B,\ \tilde{A}\Delta\tilde{B} = A\Delta B \subset A\cup B.$

The symmetric difference operator is commutative and associative:

(2.4) $$A\Delta B = B\Delta A,\quad A\Delta(B\Delta C) = (A\Delta B)\Delta C,$$

and therefore parentheses can be omitted in expressions of the form $A\Delta B\Delta C\Delta\cdots$. The equality $A\Delta C = A\Delta B\Delta B\Delta C$ yields the useful triangle inclusion relation

(2.5) $$A\Delta C \subset (A\Delta B)\cup(B\Delta C).$$

The symmetric difference operator satisfies

(2.6) $$(A\Delta B)\cap C = (A\cap C)\Delta(B\cap C),\ (A\Delta B)\cup C = (A\cup C)\Delta(B\cap\tilde{C}),$$

and if $A_\bullet$ and $B_\bullet$ are collections of subsets of S,

(2.7) $$(\cup_s A_s)\,\Delta\,(\cup_t B_t) \subset \cup(A_\bullet\Delta B_\bullet),\quad \cap(A_\bullet\Delta B_\bullet) \subset (\cap_s A_s)\,\Delta\,(\cup_t B_t),$$

(2.8) $$(\cap_s A_s)\,\Delta\,(\cap_t B_t) \subset \cup(A_\bullet\Delta B_\bullet).$$

If $A_1,\ldots,A_n$ are subsets of S,

(2.9)
$$1_{\cup A_\bullet} = \sum_{i\ge 1} 1_{A_i} - \sum_{i<j} 1_{A_i\cap A_j} + \cdots + (-1)^{n-1} 1_{A_1\cap\ldots\cap A_n},$$
$$1_{\cap A_\bullet} = \sum_{i\ge 1} 1_{A_i} - \sum_{i<j} 1_{A_i\cup A_j} + \cdots + (-1)^{n-1} 1_{A_1\cup\ldots\cup A_n}.$$

When $n = 2$, both equalities reduce to (1.3). Each equality can be proved by induction, or, more directly, by checking it at those points in A_j for exactly m values of j, for $m = 0,\ldots,n$. Each equality reduces to the other when the sets involved are replaced by their complements.

3. Limit operations on set sequences

If $A_\bullet$ is a sequence of subsets of a space S, define

(3.1) $$\limsup A_\bullet = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j,\ \liminf A_\bullet = \bigcup_{k=1}^{\infty} \bigcap_{j=k}^{\infty} A_j.$$

The superior limit is the set of those points in A_n for infinitely many values of n; the inferior limit is the set of those points in A_n for all but finitely many values of n. The inferior limit is a subset of the superior limit, and if there is equality with common limit set A, the sequence $A_\bullet$ *converges to* A, written $\lim A_\bullet = A$. The following limit properties for sets are analogous to those for numbers, because they correspond exactly to those for indicator functions, written at the end of this section.

(a) *A monotone increasing sequence of sets converges to the union of the sets; a monotone decreasing sequence of sets converges to the intersection of the sets.*

(b) *If $B_\bullet$ is a subsequence of $A_\bullet$, then $B_\bullet$ converges whenever $A_\bullet$ does,* because

(3.2) $$\liminf A_\bullet \subset \liminf B_\bullet \subset \limsup B_\bullet \subset \limsup A_\bullet.$$

Since

(3.3) $$\liminf \tilde{A}_\bullet = (\limsup A_\bullet)^{\sim},$$

(c) *the sequence $\tilde{A}_\bullet$ converges to $\tilde{A}$ when $A_\bullet$ converges to A.* Furthermore, for sequences $A_\bullet$, $B_\bullet$ of sets,

(3.4) $$\begin{aligned}&\liminf (A_\bullet \cup B_\bullet) \supset (\liminf A_\bullet) \cup (\liminf B_\bullet)\\ &\limsup (A_\bullet \cup B_\bullet) = (\limsup A_\bullet) \cup (\limsup B_\bullet),\end{aligned}$$

(3.5) $$\begin{aligned}&\liminf (A_\bullet \cap B_\bullet) = (\liminf A_\bullet) \cap (\liminf B_\bullet),\\ &\limsup (A_\bullet \cap B_\bullet) \subset (\limsup A_\bullet) \cap (\limsup B_\bullet).\end{aligned}$$

Hence,

(d) *whenever sequences $A_\bullet$ and $B_\bullet$ converge respectively to A and B, the sequences $A_\bullet \cup B_\bullet$* and *$A_\bullet \cap B_\bullet$ converge respectively to $A \cup B$ and $A \cap B$.*

The equality

(3.6) $$\bigcup A_\bullet - \bigcap A_\bullet = \bigcup_n (A_n \,\Delta\, A_{n+1}),$$

for a sequence $A_\bullet$ of sets, is useful in convergence studies, because (3.6) implies

(3.7) $$\limsup A_\bullet - \liminf A_\bullet = \limsup_{n\to\infty} (A_n \Delta A_{n+1}).$$

Set sequence limit properties in terms of indicator functions. If $A_\bullet$ is a sequence of sets, the functions lim sup $1_{A_\bullet}$ and lim inf $1_{A_\bullet}$ are respectively the indicator functions of the sets lim sup $A_\bullet$ and lim inf $A_\bullet$. Thus *the sequence* $A_\bullet$ *converges to* A *if and only if the corresponding sequence of indicator functions converges to* 1_A.

4. Probabilistic interpretation of sets and operations on them

In the application of mathematical probability to nonmathematical contexts, a space of points corresponds to a class of possible observations made in some real context, for example, heights of humans in a specified country, positions of stars, possible outcomes of tossing a coin twice, times of auto accidents on a specified highway. The subsets of the space, *events* in the applications, are determined by conditions in the real contexts. For example, in the last mentioned application, one event is the class of accident times during the hours of daylight. The union operation on sets corresponds to *or* for events; the intersection operation for sets corresponds to *and*. It will be seen in later chapters that mathematical probability (which must be distinguished from the nonmathematical variety) is a certain specialization of measure theory, distinguished by its own terminology and its field of nonmathematical applications. On the one hand, mathematicians were computing probabilities and expectations, on the other hand mathematicians were computing volumes and masses, and the two fields did not come together until this century. In fact some probabilists resented the invasion of their juicy domain by dry mathematical rigor, and even now almost all probabilists write in the traditional dialect of their subject.

II
Classes of Subsets of a Space

1. Set algebras

Definition. A class **S** of subsets of a space S is an *algebra* if the following conditions are satisfied.

(a) $\emptyset \in \mathbf{S}$.
(b) The class **S** is *closed under complementation*: if $A \in \mathbf{S}$ then $\tilde{A} \in \mathbf{S}$.
(c) The class **S** is *closed under finite unions*: finite unions of sets in **S** are in **S**.
(c') The class **S** is *closed under finite intersections*: finite intersections of sets in **S** are in **S**.

Under (b), conditions (c) and (c') are equivalent, in view of Equation I(1.1). If $A_\bullet$ is a finite or infinite sequence of sets in an algebra **S**, their union, which may or may not be in the algebra if the sequence is infinite, can be expressed as the disjunct union of a sequence of sets in **S**, each of which is a subset of the corresponding term of $A_\bullet$:

$$A_1 \cup A_2 \cup \cdots = A_1 \cup (A_2 - A_1) \cup [A_3 - (A_1 \cup A_2)] \cup \cdots . \tag{1.1}$$

Definition. An algebra **S** of subsets of a space S is a σ *algebra* if **S** contains the limit of every monotone sequence of its sets. The pair $(S,\mathbf{S})$ is then a *measurable space*, and the sets in **S** are *measurable.*

Application of complementation shows that this defining condition of a σ algebra, as distinguished from an algebra, is fulfilled even if it is specified as fulfilled only for increasing (or only for decreasing) set sequences. If $A_\bullet$ is a sequence of sets in a σ algebra, the limit sets lim sup $A_\bullet$ and lim inf $A_\bullet$ are also in the σ algebra..

The smallest algebra of subsets of a space S is the pair of sets $(\emptyset, S)$; the largest algebra is 2^S. Both these algebras are σ algebras.

2. Examples

(a) **Finite unions of right semiclosed intervals of $\mathbf{R}^N$.** A *right semiclosed interval of* **R** is either the empty set or a subset of **R** of the form

$$(2.1) \qquad \{s \in \mathbf{R}: a < s \le b\} \qquad (-\infty \le a < b \le +\infty).$$

The complement of such an interval is either a right semiclosed interval or the disjunct union of two such intervals, and the intersection of two such intervals is another one. The class of finite unions of these intervals is therefore an algebra. This algebra is not a σ algebra because, for example, it does not contain the open interval $(0,1) = \bigcup_1^\infty (0,1\text{-}1/n]$.1

The right semiclosed intervals of $\mathbf{R}^N$ for $N > 1$ are defined as the N-fold products of right semiclosed intervals of **R.** For $N \ge 1$ the class of finite unions of these intervals is an algebra, but not a σ algebra.

(a') In Example (a), replace **R** by the set of rational numbers. With this choice instead of **R** in (2.1), the class of finite unions of these intervals is still an algebra but not a σ algebra.

(b) **Classes of** 0,1 **sequences.** For $n = 1,2,...$ let S_n be the space of n–tuples of 0's and 1's, and define $S = S_1 \times S_1 \times \cdots$, the space of infinite sequences of 0's and 1's. Let x_n be the nth coordinate function of S. Under the map taking a point of S_n into the subset of S with that point as initial n-tuple, the algebra $\mathbf{S}_n$ of all subsets of S_n maps into a set algebra $\mathbf{S}_n'$ of subsets of S. The union $\bigcup \mathbf{S}_\bullet'$ of all these algebras is itself an algebra $\mathbf{S}'$ of subsets of S. The algebra $\mathbf{S}_n'$ is the algebra of sets specified by conditions on $x_1,...,x_n$; the algebra $\mathbf{S}'$ is the algebra of sets specified by conditions on finitely many coordinate functions of S. The algebra $\mathbf{S}'$ is not a σ algebra because, for example, $A_n = \{x_n{=}1\} \in \mathbf{S}_n' \subset \mathbf{S}'$, but $\bigcup A_\bullet$ is not in $\mathbf{S}'$. The set algebra $\mathbf{S}'$ has the property, to be applied in Section IV.14, that *if A in $\mathbf{S}'$ is a disjunct countable union $\bigcup A_\bullet$ of sets in $\mathbf{S}'$, then all but a finite number of the summands are empty.* Equivalently, phrased in terms of the remainder sequence $\{A - \bigcup_1^n A_\bullet,\ n \ge 1\}$, *a decreasing sequence $B_\bullet$ of nonempty sets in $\mathbf{S}'$ has a nonempty limit.* To prove this assertion about decreasing sequences, observe that by hypothesis each set B_n is specified by conditions on a finite number of coordinates, say the first a_n coordinates. (Note that if B_n is specified by conditions on the first a_n coordinates then B_n can also be specified by conditions on the first a_n' coordinates for $a_n' > a_n$.) The assertion to be proved is trivial if the sequence $a_\bullet$ is bounded. If this sequence is not bounded, it can be supposed that the sequence is monotone increasing - if necessary replace each value a_n by $a_1 \vee \cdots \vee a_n$. For each k, the set of initial a_k-tuples of points of B_k is not empty and decreases as n increases, to some nonempty set C_k of a_k-tuples. Moreover the a_k-tuples in C_k are the initial a_k-tuples of C_m for $m > k$. Thus the sequence $C_\bullet$ determines a nonempty set that is a subset of every set B_k, that is, $\bigcap B_\bullet$ is not empty, as was to be proved.

Observation for later use. If, more generally, the space S_1 is a metric space, if $S = S_1 \times S_1 \times \cdots$, and if x_n is the nth coordinate function of S, a trivial adaptation of the argument just used yields the following: *if $B_\bullet$ is a decreasing sequence of nonempty subsets of S, with $B_n = \{(x_1,\ldots,x_{a_n}) \in B_n'\}$, where B_n' is a compact subset of $S_1^{a_n}$*, with *$a_\bullet$ some sequence of* positive integers, *then $\cap B_\bullet$ is not empty*. This result is trivial unless the sequence $a_\bullet$ is unbounded. It can be assumed that $a_\bullet$ is an unbounded increasing sequence (if not already increasing, choose a subsequence of $B_\bullet$ for which the corresponding subsequence of $a_\bullet$ is increasing), and the rest of the argument for the special case is carried through without change.

3. The generation of set algebras

Let $\mathbf{S}_0$ be a class of subsets of a space S, and let Γ be the class of those algebras of subsets of S that include $\mathbf{S}_0$. Denote by $\sigma_0(\mathbf{S}_0)$ the class of sets in every algebra in the class Γ. Then $\sigma_0(\mathbf{S}_0)$ is an algebra, the smallest one including all the sets of $\mathbf{S}_0$. Similarly there is a smallest σ algebra $\sigma(\mathbf{S}_0)$ including all the sets of $\mathbf{S}_0$, the intersection of all such σ algebras. The algebras $\sigma_0(\mathbf{S}_0)$ and $\sigma(\mathbf{S}_0)$ are *generated* by $\mathbf{S}_0$. Obviously

$$\sigma[\sigma_0(\mathbf{S}_0)] = \sigma(\mathbf{S}_0) = \sigma[\sigma(\mathbf{S}_0)], \quad \sigma_0[\sigma_0(\mathbf{S}_0)] = \sigma_0(\mathbf{S}_0).$$

If $A_1,\ldots,A_n$ are subsets of a space S, they generate a partition of S into 2^n pairwise disjoint possibly empty cells, the intersections $B_1 \cap \cdots \cap B_n$, where each set B_j is either A_j or $\tilde{A}_j$. The algebra $\sigma_0(A_\bullet)$ is the class of finite unions of these cells. In general, if $\mathbf{S}_0$ is an arbitrary class of subsets of S, the algebra $\sigma_0(\mathbf{S}_0)$ is the class of finite unions of finite intersections of the members and complements of members of $\mathbf{S}_0$. There is no such simple representation of $\sigma(\mathbf{S}_0)$.

4. The Borel sets of a metric space

A metric space is a pair (S,d) consisting of a space S and a distance function d. The specification of d is usually omitted if it is irrelevant to the discussion or obvious from the context. The distance function for the product of finitely many metric spaces is to be understood to be defined by the Euclidean formula: square root of the sum of squared distances for the factor spaces.

Every closed set in a metric space S is a countable intersection of open sets: $\mathbf{F} \subset \mathbf{G}_\delta$. In fact if A is closed, the set $\{s \in S: d(s,A) < 1/n\}$ is open and

$$A = \bigcap_1^\infty \{s \in S: d(s,A) < 1/n\}. \tag{4.1}$$

Complementation yields the fact that $\mathbb{G} \subset \mathbb{F}_\sigma$, that is, every open set in a metric space is a countable union of closed sets. These two inclusions imply first, that $\sigma(\mathbb{G}) \supset \mathbb{F}$ and therefore $\sigma(\mathbb{G}) \supset \sigma(\mathbb{F})$, and next that $\sigma(\mathbb{F}) \supset \mathbb{G}$ and therefore that $\sigma(\mathbb{F}) \supset \sigma(\mathbb{G})$. Hence $\sigma(\mathbb{F}) = \sigma(\mathbb{G})$.

Definition. The class $\mathbf{B}(S)$ of *Borel subsets* of a metric space S is the σ algebra $\sigma(\mathbb{G})$ $(= \sigma(\mathbb{F}))$.

In dealing with a measurable space $(S,\mathbf{S})$ for which S is a metric space it will always be assumed, unless stated otherwise, that $\mathbf{S}=\mathbf{B}(S)$. The reasoning that led to the equality $\sigma(\mathbb{F}) = \sigma(\mathbb{G})$ for a metric space S shows that if $\mathbf{S}$ is so large a class of Borel subsets of S that $\sigma(\mathbf{S})$ includes $\mathbb{F}$ or $\mathbb{G}$, then $\sigma(\mathbf{S}) = \mathbf{B}(S)$. For example $\mathbf{B}(\mathbf{R})$ is generated by the class of open intervals, also by the class of closed intervals, also by the class of right semiclosed intervals, also by the class of semi-infinite intervals, and so on.

Relativization of Borel sets. If A is a subset of a metric space (S,d), if A is metrized by restricting d to pairs of points of A, and if $A_0 \subset A$, *then* $A_0 \in \mathbf{B}(A)$ *if and only if* A_0 *is the intersection with* A *of a set in* $\mathbf{B}(S)$, that is, in the obvious notation, $\mathbf{B}(A) = \mathbf{B}(S) \cap A$. In fact the class of sets in $\mathbf{B}(A)$ that are intersections with A of a set in $\mathbf{B}(S)$ is a σ algebra relative to A and includes the subsets of A that are open relative to A, because these are the intersections with A of open subsets of S. Hence $\mathbf{B}(A) \subset \mathbf{B}(S) \cap A$. In the other direction, $\mathbf{B}(S) \cap A \subset \mathbf{B}(A)$ because the class of Borel subsets of S meeting A in a Borel set relative to A includes the open subsets of S, is a σ algebra, and is therefore $\mathbf{B}(S)$.

In particular, *if A is a Borel subset of S, then a subset of A is Borel relative to A if and only if it is Borel relative to S.* Thus, for example, a subset of a line in $\mathbf{R}^2$ is a Borel set relative to the line if and only if the subset is a Borel set relative to the plane.

5. Products of set algebras

For $i = 1,\dots,n$, let $\mathbf{S}_i$ be an algebra of subsets of a space S_i. Let $S=S_1\times\cdots\times S_n$ be the product of these spaces. In the following, "product set" will always mean a set in the class $\mathbf{S}_1\times\cdots\times \mathbf{S}_n$ of product sets $A_1\times\cdots\times A_n$ with A_i in $\mathbf{S}_i$. Observe that the intersection of two product sets is a product set, and that the complement of a product set is a finite disjunct union of product sets. It follows that the class of finite unions of product sets is an algebra, necessarily $\sigma_0(\mathbf{S}_1\times\cdots\times \mathbf{S}_n)$.

In particular, if each space S_i is $\mathbf{R}$ and if each algebra $\mathbf{S}_i$ is the algebra of finite unions of right semiclosed intervals of $\mathbf{R}$, then $\sigma_0(\mathbf{S}_1\times\cdots\times \mathbf{S}_n)$ is the algebra of finite unions of right semiclosed intervals of $\mathbf{R}^N$. The σ algebra $\mathbf{B}(\mathbf{R}^N)$ is generated by this algebra, also generated by the class of N fold products of the one-dimensional Borel sets, also by the class of N fold products of classes that generate $\mathbf{B}(\mathbf{R})$, for example, by the class of N-fold products of open intervals of $\mathbf{R}$, or of right semiclosed intervals of $\mathbf{R}$, and so on.

Returning to general factor spaces $S_1,\dots,S_n$, observe that

$$\sigma(\mathbf{S}_1\times\cdots\times\mathbf{S}_n) = \sigma\big(\sigma(\mathbf{S}_1)\times\cdots\times\sigma(\mathbf{S}_n)\big). \tag{5.1}$$

In fact, trivially, the right side is at least as large as the left. Conversely it is sufficient to show that the left side is at least as large as the right by showing that it includes $\sigma(\mathbf{S}_1)\times\cdots\times\sigma(\mathbf{S}_n)$. Fix A_i in $\mathbf{S}_i$ for all $i>1$. The class of sets A_1 in $\sigma(\mathbf{S}_1)$ for which the product set $A_1\times\cdots\times A_n$ is in $\sigma(\mathbf{S}_1\times\cdots\times\mathbf{S}_n)$ includes $\mathbf{S}_1$, is a σ algebra, and is therefore $\sigma(\mathbf{S}_1)$. Thus the left side of (5.1) includes $\sigma(\mathbf{S}_1)\times\mathbf{S}_2\times\cdots\mathbf{S}_n$. Go on by induction to finish the proof of the stated inclusion.

Cross sections of multidimensional sets. If A is in $\sigma(\mathbf{S}_1\times\cdots\times\mathbf{S}_n)$, denote by $A_1(s)$ the section of A with first coordinate s:

$$A_1(s) = \{(s_2,\dots,s_n): (s,s_2,\dots,s_n)\in A\}.$$

This set is in the set σ algebra of subsets of $S_2\times\cdots\times S_n$ generated by $\mathbf{S}_2\times\cdots\times\mathbf{S}_n$ because the class of sets A for which this is true contains $\mathbf{S}_1\times\cdots\times\mathbf{S}_n$ and is a σ algebra of subsets of S. The corresponding assertions are true if more than one coordinate is fixed.

Right semiclosed intervals in spaces of infinite dimensionality. Section 2, Example (a), can be extended to an arbitrary infinite (not necessarily countable) dimensionality. For every point i of an arbitrary index set I, let S_i be a copy of $\mathbf{R}$ and let $\mathbf{S}_i$ be the algebra of finite unions of right semiclosed intervals of S_i. Define the space S as the class of all functions from I into $\mathbf{R}$. Let x_i be the ith coordinate function of S. If $i_1,\dots,i_n$ are index points and if A is a right semiclosed interval of $\mathbf{R}^n$, the set $\{(x_{i_1},\dots,x_{i_n})\in A\}$ is *an n-dimensional right semiclosed interval of S*. The algebra of finite unions of all such finite dimensional intervals is the infinite dimensional version of the algebra of finite unions of right semiclosed intervals of $\mathbf{R}^N$.

6. Monotone classes of sets

Monotone class definition. A class $\mathbf{S}$ of subsets of a space S is a *monotone class* if $\mathbf{S}$ contains the limit of every monotone sequence of its sets.

To each class $\mathbf{S}$ of subsets of a space corresponds a smallest monotone class $\mathbf{M}(\mathbf{S})$ containing $\mathbf{S}$ (cf. the corresponding proof for algebras in Section 3). The class $\mathbf{S}$ *generates* $\mathbf{M}(\mathbf{S})$.

Theorem. *Let* $\mathbf{S}$ *be a class of subsets of a space. Suppose that* $\mathbf{M}(\mathbf{S})$ *includes* $\tilde{\mathbf{S}}$ *and includes either the finite unions or the finite intersections of members of* $\mathbf{S}$. *Then* $\mathbf{M}(\mathbf{S}) = \sigma(\mathbf{S})$. *In particular,* $\mathbf{M}(\mathbf{S}) = \sigma(\mathbf{S})$ *if* $\mathbf{S}$ *is an algebra.*

Proof. Under the hypotheses of the theorem, the class $\mathbf{M}(\mathbf{S})$ contains the complements of its sets, because the class of sets in $\mathbf{M}(\mathbf{S})$ whose complements are in $\mathbf{M}(\mathbf{S})$ is a monotone class containing $\mathbf{S}$ and therefore must be $\mathbf{M}(\mathbf{S})$. To prove that $\mathbf{M}(\mathbf{S})$ is closed under finite unions if $\mathbf{M}(\mathbf{S})$ contains the finite unions of sets in $\mathbf{S}$, let B be in $\mathbf{S}$. The class Γ_B of sets A in $\mathbf{M}(\mathbf{S})$ for which $A \cup B$ is in $\mathbf{M}(\mathbf{S})$, contains $\mathbf{S}$, and is a monotone class. Hence $\Gamma_B = \mathbf{M}(\mathbf{S})$. Furthermore, the class of sets in $\mathbf{M}(\mathbf{S})$, whose union with each set in $\mathbf{M}(\mathbf{S})$ is in $\mathbf{M}(\mathbf{S})$, was just proved to contain $\mathbf{S}$, and is a monotone class, so is $\mathbf{M}(\mathbf{S})$. Thus $\mathbf{M}(\mathbf{S})$ is an algebra, necessarily a σ algebra because of the monotone class property, and therefore $\mathbf{M}(\mathbf{S}) = \sigma(\mathbf{S})$. This conclusion follows in the same way if it is supposed that $\mathbf{M}(\mathbf{S})$ contains the finite intersections rather than finite unions of members of $\mathbf{S}$.

Generation of the Borel sets by monotone sequential limits. According to Theorem 6, the class of Borel sets of a metric space S is the monotone class generated by the open sets, equivalently the monotone class generated by the closed sets.

The classes in the inclusion relations

$$\mathbf{G} \subset \mathbf{G}_\delta \subset \mathbf{G}_{\delta\sigma} \subset \mathbf{G}_{\delta\sigma\delta} \subset \cdots \tag{6.1}$$

are all Borel sets but in most applications the union Γ of these classes does not contain all the Borel sets.

Example: $S = \mathbf{R}$. In this case it can be shown that the monotone sequence (6.1) is strictly monotone and that Γ is a strict subclass of $\mathbf{B}(\mathbf{R})$. Moreover, it can be shown that the monotone sequence

$$\Gamma \subset \Gamma_\delta \subset \Gamma_{\delta\sigma} \subset \Gamma_{\delta\sigma\delta} \subset \cdots \tag{6.2}$$

is strictly monotone and that the union of these classes is a strict subclass of $\mathbf{B}(\mathbf{R})$. This procedure can be continued (*transfinite induction*) to obtain a *well-ordered* uncountable strictly increasing succession of classes of Borel sets containing all the Borel sets of $\mathbf{R}$. (A corresponding approach starts with the sequence

$$\mathbf{F} \subset \mathbf{F}_\sigma \subset \mathbf{F}_{\sigma\delta} \subset \mathbf{F}_{\sigma\delta\sigma} \subset \cdots \tag{6.3}$$

instead of (6.1).) This analysis of Borel sets will not be used in this book.

III
Set Functions

The point of this book is the study of countably additive set functions, and the preceding chapters have set up the appropriate context by providing an introductory analysis of classes of subsets of an abstract space. This chapter introduces the set functions to be studied.

1. Set function definitions

Let λ be a function from some class $\mathbf{S}$ of subsets of a space S into $\bar{\mathbf{R}}$.

(a) λ is *monotone increasing* [*decreasing*] if $\lambda(A) \le \lambda(B)$ [$\lambda(A) \ge \lambda(B)$], whenever $A \subset B$ and both sets are in $\mathbf{S}$.

In (b) and (c) it is supposed that $\emptyset \in \mathbf{S}$ and that $\lambda(\emptyset) = 0$.

(b) λ is *finitely* [*countably*] *subadditive* if

$$\lambda(\bigcup A_\bullet) \le \sum \lambda(A_\bullet) \tag{1.1}$$

whenever $A_\bullet$ is a finite [infinite] sequence of sets, that, together with their union, are in $\mathbf{S}$, and $-\infty$ and $+\infty$ do not both appear in the summands.

(c) λ is *finitely* [*countably*] *additive* if (1.1) is true with equality whenever $A_\bullet$ is a disjunct finite [infinite] sequence of sets that, together with their union, are in $\mathbf{S}$, and $-\infty$ and $+\infty$ do not both appear in the summands.

In checking finite additivity or subadditivity, it is sufficient to consider unions of only two sets.

Measures and signed measures. A countably additive set function from an algebra into either $[-\infty,+\infty)$ or $(-\infty,+\infty]$ is a *signed measure*, a *measure* if the range space is $\bar{\mathbf{R}}^+$. If λ is a measure defined on a σ algebra $\mathbf{S}$ of subsets of S, the triple $(S,\mathbf{S},\lambda)$ is a *measure space*, and the sets in $\mathbf{S}$ are *measurable*, or λ *measurable* if it is necessary to identify the measure. In particular, if $\lambda(S)=1$, a measure space is a *probability space*, and λ is a *probability measure*. In probability contexts, the measurable sets are sometimes called *events*.

A measure space S and its measure λ are *finite* if $\lambda(S) < +\infty$, and are *σ finite* if S is a countable union of sets of finite measure. In view of the representation II(1.1) of a countable union as a disjunct countable union, it is no further restriction on the condition for σ finiteness to demand that the union be a disjunct union.

Null sets, carriers, and supports. A measurable set of measure 0 is *null* or, more specifically, λ *null*. An assertion about points of a measure space holds *almost surely*, or *almost everywhere*, on the space, if true up to a null set, in the sense that the set where the assertion is false is a null set. A subset of a null set may not be measurable and therefore may not be a null set but (see Section IV.1) the domain of definition of a measure can be extended to remove this somewhat awkward complication. A measure is *carried* by a set if the set is measurable and has a null complement.

Borel measures. A *Borel measure* is a measure λ defined on the class of Borel subsets of a metric space. If the space is separable there is a largest open λ null set, the union of the λ null balls having centers at the points of a countable dense set and having rational radii. The complement of this open set is the smallest closed carrier of λ. This uniquely defined closed carrier is *the closed support* of λ.

Monotonicity and subadditivity. Finite additivity of a positive set function λ, defined on a set algebra **S**, implies that λ is monotone increasing, because if A and B are sets in **S** and if $A \subset B$,

$$\lambda(B) = \lambda(A) + \lambda(B-A) \geq \lambda(A). \tag{1.2}$$

Furthermore this set function λ is finitely subadditive, because if sets C and D are in **S**,

$$\lambda(C \cup D) = \lambda(C) + \lambda(D-C) \leq \lambda(C) + \lambda(D). \tag{1.3}$$

A slight extension of the argument, applying equality II(1.1), shows that *a measure on a set algebra is countably subadditive.*

Countable additivity. The added condition of countable additivity imposed on a finite valued finitely additive set function λ, defined on an algebra **S**, can be given the following equivalent forms.

(a) For a disjunct sequence $A_\bullet$ of sets in **S**, with union in **S**, (1.1) is true with equality.

(b) For an increasing sequence $B_\bullet$ of sets in **S** with limit B in **S**, $\lim \lambda(B_\bullet) = \lambda(B)$.

(c) For a decreasing sequence $B_\bullet$ of sets in **S**, with limit $\emptyset$, $\lim \lambda(B_\bullet) = 0$.

For example, to see that (a) implies (b), write B as a union:

$$B = B_1 \cup \bigcup_1^\infty (B_{n+1} - B_n).$$

Conversely (b) implies (a) because a countable union is the limit of the monotone increasing sequence of partial unions.

The added condition of countable additivity imposed on a finitely additive, not necessarily finite valued positive set function, defined on an algebra **S**, can be given the following equivalent forms: (a) and (b) as above, but (c) is replaced by

(c') For a decreasing sequence $B_\bullet$ of sets in **S** with limit $\emptyset$, $\lim \lambda(B_\bullet) = 0$ if $\lambda(B_1) < +\infty$.

2. Extension of a finitely additive set function

The following lemma will be useful in the construction of product measures on product spaces. The properties of $\mathbf{S}_0$ in the lemma are modeled on the properties of classes of product subsets of the product of a finite number of spaces.

Lemma. *Let $\mathbf{S}_0$ be a collection of subsets of a space S. Suppose that the intersection of two (and therefore every finite number) of sets in $\mathbf{S}_0$ is in $\mathbf{S}_0$, and suppose that the complement of a set in $\mathbf{S}_0$ is a finite disjunct union of sets in $\mathbf{S}_0$, so that $\sigma_0(\mathbf{S}_0)$ is the class of finite unions of sets in $\mathbf{S}_0$. Let λ_0 be a finitely additive set function on $\mathbf{S}_0$, with values in either $(-\infty,+\infty]$ or $[-\infty,+\infty)$. There is then a unique finitely additive extension of λ_0 to $\sigma_0(\mathbf{S}_0)$.*

Proof. If A is a finite union of sets in $\mathbf{S}_0$, A can be expressed as a finite disjunct union of sets in $\mathbf{S}_0$, say $A = \cup A_\bullet$. Define $\lambda(A) = \sum \lambda_0(A_\bullet)$. To prove that λ as so defined is independent of the choice of representation of A as a finite disjunct union of sets in $\mathbf{S}_0$, suppose that $\cup B_\bullet$ is another finite disjunct union of sets in $\mathbf{S}_0$ with union A. Then

$$A = \cup A_\bullet = \cup B_\bullet = \cup_{j,k} A_j \cap B_k,$$

and therefore

$$\sum \lambda_0(A_\bullet) = \sum_j \sum_k \lambda_0(A_j \cap B_k) = \sum_k \sum_j \lambda_0(A_j \cap B_k) = \sum \lambda_0(B_\bullet),$$

as was to be proved. Thus λ_0 has been given the required extension, obviously finitely additive.

3. Products of set functions

The following theorem will be useful, for example, in developing area in two dimensions from length in one dimension.

Theorem. *For* $i=1,\ldots,n$, let $\mathbf{S}_i$ *be an algebra of subsets of a space* S_i, *and define* $\mathbf{S}_0 = \mathbf{S}_1\times\cdots\times\mathbf{S}_n$, $\mathbf{S}=\sigma_0(\mathbf{S}_0)$. *If* λ_i *is a finitely additive positive set function on* $\mathbf{S}_i$, *for* $i = 1,\ldots,n$, *then there is a finitely additive set function* λ *on* $\mathbf{S}$ *for which*

$$(3.1) \qquad \lambda(A_1\times\cdots\times A_n) = \prod_1^n \lambda_\bullet(A_\bullet) \quad (A_i \in \mathbf{S}_i,\ i = 1,\ldots,n).$$

Proof. Define λ by (3.1) on $\mathbf{S}$. According to Section **II**.5, each set in $\mathbf{S}$ can be expressed as a finite disjunct union of sets in $\mathbf{S}_0$. Define λ on such a union by additivity. The only question is whether this definition gives a value independent of the representation of the given set as a disjunct product set union. In proving the desired independence, it is sufficient, according to Lemma 2, to prove this independence for product set unions in $\mathbf{S}_0$. The proof is by induction. The independence is trivial when $n=1$. If $n > 1$, suppose independence has been proved for $n-1$, and suppose that

$$(3.2) \qquad A_1\times\cdots\times A_n = \bigcup_1^k (B_\bullet\times C_\bullet),$$

where $B_i \in \mathbf{S}_1\times\cdots\times\mathbf{S}_{n-1}$, $C_i \in \mathbf{S}_n$ for $i=1,\ldots,k$ and the union is disjunct. According to the induction hypothesis, there is a finitely additive set function ν on $\sigma_0(\mathbf{S}_1\times\cdots\times\mathbf{S}_{n-1})$ satisfying

$$(3.3) \qquad \nu(D_1\times\cdots\times D_{n-1}) = \prod_1^{n-1} \lambda_\bullet(D_\bullet), \quad (D_i \in \mathbf{S}_i,\ i=1,\ldots,n-1).$$

It is to be proved that

$$(3.4) \qquad \nu(A_1\times\cdots\times A_{n-1})\,\lambda_n(A_n) = \sum_1^k \nu(B_i)\,\lambda_n(C_i).$$

According to Section **II**.3, there are 2^k pairwise disjoint sets in $\mathbf{S}_n$, with the property that each set C_i is a disjunct union of some of these sets. If in each term $B_i\times C_i$ in (3.2), the set C_i is expressed in terms of these sets, $B_i\times C_i$ is thereby expanded into a disjunct union of product sets with common first factor set B_i. The ith summand in (3.4) is thereby expanded into several summands that have sum $\nu(B_i)\lambda_n(C_i)$, because λ_n is additive. If this expansion is carried through for all i, the value of the sum in (3.4) is not changed. Suppose then that this expansion has already been carried through, yielding a union in (3.2) in which two sets C_r and C_s are either identical or nonintersecting. If identical, the terms

$B_r \times C_r$, $B_s \times C_s$ can be combined into a single product set $(B_r \cup B_s) \times C_r$. When the terms in (3.2) are combined in this way, the sum in (3.4) is unchanged, because ν is additive. After having made these changes, $C_1, C_2, \ldots$ are pairwise disjoint and their union must be A_n, whereas B_i must be $A_1 \times \cdots \times A_{n-1}$ for all i. The right side of (3.2) has become $\cup A_1 \times \cdots \times A_{n-1} \times C_\bullet$, and (3.4) is now trivial.

4. Heuristics on σ algebras and integration

Let $I_1, \ldots, I_n$ be pairwise disjoint intervals of **R** with union an interval I. Let f be a function from I into **R**, with value a_j on I_j. The Riemann integral of f on I is Σ_j $a_j\ \lambda(A_j)$, where $\lambda(A_j)$ is the absolute value of the difference between the coordinates of the endpoints of A_j. Riemann integration theory on **R** is based on this integration of functions constant on intervals. In fact the Darboux upper and lower sums for a function g (see Section VI.20), which approximate the Riemann integral of g, are the Riemann integrals of functions constant on intervals. Integration in the context of measure theory involves analogous sums, but is based not on functions constant on intervals, but on functions constant on sets of some σ algebra of sets. The details of this integration will be given later, but in this chapter preliminary definitions of integrals will be given in special contexts to clarify the general case.

5. Measures and integrals on a countable space

Suppose that S is a countable space, written as a finite or infinite sequence $s_\bullet$, and define $\mathbf{S}=2^S$. A measure λ on **S** is determined by its values on singletons: if $\lambda(\{s_i\})=p_i$ then

$$\lambda(A) = \sum_{s_i \in A} p_i.$$

Observe that if $\Sigma p_\bullet = +\infty$, but if each summand is finite, the sequence $B_\bullet$ with $B_n = \{s_n, s_{n+1}, \ldots\}$ is a decreasing sequence of sets with limit $\emptyset$, even though $\lambda(B_n) = +\infty$ for all n. This example justifies the Section 1(c') finiteness condition.

If f is a function from this countable space S into **R**, it is natural to define the integral of f on S as $\Sigma f(s_\bullet)p_\bullet$ if the sum converges absolutely, and this in fact is a special case of the final definition of an integral to be given in Section VI.4.

Adaptation of integrands to σ algebras. Let S be the finite or infinite sequence $s_\bullet$ with at least two points, define **S** as the σ algebra of those subsets of S that contain either both or neither of the two points s_1, s_2, and let $p_2, p_3, \ldots$ be positive but not necessarily finite numbers. Define $\lambda(\{s_j\}) = p_j$ for $j > 2$, and define λ on the two-point set $\{s_1, s_2\}$ as p_2. These definitions, together with countable additivity, determine λ on **S**. If f is a function from S into $\bar{\mathbf{R}}^+$, and if $f(s_1) \neq f(s_2)$, there is no natural definition of the integral of f with respect to λ on S, because λ is not defined on the singletons $\{s_1\}$ and $\{s_2\}$. The difficulty is that, as far as **S** is concerned, the point pair $\{s_1, s_2\}$ is an indivisible atom of the measure space. Thus integration theory in this context is forced to consider only those integrands f with $f(s_1) = f(s_2)$; for such a function, the natural definition of the integral is

$$\int_S f\,d\lambda = f(s_1)p_2 + \sum_{j>2} f(s_j)p_j, \tag{5.1}$$

when the series converges absolutely. The point is that an integrand must assume each of its values on a measurable set. This fact leads to the general concept of a function adapted to the class of measurable sets, a *measurable* function, to be defined and discussed in Section V.1. At the present stage the following definition is adequate.

Measurability definition for a function with a countable range space. Let $(S', 2^{S'})$ be a countable measurable space, and consider functions from a measurable space $(S, \mathbf{S})$ into S'. Such a function y is *measurable* if it assumes each of its values on a measurable set, that is, if a' is a point of S' then $\{y = a'\} \in \mathbf{S}$, equivalently, $\{y \in A'\} \in \mathbf{S}$ whenever A' is a subset of S'. The function f in the preceding paragraph, from S into $\bar{\mathbf{R}}^+$, is measurable if and only if $f(s_1) = f(s_2)$.

If $(S, \mathbf{S})$ is provided with a probability measure, a measurable function is a *random variable* in probability terminology.

6. Independence and conditional probability (preliminary discussion)

Let $(S, \mathbf{S}, P)$ be an arbitrary probability space. All subsets of S considered below are in **S**, that is, are measurable.

Independence of sets. Sets $A_1, \ldots, A_n$ in **S** are *mutually independent* if

$$P\{B_1 \cap \cdots \cap B_n\} = P\{B_1\} \cdots P\{B_n\}, \tag{6.1}$$

for every one of the 2^n choices of the n-tuple $B_1, \ldots, B_n$, where each set B_j is either A_j or $\tilde{A}_j$.

This mutual independence implies that for each choice of $B_\bullet$, these sets are also mutually independent. Moreover the sets of any subcollection of $A_\bullet$ are mutually independent. (For example, write (6.1) with B_n replaced by its complement, and then add the new equation to the original one, to find that $A_1, \ldots, A_{n-1}$ are mutually independent.). In particular, sets A_1 and A_2 are mutually independent if $P\{A_1 \cap A_2\} = P\{A_1\}P\{A_2\}$ because in this special case trivial evaluations show that the pairs $(A_1, \tilde{A}_2)$, $(\tilde{A}_1, A_2)$, and $(\tilde{A}_1, \tilde{A}_2)$ also satisfy this product relation. A null set is independent of every set, as is also the complement of a null set.

Infinitely many sets are *mutually independent* if the sets of every finite subcollection are mutually independent.

Mutual independence of σ algebras. The σ algebras of a collection of σ algebras of measurable sets are *mutually independent* if, whenever a set is chosen from each σ algebra, these sets are mutually independent. Let $\mathbf{S}_1, \ldots, \mathbf{S}_4$ be mutually independent σ algebras of measurable sets. Then $\sigma(\mathbf{S}_1, \mathbf{S}_2)$ *and* $\sigma(\mathbf{S}_3, \mathbf{S}_4)$ *are mutually independent σ algebras.* To see this, let B be the intersection of a set in $\mathbf{S}_3$ with one in $\mathbf{S}_4$. The class Γ of sets in $\sigma(\mathbf{S}_1, \mathbf{S}_2)$ independent of B is a monotone class closed under finite disjunct unions and complementation, and Γ includes every intersection of a set in $\mathbf{S}_1$ with one in $\mathbf{S}_2$ Since finite unions of such intersections can be written as disjunct unions of the same type, and in fact constitute a set algebra, Γ must be $\sigma(\mathbf{S}_1, \mathbf{S}_2)$. Thus each set in $\sigma(\mathbf{S}_1, \mathbf{S}_2)$ is independent of B. An application to $\sigma(\mathbf{S}_3, \mathbf{S}_4)$ of the reasoning just used shows that every set in $\sigma(\mathbf{S}_3, \mathbf{S}_4)$ is independent of every set in $\sigma(\mathbf{S}_1, \mathbf{S}_2)$, as was to be proved. More generally, an obvious further elaboration of this proof shows that if $\{\mathbf{S}_i,\ i \in I\}$ is a family of mutually independent σ algebras, and if $\{I_\alpha,\ \alpha \in \Xi\}$ are disjoint subsets of the index set I, then $\left\{\sigma\{\mathbf{S}_i,\ i \in I_\alpha\},\ \alpha \in \Xi\right\}$ are mutually independent σ algebras.

Independence of random variables. In particular in this discussion let S' be a countable space, and consider random variables (= measurable functions) from S into S' as defined in Section 5. The random variables of a collection of these random variables are *mutually independent* if, whenever $y_1, \ldots, y_n$ are finitely many random variables in the collection and $a_1', \ldots, a_n'$ are points of S', the sets $\{y_1 = a_1'\}, \ldots, \{y_n = a_n'\}$ are mutually independent. This condition implies that if $A_1', \ldots, A_n'$ are subsets of S' the sets $\{y_1 \in A_1'\}, \ldots, \{y_n \in A_n'\}$ are mutually independent. The general definition that underlies these special cases (keeping S' countable at this stage, however) is the following. If $y_\bullet$ is any collection of random variables, measurable sets of the form $\{y_t \in A'\}$, with y_t in the collection, and A' a subset of S', generate a σ algebra, denoted by $\sigma(y_\bullet)$, and all questions of independence of random variables are referred to the corresponding σ algebras. Thus two random variables y and z are mutually independent if and only if $\sigma(y)$ and $\sigma(z)$ are mutually independent σ algebras; similarly two families $\{y_\bullet\}$ and $\{z_\bullet\}$ of random variables are mutually independent if and only if $\sigma(y_\bullet)$ and $\sigma(z_\bullet)$ are independent σ algebras, and so on. In particular, the sets of a collection of measurable sets are mutually independent if and only if their indicator functions are mutually independent.

Independent events. Recall that in probability applications, measurable sets are sometimes called "events." Nonmathematical events that are independent of each other in a nonmathematical sense correspond in mathematical models to mathematically independent measurable sets. For example, in the coin tossing analysis to be given in Section 9, the events *heads on the first toss* and *tails on the third toss* are thought of as independent real-world events, and the corresponding measurable sets in the mathematical model are mathematically independent.

Conditional probability. Let $(S,\mathbf{S},P)$ be a probability space, and let A be a measurable nonnull set. A new probability measure $B \to P\{B|A\}$ (read "the conditional probability of B given A") is defined by

$$P\{B|A\} = P\{B \cap A\}/P\{A\}. \tag{6.2}$$

In simple contexts one can interpret such conditional probabilities for fixed A as defining a new context, based on replacing S by A, replacing $\mathbf{S}$ by the class of measurable subsets of A, and replacing P by the restriction of P to this class, normalized to make the restriction a probability measure of sets B. However, in most contexts it is preferable to keep S and $\mathbf{S}$, so that (6.2) defines a new probability measure on $(S,\mathbf{S})$, carried by A. Observe that sets A and B are mutually independent if and only if either A is null, or A is not null and $P\{B|A\}=P\{B\}$. The innocent looking conditional probability concept, when formulated in a more general context (see Section XI.2), has had a profound influence and unexpected mathematical applications, both inside and outside probability theory.

Expectation and conditional expectation. If $S=\{s_1,s_2,...\}$ is a countable space, and if a probability measure is defined on the σ algebra 2^S by setting $P\{s_i\}=p_i$ with $p_i \geq 0$ and $\Sigma p_\bullet = 1$, the integral of a numerically valued function f on S, defined in Section 5, is commonly written $E\{f\}$ by probabilists (read "expectation of f"). If this expectation exists, and if $P\{A\}>0$, the integral of f with respect to the conditional measure $P\{\bullet|A\}$ is written $E\{f|A\}$ (read "expectation of f given A").

7. Dependence examples

Let S_1 be the set $1,...,N$ of integers, define $S=S_1{}^m$ as the space of m-tuples of points of S_1, and let x_k be the kth coordinate function of S. The following lists several ways of assigning each singleton of S a measure value, in order to define a probability measure P on 2^S.

(a) Let $P^{(k)}$ be a measure on the kth factor space, say $P^{(k)}\{j\} = q_j^{(k)}$, where $q_j^{(k)} \geq 0$ and $\Sigma\, q_\bullet^{(k)} = 1$, and assign to the singleton $(j_1,...,j_m)$ of S the measure $\prod_k q_{j_k}^{(k)}$. For each pair k, j, the subset $\{x_k=j\}$ of S contains N^{m-1} points and its probability is defined by

(7.1) $$P\{x_k=j\}\} = q_j^{(k)}.$$

The sets $\{x_1=j_1\},...,\{x_m=j_m\}$ are mutually independent subsets of S for every choice of $j_1,...j_m$, and corresponding to this fact, the measure P on S is what will be defined in Section IV.11 as the *product measure* of the measures $P_1,...,P_m$. 'As the following examples illustrate, product measures are not the only way to define measures on product spaces.

(b) **Transition probabilities and stochastic matrices.** A matrix of positive elements with row sums 1 is a *stochastic matrix.* One kind of random variable dependence, *Markov dependence,* is characterized by the fact that in a sequence of random variables, probabilities for the nth, conditioned by the values of all the preceding random variables, actually depend only on the last preceding random variable, not on those farther back. More precisely, in the discrete context of (a), choose N positive numbers $p_\bullet$, with sum 1, as *initial probabilities,* setting

(7.2) $$P\{x_1=i\} = p_i, \quad i = 1,...,N.$$

Next choose an $N\times N$ stochastic matrix $(p_{ij}^{(1)})$, the matrix of first step *transition probabilities*, that is, define $\{x_1,x_2\}$ probabilities by

(7.3) $$P\{x_1=i, x_2=j\} = p_i p_{ij}^{(1)}.$$

Observe that summing over j in (7.3) yields (7.2) and that

(7.4) $$P\{x_2=j|x_1=i\} = p_{ij}^{(1)}$$

when $p_i > 0$. If $m = 2$, (7.3) provides the most general probability measure on S. If $m > 2$, choose another stochastic matrix $(p_{jk}^{(2)})$ as matrix of second step probabilities, setting

(7.5) $$P\{x_1=i, x_2=j, x_3=k\} = p_i p_{ij}^{(1)} p_{jk}^{(2)}.$$

Observe that summing over k in (7.5) yields (7.3), and observe the Markov property of the transition probabilities:

(7.6) $$P\{x_3=k|x_1=i, x_2=j\} = P\{x_3=k|x_2=j\} = p_{jk}^{(2)}$$

when $P\{x_1=i, x_2=j\} > 0$. The point of the Markov property is that the first conditional probability in (7.6) does not depend on i. If $m > 3$, go on with further transition matrices. When $m > 2$, this procedure does not furnish the most general probability measure on S. In fact to obtain the most general probability measure when $m = 3$, replace the transition matrix $(p_{jk}^{(2)})$ by a stochastic matrix $(p_{i,jk})$ (stochastic matrix in j,k for each i), which takes into account the value two steps back, thereby replacing (7.5) by

$$P\{x_1=i, x_2=j, x_3=k\} = p_i \, p_{ij}^{(1)} \, p_{i,jk}. \tag{7.7}$$

Summing over k in (7.7) yields (7.3). Equation (7.6) is replaced by

$$P\{x_3=k \mid x_1=i, x_2=j\} = p_{i,jk}, \tag{7.8}$$

when $P\{x_1=i, x_2=j\} > 0$. The Markov property is lost unless $p_{i,jk}$ does not depend on i.. This property will be defined in a more general context in Section **XI**.4.

Although the preceding discussion was based on random variables that were the coordinate functions of a product space, this special context was irrelevant. The important point was that Equations (7.2), (7.3), and so on were satisfied on whatever probability space the random variables $x_\bullet$ were defined. It is a typical feature of the special point of view of probability theory in measure theory that the probability relations between random variables define the context; the space on which the random variables are defined is irrelevant.

8. Inferior and superior limits of sequences of measurable sets

The combination of parts (a) and (c) of the following theorem is the "Borel-Cantelli Theorem." It is a historical accident that part (a) is usually stated only in probabilistic contexts.

Theorem. *Let $A_\bullet$ be a sequence of measurable sets of a measure space $(S,\mathbf{S},\lambda)$.*

(a) **(Cantelli)** *If* $\Sigma\, \lambda(A_\bullet) < +\infty$, *then*

$$\lambda\{\limsup A_\bullet\} = 0. \tag{8.1}$$

(b) $\lambda\{\liminf A_\bullet\} \le \liminf \lambda(A_\bullet) \le \limsup \lambda(A_\bullet)$, *and, if λ is a finite measure, the last term is at most* $\lambda\{\limsup A_\bullet\}$.

(c) **(Borel)** *If $(S,\mathbf{S},\lambda)$ is a probability space, and if $A_\bullet$ is a sequence of mutually* *independent measurable sets, then the condition* $\Sigma\,\lambda(A_\bullet) = +\infty$ *implies that* $\lambda\{\limsup A_\bullet\} = 1$, *and the condition* $\Sigma\,\lambda(A_\bullet) < +\infty$ *implies that* $\lambda\{\limsup A_\bullet\} = 0$.

Probability context. In the colorful language of probability, in which measurable sets are "events," (a) states for probability contexts a condition that (almost surely) an event occurs only finitely often and (c) states that, in the independence case, if the condition is not satisfied, the event is almost sure to occur infinitely often. Observe, however, that it has not yet been shown that (c) is a useful result, because no nontrivial example of an infinite sequence of mutually independent measurable sets has been exhibited. The set sequence $\emptyset, \emptyset, \ldots$ is an uninteresting trivial example that shows that (c) is not vacuous! A nontrivial example inspired by coin tossing and number theory will be exhibited in Section 9, but not justified until Section IV.14.

Proof. The definition of superior limit of a sequence of sets implies that, for all k,

$$\lambda\{\limsup A_\bullet\} \le \lambda\{\textstyle\bigcup_k^\infty A_\bullet\} \le \textstyle\sum_k^\infty \lambda(A_\bullet), \tag{8.2}$$

from which (a) is immediate. Similarly, (b) follows directly from the definitions of the inferior and superior limits of a sequence of sets. (The first inequality in (b) is a special case of Fatou's integration limit inequality, which will be proved in Section VI.8.) The second part of (c) is a special case of (a) but the following direct proof of (c) does not use (a). The probability that the event A_n does not occur for $n \ge k$, that is, the measure of the intersection $\bigcap_k^\infty \tilde{A}_\bullet$, is

$$\lim_{n\to\infty} [\lambda(\tilde{A}_k)\cdots\lambda(\tilde{A}_n)] = \textstyle\prod_k^\infty \big[1 - \lambda(A_\bullet)\big],$$

and the probability that the event occurs only finitely often is therefore

$$\lim_{k\to\infty} \textstyle\prod_k^\infty \big[1 - \lambda(A_\bullet)\big]. \tag{8.3}$$

Part (c) follows from the theory of infinite products:

(i) this infinite product converges if and only if $\Sigma\lambda(A_\bullet) < +\infty$;

(ii) if the product converges, the limit in (8.3) is 1;

(iii) if the product diverges, this limit is 0.

9. Mathematical counterparts of coin tossing

Coin tossing is not mathematics. A genuine human being of some sex, color, creed, and national origin tosses a piece of metal, giving it certain initial conditions and thereafter letting nature take its course (for which Newton devised a mathematical model). The coin comes to rest with either heads or tails showing, and the tosser, enslaved by mathematical notation, registers x_j as the result of the jth toss, setting $x_j = 1$ for heads, $x_j = 0$ for tails. She or he observes that $(x_1+\cdots+x_m)/m$ is usually close to 1/2 when m is large, and, more generally, observes that when m is large, and α is a specified n-tuple of 1's and 0's,

(number of times α appears in m successive n-tuples of tosses)/m

is usually close to 2^{-n}. The words "when m is large" suggest that, in a mathematical model of these observations, there is a limit theorem. In fact a Bernoulli proved such a limit theorem of course without the measure theoretic mathematics now available (which some old-fashioned probabilists are convinced only beclouds the context), about three hundred years ago. These observations suggest that in any mathematical model for coin tossing, whatever corresponds to a specified succession of n heads and tails at specified times should be assigned the measure 2^{-n}.

Desert now the interesting but imprecise real world in favor of the duller but more precise mathematical world, and construct n functions $x_1,\ldots,x_n$ on a probability measure space, imposing the following conditions: x_j is to have only two possible values 0 and 1; the measure of the set on which these functions take on any specified n–tuple of 0's and 1's is 2^{-n}. A trivial computation (addition) shows that then, for example, for each j, the measure of the set on which $x_j = 1$ is 1/2. There are many ways such a mathematical context can be constructed. Two important ones will be exhibited in this section.

First mathematical coin tossing model. This model will look simpler and be more interesting after Lebesgue measure is defined in Section **IV**.8. Every number s in the interval $(0,1]$ of $\mathbf{R}$ has a dyadic expansion $s = .x_1x_2,\ldots$, that is,

$$s = x_1 2^{-1} + x_2 2^{-2} + \cdots . \tag{9.1}$$

Here x_j is a function of s, with the possible values 1 and 0, made single valued by choosing the representation of s ending in a sequence of 1's rather than 0's at the dyadic points. Thus $\{x_1=0\}$ is the interval $(0,1/2]$, and, more generally, if values are assigned to $x_1,\ldots,x_n$, these functions will have those values on a right semiclosed interval of length 2^{-n}. If n is fixed, and if this length is assigned as measure to each of these intervals, additivity determines the measures assigned to the unions of these intervals. In this way, for each value of n, a probability measure space $(S,\mathbf{S}_n,P_n)$ has been defined, consisting of the interval $S = (0,1]$, together with the σ algebra $\mathbf{S}_n$ of unions of right semiclosed intervals of the form $((j-1)2^{-n}, j2^{-n}]$ for $j = 1,\ldots,2^n$, with measure P_n given by ordinary length. The functions $x_1,\ldots,x_n$ have the required properties and are mutually independent, corresponding to the notion of nonmathematical independence in actual coin tossing. Statements about the results of actual coin tossing can be translated into statements about this dyadic representation. The measures $P_\bullet$ are mutually consistent, in the sense that if say $m < n$, then $\mathbf{S}_m \subset \mathbf{S}_n$ and $P_m = P_n$ on $\mathbf{S}_m$. (This equality is trivial when $n=m+1$, and induction yields the general case.) Define P_∞, an additive set function on the algebra $S_\infty = \cup \mathbf{S}_\bullet$ of finite unions of dyadic right semiclosed subintervals of (0,1], by assigning its length to each subinterval, so that $P_\infty = P_m$ on $\mathbf{S}_m$. The set function P_∞ is finitely additive because each measure P_n is additive. (Actually P_∞ is countably additive but the proof is deferred until Section IV.14.) The space $(S, \mathbf{S}_\infty, P_\infty)$ is not a probability space, because $\mathbf{S}_\infty$ is not a σ algebra. Such probabilities as

$$P\{x_1 + \cdots + x_n \le cn^{1/2}\} \tag{9.2}$$

can be evaluated for all n, and the *central limit theorem*, which describes the limit of this probability when $n \to \infty$, can be proved, but probabilities of the two sets

$$\cup\{x_\bullet = 1\}, \qquad \left\{\lim\,(x_1 + \cdots + x_n)/n = 1/2\right\} \tag{9.3}$$

are not defined because these sets are not in $\mathbf{S}_\infty$. The sequence $\{x_\bullet = 1\}$ of sets is not a sequence of mutually independent sets in a probability space because $\mathbf{S}_\infty$ is not a σ algebra.

In nonmathematical probability language, the first event in (9.3) is that in an infinite sequence of tosses, heads occurs at least once; the second is that, in such an infinite sequence,

$$(\text{number of heads in } n \text{ tosses})/n$$

has limit 1/2. Although neither of these events is meaningful in actual coin tossing in the real world, because infinitely many tosses cannot be performed, a further development of the mathematical model makes the sets in (9.3) measurable, with probability 1 for both. More precisely, Lebesgue measure, developed in Section IV.9, makes it possible to extend P_∞ to a probability measure P on $\mathbf{S} = \sigma(\mathbf{S}_\infty)$. The Borel-Cantelli theorem can then be applied to the probability space $(S, \mathbf{S}, P)$ to obtain 1 for the probability that heads occurs infinitely often. The *strong law of large numbers* in Section XI.19, when applied to this probability space, yields 1 for the probability of the second set in (9.3), but far more elementary proofs yield this special result.

In 1909 Borel stressed the significance of such mathematical results in an influential paper (whose proofs were, however, defective even for that era).

Second mathematical coin tossing model. This model is more direct than the first. The notation used corresponds to that in the first model. In this model, let S_n be the space of the 2^n n-tuples of 0's and 1's, and determine the discrete probability measure P_n on the subsets of S_n by defining the measure of each singleton as 2^{-n}. The probability measure space $(S_n, 2^{S_n}, P_n)$ is a mathematical model for tossing a coin n times. Each succession of n tosses corresponds to a point of this model, whereas the succession of tosses corresponded to a dyadic interval in the first model. The space S_n is the n-fold product space $S_1{}^n$, and the measure P_n is the corresponding n-fold product measure, a simple special case of the set functions considered in Theorem 3. The jth coordinate function $x_j^{(n)}$ of S_n is the measurable function (alias *random variable*) corresponding to the result of the jth toss. It is perhaps a bit more obvious in this model than in the first that probability calculations in this simple context are counting problems: how many points of S_n have the property whose probability is to be calculated? The required probability is the number of those points multiplied by 2^{-n}. In order to define a model adapted simultaneously to all values of n, consider the space S of infinite sequences of 0's and 1's. Let x_j be the jth coordinate function of S, and define an additive set function P_∞ by setting, for each value of n, 2^{-n} as the probability of the subset of S whose first n coordinates form a specified n-tuple of 0's and 1's. The finite unions of these sets form an algebra $\mathbf{S}_\infty$, and probability is defined on this algebra by additivity. Note the parallelism between this model and the first one, which was based on dyadic expansions. This model has the same defect as the first one, in that probabilities like those in (9.2) are accessible, but not those in (9.3). In both models, in order to go on, an additive function on a set algebra must be extended to a measure on the generated σ algebra. This will be done in Chapter **IV**.

Finally, it is important to remember to keep mathematics and real life apart. It is an interesting facet of human behavior that, even when actual coin tossing is analyzed, the analysis has almost always been philosophical, ignoring the laws of mechanics, which quite unphilosophically govern the motion of real-world coins, under initial conditions imposed by real-world humans, and thereafter subject to the laws of motion of a real body falling under the influence of real gravity. The point is that the impossible-to-make-precise description of the actual results of coin tossing has a precise mathematical counterpart, in which mathematical theorems can be proved, some of which suggest real-world observational results.

10. Setwise convergence of measure sequences

Let $(S,\mathbf{S})$ be a measurable space, and let $\lambda_\bullet$ be a sequence of measures defined on $\mathbf{S}$. If $\lim \lambda_\bullet(A) = \lambda(A)$ exists for every measurable set A, then $\lambda_\bullet$ *converges setwise* to λ. Under certain added hypotheses stated in the following theorem, the limit set function λ is a measure. Part (b) is generalized to signed measures in Section **IX**.11, using a quite different type of proof.

Theorem. *Let $(S,\mathbf{S})$ be a measurable space, and let $\lambda_\bullet$ be a sequence of measures on $\mathbf{S}$, converging setwise to λ. Then λ is a measure if either of the following conditions is satisfied.*

(a) *$\lambda_\bullet$ is an increasing sequence.*

(b) **(Vitali-Hahn-Saks)** λ is *finite valued.*

Proof of (a). The limit set function λ is obviously finitely additive. If $A_\bullet$ is a disjunct sequence of measurable sets, with union A, monotonicity and finite additivity of λ imply

$$\lambda(A) \geq \sum_1^n \lambda(A_\bullet) \tag{10.1}$$

for all n, and therefore

$$\lambda(A) \geq \sum \lambda(A_\bullet). \tag{10.2}$$

On the other hand, if $c < \lambda(A)$, and if k is sufficiently large,

$$c < \lambda_k(A) = \sum \lambda_k(A_\bullet) \leq \sum \lambda(A_\bullet). \tag{10.3}$$

Therefore (10.2) is also true with the inequality reversed, that is, λ is a measure.

Proof of (b). The limit set function is again obviously finitely additive. If λ is not countably additive, there is a decreasing sequence $A_\bullet$ of measurable sets, with empty intersection, but with $\lim \lambda(A_\bullet) = \varepsilon > 0$. Define $\alpha_1 = \beta_1 = 1$; if α_j and β_j have been defined for $j \leq n$ choose α_{n+1} so large that $\alpha_{n+1} > \alpha_n$ and that

$$+\infty > \lambda_{\alpha_{n+1}}(A_{\beta_n}) \geq 7\varepsilon/8, \tag{10.4}$$

and then choose β_{n+1} so large that $\beta_{n+1} > \beta_n$ and that

$$\lambda_{\alpha_{n+1}}(A_{\beta_{n+1}}) \leq \varepsilon/8. \tag{10.5}$$

Define $B_n = A_{\beta_n} - A_{\beta_{n+1}}$. Then $\lambda_{\alpha_{n+1}}(B_n) \geq 3\varepsilon/4$, and it follows that, for $k \geq 1$,

$$\lambda_{\alpha_j}\left(\bigcup\{B_n : n \text{ even}, n \geq k\}\right) \geq 3\varepsilon/4 \qquad (j \text{ odd}, j > k). \tag{10.6}$$

Hence

$$\lambda\left(\bigcup\{B_n : n \text{ even}, n \geq k\}\right) \geq 3\varepsilon/4 \qquad (k \geq 1). \tag{10.7}$$

Similarly, (10.7) is true if the union is over odd values of n. Add these inequalities for even and odd values to obtain, since $B_\bullet$ is a disjunct sequence,

$$\lambda(A_{\beta_k}) = \lambda\left(\bigcup_k^\infty B_\bullet\right) \geq 3\varepsilon/2 \tag{10.8}$$

for all k. This inequality contradicts the definition of ε and thereby implies the truth of (b).

Observation. A glance at the proof of (b) shows that what has been proved, in order to prove countable additivity of λ, is that *if $\lambda_\bullet$ is a sequence of measures, with finite valued setwise limit λ, then* $\lim \lambda(A_\bullet) = 0$ *implies that* $\lim \lambda_\bullet(A_\bullet) = 0$. This result expresses a kind of uniformity of the setwise convergence, to be exploited in the proof of Theorem **IX**.10.

11. Outer measure

Outer measures, set functions for which the countable additivity measure hypothesis is weakened to an inequality, are fundamental in the analysis and construction of measures.

Definition. An *outer measure* on a space S is a function λ^* from 2^S into $\bar{\mathbf{R}}^+$ satisfying the following conditions:

(a) $\lambda^*(\emptyset) = 0$;
(b) λ^* is monotone increasing;
(c) λ^* is countably subadditive.

A set for which λ^* vanishes is *null*, or, more specifically, λ^* *null*. Condition (b) implies that a subset of a λ^* null set is also λ^* null, and condition (c) implies that a countable union of λ^* null sets is λ^* null.

Observe that if λ^* is an outer measure, and if c is a positive constant, the set function $\lambda^* \wedge c$ is also an outer measure. It will be useful later to modify a possibly infinite valued outer measure in this way.

Generation of an outer measure. The most common way of obtaining an outer measure is the following. Let $\mathbf{A}$ be an arbitrary collection of subsets of a space S, containing $\emptyset$, and let ϕ be a function from $\mathbf{A}$ into $\bar{\mathbf{R}}^+$ with infimum 0. If B is a subset of S, define $\lambda^*(B) = +\infty$ if B cannot be covered by a countable union of members of $\mathbf{A}$, and otherwise define $\lambda^*(B)$ by

$$\lambda^*(B) = \inf\left\{\sum \phi(B_\bullet):\ B \subset \cup B_\bullet,\ B_n \in \mathbf{A}\ \ (n \ge 1)\right\}. \tag{11.1}$$

It will now be shown that λ^* is an outer measure. Conditions (a) and (b) are obviously satisfied. To verify the countable subadditivity inequality (1.1), observe that (1.1) is trivial unless $\sum \lambda^*(A_\bullet) < +\infty$. In the latter case, choose $\varepsilon > 0$, and, for $n \ge 1$, choose a countable union $\bigcup A_{n\bullet}$ of sets in $\mathbf{A}$ in such a way that $A_n \subset \bigcup A_{n\bullet}$ and that

$$\sum \phi(A_{n\bullet}) \le \lambda^*(A_n) + \varepsilon 2^{-n}.$$

Then $\cup A_\bullet \subset \cup_{nj} A_{nj}$ and (c) is satisfied, because

$$\lambda^*(\bigcup A_\bullet) \le \Sigma_{nj}\,\phi(A_{nj}) \le \Sigma\,\lambda^*(A_\bullet) + \varepsilon. \tag{11.2}$$

The outer measure obtained in this way is the *outer measure generated by* **A** *and* ϕ. For example, if **A** consists only of the empty set, with $\phi(\emptyset) = 0$, the generated outer measure has value $+\infty$ for every other set.

The outer measure generated by a measure on an algebra. Suppose in (11.1) that **A** is an algebra, and that $\phi = \lambda$ is a measure on the algebra. In this case *the generated outer measure coincides with* λ *on* **A**. In fact, if $B \in$ **A**, the sum in (11.1) can only be decreased if each summand set B_n is decreased by making the covering sequence disjunct, and can possibly be further decreased by substituting $B_n \cap B$ for B_n. With these changes $B_\bullet$ becomes a disjunct sequence with union B, and the infimum in (11.1) is therefore $\lambda(B)$.

12. Outer measures of countable subsets of R

Let λ^* be the outer measure on **R** generated by the class of bounded open intervals together with the function ϕ with value $b\text{-}a$ on the interval (a,b). With this definition, every countable set $a_1,a_2,\ldots$ is null, contrary to unsophisticated intuition, because if $\varepsilon > 0$, and if B_n is an open interval containing a_n, of length $\varepsilon 2^{-n}$, then $B_\bullet$ covers the set $a_\bullet$, and $\Sigma\ \phi(B_\bullet) = \varepsilon$. More generally, a trivial modification of this proof shows that if $\phi\big((a,b)\big) = F(b)\text{-}F(a)$, where F is a monotone increasing function on **R**, then the outer measure generated using this choice of ϕ is 0 for every countable set of continuity points of F.

If instead of **R**, the space is the set S of rational numbers, and if λ^* is generated by the bounded "intervals" of the form $\{r \text{ rational}: a < r < b\}$, with ϕ defined on this interval as $b\text{-}a$, then $\lambda^*(S) = 0$. Thus it is not necessary that the generated outer measure majorizes ϕ on the sets where ϕ is defined.

13. Distance on a set algebra defined by a subadditive set function

If λ is a finitely subadditive function from a collection **S** of subsets of a space S into $\bar{\mathbf{R}}^+$, with $\lambda(\emptyset) = 0$, define two distances between sets A and B in **S** by

$$d_\lambda(A,B) = \lambda(A\,\Delta B), \qquad d_\lambda'(A,B) = \lambda(A\Delta B)\wedge 1. \tag{13.1}$$

Recall that if λ is subadditive, the function $\lambda\wedge 1$ is also subadditive, and that if d is a pseudometric, then $d\wedge 1$ is also. Each of these distance functions is positive, vanishes if its two arguments are the same, and satisfies the distance triangle inequality because the symmetric difference operator Δ satisfies the triangle inclusion relation I(2.5). Thus $(\mathbf{S},d_\lambda)$ is a pseudometric space if λ is finite valued, and $(\mathbf{S},d_\lambda')$ is a pseudometric space even without this finiteness condition. These spaces are metric if every λ null set is empty. The choice of the number 1 in (13.1) is arbitrary in the sense that the topology defined by $d_\lambda \wedge c$ is independent of the choice of the strictly positive constant c, and, if λ is finite valued, is the same as that defined by d_λ.

[The following standard procedure, noted for clarification, can be applied to obtain a metric space from $(\mathbf{S},d_\lambda)$ but this device will not be used in this book except in the discussion of L^2 as a Hilbert space. Let λ be a finite valued outer measure on the subsets of a space S and let $\mathbf{S}$ be the space of equivalence classes of subsets of S, putting two subsets in the same equivalence class when the distance between them is 0, that is, when they differ by a λ null set. The space of equivalence classes becomes a metric space if the distance between the equivalence class containing a set A and the equivalence class containing a set B is defined as $d_\lambda(A,B)$. The corresponding procedure is applicable to d_λ' and a not necessarily finite outer measure.]

d_λ' continuity of basic functions. In a pseudometric space, the pseudometric distance function is uniformly continuous from $\mathbf{S}\times\mathbf{S}$ into $\mathbf{R}^+$, as exhibited by an application of the triangle inequality (A to A_0 to B_0 to B):

$$|d_\lambda'(A,B)-d_\lambda'(A_0,B_0))| \le d_\lambda'(A,A_0)+d_\lambda'(B,B_0), \tag{13.2}$$

and the primes can be omitted if λ is finite valued.

The function $\lambda\wedge 1$ from $\mathbf{S}$ into $\mathbf{R}^+$ is uniformly continuous, because (13.2) reduces to the inequality $|(\lambda(A)\wedge 1)-(\lambda(A_0)\wedge 1)| \le d_\lambda'(A,A_0)$ when $B=B_0=\emptyset$. A trivial modification of the discussion yields uniform continuity of $\lambda\wedge c$ for an arbitrary strictly positive constant c, and continuity of (possibly infinite valued) λ.

Apply I(2.7) and I(2.8) to prove that the union and intersection operations from $\mathbf{S}\times\mathbf{S}$ into $\mathbf{S}$ are d_λ' uniformly continuous:

$$\begin{aligned} d_\lambda'(A\cup B,A_0\cup B_0) &\le d_\lambda'(A,A_0)+d_\lambda'(B,B_0), \\ d_\lambda'(A\cap B,A_0\cap B_0) &\le d_\lambda'(A,A_0)+d_\lambda'(B,B_0). \end{aligned} \tag{13.3}$$

The primes can be omitted if λ is finite valued.

14. The pseudometric space defined by an outer measure

The following theorem suggests that outer measures and measures endow their domains with useful topologies. These topologies will be exploited in Chapter IV.

Theorem. *Suppose either that*
(i) $(S,\mathbf{S},\lambda)$ *is a space S, with* $\mathbf{S} = 2^S$ *and* λ *an outer measure, or that*
(ii) $(S,\mathbf{S},\lambda)$ *is a measure space.*

Let $\mathbf{S}_f$ *be the subset of* $\mathbf{S}$ *on which* λ *is finite valued and let* $\mathbf{S}_0$ *be a subclass of* $\mathbf{S}$*, with* d_λ' *closure* $\bar{\mathbf{S}}_0$ (= d_λ *closure if* $\mathbf{S} = \mathbf{S}_f$). *Then under either* (i) *or* (ii):

(a) *The pseudometric space* $(\mathbf{S},d_\lambda')$ *is complete, and the class* $\mathbf{S}_f$ *is a closed subset of* $\mathbf{S}$*, at distance* 1 *from* $\mathbf{S}-\mathbf{S}_f$.

(b) *The* d_λ' *limit A of a* d_λ' *convergent sequence is in* $\mathbf{S}_f$, $[\mathbf{S}-\mathbf{S}_f]$ *if and only if all but a finite number of members of the sequence are in* $\mathbf{S}_f$, $[\mathbf{S}-\mathbf{S}_f]$; *up to a null set, there is a subsequence with limit A in the convergence sense of Section* I.3.

This theorem implies that the pseudometric space $(\mathbf{S}_f,d_\lambda)$ is complete.

Proof of (a). Let $A_\bullet$ be a d_λ' Cauchy sequence of sets in $\mathbf{S}$. Choose strictly positive integers $\alpha_1 < \alpha_2 < \cdots$ that are so large that $\lambda(A_m \Delta A_{\alpha_n}) = d_\lambda(A_m,A_{\alpha_n}) < 2^{-n}$ when $m > \alpha_n$. Then (see I(3.7)),

$$\text{(14.1)}\quad \lambda\big[\limsup A_{\alpha_\bullet} - \liminf A_{\alpha_\bullet}\big] = \lambda\big[\limsup_{n\to\infty}(A_{\alpha_n}\Delta A_{\alpha_{n+1}})\big] \le \lim_{k\to\infty}\sum_{n=k}^{\infty}\lambda(A_{\alpha_n}\Delta A_{\alpha_{n+1}}) = 0.$$

Define A as the superior or inferior limit (defined in Section I.3) of the sequence $A_{\alpha_\bullet}$, or any set in $\mathbf{S}$ and between these limits, so that

$$\text{(14.2)}\quad \bigcap_n^\infty A_{\alpha_\bullet} \subset A \subset \bigcup_n^\infty A_{\alpha_\bullet} = \Big(\bigcap_n^\infty A_{\alpha_\bullet}\Big) \cup \bigcup_{k=n}^\infty (A_{\alpha_k}\Delta A_{\alpha_{k+1}}).$$

Since (14.2) remains true if A is replaced by A_{α_n}, $d_\lambda'(A,A_{\alpha_n}) = \lambda(A\Delta A_{\alpha_n}) < 2^{-n+1}$, and

$$\text{(14.3)}\qquad d_\lambda'(A,A_m) \le d_\lambda'(A,A_{\alpha_n}) + d_\lambda'(A_{\alpha_n},A_m) < 2^{-n+2}, \qquad (m > \alpha_n).$$

Thus A is a d_λ' limit (a d_λ limit if $\mathbf{S} = \mathbf{S}_f$) of the Cauchy sequence $A_\bullet$, and therefore $(\mathbf{S},d_\lambda')$ is a complete pseudometric space. It is trivial that every set in $\mathbf{S}_f$ is at d_λ' distance 1 from every set in $\mathbf{S}-\mathbf{S}_f$, and that therefore $\mathbf{S}_f$ is d_λ' closed.

Proof of (b). The assertions in (b) are trivial in the light of (a) and the set convergence definition in Section I.3.

15. Nonadditive set functions

This book is devoted to additive set functions and their application to integration. Subadditive outer measures are introduced only to derive measures. Nevertheless, it is important to realize that nonadditive set functions are intrinsic in some contexts, for example, in classical and probabilistic potential theory. The following is a deceptively simple example of how a nonadditive set function can arise. Let $(S,\mathbb{S},P)$ be a probability space, let $f_\bullet$ be a sequence of functions from S into a space S' and let A' be a subset of S'. Define the function ϕ on certain subsets of S' (all hypotheses of function and set measurability are omitted here) by

$$(15.1)\quad \begin{aligned} \phi(A') &= P\Big\{ \bigcup \{f_\bullet \in A'\} \Big\} \\ &= P\{\text{at least one function of the sequence takes on a value in } A'\}. \end{aligned}$$

This function becomes more interesting if the context is glamorized! At each point of S the corresponding sequence of values of $f_\bullet$ is a *trajectory*. The value $\phi(A')$ is *the probability that a trajectory hits* A'. The set function ϕ is additive only in trivial contexts, for example, if the functions $f_1, f_2, \ldots$ are identical, but ϕ is subadditive in a strong sense which will not be discussed here.

IV
Measure Spaces

1. Completion of a measure space $(S,\mathbb{S},\lambda)$

The measure λ and its measure space are *complete* if subsets of λ null sets in **S** are also in **S**; if so, the subsets are λ null. According to the following theorem, if $(S,\mathbf{S},\lambda)$ is not complete it can be completed, that is, **S** can be enlarged to obtain a complete measure space.

Theorem. *There is a smallest σ algebra $\mathbf{S}^*$ satisfying the following conditions:*

(i) $\mathbf{S}^* \supset \mathbf{S}$;
(ii) *there is a complete measure on $\mathbf{S}^*$ whose restriction to* **S** *is* λ. $\mathbf{S}^*$ *consists of those sets A for which there are sets B and C in* **S** *satisfying the conditions*

(1.1) $$B \subset A \subset C, \quad \lambda(C - B) = 0.$$

This extension of λ is the *completion* of λ.

Proof. Let $\mathbf{S}^*$ be the class of subsets A of S for which there are sets B and C as described in the theorem. Then $\lambda(B) = \lambda(C)$. Morover if the pair of sets (B',C') has the same properties as the pair (B,C), then $\lambda(B') = \lambda(C') = \lambda(B) = \lambda(C)$ because $C - B' \in \mathbf{S}$, and this difference is a subset of the λ null set $(C-B)\cup(C'-B')$. Define $\lambda^*(A) = \lambda(B)$, a definition just proved to be independent of the choice of B and C in (1.1). The class $\mathbf{S}^*$ includes **S**, and $\lambda^* = \lambda$ on **S**. The class $\mathbf{S}^*$ is closed under complementation, because if B and C are as above, $\tilde{C} \subset \tilde{A} \subset \tilde{B}$ and $\tilde{B}-\tilde{C} = C-B$ is λ null. The class $\mathbf{S}^*$ is closed under countable unions, because if $A_\bullet$ is a sequence of sets in $\mathbf{S}^*$, and if, for $n \geq 1$, $B_n \subset A_n \subset C_n$, with B_n and C_n in **S**, and $\lambda(C_n - B_n) = 0$, then $\cup B_\bullet \subset \cup A_\bullet \subset \cup C_\bullet$, and

$$\lambda(\cup C_\bullet - \cup B_\bullet) \leq \lambda[\cup(C_\bullet - B_\bullet)] = 0.$$

Thus $\mathbf{S}^*$ is a σ algebra. Moreover λ^* is countably additive on $\mathbf{S}^*$, because if $A_\bullet$, $B_\bullet$, and $C_\bullet$ are as above, and if $A_\bullet$ is a disjunct sequence, then $B_\bullet$ is a disjunct sequence, and therefore

$$(1.2) \qquad \lambda^*(\cup A_\bullet) = \lambda(\cup B_\bullet) = \sum \lambda(B_\bullet) = \sum \lambda^*(A_\bullet).$$

Finally, if λ' is an arbitrary complete measure extension of λ, defined on a σ algebra $\mathbf{S}'$, $\mathbf{S}'$ must include the subsets of λ null sets and therefore must include A in (1.1). Hence $\mathbf{S}' \supset \mathbf{S}^*$, that is, λ^* is the minimal complete measure extension of λ.

2. Generalization of length on R

Consider the problem of defining the length of a subset of $\mathbf{R}$. To avoid problems connected with infinite length, consider only subsets of some closed finite interval J. Borel proposed the following procedure to extend the definition of length to a wide class of subsets of J. For a closed subinterval I of J, define $\lambda(I)$ as the positive difference between the coordinates of its endpoints. Next define λ for finite disjunct unions of intervals by additivity, next define λ by continuity successively on the class of sets that are limits of increasing sequences of sets on which λ is already defined, the class of sets that are limits of decreasing sequences on which λ is already defined, and so on, alternating between increasing and decreasing sequences. The point of this procedure was to define λ as a measure on the σ algebra $\mathbf{B}(J)$ of what are now called Borel sets, but the procedure proved to be impractical, and it was Lebesgue who first devised a procedure to extend length to the class $\mathbf{B}(J)$ and to $\mathbf{B}(\mathbf{R})$.

3. A general extension problem

A common measure theoretic context is the following. Let λ be a finite measure defined on an algebra $\mathbf{S}_0$ of subsets of some space, but suppose that $\mathbf{S}_0$ is not a σ algebra. If one wishes to treat problems involving repeated applications of countable unions and intersections of sets without going beyond the domain of λ, this measure must be extended to a measure on $\sigma(\mathbf{S}_0)$. Carrying through Borel's idea of extending the definition of λ first to $\mathbf{S}_{0\sigma}$, and then to $\mathbf{S}_{0\sigma\delta}$,... and so on (the last three words stand for transfinite procedures) to reach $\sigma(\mathbf{S}_0)$ would be difficult, but is unnecessary because, according to Sections 3-5, the sets in $\sigma(\mathbf{S}_0)$ are close to those in $\mathbf{S}_0$ in a sense that makes the extension easy. In fact this extension will be formulated in Theorems 3 and 4 as the extension of λ from a subset of a certain pseudometric space into the closure of the subset.

Theorem. *Let* $(S,\mathbf{S},\lambda)$ *be a finite measure space, let* $\mathbf{S}^\lambda$ *be the domain of the completion of* λ, *and suppose that* $\mathbf{S} = \sigma(\mathbf{S}_0)$, *where* $\mathbf{S}_0$ *is a set algebra. Then:*

(a) $\mathbf{S}^\lambda$ *is the* d_λ *closure of* $\mathbf{S}_0$.

(b) *If* $\varepsilon > 0$, *and if* A *is in* $\mathbf{S}^\lambda$, *there are sets* $A'(\varepsilon)$ *and* $A''(\varepsilon)$, *with the following properties:*

$$(3.1)\qquad A'(\varepsilon)\subset A\subset A''(\varepsilon),\ A'(\varepsilon)\in \mathbf{S}_{0\delta},\ \lambda\big(A-A'(\varepsilon)\big)<\varepsilon;$$

$$A''(\varepsilon)\in \mathbf{S}_{0\sigma},\ \lambda\big(A''(\varepsilon)-A(\varepsilon)\big)<\varepsilon.$$

(c) *The space* $(\mathbf{S}^\lambda, d_\lambda)$ *is a complete pseudometric space. If the* σ *algebra* $\mathbf{S}$ *is generated up to null sets by a countable collection of sets, then this pseudometric space is separable.*

If the measure λ *is not supposed finite, but* S *is a countable union of sets in* $\mathbf{S}_0$ *of finite measure, then* (3.1) *is still true.*

The meaning of the countable generation hypothesis in (c) is that there is a countable collection $\mathbf{S}_1$ of sets in $\mathbf{S}$ for which every set in $\mathbf{S}$, equivalently every set in $\mathbf{S}^\lambda$, differs by an $\mathbf{S}^\lambda$ null set from some set in $\sigma(\mathbf{S}_1)$. For example, this hypothesis is satisfied if there is a countable collection $\mathbf{S}_2$ of sets in $\mathbf{S}$ for which $\sigma(\mathbf{S}_2)=\mathbf{S}$.

It will be seen that the separability assertion in (c) is false for infinite valued measures (with d_λ repaced by d_λ') even if λ is σ finite.

It was pointed out in Section **III.13** that λ is a d_λ' uniformly continuous function from $\mathbf{S}$ into $\mathbf{R}$. Thus in going from $\mathbf{S}_0$ to $\mathbf{S}^\lambda$, the domain of λ is changed from a set algebra to its d_λ' closure, and the function λ is extended by continuity.

Proof of (a). The d_λ closure $\bar{\mathbf{S}}_0$ of $\mathbf{S}_0$ is a d_λ closed subset of the complete pseudometric space $(S,\mathbf{S}^\lambda,d_\lambda)$. The class $\bar{\mathbf{S}}_0$ is closed under complementation, because if $A_\bullet$ is a sequence in $\mathbf{S}_0$ with d_λ limit A, then $\tilde{A}_\bullet$ is a sequence in $\mathbf{S}_0$ with d_λ limit $\tilde{A}$. The class $\bar{\mathbf{S}}_0$ is closed under finite unions and intersections, because (Section III.13) if $A_\bullet$ and $B_\bullet$ are sequences with respective pseudometric limits A and B, then $A_\bullet\cup B_\bullet$ and $A_\bullet\cap B_\bullet$ are sequences with respective pseudometric limits $A\cup B$ and $A\cap B$. Thus $\bar{\mathbf{S}}_0$ is an algebra, even a σ algebra, because a countable union of measurable sets is the d_λ limit of the partial unions. Hence $\bar{\mathbf{S}}_0\supset\mathbf{S}$, and finally $\bar{\mathbf{S}}_0=\mathbf{S}^\lambda$ because the sets of $\mathbf{S}^\lambda$ differ from those of $\mathbf{S}$ by null sets.

Proof of (b). It is sufficient to show that $A''(\varepsilon)$ exists, because application of this result to $\tilde{A}$ yields $A'(\varepsilon)$, on complementation. Let λ^* be the outer measure generated by $\mathbf{S}_0$ and the restriction of λ to $\mathbf{S}_0$. Then $\lambda=\lambda^*$ on $\mathbf{S}_0$, and the existence of $A''(\varepsilon)$ is equivalent to the statement that $\lambda^*=\lambda$ on $\mathbf{S}^\lambda$. The outer measure λ^* is finitely additive on $\mathbf{S}^\lambda$ because if $A_\bullet$ and $B_\bullet$ are sequences in $\mathbf{S}_0$ with respective d_λ limits the disjunct pair of sets A, B, then the d_λ pseudometric continuity properties yield

$$(3.2)\qquad \lambda^*(A\cup B)=\lim\lambda^*(A_\bullet\cup B_\bullet)=\lim\,[\lambda^*(A_\bullet)+\lambda^*(B_\bullet-A_\bullet)]=\lambda^*(A)+\lambda^*(B).$$

The outer measure is countably additive on $\mathbf{S}^\lambda$ because, on the one hand, as an outer measure it is countably subadditive, and on the other hand, if $A_\bullet$ is a disjunct sequence of sets in $\mathbf{S}^\lambda$ with union A, then $\lambda^*(A) \geq \Sigma_1^n \lambda^*(A_\bullet)$, and therefore $\lambda^*(A) \geq \Sigma\lambda^*(A_\bullet)$. Thus, on $\mathbf{S}$, the two set functions λ and λ^* are measures that are equal on $\mathbf{S}_0$ and therefore equal on $\mathbf{S}$ because they are the unique d_λ continuous extensions of their common restriction to $\mathbf{S}_0$.

Proof of (c). According to Theorem **III.**13, $(\mathbf{S}^\lambda, d_\lambda)$ is a complete pseudometric space. Suppose that up to null sets $\mathbf{S}$ is countably generated by a sequence $B_\bullet$ of sets in $\mathbf{S}$. The algebra $\sigma_0(B_\bullet)$ is countable because it consists of finite unions of finite intersections of the members of $B_\bullet$ and $\tilde{B}_\bullet$. It is therefore sufficient to prove that the d_λ closure of $\sigma_0(B_\bullet)$ includes $\mathbf{S}^\lambda$. But according to what has just been proved, such a closure is $\mathbf{S}^\lambda$.

Proof of the last assertion of the theorem. It is no further restriction than that stated in the theorem to assume that $S = \cup S_\bullet$ is a disjunct countable union of sets in $\mathbf{S}_0$ of finite measure. Apply the theorem for finite λ separately to each measure $B \rightarrow \lambda(B \cap S_n)$ on the class of intersections with S_n of the sets in $\mathbf{S}^\lambda$, with $\mathbf{S}_0$ replaced by the algebra of intersections with S_n of the sets in $\mathbf{S}_0$. If A is in $\mathbf{S}^\lambda$ and if $\varepsilon > 0$, there is a set C_n satisfying the conditions

$$A \cap S_n \subset C_n \subset S_n, \quad C_n \in \mathbf{S}_{0\sigma}, \quad \lambda(C_n - A) < 2^{-n}\varepsilon,$$

then $\cup C_\bullet$ is a superset of A, in $\mathbf{S}_{0\sigma}$, and $\lambda(\cup C_\bullet - A) < \varepsilon$. Thus the part of (3.1) involving a superset of A in $\mathbf{S}_{0\sigma}$ is true. Application of this result to $\tilde{A}$, together with complementation, yield the other part of (3.1).

4. Extension of a measure defined on a set algebra

Theorem 3 shows how close the measurable sets of a finite measure space are to the sets of an algebra that generates the class of measurable sets. The following theorem uses this fact to show, under appropriate hypotheses, that a measure on an algebra can be extended to a measure on the generated σ algebra.

Theorem. (**Hahn-Kolmogorov**). *A σ finite measure λ_0 on an algebra $\mathbf{S}_0$ of subsets of a space has a unique extension to a σ finite measure on $\sigma(\mathbf{S}_0)$.*

Proof when λ_0 is finite valued. Define $\mathbf{S} = \sigma(\mathbf{S}_0)$ and define λ^* as the outer measure generated by $\mathbf{S}_0$ and λ. Then $\lambda = \lambda^*$ on $\mathbf{S}_0$. It will be shown that the d_{λ^*} closure $\bar{\mathbf{S}}_0$ of $\mathbf{S}_0$ includes $\mathbf{S}$ and that the restriction of λ^* to $\mathbf{S}$ is the desired extension of λ. In other words, the situation will be brought into the context of Theorem 3. The class $\bar{\mathbf{S}}_0$ is an algebra, according to the argument used in the proof of Theorem 3, except that λ^* takes the place of λ in the pseudometric. Moreover λ^* is countably additive on $\bar{\mathbf{S}}_0$ by the proof of countable additivity of λ^* in the proof of Theorem 3. It follows that $\bar{\mathbf{S}}_0$ is a σ algebra. Thus λ^* offers the desired extension of λ and is unique because, as remarked in Section 3, the extension from $\mathbf{S}_0$ to $\mathbf{S}$ necessarily extends λ by continuity.

Proof when λ_0 is not finite. If $S = \cup S_\bullet$ is a disjunct countable union of sets in $\mathbf{S}_0$ of finite measure, apply the theorem for spaces of finite measure separately to each measure $B \to \lambda_0(B \cap S_n)$ on the class of intersections with S_n of the sets in $\mathbf{S}_0$ to obtain an extension of λ_0 to $\sigma(\mathbf{S}_0)$. Since the separate extensions are unique, the overall extension is unique.

5. Application to Borel measures

Theorem. *Let S be a metric space, let λ be a measure extended by completion from $\mathbf{B}(S)$ to $\mathbf{B}^\lambda(S)$, and suppose that S is a countable union of open sets of finite measure. Then if A is a measurable set and $\varepsilon > 0$, there is a closed subset $A'(\varepsilon)$ of A and an open superset $A''(\varepsilon)$ satisfying the conditions* $\lambda\big(A - A'(\varepsilon)\big) < \varepsilon$, $\lambda\big(A''(\varepsilon) - A\big) < \varepsilon$.

Proof when λ is finite valued. Apply Theorem 3 with $\mathbf{S}_0$ the set algebra generated by the class of open subsets of S. According to Theorem 3, the assertions involving $A'(\varepsilon)$ and $A''(\varepsilon)$ are true except that unfortunately these sets, described in Theorem 3, are not respectively closed and open in the present context. To get an open version of $A''(\varepsilon)$ it is sufficient to show that if A is in $\mathbf{S}_{0\sigma}$ there is an open superset of A that can be chosen to have measure arbitrarily near that of A. Now the sets of $\mathbf{S}_{0\sigma}$ are disjunct countable unions of subsets of S of the form $B \cap C$, where B is open and C is closed. It is therefore sufficient to show that a closed set C has open supersets of measure arbitrarily close to that of C, and since a closed set C is a countable intersection of open supersets, this assertion is true. To get a closed version of $A'(\varepsilon)$, apply the result just obtained to $\tilde{A}$. Thus the theorem is true when λ is a finite measure.

Proof when λ is not finite valued. If B is an open subset of S of finite measure apply the present theorem for finite measures to the restriction of λ to subsets of B and thereby find that if A is a measurable subset of B, then there are open subsets of B that are supersets of A of measure arbitrarily close to $\lambda(A)$. This fact will now be applied in the present context, in which $S = \cup S_\bullet$ is a countable union of open sets of finite measure, to find a set $A''(\varepsilon)$ with the desired properties. If A is a measurable subset of S, and if $\varepsilon > 0$, there is an open subset A_n'' of S_n for which $A \cap S_n \subset A_n''$ and $\lambda(A_n'' - A \cap A_n) < \varepsilon 2^{-n}$. The set $\cup A_n''$ satisfies the conditions for $A''(\varepsilon)$. Apply this result to $\tilde{A}$ to find a closed subset of A satisfying the conditions for $A'(\varepsilon)$.

6. Strengthening of Theorem 5 when the metric space S is complete and separable

The following theorem strengthens Theorem 5 in the more restrictive context of a complete separable metric space.

Theorem. *Let S be a separable metric space and let λ be a finite measure extended by completion from* $\mathbf{B}(S)$ *to* $\mathbf{B}^*(S)$.

(a) *The space* $(\mathbf{B}^*(S), d_\lambda)$ *is a complete separable pseudometric space.*

(b) (**Prohorov**) *If S is complete then Theorem 5 is true with $A'(\varepsilon)$ compact, that is, the measure of a measurable set is the supremum of the measures of its compact subsets.*

Proof of (a). The class $\mathbf{S}_0$ of finite unions of balls with rational radii, with centers a countable dense subset of S, is countable. Each open subset of S is the limit of a monotone increasing sequence of sets in $\mathbf{S}_0$, and therefore is a d_λ limit of $\mathbf{S}_0$. Hence the class of open subsets of S is separable in the d_λ pseudometric. According to Theorem 5, the class of open sets is d_λ dense in the class of measurable sets, and therefore the latter class is d_λ separable.

Proof of (b). In view of Theorem 5, it is sufficient to prove that if A is a closed subset of S, then

$$\lambda(A) = \sup\{\lambda(F): F \subset A, F \text{ compact}\}. \tag{6.1}$$

In fact it is sufficient to prove (6.1) with $A = S$, because A provided with the S metric is itself a complete separable space, and $\mathbf{B}(A)$ is the class of those Borel sets relative to S that are subsets of A. If $\lambda(S) = 0$ the theorem is trivial. If $\lambda(S) > 0$, choose $c < \lambda(S)$. The space S is the union $\bigcup B_{1\bullet}$ of countably many closed sets of diameter ≤ 1 (say the closed balls of diameter 1 with centers at the points of a countable dense subset of S). Choose enough sets from $B_{1\bullet}$, with union B_1, to get $\lambda(B_1) > c$. If closed sets $B_1, \ldots, B_{n-1}$ have been defined, with each set B_j a finite union of closed sets of diameter $\leq 1/j$ and $\lambda(B_1 \cap \cdots \cap B_{n-1}) > c$, go on to choose finitely many closed sets $B_{n\bullet}$ of diameter $\leq 1/n$ and union B_n, in such a way that $\lambda(B_1 \cap \cdots \cap B_n) > c$. The closed set $\bigcap B_\bullet$ has measure at least c and is compact because the set has the property that, for every strictly positive integer n, the set B can be covered by finitely many closed sets of diameter $\leq 1/n$. Hence (b) is true. (This covering property is a standard compactness criterion: the set B is compact because if C is an infinite subset of B, there must be an infinite subset of C in a closed set C_1 of diameter ≤ 1, an infinite subset of C_1 in a closed subset C_2 of C_1 of diameter $\leq 1/2$, and so on. The intersection $\bigcap C_\bullet$ is a limit point of C.)

7. Continuity properties of monotone functions

Recall that a monotone increasing function F from $\mathbf{R}$ into $\mathbf{R}$, has left and right limits at each point s of $\mathbf{R}$,

$$F(s-) = \sup_{t<s} F(t), \quad F(s+) = \inf_{t>s} F(t),$$

and that $F(s-) \leq F(s) \leq F(s+)$. If the first [second] inequality is actually an equality, F is left [right] continuous at s; if both inequalities are actually equalities, F is continuous at s. The function F can have at most countably many discontinuities, because at each discontinuity point s there is a rational number strictly between the left and right limits at s, and different discontinuity points correspond to different rational numbers. At a discontinuity point s, the difference $F(s+)-F(s-)$ is the *jump* of F at s. The left [right] limit function $s \to F(s-)$ [$s \to F(s+)$] is a left [right] continuous monotone increasing function with itself as left [right] limit function, and a continuity point of F or of its left or right limit function is necessarily a continuity point of all three functions.

More generally, if F is a monotone increasing function from a dense subset of $\mathbf{R}$ into $\mathbf{R}$, one sided limits $F(s-)$ and $F(s+)$ exist at every point s of $\mathbf{R}$, and $F(s-) \leq F(s) \leq F(s+)$ whenever F is defined at s. The left and right limit functions are respectively left and right continuous monotone increasing functions on $\mathbf{R}$, a continuity point of either is a continuity point of the other, and F has a limit at such a point. The set of points of $\mathbf{R}$ at which F does not have a limit is countable. An extension of F with domain $\mathbf{R}$ is monotone if and only if the extension lies between the left and right limit functions of F. A monotone extension is therefore uniquely determined at all the continuity points of these left and right limit functions.

The necessary changes in the preceding discussion if the domain of F is a subinterval of $\bar{\mathbf{R}}$ or a set dense in such an interval are immediate.

8. The correspondence between monotone increasing functions on R and measures on B(R)

The class of monotone increasing functions F on $\mathbf{R}$ corresponds to the class of measures λ on $\mathbf{B}(\mathbf{R})$ by way of the fact that the λ measure of a Borel set is the increase of F on the set. A precise statement of this correspondence is the content of the following theorem.

Theorem. *Let F be a finite valued monotone increasing right continuous function on* $\mathbf{R}$. *Define $F(-\infty)$ as the right limit ($\geq -\infty$) of F at $-\infty$, and define $F(+\infty)$ as the left limit ($\leq +\infty$) of F at $+\infty$.*

(a) *Define a set function λ_{0F} on each right semiclosed interval of* $\mathbf{R}$ *by setting*

$$\lambda_{0F}\big((a,b]\cap\mathbf{R}\big) = F(b)-F(a), \qquad (-\infty \leq a < b \leq +\infty). \tag{8.1}$$

Then λ_{0F} has a unique extension to a measure λ_F on **B(R)**; *this extension is finite on compact sets. Denote by $\lambda_F{}^*$ the completion of this measure and choose $\varepsilon > 0$. Each set A in the domain of $\lambda_F{}^*$ lies between a closed subset and an open superset whose difference has $\lambda_F{}^*$ measure at most ε. Moreover, if $\lambda_F{}^*(A) < +\infty$, there is a finite union B of open intervals for which $\lambda_F{}^*(A \Delta B) < \varepsilon$.*

(b) *Conversely, if λ is a measure on* **B(R)**, *finite on compact sets, there is a finite valued monotone increasing right continuous function F_λ on* **R**, *uniquely determined up to an additive constant by the condition*

$$F_\lambda(b) - F_\lambda(a) = \lambda\big((a,b]\big), \qquad -\infty < a < b < +\infty. \tag{8.2}$$

(c) *One monotone function determined in accordance with* (b) *by the measure λ_F in* (a) *is F, and the measure determined in accordance with* (a) *by the monotone function F_λ in* (b) *is λ.*

After this theorem has been proved the asterisk will be dropped, that is, λ_F will be written instead of $\lambda_F{}^*$.

Proof that λ_{0F} defined by (8.1) has an additive extension to the set algebra $\mathbf{S}_0$ of finite unions of right semiclosed intervals. It is trivial that λ_{0F} is finitely additive on the class of right semiclosed intervals of **R**. According to Lemma III.2, it follows that this set function has a unique finitely additive extension to the algebra $\mathbf{S}_0$. From now on, "λ_{0F}" refers to this extension.

Proof that λ_{0F} is a measure on $\mathbf{S}_0$. The monotonicity and right continuity of F, not yet used, are needed to prove that λ_{0F} is a measure on $\mathbf{S}_0$. The fact that F is monotone increasing makes λ_{0F} positive. To prove countable additivity, it must be shown that if I is in $\mathbf{S}_0$, that is, if I is a finite disjunct union of right semiclosed intervals, and if $I = \bigcup I_\bullet$ is a disjunct countable union of members of $\mathbf{S}_0$, then

$$\lambda_{0F}(I) = \sum \lambda_{0F}(I_\bullet). \tag{8.3}$$

Since λ_{0F} is additive, each member of $I_\bullet$ can be replaced by its component intervals, and therefore it can be supposed that each set I_j is a right semiclosed interval. Since each component interval of I is the countable union of its intersections with the members of $I_\bullet$, it is sufficient to prove (8.3) for I a right semiclosed interval. Thus from now in it will be supposed that the sets in (8.3) are all right semiclosed intervals.

Since $I \supset \bigcup_1^n I_\bullet$ for all n, and since λ_{0F} is monotone and finitely additive on $\mathbf{S}_0$,

$$\lambda_{0F}(I) \geq \sum_1^n \lambda_{0F}(I_j)$$

for all n, and therefore

(8.4) $$\lambda_{0F}(I) \geq \sum \lambda_{0F}(I_\bullet).$$

The reverse (subadditivity) inequality is trivial if the sum is $+\infty$. It is therefore sufficient to prove subadditivity when each summand in (8.3) is finite. Let J be a right semiclosed interval with compact closure J'', a subset of I. Choose $\varepsilon > 0$. Let I_j' be a right semiclosed interval, with the same left-hand endpoint as I_j and the same right-hand endpoint if that endpoint is $+\infty$, but otherwise with right-hand endpoint to the right of that of I_j but so close that

$$\lambda_{0F}(I_j') \leq \lambda_{0F}(I_j) + \varepsilon 2^{-j}.$$

Let I_j^{o} be the interior of I_j'. The compact interval J' is covered by $\bigcup I_\bullet^{\mathrm{o}}$. Apply the Heine-Borel theorem to find that

$$J \subset J' \subset \bigcup_1^k I_\bullet^{\mathrm{o}} \subset \bigcup_1^k I_\bullet'$$

for sufficiently large k, from which it follows, in view of the monotonicity and finite subadditivity of λ_{0F}, that

(8.5) $$\lambda_{0F}(J) \leq \lambda_{0F}(J') \leq \sum_1^k \lambda_{0F}(I_\bullet') \leq \sum \lambda_{0F}(I_\bullet) + \varepsilon.$$

When ε tends to 0 and J increases to I, (8.5) yields the desired countable subadditivity, and therefore the countable additivity, of λ_F on $\mathbf{S}_0$.

Proof of (a). According to Theorem 4, the measure λ_{0F} has a unique extension to a measure on $\sigma(\mathbf{S}_0)=\mathbf{B}(\mathbf{R})$, and according to Theorem 1 this measure has a unique completion. According to Theorem 5, a set A in the domain of the completion lies between a closed subset and an open superset B' with an arbitrarily small difference set measure. If $\lambda_F(A) < +\infty$ and $\varepsilon > 0$, the last assertion of (a) is proved by choosing B' to make the difference set measure at most $\varepsilon/2$, and then choosing a large enough number of the pairwise disjoint intervals making up B' to be within $\varepsilon/2$ of the measure of B'. The measure λ_F is uniquely determined by F, because the values of λ_F on right semiclosed intervals, and therefore on the sets in $\mathbf{S}_0$, and finally on the sets of $\mathbf{B}(\mathbf{R})$, which are all in the d_λ closure of $\mathbf{S}_0$, are uniquely determined by F.

Proof of (b). If λ is a measure on the σ algebra $\mathbf{B}(\mathbf{R})$, finite on compact sets, there is a monotone increasing right continuous function F_λ on $\mathbf{R}$, determined up to an additive constant by its increase on right semiclosed intervals. For example, the monotone increasing right continuous function defined by

(8.6)
$$\begin{aligned} F_\lambda(b) &= \lambda\big((0,b]\big) && \text{if } b > 0, \\ &= 0 && \text{if } b = 0, \\ &= -\lambda\big((b,0]\big) && \text{if } b < 0 \end{aligned}$$

satisfies (8.2). If $\lambda((-\infty,b])$ is finite for some b, and therefore for all b, F_λ is usually defined by

$$F_\lambda(b) = \lambda((-\infty,b]) \tag{8.7}$$

for all b, to make $F_\lambda(-\infty+) = 0$.

Proof of (c). What has been proved under (a) and (b) is that certain functions and measures are paired: if a measure λ and a monotone function F are paired, the increase in F on a right semiclosed interval is the λ measure of the interval. This is the content of Equations (8.1) and (8.2). When λ and F are paired, λ is written as λ_F or F is written as F_λ to stress the pairing. Part (c) of the theorem is thus trivial.

Terminology. When $F(s) = s$ for all s, the measure λ_F is *Lebesgue measure*, named after the mathematician who inaugurated modern measure theory by defining this measure. In the general case λ_F is called, depending on the context and the predilections of the caller, a *Lebesgue-Stieltjes measure on* **R**, a *Radon measure on* **R**, or a *distribution on* **R**, or if $\lambda(\mathbf{R}) = 1$, a *probability measure* or *probability distribution on* **R**. In the last case, in which F is normalized by setting $F(-\infty+) = 0$ and therefore $F(+\infty-) = 1$, F is a *probability distribution function on* **R**. The terms "Lebesgue measure" and "Lebesgue-Stieltjes measure" usually refer to the completed measures of Borel sets. The general theory of measure and integration studied in this book is sometimes referred to as *Lebesgue measure theory.*

Modification for intervals of $\bar{\mathbf{R}}$. It is obvious how to adapt the preceding discussion to define measure on an open subinterval I of **R**: one simply starts with a monotone function on I instead of a monotone function on **R**. There is no added complication if I contains its right endpoint. There is, however, a slight complication if I contains its left endpoint, in that the monotone function on I must be allowed to have the left endpoint as a right discontinuity. Let I be an interval containing its left endpoint a, and let F be a finite valued monotone increasing function on I, right continuous except possibly at a, with $F_\lambda(a) = 0$. The discussion in this section, as adapted to I, leads to a measure λ_F on $\mathbf{B}(I)$, finite on compact sets, determined by setting $F(a) = 0$ and

$$\lambda_F([a,b]) = F(b) - F(a) \qquad (b > a). \tag{8.8}$$

The singleton $\{a\}$ has measure $F(a+)$. Finally, the adaptation of the discussion to intervals of $\bar{\mathbf{R}}$ is now trivial: map such an interval onto a subinterval of **R**.

Example. Let λ be Lebesgue measure on **R**, and consider the class of countable unions of open intervals $(n,n+1)$ with n an arbitrary integer. If A and B are two such unions, not identical, then they differ by at least one set of measure 1.

Hence $d_{\lambda}'(A,B) = 1$. Since there are an uncountable number of such unions, the class of Borel sets is not separable in the d_{λ}' metric, even though the class is the σ algebra generated by a countable collection of sets, for example by the open intervals with rational endpoints.

9. Discrete and continuous distributions on R

Let F be a monotone increasing right continuous function on **R**. If F has a jump at a point s, $\lambda_F(\{s\}) = F(s)-F(s-) > 0$. Every singleton is λ_F null if and only if F is continuous, and in that case λ_F is a *continuous distribution*. If F increases only in jumps, that is, if $F(b)-F(c)$ is the sum of the jumps at points in $(a,b]$, for every right semiclosed interval $(a,b]$, F is a *jump function*, and λ_F is a *discrete* distribution. For example, if the sequence $r_\bullet$ is dense in **R**, the function F defined by

$$F(s) = \sum_{r_n \leq s} 2^{-n} \tag{9.1}$$

is a jump function, with jump of 2^{-n} at r_n, and F is continuous except at the points of $r_\bullet$.

10. Lebesgue-Stieltjes measures on $\mathbf{R}^N$ and their corresponding monotone functions

If F is a finite valued function from $\mathbf{R}^N$ into **R**, and if $a < b$, define the difference operator $\mathbf{D}_j(a,b)$ acting on F by

$$\begin{aligned} &\big(\mathbf{D}_j(a,b)F\big)(s_1,\ldots,s_N) \\ &= F(s_1,\ldots,s_{j-1},b,s_{j+1},\ldots,s_N)-F(s_1,\ldots,s_{j-1},a,s_{j+1},\ldots,s_N), \end{aligned} \tag{10.1}$$

with the obvious conventions when $j = 1$ and $j = N$. The N operators defined in this way commute with each other. In the present context, the appropriate definition of a right continuous monotone increasing function is that it is a function F from $\mathbf{R}^N$ into **R** which satisfies the following two conditions:

(a) F is right continuous in each variable when the others are fixed.

(b) If $(a_1,\ldots,a_N)$ and $(b_1,\ldots,b_N)$ are points of $\mathbf{R}^N$, with $a_j < b_j$ for all j, then

$$\Big(\prod_1^N \mathbf{D}_\bullet(a_\bullet,b_\bullet)\Big)F \geq 0. \tag{10.2}$$

In particular, if F tends to 0 when at least one of its arguments tends to $-\infty$, and tends to 1 when all its arguments tend to $+\infty$, F is a *probability distribution function on* $\mathbf{R}^N$. Let $\mathbf{S}_0$ be the algebra of finite unions of right semiclosed intervals of $\mathbf{R}^N$. If I is the bounded right semiclosed interval $(a_1,b_1]\times\cdots\times(a_N,b_N]$, define $\lambda_{0F}(I)$ as the left side of (10.2). If I is not bounded, define $\lambda_{0F}(I)$ as the obvious limit of λ_F on bounded intervals. The argument in Section 8, for $N = 1$ – that λ_{0F} is finitely additive on the class of right semiclosed intervals and therefore has a finitely additive extension to the algebra $\mathbf{S}_0$ of finite unions of these intervals, that this extension is a measure on $\mathbf{S}_0$ and therefore can be extended to a measure on $\sigma(\mathbf{S}_0) = \mathbf{B}(\mathbf{R}^N)$, finite valued on compact sets, and then can be completed – is only slightly more complex when $N>1$, and the details of this generalization will be omitted. Theorem 8 is therefore true for measures on $\mathbf{R}^N$, with essentially the same proof as given when $N=1$. In particular, if F_λ is a probability distribution function on $\mathbf{R}^N$ then $\lambda(\mathbf{R}^N) = 1$, and λ_F is a *probability measure*, or *probability distribution on* $\mathbf{R}^N$.

Conversely suppose that λ is a measure on $\mathbf{B}(\mathbf{R}^N)$, finite on compact sets. There is then a monotone increasing right continuous function F_λ on $\mathbf{R}^N$, satisfying the N-dimensional version of (8.1), that is, the left side of (10.2) is $\lambda_F\big((a_1,b_1]\times\cdots\times(a_N,b_N]\big)$. The function can be normalized, say by defining it as 0 at the origin. The measure defined by F_λ, following the procedure in the first part of this section, is then λ. In particular, if $\lambda\big((-\infty,0]\times\cdots\times(-\infty,0]\big)$ is finite, F_λ can be defined by

$$F_\lambda(s_1,\ldots,s_N) = \lambda\big((-\infty,s_1]\times\cdots\times(-\infty,s_N]\big). \tag{10.3}$$

11. Product measures

Note on the construction of product measures. If integration is introduced before product measures, product measures can be defined directly, using certain integrals, thus avoiding repetition of some of the arguments in the proof of the following theorem. For further details on product measures defined in terms of integrals, see **Note on the construction of product measures** in Section VI.10.

Theorem. *For $i = 1,\ldots,N$ let $(S_i,\mathbf{S}_i,\lambda_i)$ be a σ finite measure space, define*

$$S = S_1\times\cdots\times S_N,\quad \mathbf{S}' = \mathbf{S}_1\times\cdots\times \mathbf{S}_N,$$

and define λ on $\mathbf{S}'$ by

$$\lambda(A_1\times\cdots\times A_N) = \prod_1^N \lambda_\bullet(A_\bullet) \qquad (A_i \in \mathbf{S}_i,\ i = 1,\ldots, N). \tag{11.1}$$

Then λ can be extended uniquely to a measure on $\sigma(\mathbf{S}')$.

Proof. According to Theorem III.3, the set function λ has a unique finitely additive extension to a set function on $\sigma_0(\mathbf{S}')$. To prove the theorem, it is sufficient, according to Theorem 4, to prove that this extension is a measure on $\sigma_0(\mathbf{S})$. Suppose first that N=2 and that λ_1 and λ_2 are finite valued. To prove that λ is a measure on $\sigma_0(\mathbf{S}')$ it is sufficient to prove that if $A_\bullet$ is a decreasing sequence of sets in $\sigma_0(\mathbf{S}')$, given for $n \ngeq 1$ by

$$(11.2) \qquad A_n = \bigcup B_{n\bullet} \times C_{n\bullet} \qquad (B_{nk} \in \mathbf{S}_1, C_{nk} \in \mathbf{S}_2),$$

where the union is finite and disjunct, and if $\bigcap A_\bullet = \emptyset$, then $\lim \lambda(A_\bullet) = 0$. Without loss of generality, the summand sets in (11.2) can be partitioned for each value of n if necessary (see Section II.5) to make the members of $B_{n\bullet}$ mutually disjoint. Define a function f_n on S_1 by

$$f_n(s) = \lambda_2(C_{nk}) \text{ if } s \in B_{nk} \qquad (k \geq 1),$$

$$= 0 \qquad \text{if } s \notin \bigcup B_{n\bullet}.$$

The sequence $f_\bullet$ is a decreasing sequence of functions, with limit 0 because, for each s in S_1, the set $\{t \in S_2\colon (s,t) \in A_n\}$ decreases monotonely when $n \to \infty$, with limit the empty set. If $\varepsilon > 0$, the set $\{s \in S_1 : f_n(s) > \varepsilon\}$ is a subunion of $\bigcup B_{n\bullet}$ and decreases monotonely when $n \to \infty$, with limit the empty set . The λ_1 measure of this subunion therefore decreases monotonely, with limit 0 when $n \to \infty$. Hence

$$(11.3) \qquad \lambda(A_n) = \sum \lambda_1(B_{n\bullet})\lambda_2(C_{n\bullet}) \leq \lambda_1\{s : f_n(s) > \varepsilon\}\lambda_2(S_2)$$

$$+ \varepsilon\lambda_1(S_1) \to \varepsilon\lambda_1(S_1) \qquad (n \to \infty).$$

It follows that $\lim \lambda(A_\bullet) = 0$, as was to be proved. If λ_1 and λ_2 are not necessarily finite valued but if, for i = 1,2, $S_i' \in \mathbf{S}_i$, and $\lambda(S_i') < +\infty$, then the result just obtained is applicable to $S_1' \times S_2'$. It follows that the theorem covers σ finite measures, as stated. If N = 3, the space $S_1 \times S_2 \times S_3$ can be written in the form $(S_1 \times S_2) \times S_3$ and the theorem for N=3 is thereby reduced to the case N=2. The induction proof for general N is now obvious.

In Theorem 11 the measure λ on $\sigma(\mathbf{S}')$ is the *product measure*, written $\lambda_1 \times \cdots \times \lambda_N$, of the *factor measures* $\lambda_1, \ldots, \lambda_N$.

12. Examples of measures on $\mathbf{R}^N$

Example (a). For i=1,...,N let F_i be a monotone increasing right continuous function on $\mathbf{R}$ and define the monotone increasing function F on $\mathbf{R}^N$ by

$$F(s_1,\dots,s_N) = F_1(s_1)\cdots F_N(s_N).$$

The Lebesgue-Stieltjes measure λ_F on $\mathbf{R}^N$ is then the product measure of the factor measures $\lambda_{F_1},\dots,\lambda_{F_N}$ In particular, if $F_i(s) = s$ for all i, λ_F is *N-dimensional Lebesgue measure* on $\mathbf{B}(\mathbf{R}^N)$, the extension of N-dimensional volume to the class of Borel sets. As always, this measure can be completed.

Example (b). If λ is a finite valued measure of Borel subsets of the unit square $S=[0,1]\times[0,1]$, it is the restriction to subsets of S of a Lebesgue-Stieltjes measure on $\mathbf{R}^2$ for which $\mathbf{R}^2$-S is null.

Example (c). In Example (b), let D be the diagonal of S through the origin and let ν be a probability measure of Borel subsets of D. The measure ν can be considered to be the restriction to subsets of D of a probability measure λ on the Borel subsets of S, with $\lambda(S-D)=0$.

13. Marginal measures

Let $(S_1,\mathbf{S}_1)$ and $(S_2,\mathbf{S}_2)$ be measurable spaces, and define $S=S_1\times S_2$, $\mathbf{S}=\mathbf{S}_1\times\mathbf{S}_2$. Then a measure λ on $\sigma(\mathbf{S})$ induces *marginal measures* λ_1 on $\mathbf{S}_1$ and λ_2 on $\mathbf{S}_2$: $\lambda_1(A_1) = \lambda(A_1\times S_2)$ and $\lambda_2(A_2)=\lambda(S_1\times A_2)$ for A_1 in $\mathbf{S}_1$and A_2 in $\mathbf{S}_2$.

Example (a). If $\lambda = \lambda_1\times\lambda_2$ is a product measure with finite valued factor measures λ_1 and λ_2, the marginal measures are $\lambda_2(S_2)\lambda_1$ and $\lambda_1(S_1)\lambda_2$. For instance, in Section 12 Example (b), if λ is two dimensional Lebesgue measure on the square S, the marginal measures are both one dimensional Lebesgue measure on the unit interval $[0,1]$.

Example (b). In Section 12 Example (c), if ν is $2^{-1/2}$ times one-dimensional Lebesgue measure on D, the two marginal measures of λ are again one-dimensional Lebesgue measure on $[0,1]$.

As these two examples show, marginal measures by no means determine the measure of which they are marginal.

14. Coin tossing (Continuation of Section III.9)

First mathematical model. In the discussion of this model in Section III.9, a finitely additive set function P was defined on the σ algebra $\mathbf{S}_\infty$ of finite unions of dyadic right semiclosed subintervals of $(0,1]$. Functions $x_1,x_2,\dots$ were defined as the successive digits in the dyadic representation of a point of $(0,1]$. Unfortunately the sets III(9.3) are not in $\mathbf{S}_\infty$ and therefore their probabilities cannot be defined until P is defined on $\sigma(\mathbf{S}_\infty)=\mathbf{B}\big((0,1]\big)$. Lebesgue measure on

(0,1] provides the necessary extension of P. Under this extension, the first set $A=\bigcup\{x_\bullet=1\}$ in **III**(9.3) has a well defined probability:

$$(14.1) \qquad P\{A\} = 1 - P\{\tilde{A}\} = 1 - P\Big\{\bigcap\{x_\bullet=0\}\Big\} = 1 - \lim_{n\to\infty} 2^{-n} = 1.$$

This result is a special case of the Borel-Cantelli theorem. It will be proved in Section **XI**.19 that the probability of the second set in **III**(9.3) is also 1.

Second mathematical model. In the second mathematical model, P is a finitely additive set function defined on the algebra $\mathbf{S}_\infty$ of subsets of the space S of infinite sequences of 1's and 0's. This algebra is the class of finite unions of sets determined by fixing a finite number of coordinates of S. The extension of the domain of P from $\mathbf{S}_\infty$ to $\sigma(\mathbf{S}_\infty)$ can be made by mapping this model into the first model, that is, the conditions $x_1=a_1,\ldots,x_n=a_n$ in the second model define a subset of (0,1] in the first model, and probabilities are thereby referred from the second model to the first. A more direct approach is to prove that, in the second model, P is countably additive on $\mathbf{S}_\infty$ and therefore has an extension to a measure on $\sigma(\mathbf{S}_\infty)$. To prove countable additivity of P on $\mathbf{S}_\infty$ it need only be remarked that, according to Section **II**.2 Example (b), if a countable union of sets in $\mathbf{S}_\infty$ is itself in $\mathbf{S}_\infty$ then only a finite number of summands are nonempty. In other words, the union is effectively a finite union. Thus, countable additivity is trivially the same as finite additivity in this case. Alternatively, the Hahn-Kolmogorov theorem can be invoked to prove that P has a measure extension to $\sigma(\mathbf{S}_\infty)$.

15. The Carathéodory measurability criterion

Let $(S,\mathbf{S},\lambda)$ be a measure space, and let λ^* be the outer measure generated by $\mathbf{S}$ and λ. Then (Section **III**.11) λ^* is a countably subadditive set function, equal to λ on $\mathbf{S}$.

Theorem. *If A is a subset of S, and if $B_\bullet$ is a finite or infinite disjunct sequence of measurable sets, with union B, then*

$$(15.1) \qquad \lambda^*(A\cap B) = \Sigma\, \lambda^*(A\cap B_\bullet).$$

Proof. Since λ^* is countably subadditive, it is sufficient to prove (15.1) with "≥" instead of "=". Let C be a measurable superset of $A\cap B$. Then

$$(15.2) \qquad \lambda(C) \ge \lambda(C\cap B) = \Sigma\, \lambda(C\cap B_\bullet) \ge \Sigma\, \lambda^*(A\cap B_\bullet).$$

By definition of λ^*, the set C can be chosen to make $\lambda(C)$ arbitrarily close to $\lambda^*(A\cap B)$, thereby yielding the desired inequality.

The Carathéodory approach to measure theory starts with an outer measure and *defines* a set B to be measurable if (15.1) is satisfied with $B = S$ and only two summands.

16. Measure hulls

In this section the subsets of a σ finite measure space $(S,\mathbf{S},\lambda)$ are treated, and the outer measure λ^* is the outer measure generated by $\mathbf{S}$ and λ. If A is a subset of S of finite outer measure, a set A^* is a *measure hull* of A if A^* is a measurable superset of A and if $\lambda(A^*) = \lambda^*(A)$. A measure hull A^* is determined uniquely up to null sets because if $A_1{}^*$ and $A_2{}^*$ are measure hulls of A then $A_1{}^*\cap A_2{}^*$ is also a measure hull of A, and

$$\lambda^*(A) = \lambda(A_1{}^*) = \lambda(A_1{}^*\cap A_2{}^*) = \lambda(A_2{}^*).$$

Every set A of finite outer measure has a measure hull, because if $A_\bullet$ is a sequence of measurable supersets of A with $\lim \lambda(A_\bullet) = \lambda^*(A)$, then $\cap A_\bullet$ is a measure hull of A.

If A has measure hull A^* and B is measurable, then $A\cap B$ has measure hull $A^*\cap B$ because there must be equality throughout the following string of equalities and inequalities:

$$(16.1)\quad \lambda(A^*\cap B) \geq \lambda^*(A\cap B) = \lambda^*(A) - \lambda^*(A\cap \tilde{B}) \geq \lambda(A^*) - \lambda(A^*\cap \tilde{B})$$
$$= \lambda(A^*\cap B).$$

Here the first equality is a special case of (15.1). This hereditary character of the measure hull justifies the definition that, for an arbitrary subset A of S, its measure hull is defined as a superset A^* of A with the property that if B is a measurable set of finite measure, then $A^*\cap B$ is a measure hull for $A\cap B$. Every set A has a measure hull, because if $S = \cup S_\bullet$ is a representation of S as a countable union of measurable sets of finite measure, and if $A\cap S_n$ has measure hull $A_n{}^*$, then $\cup A_\bullet{}^*$ is a measure hull for A. The reader is invited to verify that a measurable set A^* is a measure hull of A if and only if the difference A^*-A has no non null measurable subset. In analysis involving arbitrary subsets of S it is frequently advantageous to replace sets by their measure hulls.

V
Measurable Functions

1. Function measurability

In the operations of analysis, it is desirable to work in a class of admissible objects that does not have to be enlarged as the work progresses. For example, in real analysis the basic set of admissible numbers is **R**, sometimes enlarged to $\bar{\mathbf{R}}$. The class of rational numbers is too small because it is not closed under limit operations. Similarly, in studying measures, a natural class of admissible sets is a σ algebra, because closure under the operations of complementation and the forming of countable unions and intersections are needed.

In this chapter, functions from a space S into a space S' will be studied, and the first decision to be made is the choice of admissible functions. Again, it is desirable to choose a class that need not be enlarged as the work progresses, and if S is coupled with a σ algebra $\mathbf{S}$ of its subsets to form a measurable space $(S,\mathbf{S})$, it is to be expected that the chosen class of functions will depend on $\mathbf{S}$. For example, it is desirable that the indicator functions of sets in $\mathbf{S}$ be in the class of admissible functions. The following are (interrelated) reasonable requirements.

(a) The class should be closed under the operations of taking linear combinations, products, and limits, if such operations are meaningful for S'. If $S'=\mathbf{R}$, the class of continuous functions is not large enough to satisfy this condition because the limit of a convergent sequence of continuous functions need not be continuous.

(b) If f is an admissible function from S into a space S' and g is an admissible function from S' into a space S'', then $g(f)$ should be admissible as a function from S into S''.

(c) If f is an admissible function from S into a space S', the set of points of S at which f satisfies reasonable conditions, say that the set of points of S at which the values of f lie in an admissible subset of S', should be an admissible subset of S.

Condition (c) leads to the concept of measurability of f, as formulated in the next paragraph, and will be seen to imply conditions (a) and (b).

Let $(S,\mathbf{S})$ and $(S',\mathbf{S}')$ be measurable spaces, and let f be a function from S into S'. It is convenient to call such a function a function from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$ as well as from S into S'. The range space S' of the function is commonly, especially in probability contexts, called the *state space* of the function. The inverse function f^{-1} takes complements relative to S' into complements relative to S, unions (even uncountable ones) in S' into the corresponding unions in S, intersections (even uncountable ones) in S' into the corresponding intersections in S. That is, for example, if $A'_\bullet$ is a family of subsets of S', then

$$f^{-1}(\cup_\bullet A'_\bullet) = \cup f^{-1}(A'_\bullet).$$

Hence $f^{-1}(\mathbf{S}')$, the class of inverse images of the sets in $\mathbf{S}'$, is a σ algebra; it will be denoted by $\sigma(f)$. This σ algebra is, for given $(S',\mathbf{S}')$, *the class of subsets of S determined by measurable conditions on f*. If $\sigma(f) \subset \mathbf{S}$, that is, if the inverse image of a measurable set in the range space of f is a measurable set in the domain space, the function f is *measurable from* $(S,\mathbf{S})$ *into* $(S',\mathbf{S}')$. It is immediate that the transitivity condition (b) is satisfied: if f is a measurable function from a measurable space into a second one, and if g is a measurable function from the second space into a third, then $g(f)$ is measurable from the first into the third, and $\sigma[g(f)] \subset \sigma(f)$.

Example. Given a space S, a measurable space $(S',\mathbf{S}')$, and a function f from S into S', one choice of σ algebra $\mathbf{S}$ of subsets of S making f measurable from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$ is $\mathbf{S} = 2^S$. The smallest choice of $\mathbf{S}$ making f measurable from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$ is $\sigma(f)$. In particular, if S' is countable and $\mathbf{S}' = 2^{S'}$, f is measurable if and only if the inverse image of every S' singleton is in $\mathbf{S}$. This is the definition given in Section III.5 in studying discrete state spaces.

A function f may be described as $\mathbf{S}$ *measurable*, or *measurable with respect to* $\mathbf{S}$, or *measurable from S into S'*, or simply *measurable*, if the relevant spaces or σ algebras missing from the description have been specified or if the context is so general that full measure space identification is not needed. Thus a function identically equal to a real number is a measurable function from an arbitrary measurable space into $\mathbf{R}$, that is, into $(\mathbf{R},\mathbf{B}(\mathbf{R}))$. More generally, the indicator function of a subset of a measurable space is a measurable function from the space into $\mathbf{R}$ if and only if the subset is measurable. In probability contexts, a measurable function is given the alias *random variable*.

Testing for measurability. In testing for measurability of a function f from a measurable space $(S,\mathbf{S})$ into a measurable space $(S',\mathbf{S}')$, the fact that $\mathbf{S}'$ and $f^{-1}(\mathbf{S}')$ are σ algebras implies that it is sufficient for measurability that the condition $f^{-1}(\mathbf{S}_0') \subset \mathbf{S}$ be satisfied for a subclass $\mathbf{S}_0'$ of $\mathbf{S}'$ large enough to generate the σ algebra $\mathbf{S}'$, that is, large enough to make $\mathbf{S}' = \sigma(\mathbf{S}_0')$. In particular if $(S',\mathbf{S}') = (\mathbf{R},\mathbf{B}(\mathbf{R}))$, the real valued function f is measurable if $f^{-1}(A') \in \mathbf{S}$ for every interval A' of the form $(-\infty,b)$, that is, if $\{f<b\} \in \mathbf{S}$ for all b. In fact, it was pointed out in Section **II**.4 that the class of these intervals generates $\mathbf{B}(\mathbf{R})$. Other sufficiently large classes $\mathbf{S}_0'$ are the classes of intervals of the form $(-\infty,b]$ or of the form $(b,+\infty)$, and so on. A dense set of values of b yields a sufficiently large class of these intervals. The definition of measurability of a real valued function is frequently given using one of these classes of intervals instead of the full class $\mathbf{S}'=\mathbf{B}(\mathbf{R})$.

Vector functions. If $f_1,\ldots,f_n$ are functions from S into S', the vector function $f\colon s \to [f_1(s),\ldots,f_n(s)]$ from S into S'^n is measurable from $(S,\mathbf{S})$ into $(S'^n,\sigma(\mathbf{S}'^n))$ if and only if each function f_j is measurable from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$, because the product sets $A_1'\times\cdots\times A_n'$ with factor sets in $\mathbf{S}'$ generate the σ algebra $\sigma(\mathbf{S}'^n)$, and

$$f^{-1}(A_1'\times\cdots\times A_n') = f_1^{-1}(A_1')\cap\cdots\cap f_n^{-1}(A_n').$$

The class of sets determined by measurable conditions on functions. If $\{f_t, t\in I\}$ is a collection of measurable functions from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$, with I an arbitrary index set, the σ algebra $\sigma(f_t, t \in I)$ of subsets of S *determined by measurable conditions on* $f_\bullet$ is the σ algebra generated by the sets of the form $\{f_t \in A'\}$, for t in I and A' in $\mathbf{S}'$. This σ algebra is the smallest σ algebra of subsets of S making each of the given functions measurable. In particular, when $I = 1,\ldots,n$, the σ algebra $\sigma(f_1,\ldots,f_n)$ is the σ algebra of subsets of S of the form $\{(f_1,\ldots,f_n) \in A'\}$ for A' in $\sigma(\mathbf{S}'^n)$.

Let $(S,\mathbf{S})$, $(S',\mathbf{S}')$, and $(S'',\mathbf{S}'')$ be measurable spaces, and let f be a measurable function from the first space into the second. By definition of $\sigma(f)$, f is not only measurable from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$, but even measurable from $(S,\sigma(f))$ into $(S',\mathbf{S}')$. Thus if g is a measurable function from $(S',\mathbf{S}')$ into $(S'',\mathbf{S}'')$, then $h = g(f)$ is not only measurable from $(S,\mathbf{S})$ into $(S'',\mathbf{S}'')$ but even measurable from $(S,\sigma(f))$ into $(S'',\mathbf{S}'')$. This restrictive measurability condition on a measurable function h from $(S,\mathbf{S})$ into $(S'',\mathbf{S}'')$ is not only necessary but, under certain conditions on the spaces, is sufficient to ensure, for given f, measurable from $(S,\mathbf{S})$ into $(S',\mathbf{S}')$, and given h, measurable from $(S,\mathbf{S})$ into $(S'',\mathbf{S}'')$, that h can be written in the form $g(f)$, with g measurable from $(S',\mathbf{S}')$ into $(S'',\mathbf{S}'')$. This fact will not be needed explicitly but gives intuitive content to later definitions of conditional expectations and probabilities. It will be proved, to exhibit the principle involved, when $\mathbf{S}''$ contains the singletons and h is a function from S into S'', taking on only countably many values. Suppose then that h takes on the values in the sequence $a_\bullet$, and $\{h = a_n\} \in \sigma(f)$. Then $\{h = a_n\} = \{f \in A_n\}$ for some set A_n in $\mathbf{S}'$. Define $g = a_n$ on A_n to obtain the representation $h = g(f)$, with g measurable from $(S',\mathbf{S}')$ into $(S'',\mathbf{S}'')$.

These simple remarks suggest that, whenever $\{f_t, t\in I\}$ is a collection of measurable functions from a measurable space $(S,\mathbf{S})$ into the measurable space $(S',\mathbf{S}')$, and a reasonable definition is needed of a measurable function g of all of these functions into some measurable space, one reasonable definition is that g be measurable from $(S, \sigma(f_t, t \in I))$ into the prescribed space.

Measurability of a function defined on a subset of a space. If $(S,\mathbb{S})$ is a measurable space, a function f from a set A in $\mathbb{S}$ into the measurable space $(S',\mathbb{S}')$ is *measurable* if $f^{-1}(\mathbb{S}') \subset \mathbb{S}$. Equivalently, denoting by $\mathbb{S}_A$ the class of subsets of A in $\mathbb{S}$, the function f is measurable if and only if the function, when considered as a function from $(A,\mathbb{S}_A)$ into $(S',\mathbb{S}')$, is measurable. In particular, the restriction to A of a measurable function from $(S,\mathbb{S})$ into $(S',\mathbb{S}')$ is measurable.

Borel measurable functions. A measurable function from one metric space into a second is *Borel measurable*. A continuous function from one metric space into a second is Borel measurable, because the inverse image of an open state space set is open and the open state space sets generate the σ algebra of Borel state space sets.

Approximation of measurable functions by step functions. A *step function* from a measurable space $(S,\mathbb{S})$ into $\mathbf{R}$ is a finite linear combination, with coefficients in $\mathbf{R}$, of indicator functions of sets in $\mathbb{S}$. A step function is a simple example of a measurable function from the measurable space into $\mathbf{R}$. An essential tool in the study of measurable functions from $(S,\mathbb{S})$ into $\mathbf{R}$ is the fact that a measurable function f from S into $\bar{\mathbf{R}}^+$ is the limit of a monotone increasing sequence $f_\bullet$ of positive step functions. For example, define

$$
\begin{aligned}
f_n &= (j-1)2^{-n} &&\text{on} \quad \{(j-1)2^{-n} \le f < j2^{-n}\} \qquad (1 \le j \le 4^n),\\
&= 2^n &&\text{on} \quad \{f \ge 2^n\}.
\end{aligned}
\tag{1.1}
$$

If f is a measurable function from $(S,\mathbb{S})$ into $\bar{\mathbf{R}}$, f is still the limit of a sequence of step functions by way of the definition

$$
\begin{aligned}
f_n &= (j-1)2^{-n} &&\text{on} \quad \{(j-1)2^{-n} \le f < j2^{-n}\} \qquad (|j| \le 4^n)\\
&= -2^n &&\text{on} \quad \{f < -2^n - 2^{-n}\}\\
&= 2^n &&\text{on} \quad \{f \ge 2^n\}.
\end{aligned}
\tag{1.2}
$$

Under this definition, the sequence $f_\bullet$ is monotone increasing, neglecting a finite number of terms, if f is lower bounded.

2. Function measurability properties

(a) **Applications of transitivity**. If $f_1,\ldots,f_n$ are measurable from $(S,\mathbb{S})$ into a metric space S', and if f is measurable from $\big(S'^n,\mathbf{B}(S'^n)\big)$ into $\mathbf{R}$, then $f(f_1,\ldots,f_n)$ is measurable from $(S,\mathbf{S})$ into $\mathbf{R}$. For example, if $S=\mathbf{R}$ it follows that $|f_1|$, cf_1 (for c a constant), $1/f_1$ (if f_1 never vanishes), $\Sigma f_\bullet$, $\prod f_\bullet$, the pointwise maximum $f_1\vee\cdots\vee f_n$, the pointwise minimum $f_1\wedge\cdots\wedge f_n$ are measurable whenever each function f_i is. The last two are also measurable when the functions are extended real valued. It can argued more directly that the pointwise supremum $f = \sup f_\bullet$ is measurable for a finite or countably infinite sequence $f_\bullet$ of extended real valued measurable functions, by noting that $\{f>c\} = \bigcup\{f_\bullet>c\}$. This assertion of measurability is incorrect for uncountable collections of functions.

(b) **Sets defined by inequalities between extended real valued measurable functions.** If f_1 and f_2 are measurable functions from a measurable space into $\overline{\mathbf{R}}$ the sets $\{f_1>f_2\}$, $\{f_1\leq f_2\}$, and $\{f_1=f_2\}$ are measurable because (first set)

$$\{f_1>f_2\} = \bigcup_{r \text{ rational}} \left[\{f_1>r\}\cap\{f_2<r\}\right],$$

the second set is the complement of the first, and the third set is $\{f_2\leq f_1\}\cap\{f_1\leq f_2\}$. A somewhat more sophisticated proof of measurability of these sets applies (a).

(c) **Completeness of a measure and function measurability.** If f and g are functions from a complete measure space into a measurable space, and if $f=g$ almost everywhere, then if one of the functions is measurable, the other is also. Less trivial than this is the fact that if $(S,\mathbf{S},\lambda)$ is a σ finite measure space and if $(S,\mathbf{S}^*,\lambda^*)$ is the completed measure space, then if f is a measurable function from $(S,\mathbf{S}^*)$ into $\overline{\mathbf{R}}$, there is a function g, measurable from $(S,\mathbf{S})$ into $\overline{\mathbf{R}}$ and λ^* almost everywhere equal to f, that is equal to f except on a subset of a λ null set. It is sufficient to prove this assertion for f positive, because it is then trivial that the result is true for f negative, and the two results combine to give the result for arbitrary f. There is a function g as stated when f is the indicator function of a set in $\mathbf{S}^*$, because if $A \in \mathbf{S}^*$ there is a subset A_0 of A, in $\mathbf{S}$, and differing from A by a λ^* null set. It follows that the assertion is true for an $(S,\mathbf{S}^*)$ step function. It was pointed out above, see (1.1), that, in the general case of a positive $\mathbf{S}^*$ measurable function f, f is the limit of an increasing sequence of $(S,\mathbf{S}^*)$ step functions. The desired function g is the pointwise supremum of the corresponding sequence of $(S,\mathbf{S})$ step functions.

(d) **Measurability of functions of several variables if one is fixed.** If $(S_1,\mathbf{S}_1)$ and $(S_2,\mathbf{S}_2)$ are measurable spaces, if $S = S_1\times S_2$, if $\mathbf{S} = \sigma(\mathbf{S}_1\times\mathbf{S}_2)$, and if $f\colon (s_1,s_2)\mapsto f(s_1,s_2)$ is a measurable function from $(S,\mathbf{S})$ into $\overline{\mathbf{R}}$, then for each point s_1 of S_1, the function $f(s_1,\bullet)$ is a measurable function from $(S_2,\mathbf{S}_2)$ into $\overline{\mathbf{R}}$. In fact, this is true when f is the indicator function of a set in $\mathbf{S}$ according to Section II.5, and therefore it is true when f is a step function. It is sufficient to prove measurability for positive f; in that case apply the representation in (1.1) of f as the limit of an increasing sequence of step functions.

3. Measurability and sequential convergence

If $f_\bullet$ is a sequence of measurable functions from a measurable space $(S,\mathbf{S})$ into $\bar{\mathbf{R}}$, it was pointed out in Section 2 that $\inf f_\bullet$ and $\sup f_\bullet$ are measurable. It follows that the functions

$$\limsup f_\bullet = \inf_{j>1} \sup_{n>j} f_n, \qquad \liminf f_\bullet = \sup_{j>1} \inf_{n>j} f_n$$

are measurable, the convergence set is measurable, and the restriction of the limit function to the convergence set is measurable on that set.

More generally, it will now be proved that the last two assertions in the preceding sentence are true if the range space of the sequence $f_\bullet$ is a complete metric space (S',d'). For fixed m and n, the function $s \to [f_n(s), f_m(s)]$ is a measurable function from S into the metric product space S'^2. Since a metric space distance function is continuous, the function $s \to d'(f_n(s), f_m(s))$ from $(S,\mathbf{S})$ into $\mathbf{R}$ is measurable, and therefore the supremum h_j of these functions for n and m at least j is a measurable function. The convergence set C of the sequence $f_\bullet$ is measurable because C is the set on which the sequence $h_\bullet$ has limit 0. To prove that on C the limit function f is measurable, it is sufficient to show that the set $\{s \in C: f(s) \in A'\}$ is measurable whenever A' is a closed subset of S'. This measurability follows from the evaluation

$$\{s \in C: f(s) \in A'\} = \{s \in C: d'(f(s), A') = 0\}, \tag{3.1}$$

since the distance from a point of S' to A' is a continuous function of the point and vanishes if and only if the point is in A'.

4. Baire functions

If $\mathbf{B}_1$ is the class of Borel measurable functions from a metric space (S,d) into $\mathbf{R}$, then

(a) $\mathbf{B}_1$ contains the continuous functions, and

(b) $\mathbf{B}_1$ is closed under sequential convergence,

that is, the limit of a convergent (everywhere on S) sequence of functions in the class is itself in the class. Consider the classes of functions from S into $\mathbf{R}$ satisfying conditions (a) and (b). The intersection $\mathbf{B}_2$ of all these classes is a class satisfying conditions (a) and (b) and is the smallest such class; its members are the *Baire functions*. According to the following theorem, $\mathbf{B}_1 = \mathbf{B}_2$.

Theorem. *A function from a metric space into* **R** *is a Baire function if and only if the function is Borel measurable.*

Proof. (The notation $\mathbf{B}_1, \mathbf{B}_2$, will be used as just defined.)

(a) $\mathbf{B}_2 \subset \mathbf{B}_1$, because $\mathbf{B}_1$ satisfies conditions (a) and (b) defining $\mathbf{B}_2$, and $\mathbf{B}_2$ is the minimum class satisfying these conditions.

The converse will be proved in several steps.

(b) *If ϕ is a continuous function from* $\mathbf{R}^2$ *into* **R**, *f and g are in* $\mathbf{B}_2$, *and f is continuous, then $\phi(f,g)$ is in* $\mathbf{B}_2$, because the class of functions g for which this assertion is true contains the continuous functions and is closed under sequential convergence.

(c) *If ϕ is a continuous function from* $\mathbf{R}^2$ into **R**, *and f and g are in* $\mathbf{B}_2$, *then $\phi(f,g)$ is in* $\mathbf{B}_2$ because the class of functions f for which this is true contains the continuous functions according to (a) and is closed under sequential convergence. In particular if ϕ is a continuous function from **R** into **R**, and if f is in $\mathbf{B}_2$ then $\phi(f)$ is in $\mathbf{B}_2$.

(d) The obvious induction proof shows that if ϕ is continuous from **R** into **R**, and if $f_1,\ldots,f_n$ are in $\mathbf{B}_2$, then $\phi(f_1,\ldots,f_n)$ is in $\mathbf{B}_2$. Furthermore, the latter function is in $\mathbf{B}_2$ not only when ϕ is continuous but even for ϕ a Baire function from $\mathbf{R}^n$ into **R**, because the class of functions for which the assertion is true was just proved to contain the continuous functions and is closed under sequential convergence.

(e) *The class* $\mathbf{B}_3$ *of subsets of S whose indicator functions are in* $\mathbf{B}_2$ *is* $\mathbf{B}(S)$. If A is a closed subset of S, the continuous function f_n: $s \to \exp[-nd(s,A)]$ is in $\mathbf{B}_2$, and the sequence $f_{\cdot}$ has limit $\mathbf{1}_A$. Thus $\mathbf{B}_3$ includes the closed sets. Moreover the class $\mathbf{B}_3$ is closed under monotone convergence and therefore is $\mathbf{B}(S)$.

(f) $\mathbf{B}_1 = \mathbf{B}_2$. If $g \in \mathrm{B}_1$, consider the step function g_n from S into **R** defined by

$$g_n(s) = (j-1)2^{-n} \quad \text{on } \{s: (j-1)2^{-n} \le f(s) < j2^{-n}\} \qquad (|j|) \le 4^n)$$

(4.1)

$$= 0 \quad \text{elsewhere.}$$

The function g_n is in $\mathbf{B}_2$ because it is a linear combination of a finite number of indicator functions of Borel subsets of S and therefore is a continuous function of these indicator functions. Since the sequence $g_{\cdot}$ converges to g, the function g is in $\mathbf{B}_2$, and therefore $\mathbf{B}_1 \subset \mathbf{B}_2$. The reverse inclusion was proved in (a).

5. Joint distributions

Let $(S,\mathbf{S},\lambda)$ be a finite measure space and let $x_1,\ldots,x_N$ be measurable functions from this space into a measurable space $(S',\mathbf{S}')$. These functions determine a measure λ' on $\sigma(\mathbf{S}'^N)$ by

$$\text{(5.1)}\qquad \lambda'(\underline{A}') = \lambda\left\{ s: [(x_1(s),\ldots,x_N(s)] \in \underline{A}' \right\}, \qquad \left(\underline{A}' \in \sigma(\mathbf{S}'^N)\right).$$

(This measure, the *joint distribution* of the given N functions, is almost exclusively applied in probabilistic contexts, in most of which λ is a probability measure and $S' = \mathbf{R}$.) In particular, the (one-dimensional marginal) distribution of x_j is given by

$$\text{(5.2)}\qquad \lambda_j'(A') = \lambda\{ s: x_j(s) \in A'\} \qquad (A' \in \mathbf{S}').$$

Representations of sets of measurable functions. Let $x_1,\ldots,x_N$ be as above, but suppose for simplicity that $S' = \mathbf{R}$. The distribution of $x_1,\ldots,x_N$ is a Lebesgue-Stieltjes measure on $\mathbf{R}^N$ and, as determined by this measure, the coordinate functions on $\mathbf{R}^N$ have the same joint distribution on $\mathbf{R}^N$ as the given functions on S. In investigations in which only joint distributions of functions are involved, it is sometimes convenient to use these N coordinate functions on $\mathbf{R}^N$ instead of the given N functions on S.

6. Measures on function (coordinate) space

Let U be a complete separable metric space, I be an arbitrary infinite set to be used as an index set and S be the space of all functions ω from I into U. The space S is a coordinate space of dimensionality the cardinality of I. Denote by x_i the ith coordinate function, a function from S into U, defined by setting $x_i(\omega) = \omega(i)$. For example, if I is the set of strictly positive integers, and if $U = \mathbf{R}$, the space S is countably infinite dimensional Euclidean space. The following discussion would not be simplified by supposing I to be only countably infinite, and it is important that this restriction not be imposed, because in the probability context of continuous parameter stochastic process theory the index set is commonly an interval of $\mathbf{R}$. Call a subset $\underline{S}$ of S a *finite dimensional measurable set* based on the finite index set $(i_1,\ldots,i_n)$ if

$$\text{(6.1)}\qquad \underline{S} = \left\{\omega: [x_{i_1}(\omega),\ldots,x_{i_n}(\omega)] \in \underline{A}'\right\},$$

where $\underline{A}' \in \mathbf{B}(U^n)$. The standard abbreviation will be used below, in which the notation for the set in (6.1) is shortened to $\{(x_{i_1},\ldots,x_{i_n}) \in \underline{A}'\}$. The class of subsets of S obtained when $i_1,\ldots,i_n$ are specified, but $\underline{A}'$ is allowed to vary in $\mathbf{B}(U^n)$, is $\sigma(x_{i_1},\ldots,x_{i_n})$. Denote by $\mathbf{S}_0$ the union of all these algebras of subsets of S, for all finite index sets, that is, the algebra of subsets of S determined by measurable conditions on finitely many coordinates. The class $\mathbf{S}_0$ is an algebra, but not a σ algebra unless U is a singleton.

Theorem (Kolmogorov). *(Separable complete metric state space U, I an arbitrary infinite index set, S the space of functions from I into U.) Let λ be a positive finite valued set function, defined on the algebra $\mathbf{S}_0$ of finite dimensional measurable subsets of S, and suppose that λ is countably additive on each σ algebra of finite dimensional measurable sets based on a specified finite coordinate set. Then λ is a measure on $\mathbf{S}_0$, and therefore has an extension to a measure on $\sigma(\mathbf{S}_0) = \sigma(x_i, i \in I)$.*

The hypotheses imply that λ is finitely additive on $\mathbf{S}_0$.

The context of this theorem is a generalization of that encountered in the second mathematical model for coin tossing studied in Sections **III.9** and **IV.14**, in which case U consisted of two points.

Proof. The fact that each product space U^n is a complete separable metric space and that therefore (Prohorov theorem) every finite measure on U^n has the property that the measure of a Borel subset is the supremum of the measures of its compact subsets will be used. To show that λ is a measure on $\mathbf{S}_0$, it suffices to show that if $S_\bullet$ is a decreasing sequence of sets in $\mathbf{S}_0$, with empty intersection, then $\lim \lambda(S_\bullet) = 0$. This will be shown by showing that if $S_\bullet$ is a decreasing sequence of sets in $\mathbf{S}_0$ and $\lambda(S_n) > \varepsilon > 0$ for all n, then the sequence $S_\bullet$ must have a nonempty intersection. By hypothesis S_n is defined by conditions on coördinates with some finite index set, say $S_n = \{(x_i, i \in I_n) \in \Delta'_n\}$, where I_n is an index set containing a_n points and $\Delta'_n \in \mathbf{B}(U^{a_n})$. The distribution of $\{x_i, i \in I_n\}$ is a Borel measure on U^{a_n} and therefore there is a compact subset Δ''_n of Δ'_n for which

$$\lambda\{(x_i, i \in I_n) \in \Delta''_n\} > \varepsilon(1-3^{-n}). \tag{6.2}$$

Define S_{n1} as the subset of S_n on the left in (6.2), and define $S_{n2} = S_{11} \cap \cdots \cap S_{n1}$. Then $\lambda(S_{n2}) > \varepsilon/2$ and $S_{\bullet 2}$ is a decreasing sequence of nonempty sets in $\mathbf{S}_0$ determined by conditions on values of the functions in S at compact subsets of powers of U. This is precisely the context discussed in Section **II.2** (**Observation**), where it was shown that the sequence $S_{\bullet 2}$ must have a nonempty intersection. Hence the sequence $S_\bullet$ has a nonempty intersection, as was to be proved.

7. Applications of coordinate space measures

To a distribution on $\mathbf{R}^N$, that is, to a Lebesgue-Stieltjes measure on $\mathbf{B}(\mathbf{R}^N)$, correspond N functions, the coordinate functions of $\mathbf{R}^N$, with that joint distribution. In other words, the statement "let $x_1,\ldots,x_N$ be measurable functions with distribution ⋯" is never vacuous; the situation can be realized by coordinate functions on $\mathbf{R}^N$. Theorem 6 justifies the corresponding statement for infinitely many functions with a complete metric separable state space in the following sense. Suppose a finite measure space, together with an infinite family of measurable functions from that space into a complete separable metric space U, is to be constructed, and that each finite set of the functions is to have a prescribed joint distribution. According to Theorem 6, such a family of functions can be realized as the family of coordinate functions on a coordinate space if the prescribed joint distributions are mutually consistent. "Mutually consistent" means that the joint distributions of finite sets of the functions have the property that if finitely many coordinate functions $f_\bullet$ have prescribed joint distribution ν, the joint distribution prescribed for a subset of these functions is the corresponding marginal distribution of ν. In fact, if this is so, these prescribed distributions define a set function λ on the σ algebra $\mathbf{S}_0$ in Theorem 6, with the properties stated in that theorem, and this set function is then a measure, which can be extended to a measure on $\sigma(\mathbf{S}_0)$. The coordinate functions of the measure space obtained in this way have the prescribed joint distributions.

Observe that all these finite dimensional distributions need not be defined explicitly. For example, if the index set is the set of strictly positive integers, the state space is $\mathbf{R}$, and $x_\bullet$ is the sequence of coordinate functions of $\mathbf{R}\times\mathbf{R}\times\cdots$, it is sufficient to prescribe, for $n \geq 1$, the distribution of $x_1,\ldots,x_n$, prescribing this distribution in such a way that it induces as marginal distribution the prescribed distribution of $x_1,\ldots,x_{n-1}$. The distribution prescribed for an arbitrary k-tuple of the coordinate functions is then to be the corresponding k-dimensional marginal distribution of $x_1,\ldots,x_n$, for n so large that the largest of the k-tuple of indices is at most n.

Example (a). (Arbitrary index set) If the state space U is the interval $[0,1]$ of $\mathbf{R}$ and if, for $n \geq 1$, the specified distribution of every n-tuple of coordinate functions is n-dimensional Lebesgue measure on $[0,1]^n$, then these finite dimensional measures are mutually consistent, and Theorem 6 yields Lebesgue measure on the unit "cube" of dimensionality the (not necessarily countable) cardinality of I.

Example (b). Let the index set I be the set of strictly positive integers, N be a strictly positive integer, the state space U be the set $1,\ldots,N$, $p_1,\ldots,p_N$ be positive numbers with sum 1, and (p_{ij}) be an $N\times N$ stochastic matrix. According to Theorem 6 there is a probability measure on the space S of infinite sequences of the integers $1,\ldots,N$, determined by (see Section III.7(b))

$$\lambda\{x_1 = a_1,\ldots,x_n = a_n\} = p_{a_1}\,p_{a_1 a_2}\cdots p_{a_{n-1}a_n}. \tag{7.1}$$

8. Mutually independent random variables on a probability space

Mutual independence of measurable sets and of σ algebras of measurable sets was defined in Section III.6, and independence relations involving random variables will now be reduced to independence relations between σ algebras of sets. Each function x from a probability space into a measurable space determines the σ algebra $\sigma(x)$ defined in Section 1, and, more generally, a family $x_{\bullet}$ of such functions determines a σ algebra $\sigma(x_{\bullet})$. An independence statement involving random variables is to be interpreted as that statement with the random variables replaced by the corresponding σ algebras generated by the random variables. Thus, families $x_{\bullet}$ and $y_{\bullet}$ of random variables are independent of each other if that is true of their σ algebras $\sigma(x_{\bullet})$ and $\sigma(y_{\bullet})$, and so on. If $x_1,\ldots,x_n$ are measurable functions from a probability space $(S,\mathbf{S},P)$ into a measurable space $(S',\mathbf{S}')$, these functions are mutually independent, by definition, if and only if the σ algebras $\sigma(x_1),\ldots,\sigma(x_n)$ are mutually independent, that is, if and only if

$$(8.1)\qquad P\{x_1 \in A'_1,\ldots,x_n \in A'_n\} = P\{x_1 \in A'_1\}\cdots P\{x_n \in A'_n\} \quad (A'_i \in \mathbf{S}',\ i=1,\ldots,n).$$

In particular, measurable sets are mutually independent if and only if their indicator functions are mutually independent.

The condition (8.1) is satisfied if it is satisfied for sets A_i' generating the σ algebra $\mathbf{S}'$. Thus if $(S',\mathbf{S}')$ is $(\mathbf{R},\mathbf{B}(\mathbf{R}))$ in (8.1), it is sufficient if, for each i the sets A_i' run through the intervals of the form $(-\infty,b]$, and it is in this form that the independence definition is sometimes formulated.

Let $\{x_i,\ i \in I\}$ be a family of mutually independent measurable functions ("random variables") from a probability space $(S,\mathbf{S})$ into a measurable space $(S',\mathbf{S}')$. Let I_1 and I_2 be disjoint subsets of I. Then $\sigma(x_i,\ i \in I_1)$ and $\sigma(x_i,\ i \in I_2)$ are mutually independent sub σ algebras of $\mathbf{S}$. Therefore, if x and y are random variables from $(S,\mathbf{S})$ into some state space and are measurable respectively with respect to the first and second of these sub σ algebras, then these two random variables are mutually independent. This statement can be stated more intuitively (but less precisely) by stating that x and y are mutually independent because they are defined respectively in terms of the collections $(x_i,\ i \in I_1)$ and $(x_i,\ i \in I_2)$, which are independent collections.

In particular, if $x_{\bullet}$ is a sequence of mutually independent random variables with state space $\mathbf{R}$, y is a Borel measurable function of some of these random variables, and z is a Borel measurable function of others, it follows that y and z are mutually independent.

If $x_1,\ldots,x_N$ are mutually independent, their joint distribution is the product of the measures of the individual distributions. This property is the content of (8.1).

9. Application of independence: the 0-1 law

The following two elementary facts about σ algebras of measurable sets of a probability space $(S,\mathbf{S},P)$ will be needed.

(a) *A sub σ algebra of* $\mathbf{S}$ *is independent of itself if and only if each of its sets is either null or the complement of a null set.* In fact, a measurable set A is independent of itself if and only if $P\{A\} = P\{A\}^2$, that is, if and only if $P\{A\}$ is 0 or 1, and two sets, each independent of itself, are mutually independent.

(b) *If* $\mathbf{B}$ *is a sub σ algebra of* $\mathbf{S}$ *containing only null sets and their complements, and if x is a random variable measurable from* $(S,\mathbf{B},P)$ into $\bar{\mathbf{R}}$, *then x is equal almost everywhere to a constant.* In fact, if c is a constant the set $\{x < c\}$ must have probability 0 or 1, and this probability is a monotone increasing function of c. If this probability is 0 for all finite c then $x = +\infty$ almost everywhere. If this probability is 1 for all finite c then $x = -\infty$ almost everywhere. Aside from these two cases there must be a point s at which the monotone function jumps from 0 to 1 and then $x = s$ almost everywhere, because $P\{x = s\} = \lim_{\varepsilon \downarrow 0} P\{s - \varepsilon < x < s + \varepsilon\}$.

Theorem. (0-1 **law**) *Let* $\mathbb{F}_\bullet$ *be an increasing sequence of σ algebras of measurable sets of a probability space* $(S,\mathbf{S},P)$. *Let* $\mathbf{G}_\bullet$ *be a decreasing sequence of σ algebras of measurable sets of the space, with* $\mathbf{G}_1 \subset \sigma(\bigcup \mathbb{F}_\bullet)$. *Suppose that, for each value of n, the two σ algebras* $\mathbb{F}_n$ *and* $\mathbf{G}_n$ *are mutually independent. Then* $\bigcap \mathbf{G}_\bullet$ *contains only null sets and their complements.*

In intuitive language: for each value of n, $\mathbb{F}_n$ is the past through time n, $\mathbf{G}_n$ is the future strictly after time n, and by hypothesis the two are mutually independent. The theorem asserts that in the given context, an event in the distant future is either sure to occur or sure not to occur.

Proof. If $A \in \bigcap \mathbf{G}_\bullet$ then, since $A \in \sigma(\bigcup \mathbb{F}_\bullet)$, there is (Theorem IV.3(b)), for every strictly positive integer k, a set A_k in some $\mathbb{F}_n$, depending on k, with $P\{A \Delta A_k) < 1/k$. Since the sets A_k and A are mutually independent, $P\{A \Delta A_k\} = P\{A\}P\{A_k\}$, and therefore $(k \to \infty)$, $P\{A\} = P\{A\}^2$, as was to be proved.

10. Applications of the 0-1 law

In each of the following applications, $x_\bullet$ is a sequence of mutually independent finite valued random variables, and the σ algebras

$$\mathbb{F}_n = \sigma(x_1,\ldots,x_n), \qquad \mathbf{G}_n = \sigma(x_{n+1}, x_{n+2},\ldots).$$

are therefore mutually independent. The σ algebra $\bigcap \mathbf{G}_\bullet$ is the *tail σ algebra*, or *tail* of $x_\bullet$. The hypotheses of the 0-1 law are obviously satisfied. Hence a measurable set (alias event) in the tail of the sequence must be either a null set or the complement of a null set, and a measurable function (alias random variable) measurable with respect to this tail, must be almost everywhere constant.

Application (a). The convergence of the series $\Sigma x_\bullet$ depends only on the tail of $x_\bullet$, and therefore the series converges either almost everywhere or almost nowhere on S.

Application (b). If $A_\bullet$ is an infinite sequence of Borel subsets of $\mathbf{R}$ then

$$P\{\liminf_{n\to\infty}\{x_n \in A_n\}\} = 0 \text{ or } 1, \tag{10.1}$$

$$P\{\limsup_{n\to\infty}\{x_n \in A_n\}\} = 0 \text{ or } 1,$$

because the sets in (10.1) are tail sets. In colloquial language these probabilities are respectively the probabilities that x_n enters the set $\tilde{A}_n$ only finitely often, and that x_n enters the set A_n infinitely often.

Application (c) The random variables

$$\liminf_{n\to\infty}(x_1+\cdots+x_n)/n\,, \quad \limsup_{n\to\infty}(x_1+\cdots+x_n)/n \tag{10.2}$$

are measurable with respect to the tail σ algebra because, say for the first,

$$\liminf_{n\to\infty}(x_1+\cdots+x_n)/n = \liminf_{n\to\infty}(x_m+\cdots+x_n)/n \tag{10.3}$$

for all m. Hence the inferior and superior limits in (10.2) are almost everywhere constant. The two constants are equal, with value say c, not necessarily finite, if and only if the sequence of averages converges almost surely to c. Thus the sequence of averages converges either almost everywhere on S (to a constant function) or almost nowhere on S.

11. A pseudometric for real valued measurable functions on a measure space

Let $(S,\mathbf{S},\lambda)$ be a measure space, denote by $\underline{\mathbf{S}}'$ the class of almost everywhere finite valued measurable functions from S into $\bar{\mathbf{R}}$, and denote by $\underline{\mathbf{S}}$ the subclass of functions f in $\underline{\mathbf{S}}'$ for which $\lambda\{|f| > \varepsilon\}$ is finite for sufficiently large ε, in which case this measure decreases when ε increases, with limit 0 when $\varepsilon \to +\infty$. The class $\underline{\mathbf{S}}'$ is linear in the sense that a linear combination of members of the class coincides, on the set of finiteness of those members, with another member of the class. In the same sense, the subclass $\underline{\mathbf{S}}$ is also linear; it is obviously closed under multiplication by constants, and is closed under summation because

$$\lambda\{|f+g|\geq\varepsilon\}\leq\lambda\{|f|\geq\varepsilon/2\}+\lambda\{|g|\geq\varepsilon/2\}.$$

The norm of a function in $\underline{\mathbf{S}}'$. If f is in $\underline{\mathbf{S}}$, the inequality $\lambda\{|f|>\varepsilon\}<\varepsilon$ is satisfied for sufficiently large values of ε, and if satisfied for one value of ε it is satisfied for all larger values. Define

$$\|f\|_\lambda=\inf\{\varepsilon:\lambda\{|f|>\varepsilon\}<\varepsilon\}\quad\text{if } f\in\underline{\mathbf{S}},\tag{11.1}$$

$$=+\infty\qquad\text{if } f\in\underline{\mathbf{S}}'-\underline{\mathbf{S}},$$

and define

$$d_\lambda(f,g)=\|f-g\|_\lambda,\quad d_\lambda'(f,g)=\|f-g\|_\lambda\wedge 1.\tag{11.2}$$

Here the norm $\|f\|_\lambda$ is the *convergence in measure norm* of f. The following three properties of this norm will be needed.

(a) $\lambda\{|f|>\|f\|_\lambda\}\leq\|f\|_\lambda$. In fact the inequality is trivial if the right-hand side is $+\infty$; if f has finite norm and if $\varepsilon>\|f\|_\lambda$ then $\lambda\{|f|>\varepsilon\}<\varepsilon$.

(b) $\|f\|_\lambda=0$ if and only if $f=0$ almost everywhere because, according to (a), zero norm implies that $f=0$ almost everywhere, and the converse is obvious.

(c) Finite or not, the norm is subadditive:

$$\|f+g\|_\lambda\leq\|f\|_\lambda+\|g\|_\lambda.\tag{11.3}$$

In fact this inequality is trivial unless f and g have finite norms; if they do, the inequalities $\varepsilon_f>\|f\|_\lambda$ and $\varepsilon_g>\|g\|_\lambda$ imply

$$\lambda\{|f+g|>\varepsilon_f+\varepsilon_g\}\leq\lambda\{|f|>\varepsilon_f\}+\lambda\{|g|>\varepsilon_g\}<\varepsilon_f+\varepsilon_g.\tag{11.4}$$

Hence $\|f+g\|_\lambda\leq\varepsilon_f+\varepsilon_g$ and therefore (11.3) is true. Properties (a), (b), and (c) imply that d_λ satisfies the conditions for a pseudometric on the space $\underline{\mathbf{S}}'$, aside from the fact that it may take on the value $+\infty$, and that therefore d_λ' satisfies the conditions for a pseudometric on this space. The d_λ' distance between a function of infinite norm and one of finite norm is 1.

In Section III.13, the notations d_λ' and d_λ referred to distance between sets; here the notation refers to distance between functions. According to Theorem III.14, the class **S** is a complete pseudometric space under the distance definition $d_\lambda'(A,B)=\lambda(A\Delta B)\wedge 1$, and the subclass of sets of finite measure is a closed subset of this space. Under the two uses of the notation d_λ', $d_\lambda'(A,B)=$

$d_\lambda'(1_A,1_B)$, and $d_\lambda(A,B) = d_\lambda(1_A,1_B)$ if $\lambda(A\Delta B) \le 1$. Thus it should cause no confusion if the notations d_λ' and d_λ are used both for pseudometrics on **S**' and **S**. The role of the class of sets of finite measure in the pseudometric of measurable sets is taken in the present context by the class $\underline{\mathbf{S}}$ of measurable functions of finite norm.

12. Convergence in measure (*Notation as in Section* 11)

A sequence $f_\bullet$ of functions in $\underline{\mathbf{S}}'$ *converges in measure with limit* f in $\underline{\mathbf{S}}'$ if there is convergence to f in the d_λ' norm. The limit function is in $\underline{\mathbf{S}}'-\underline{\mathbf{S}}$ [$\underline{\mathbf{S}}$] if and only if all but a finite number of the functions are in $\underline{\mathbf{S}}'-\underline{\mathbf{S}}$ [$\underline{\mathbf{S}}$]. Written out, the sequence converges in measure to f if and only if, for every strictly positive ε, $\lim \lambda\{|f-f_\bullet| > \varepsilon\} = 0$; the sequence is a Cauchy sequence for convergence in measure, that is, a d_λ' Cauchy sequence, if and only if, for every strictly positive ε, $\lim_{m,n\to+\infty}\lambda\{|f_m-f_n| > \varepsilon\} = 0$.

Theorem (*Measure space* $(S,\mathbf{S},\lambda)$). *The space* $(\underline{\mathbf{S}}',d_\lambda')$ *is a complete pseudometric space, and the subset* $\underline{\mathbf{S}}$ *is a closed subset of* $\underline{\mathbf{S}}'$ *at distance* 1 *from* $\underline{\mathbf{S}}'-\underline{\mathbf{S}}$. *The space* $(\underline{\mathbf{S}},d_\lambda)$ *is separable if* λ *is a finite measure and if the* σ *algebra* **S** *is generated up to null sets by a countable subcollection of sets.*

It makes no difference in the last assertion of the theorem whether d_λ or d_λ' is used as the pseudometric on $\underline{\mathbf{S}}$.

Proof. If $f_\bullet$ is a d_λ' Cauchy sequence, choose $a_1 = 1 < a_2 < \cdots$ successively so large that $\lambda\{|f_n-f_{a_k}| > 2^{-k}\} \le 2^{-k}$ for $n > a_k$. Then $\lambda\{|f_{a_{k+1}}-f_{a_k}| > 2^{-k}\} \le 2^{-k}$, and therefore (Cantelli's theorem), except for the points of a null set, $|f_{a_{k+1}}-f_{a_k}| \le 2^{-k}$ for sufficiently large k, depending on the point of S. Thus the sequence $f_{a_\bullet}$ is almost everywhere convergent to some function f, and $\lambda\{|f-f_{a_k}| > 2^{-k+1}\} \le 2^{-k+1}$. The sequence $f_\bullet$ converges in measure to f, because for $2^{-k+1} < \varepsilon/2$ and $n > a_k$,

$$(12.1)\quad \lambda\{|f-f_n| > \varepsilon\} \le \lambda\{|f-f_{a_k}| > \varepsilon/2\} + \lambda\{|f_{a_k}-f_n| > \varepsilon/2\} \le 2^{-k+1} + 2^{-k}.$$

Thus $(\underline{\mathbf{S}}',d_\lambda')$ is a complete pseudometric space. Since every function in $\underline{\mathbf{S}}$ is at distance 1 from $\underline{\mathbf{S}}'-\underline{\mathbf{S}}$, the class $\underline{\mathbf{S}}$ is necessarily closed and the space $\underline{\mathbf{S}}$ with pseudometric, the restriction of d_λ' (or equivalently of d_λ) is complete. In the following, restrictions of d_λ' and d_λ will not have special notation.

To prove separability of $(\underline{\mathbf{S}},d_\lambda)$ when λ is a finite measure and **S** is countably generated up to null sets, observe first that the space $(\mathbf{S},d_\lambda)$ is separable according to Theorem IV.3(d). It follows that the space of indicator functions of sets in **S** is separable in the d_λ pseudometric, and therefore the space of rational valued step functions is separable in this pseudometric. Now suppose that f is in $\underline{\mathbf{S}}$ and define f_n by (1.2). Then f_n is a rational valued step function and the sequence $f_\bullet$ converges in measure to f. It follows that $(\underline{\mathbf{S}}, d_\lambda)$ is separable.

13. Convergence in measure vs. almost everywhere convergence

The following example shows that convergence in measure does not imply almost everywhere convergence, but Theorem 13 shows that the two types of convergence are intimately related.

Example. Order the indicator functions of the right semiclosed subintervals $(j/n, (j+1)/n]$ of $\mathbf{R}$, with $j = 0, \ldots, n-1$ and $n = 1, 2, \ldots$ into a sequence $f_\bullet$. Let λ be Lebesgue measure on $\mathbf{B}((0,1])$. Then $f_\bullet$ is a bounded sequence of measurable functions from $(0,1]$ into $\mathbf{R}$ and, whatever the ordering of the indicator functions, the limits inferior and superior of $f_\bullet$ are identically 0 and 1, respectively, although the sequence $f_\bullet$ converges in λ measure to 0.

Theorem. *Let $f_\bullet$ be a sequence of measurable functions from a measure space $(S, \mathbf{S}, \lambda)$ into $\mathbf{R}$.*

(a) *If $f_\bullet$ converges in measure to a function f, then some subsequence converges almost everywhere to f.*

(b) *If λ is a finite measure and if $f_\bullet$ is almost everywhere convergent to an almost everywhere finite valued measurable function f, then $f_\bullet$ converges in measure to f.*

Proof of (a). Choose $a_1 = 1 < a_2 < \cdots$ successively so large that

$$\lambda\{\, |f_{a_k} - f| > 1/k\} \leq 2^{-k}. \tag{13.1}$$

According to Cantelli's theorem, except possibly for the points of a null set, $|f_{a_k} - f| \leq 1/k$ for all sufficiently large k, and therefore $f_{a_\bullet}$ converges to f almost everywhere.

Proof of (b). Choose $\varepsilon > 0$. In the inclusion relation

$$\{|f_n - f| > \varepsilon\} \subset \bigcup_n^\infty \{|f_\bullet - f| > \varepsilon\}, \tag{13.2}$$

the union on the right decreases to a null set when $n \to \infty$. Hence the measure of the set on the left tends to 0, as was to be proved.

14. Almost everywhere convergence vs. uniform convergence

Theorem (Egoroff). *Let $f_\bullet$ be an amost everywhere convergent sequence of measurable functions from a finite measure space $(S,\mathbf{S},\lambda)$ into a metric space (S',d'). Then to every strictly positive ε corresponds a subset A_ε of S, with $\lambda(S-A_\varepsilon)<\varepsilon$, and with the property that the sequence $f_\bullet$ is uniformly convergent on A_ε.*

Proof. Let f be an almost everywhere limit of the sequence $f_\bullet$, and for strictly positive integers m and n, define $A(m,n) = \{s\colon \sup_{k\geq m} d'(f(s),f_k(s)) > 1/n\}$. For each value of n, $A(\bullet,n)$ is a decreasing sequence of sets, with intersection a null set. Choose an increasing sequence $m_\bullet$ of integers satisfying $\lambda(A(m_n,n)) < 2^{-n}$ for $n\geq 1$, and define $B_j = \bigcup_{n=j}^{\infty} A(m_n,n)$. For each value of j, it follows that $\lambda(B_j) < 2^{-j+1}$, and on $S-B_j$ the sequence $f_\bullet$ converges uniformly, because $d'(f(s)f_k(s)) \leq 1/n$ when $k\geq m_n$ and s is in this set.

15. Function measurability vs. continuity

Theorem (Lusin). *Let f be a measurable function from a complete separable metric finite measure space $(S,\mathbf{B}(S)\lambda)$ into a separable metric space S'. If $\varepsilon>0$, there is a compact subset A_ε of S, with $\lambda(S-A_\varepsilon)<\varepsilon$, and with the property that the restriction of f to A_ε is continuous.*

Proof. Let $B'_\bullet$ be an enumeration of the balls in S' of rational radius, centered at the points of some countable dense subset of S'. The set $f^{-1}(B_n')$ is a measurable subset of S and, according to Theorem IV.4, is contained in an open subset G_n of S, satisfying the inequality $\lambda(G_n - f^{-1}(B_n')) < \varepsilon 2^{-n}$. Observe that the set $f^{-1}(B_n')$ is relatively open in every subset of $\tilde{G}_n \cup f^{-1}(B_n')$ because $f^{-1}(B_n')$ is the intersection of such a subset with G_n. Define

$$A_\varepsilon^* = S - \bigcup_1^\infty \left(G_\bullet - f^{-1}(B'_\bullet)\right) = \bigcap_1^\infty\left(\tilde{G}_\bullet \cup f^{-1}(B'_\bullet)\right).$$

Then $\lambda(S-A_\varepsilon^*)<\varepsilon$ and, according to Prohorov's theorem, a subset A_ε of A_ε^* can be chosen to be compact, with $\lambda(S-A_\varepsilon)<\varepsilon$. For every value of n, $f^{-1}(B_n')$ is relatively open in A_ε. Hence the inverse image under f of every open subset of S' is relatively open in A_ε, that is, the restriction of f to A_ε^* is continuous.

16. Measurable functions as approximated by continuous functions

Theorem. *Let* $(S,\mathbf{B}^*(S),\lambda)$ *be a metric finite measure space, with* $\mathbf{B}^*(S)$ the domain of *the completion* λ *of a measure on* $\mathbf{B}(S)$, *and let f be a measurable function from this space into* $\mathbf{R}$. *There are then*

(a) *a sequence of continuous functions from S into* $\mathbf{R}$ *with almost everywhere limit f;*

(b) *a sequence* $f_\bullet$ *of upper semicontinuous functions from S into* $\bar{\mathbf{R}}$ *for which*

$$f_1 \le f_2 \le \cdots \le f \text{ on } \mathbf{R},\ \lim f_\bullet = f \text{ a.e.};$$

(c) *a sequence* $f_\bullet''$ *of lower semicontinuous functions from S, into* $\bar{\mathbf{R}}$ *for which*

$$f_1 \ge f_2 \ge \cdots \ge f \text{ on } \mathbf{R},\ \lim f_\bullet = f \text{ a.e.}$$

This theorem generalizes Theorem IV.3(c) in the following sense. Suppose that $f = \mathbf{1}_A$ in the present theorem, let $f_\bullet$ be a decreasing sequence of lower semicontinuous functions with almost everywhere limit $\mathbf{1}_A$, and define $A_n = \{f_n + 1/n > 1\}$. Then the sequence $A_\bullet$ is a decreasing sequence of open sets, and $\lim \lambda(A_\bullet) = \lambda(A)$. Thus part (c) of the present theorem yields half of Theorem IV.3(c); part (b) of the present theorem yields the other half of Theorem IV.3(c).

Proof of (a). Let Γ be the class of measurable functions f from S into $\mathbf{R}$ for which (a) is true. This class contains the continuous functions, and it will now be shown that this class is closed under pointwise sequential convergence. To prove this closure, suppose that $g_\bullet$ is a convergent sequence of functions in Γ, with almost everywhere limit g, finite valued almost everywhere. According to Egoroff's theorem, the fact that g_n is in Γ implies that there is a continuous function h_n for which $|h_n - g_n| < 1/n$ except possibly at the points of a set A_n of measure 2^{-n}. An application of Cantelli's theorem shows that the sequence $h_\bullet$ has limit g almost everywhere on S, and therefore g is in Γ. Thus Γ is a class of functions containing the continuous functions and closed under pointwise sequential convergence. It follows that Γ contains the Baire functions, that is (Theorem 4) Γ contains the Borel measurable functions. Finally, if f is measurable from S into $\mathbf{R}$, there is (Section 2(c)) a finite valued Borel measurable function g, equal to f almost everywhere. Since g is in Γ, the function f is also in Γ, as was to be proved.

Proof of (b) **and** (c). If $f_\bullet$ is a sequence of continuous functions with almost everywhere limit f, the almost everywhere equality

$$f = \limsup f_\bullet = \lim_{n\to\infty}(f_n \vee f_{n+1} \vee \cdots)$$

exhibits f as the almost everywhere limit of a decreasing sequence of functions from S into $\bar{\mathbf{R}}^+$. Each member $g_n = f_n \vee f_{n+1} \vee \cdots$ of this decreasing sequence is itself the limit of an increasing sequence of continuous functions, the sequence of partial maxima, and is therefore lower semicontinuous. Finally let H_m be an open set of measure at most $1/m$ that is a superset of every null set $\{g_n \le f_n\}$, and define $h_m = +\infty$ on $H_1 \cap \cdots \cap H_m$, but $h_m = 0$ elsewhere on S. Then the sequence $g_\bullet + h_\bullet$ is a decreasing sequence of lower semicontinuous functions, majorizing f and with almost everywhere limit f. Thus (c) is true. To prove (b), apply (c) to $-f$.

17. Essential supremum and infimum of a measurable function

If f is a measurable function from a measure space $(S,\mathbf{S},\lambda)$ into $\bar{\mathbf{R}}$, ess $\sup_S f$, *the essential supremum of f*, also called *the supremum of f neglecting null sets*, is the supremum of constants c for which $\{f \ge c\}$ is nonnull. Thus $f \le$ ess $\sup_S f$ almost everywhere on S but this almost everywhere inequality is true of no smaller constant. The *essential infimum of f*, ess $\inf_S f$, is defined as -ess $\sup_S(-f)$.

18. Essential supremum and infimum of a collection of measurable functions

If Γ is a collection of measurable functions from a measure space into $\bar{\mathbf{R}}$, the pointwise supremum of Γ, that is, the function defined at each point s as $\sup\{f(s): f \in \Gamma\}$, need not be measurable unless Γ is countable. A looser supremum of Γ, described in the following theorem, avoids this measurability difficulty at the cost of ignoring null sets.

Theorem. *Let Γ be a class of measurable functions from a σ finite measure space $(S,\mathbf{S},\lambda)$ into $\bar{\mathbf{R}}$. There is then a measurable function g, uniquely determined up to null sets by the two properties that if f is in Γ then $g \ge f$ almost everywhere, and if h is another measurable function with this property, then $h \ge g$ almost everywhere. One choice of g is the supremum of a suitably chosen countable subset of Γ.*

The function g is the *essential supremum*, or *essential upper envelope*, or *essential order supremum* of Γ.

Proof. It is trivial that two functions with the properties ascribed to the essential supremum are equal almost everywhere and that a measurable function equal almost everywhere to a function with those properties also has the properties. Thus "*the* essential supremum" is unique only up to null sets.

In proving that there is a function with these properties, it can be assumed without loss of generality that Γ contains the pointwise maximum of every finite set of its functions, because the class of all such finite maxima will have the same essential supremum, if any, as the original class. Moreover a countable supremum of members of the new class is also a countable supremum of members of the original class. Furthermore it can be supposed that the members of Γ have a common finite constant upper bound, because the problem is an order problem, and each function u of Γ can be replaced by $\arctan u$ to obtain a common upper bound without changing the order relations of the functions. Finally, it can be supposed that λ is a finite measure, because if S is the disjunct union $\bigcup S_\bullet$ of a sequence of sets of strictly positive finite measure, the measure λ' defined by

$$\lambda'(A) = \Sigma_n \lambda(S_n \cap A)/(2^n \lambda(S_n))$$

is a finite measure with the same measurable sets and null sets as λ. Choose a member f_1 of Γ. If there is a function h_1 in Γ with $\lambda\{h_1 \geq f_1+1\} \geq 1$, define $f_2 = f_1 \vee h_1$. If there is no such function h_1, stop. If there is a function h_1, and if there is a function h_2 in Γ, with $\lambda\{h_2 \geq f_2+1\} \geq 1$, define $f_3 = f_2 \vee h_2$. If there is no such function, stop,... . This procedure must finally stop or there would be a monotone increasing (neglecting null sets) sequence of functions in Γ, each exceeding its predecessor by at least 1 on a set of measure ≥ 1. The limit superior of this sequence of sets, that is the class of points in infinitely many of the sets, has measure ≥ 1 (Theorem III.8), and at each point of this limit superior the class Γ is unbounded, contrary to fact. Hence the procedure must end with a finite monotone increasing sequence, say $f_1, \ldots, f_{a_1}$. Repeat this procedure, starting with f_{a_1}, but with the number 1 replaced by 1/2, next by 1/3, and so on, to obtain an increasing finite or infinite sequence $f_1, f_{a_1}, f_{a_2}, \ldots$. Define $a_0 = 1$. Under these definitions, $f_{a_k} \geq f_{a_{k-1}}$, and the difference is at least $1/k$ on a set of measure at least $1/k$, but there is no function in Γ exceeding f_{a_k} by at least $1/k$ on a set of measure at least $1/k$. If the procedure stops, ending with f_{a_k}, then for $n \geq 1$ there is no function in Γ exceeding f_{a_k} by at least $1/n$ on a set of measure $\geq 1/n$. Hence there is no function h in Γ strictly exceeding f_{a_k} on a set of strictly positive measure, and f_{a_k} is the required function g, the pointwise supremum of a finite subset of Γ. If the procedure never stops, the limit of the sequence $f_{a_\bullet}$ is the required function g, the pointwise supremum of a countable subset of Γ.

The *essential infimum of* Γ, also called the *essential lower envelope of* Γ, is defined as the negative of the essential upper envelope of $-\Gamma$.

VI
Integration

1. The integral of a positive step function on a measure space $(S,\mathbb{S},\lambda)$

A positive step function f can be written uniquely in the form $f = \Sigma a_\bullet \mathbf{1}_{A_\bullet}$, a linear combination, with positive pairwise distinct coefficients, of indicator functions of a finite disjunct sequence of measurable sets. Define *the integral of f on S with respect to* λ, using this representation of f, by

$$\int_S f\,d\lambda = \Sigma a_\bullet \lambda(A_\bullet). \tag{1.1}$$

Observe that if equality of coefficients is allowed in the representation of f, the value of the integral, as just defined, is still given by the sum on the right in (1.1).

If $f = \Sigma\, b_\bullet \mathbf{1}_{B_\bullet}$ is a step function, with each b_j positive, the integral of f, defined in the preceding paragraph, will now be shown to be equal to $\Sigma b_\bullet \lambda(B_\bullet)$. If $C_\bullet$ is a finite disjunct sequence of measurable sets, chosen in such a way that each set B_m is a union of sets in $C_\bullet$, then

$$f = \Sigma_j \left(\Sigma_{B_m \supset C_j} b_m\right) \mathbf{1}_{C_j}, \tag{1.2}$$

and therefore

$$\int_S f\,d\lambda = \Sigma_j \left(\Sigma_{B_m \supset C_j} b_m\right) \lambda(C_j) = \Sigma_m b_m \,\Sigma_{C_j \subset B_m} \lambda(C_j) = \Sigma_m b_m\, \lambda(B_m), \tag{1.3}$$

as asserted.

It will frequently be convenient, especially in undisplayed text, to write $\lambda[f]$ for the integral with respect to a measure λ, of a function f on a space.

The class of integrands whose integrals are defined will be extended far beyond the positive step functions, but proofs in integration theory are frequently based on properties of integrals of these very special functions. Step functions play the same role in integration with respect to measures that linear combinations of indicator functions of intervals play in the theory of the Riemann integral on intervals of **R** (see Section 20).

If f and g are positive step functions, then $f \le g$ implies the same inequality for their integrals, and if a and b are positive constants,

$$\int_S (af+bg)\, d\lambda = a \int_S f\, d\lambda + b \int_S g\, d\lambda. \tag{1.4}$$

In fact these two assertions become trivial if the partitions of S in representations of f and g are combined into a common partition.

2. The integral of a positive function

If f is a measurable function from S into $\bar{\mathbf{R}}^+$, define

$$\int_S f\, d\lambda = \sup\left\{ \int_S g\, d\lambda\colon g \le f,\ g \text{ is a positive step function}\right\}, \tag{2.1}$$

and if A is a measurable set, define *the integral of f on A* by

$$\int_A f\, d\lambda = \int_S f 1_A\, d\lambda. \tag{2.2}$$

Observe that if the measure space $(S,\mathbb{S},\lambda)$ is replaced by the measure space $(A,\mathbf{S}_A,\lambda_A)$ consisting of a set A in $\mathbf{S}$, the σ algebra $\mathbf{S}_A$ of subsets of A in $\mathbf{S}$, and the restriction λ_A of λ to $\mathbf{S}_A$, the integral on this measure space of the restriction to A of a measurable function f from A to $\bar{\mathbf{R}}^+$ is equal to the integral of f on A. Thus, for positive functions, it is no more general to discuss integration of functions on S than integration of functions on a measurable subset of S, in the sense that in both cases one can consider the discussion as one of integration of functions with domain a measure space. Since the integration of not necessarily positive functions is based on that of positive functions, there is no reason to derive integration properties for functions defined on subsets of S rather than functions defined on S.

It is trivial from (2.1) that inequality $f \le g$ for positive functions implies the same inequality for their integrals. It is almost as trivial that for given f the set function of A defined by (2.2) is additive. In fact if A and B are disjoint measurable sets with union C, let $\alpha_\bullet$, $\beta_\bullet$, and $\gamma_\bullet$ be sequences of positive step functions majorized respectively by $f1_A, f1_B$, and $f1_C$, with

$$\lim \lambda[\alpha_\bullet] = \lambda[f1_A],\ \lim \lambda[\beta_\bullet] = \lambda[f1_B],\ \lim \lambda[\gamma_\bullet] = \lambda[f1_C]. \tag{2.3}$$

Replace γ_n by $\gamma_n \vee (\alpha_n+\beta_n)$, and then replace α_n and β_n by $\gamma_n 1_A$ and $\gamma_n 1_B$ respectively. These changes can only increase the functions changed, without negating the conditions imposed on the sequences. After these changes have been made, $\gamma_n = \alpha_n+\beta_n$, and the stated additivity of the integral as a function of the integration set follows from the additivity stated in (1.4) when the integrand is a positive step function.

3. Integration to the limit for monotone increasing sequences of positive functions

The following theorem is the basic theorem on going to the limit under the sign of integration, to which the other theorems legitimizing the limit procedure will be reduced.

Theorem (**Beppo-Levi**). *Let $f_\bullet$ be an increasing sequence of measurable functions from a measure space* $(S,\mathbf{S},\lambda)$ *into* $\overline{\mathbf{R}}^+$, *with limit f. Then*

$$\lim \int_S f_\bullet \, d\lambda = \int_S f \, d\lambda. \tag{3.1}$$

Proof. Since the sequence $f_\bullet$ is monotone increasing and is majorized by f, it is trivial that the limit in (3.1) exists and is at most the value of the integral on the right. If the limit on the left side of (3.1) is $+\infty$, the theorem is therefore trivial. Thus it will be supposed, in proving (3.1), that the limit in (3.1) is finite.

(a) **Proof of** (3.1) **when f is a positive constant function** c. If $c=0$, (3.1) is trivial. If $c > 0$, choose c' with $0 < c' < c$. Then

$$\int_S f \, d\lambda \ge \int_S f_n \, d\lambda \ge \int_{\{f_n \ge c'\}} f_n \, d\lambda \ge c'\lambda\{f_n \ge c'\}. \tag{3.2}$$

Since the last term in (3.2) tends to the limit $c'\lambda\{S\}$ as n increases, and since c' can be taken arbitrarily close to c,

$$\int_S f \, d\lambda \ge \lim \int_S f_\bullet \, d\lambda \ge c\lambda(S) = \int_S f \, d\lambda, \tag{3.3}$$

as was to be proved under (a).

(b) **Proof of** (3.1) **when f is a step function.** When f is a step function, an application of the result in (a) to the restriction of f to each of the sets in a representation of f yields (3.1).

(c) **Proof in the general case.** Choose a positive step function g_n in such a way that $g_n \le f_n$ and $\lambda[f_n] \le \lambda[g_n] + 1/n$. It can be assumed that $g_\bullet$ is an increasing sequence, replacing g_n by $g_1 \vee \cdots \vee g_n$ if necessary, to achieve monotonicity. If g is a positive step function and $g \le f$, the sequence $g \wedge g_\bullet$ is an increasing sequence of positive step functions with limit the step function g, and it follows from (b) that

$$\int_S g \, d\lambda = \lim \int_S g \wedge g_\bullet \, d\lambda \le \lim \int_S g_\bullet \, d\lambda = \lim \int_S f_\bullet \, d\lambda. \tag{3.4}$$

The limit in (3.1) is therefore at least $\lambda[g]$, and therefore at least the right side of (3.1). It has already been noted that the reverse inequality is trivial.

Linearity of the function $f \mapsto \lambda[f]$ for $f \geq 0$. To prove (1.4) for positive constants a and b and positive measurable functions, it need only be remarked that in view of the Beppo-Levi theorem, if $f_\bullet$ and $g_\bullet$ are increasing sequences of positive step functions with respective limits f and g (see Section V.1), then the fact that (1.4) is true for f_n and g_n implies that the equality is true for f and g.

4. Final definition of the integral

Consider an arbitrary measurable function from $(S,\mathbf{S},\lambda)$ into $\bar{\mathbf{R}}$. The function can be written in many ways as the difference between two measurable functions from S into $\bar{\mathbf{R}}^+$, never simultaneously $+\infty$, for example $f = f^+ - f^-$ where

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

This representation of f as a difference between two positive functions never simultaneously $+\infty$ minimizes those two functions because if $f = g-h$ is such a representation, then it is trivial that $f \vee 0 \leq g$ and $(-f) \vee 0 \leq h$. If either f^+ or f^- has a finite integral, the integral of f is defined by

$$(4.1) \qquad \int_S f\,d\lambda = \int_S f^+\,d\lambda - \int_S f^-\,d\lambda.$$

If, say, f^+ has a finite integral and f_1-f_2 is a representation of f as the difference between two positive measurable functions simultaneously $+\infty$ on at most a null set, for which either f_1 or f_2 has a finite integral, then $f^+ + f_2 = f^- + f_1$ almost everywhere and therefore f_1 must have a finite integral and (4.2) is also true for the representation of f in terms of f_1 and f_2 :

$$(4.2) \qquad \int_S f\,d\lambda = \int_S f_1\,d\lambda - \int_S f_2\,d\lambda.$$

If $|f|$ has a finite integral, that is, if f^+ and f^- have finite integrals, then f is *integrable*. By definition a positive ($\leq +\infty$) measurable function can always be integrated, even though it is not called integrable unless the integral is finite.

The integral of f over a measurable set A. This integral is defined by (2.2) when the integral of $f\mathbf{1}_A$ over S is well defined. It is trivial that the integral of $f\mathbf{1}_A$ over S is well defined if and only if the integral of the restriction of f to A is well defined on the measure space $(A,\mathbf{S}_A,\lambda_A)$ (notation as in the paragraph following (2.2)), and that if these integrals are well defined they are equal.

According to the definition of an integral, if f is a function that is identically $+\infty$ on a measurable subset of S, the integral of f on that set is well defined, equal to 0 if the set is null, equal to $+\infty$ otherwise. It follows that if f has an integral over the whole space and $f=+\infty$ on a set of strictly positive measure, then the integral of f over the whole space is $+\infty$.

Integrands not defined on null sets. If the measure involved is complete, an integrand can be changed arbitrarily on a null set without affecting measurability of the integrand, its integrability, or the value of the integral if there is integrability. In view of this fact, integrands will be allowed that are not defined on some null set of the integration domain, under either the convention that the integration domain is decreased to the domain of definition, of the integrand, or that the integrand is defined (arbitrarily) where it is not already defined. If the measure is not complete and the integrand is not defined on a subset of a null set, the integration is treated as if the measure has been completed.

Notation. It is sometimes convenient to write an integral in an expanded form, in which the integrand's argument is explicit, writing

$$\int_S f(s)\,\lambda(ds)$$

instead of

$$\int_S f\,d\lambda \text{ or } \lambda[f].$$

The notation $E\{f\}$ (read "expectation of f") is commonly used for this integral when $\mu(S) = 1$, that is, when the measure space is a probability space.

Basic integration properties. All functions involved in the following list of properties are supposed measurable.

(a) *If* $f \geq 0$ *almost everywhere, then* $\lambda[f] = 0$ *implies that* $f = 0$ *almost everywhere.*

(b) *If f is integrable, then* $|f| < +\infty$ *almost everywhere, and the inequality* $|g| \leq |f|$ *implies that g is also integrable.*

(c) *If f and g are integrable and if a and b are constants, then $af+bg$ is integrable, and* (1.4) *is true.*

(d) *If f and g are integrable and* $f \leq g$ *almost everywhere, then* $\lambda[f] \leq \lambda[g]$. *In particular,*

$$\left|\int_S f\,d\lambda\right| \leq \int_S |f|\,d\lambda \tag{4.3}$$

if f is integrable.

(e) *If f is measurable and if ϕ is a monotone increasing function from* $\mathbf{R}^+$ *into* $\mathbf{R}^+$, *then*

$$\int_S \phi(|f|)\,d\lambda \geq \int_{\{|f|\geq c\}} \phi(c)\,d\lambda = \phi(c)\lambda\{|f| \geq c\}. \tag{4.4}$$

In particular, if $f \geq 0$,

(4.5)
$$\int_S f\,d\lambda \geq c\lambda\{f \geq c\}.$$

Assertion (a) is true because

$$0 = \int_S f\,d\lambda \geq \int_{\{f \geq 1/n\}} f\,d\lambda \geq \lambda\{f \geq 1/n\}/n, \text{ and } \{f > 0\} = \cup_1^{\infty} \{f \geq 1/n\}.$$

The other assertions need no comment.

Integration of complex valued functions. The definition of measurability covers functions from a measure space into the complex plane: such a function $f = f_1 + if_2$ is measurable if and only if its real part $\Re f$ and its imaginary part $\Im f$ are. The function f is defined as *integrable* when $\Re f$ and $\Im f$ are integrable, and in that case (definition),

$$\int_S f\,d\lambda = \int_S \Re f\,d\lambda + i\int_S \Im f\,d\lambda.$$

The linearity property (c) with complex constants a and b is true for complex integrands, because it is true for real integrands. Inequality (4.3) is true for complex valued functions, because if c is a complex number of modulus 1, chosen to make real and positive the product of c and the value of the integral on the left in (4.3),

(4.6)
$$\left|\int_S f\,d\lambda\right| = \int_S cf\,d\lambda = \int_S \Re(cf))\,d\lambda \leq \int_S |f|\,d\lambda.$$

In the following, integrands will be real valued unless the contrary is stated explicitly.

The class $L^p(S,\mathbf{S},\lambda)$. This class, for $1 \leq p < +\infty$, is the class of measurable functions f for which $|f|^p$ is integrable; for $p = +\infty$ the class is defined as the class of essentially bounded measurable functions, that is those measurable functions f for which there is a constant c such that $|f| \leq c$ almost everywhere. The identification of space, σ algebra and measure, will be omitted from the notation if the context is clear. The notation (read "L_p norm of f", and when $p = +\infty$, "essential norm of f")

(4.7)
$$\|f\|_p = \Big(\int_S |f|^p\,d\lambda\Big)^{1/p} \qquad (p < +\infty),$$
$$= \text{ess sup}_S |f| \qquad (p = +\infty),$$

is used for functions in L^p. A family of functions in L^p is said to be L^p *bounded* if the supremum of the set of L^p norms of the functions is finite. The class L^p is linear: it is trivial that a constant multiple of a function in the class is also in the class, and in view of the inequality

$$(4.8) \qquad (a+b)^p \leq 2^p(a^p \vee b^p) \leq 2^p(a^p+b^p)$$

for positive numbers a and b, the sum of two functions in the class is in the class.

L^p and the convergence in measure norm. If f is in L^p for some finite value of p, it is obvious that the convergence in measure norm $\|f\|_\lambda$ is finite. Thus the d_λ pseudometric is applicable to L^p.

The class L^p for complex valued functions is the class of measurable complex valued functions f with $|f|$ in the class L^p of real functions, for which it is necessary and sufficient that the real and imaginary parts of f are in real L^p because

$$(4.9) \qquad |\Re f| \vee |\Im f| \leq |f| \leq |\Re f| + |\Im f|.$$

Extension of the Beppo-Levi theorem. In the Beppo-Levi theorem, the hypothesis of positivity of $f_\bullet$ can be weakened to suppose only that f_n for some value of n majorizes a function g for which g^- is integrable. To prove the theorem under this hypothesis, apply Theorem 3 as stated to the sequence $f_\bullet + g^-$, an increasing sequence of positive functions if some initial members are omitted, and then use linearity property (c) to remove g^-.

The Young definition of an integral. Theorem V.16 suggests that one way to define the integral of a function is to define the integral of a semicontinuous function and then to define the integral of a function f as the infimum [supremum] of the integrals of larger lower [smaller upper] semicontinuous functions, if these extreme values are equal. This is the Young approach to integration.

5. An elementary application of integration

Let $(S,\mathbf{S},\lambda)$ be a finite measure space, and let $A_1,\ldots,A_n$ be measurable sets. Integration of the equations I(2.9) yields

(5.1)
$$\lambda(\cup A_\bullet) = \sum_{i\geq 1} \lambda(A_i) - \sum_{i<j} \lambda(A_i \cap A_j) + \cdots + (-1)^{n-1}\lambda(A_1 \cap \cdots \cap A_n),$$

$$\lambda(\cap A_\bullet) = \sum_{i\geq 1} \lambda(A_i) - \sum_{i<j} \lambda(A_i \cup A_j) + \cdots + (-1)^{n-1}\lambda(A_1 \cup \cdots \cup A_n).$$

For example, let S be the space of $n!$ permutations of the integers $1,\ldots,n$, and assign each singleton of S the λ measure $1/n!$. Then λ is a probability measure. Let A_i be the set of permutations in S for which the ith place is i, that is, the set of permutations matching the identity permutation in the ith place. Then $\cup A_\bullet$ is the set of permutations with at least one match with the identity permutation. According to the first equation in (5.1), this set has measure

(5.2)
$$\left[n[(n-1)!] - \frac{n(n-1)}{2!}(n-2)! + \cdots + (-1)^{n-1}\right]/n!$$
$$= 1 - \frac{1}{2!} + \cdots + \frac{(-1)^{n-1}}{n!},$$

which is close to e^{-1} when n is large. Roughly speaking, about a third of the permutations have at least one match with the identity permutation.

Card mixing interpretation. If a deck of n cards, is shuffled thoroughly, the probability that at least one card is in the same position as before the shuffling is about a third. Here *shuffling thoroughly* is translated mathematically into the assignment of probability $1/n!$ to each shuffling permutation. In an attempt to make this integration problem more interesting, succeeding in making it more confusing, the context has sometimes been changed to replace n cards by n drunken men who try to find their way home (very cleverly arriving at least finding different homes), the conclusion being that the probability is about one third that at least one will get to his own home.

6. Set functions defined by integrals

Let f be a measurable function from a measure space $(S,\mathbf{S},\lambda)$ into $\overline{\mathbf{R}}^+$. Define μ on $\mathbf{S}$ by

(6.1)
$$\mu(A) = \int_A f\, d\lambda.$$

Then μ is a measure: finite additivity follows from Section 4(c), and countable additivity then follows from the Beppo–Levi theorem formulated in terms of a sum instead of a sequence. More generally, if f need not be positive, but if either f^- or f^+ is integrable, this argument can be applied to f^- and f^+ to show that μ in (6.1) is a signed measure.

The following theorem covers a change of variable under the sign of integration.

Theorem. *If* μ *is defined by* (6.1), *with* $f \geq 0$, *and if* g *is a measurable function from* S *into* $\overline{\mathbf{R}}^+$, *then*

$$\int_S g \, d\mu = \int_S gf \, d\lambda. \tag{6.2}$$

Proof. Equation (6.2) is true, by definition of μ, when g is the indicator function of a measurable set, and the equation is therefore also true when g is a positive-coefficient linear combination of such indicator functions. In particular, if $A_{nj} = \{(j-1)2^{-n} \leq g < j2^{-n}\}$, and

$$g_n = \sum_{j=1}^{4^n} (j-1)2^{-n}\mathbf{1}_{A_{nj}} + (+\infty)\mathbf{1}_{\{g=+\infty\}},$$

then (6.2) is true for $g=g_n$. Since $g_\bullet$ is an increasing sequence with limit g, (6.2) is true as stated.

The function g in (6.2) is μ integrable if and only if gf is λ integrable. More generally, if g is a measurable function from S into $\overline{\mathbf{R}}$, the preceding discussion is applicable to g^+ and g^- and thereby shows that if $|g|$ is μ integrable, equivalently if $|g|f$ is λ integrable, then (6.2) remains true.

7. Uniform integrability test functions

A *uniform integrability test function* is a function ϕ from $\mathbf{R}^+$ into $\mathbf{R}^+$, with the property that the function $t \to \phi(t)/t$ is monotone increasing with limit $+\infty$ at $+\infty$. The function ϕ itself is then necessarily also monotone increasing. The following intuitively obvious theorem will be needed.

Theorem. *If* f *is a positive integrable function on a finite measure space* $(S,\mathbf{S},\lambda)$, *there is a convex uniform integrability test function* ϕ *for which* $\phi(f)$ *is also integrable.*

Proof. Define the measure μ by (6.1) and define $\psi_1(\alpha) = \mu\{[\alpha,+\infty)\}$ for $\alpha \geq 0$. The function ψ_1 is monotone decreasing, with limit 0 at $+\infty$. Define $\psi(n) = \psi_1\big((n-1)\vee 0\big)$ for n a positive integer, and define ψ between the positive integers by linear interpolation. Then $\psi_1 \leq \psi$, that is,

$$\int_{\{f\geq\alpha\}} f\,d\lambda \leq \psi(\alpha), \tag{7.1}$$

and ψ is a continuous monotone decreasing function on $\mathbf{R}^+$, with limit 0 at $+\infty$. If ψ ever vanishes, the function f must be essentially bounded, in which case the existence of the desired uniform integrability test function is trivial. If ψ is strictly positive, define

$$a_0 = 0, \quad a_n = \inf\{s: \psi(s) \leq 1/n\} \quad (n\geq 1); \qquad \phi(s) = s[\psi(s)]^{-1/2}.$$

The function ϕ is a uniform integrability test function, and the following inequalities show that $\phi(f)$ is integrable:

$$\int_S \phi(f)\,d\lambda = \int_{\{f>0\}} \frac{\phi(f)}{f}\,d\mu \leq \sum_{n=1}^{\infty} n^{1/2}\big[\psi_1(a_{n-1}) - \psi_1(a_n)\big] \tag{7.2}$$

$$\leq \sum_{n=1}^{\infty} \psi_1(a_n)[n^{1/2} - (n-1)^{1/2}] \leq \sum_{n=1}^{\infty} n^{-3/2} < +\infty.$$

To finish the proof it will be shown, as is sufficient, that if ϕ is a uniform integrability test function there is a convex uniform integrability test function majorized by ϕ for sufficiently large values of the argument. Choose a strictly positive number b_1 that is so large that $\phi(s) \geq s$ when $s \geq b_1$. If $b_1,\ldots,b_{n-1}$ have been chosen, choose b_n so large that $b_n > b_{n-1}$ and $\phi(s) \geq ns$ when $s \geq b_n$. Define $\phi_1(s) = b_1$ when $0 \leq s \leq b_1$ and $\phi_1(s) = (n-1)(s-b_{n-1})+\phi_1(b_{n-1})$ when $b_{n-1} \leq s \leq b_n$. Then $\phi_1 \leq \phi$ on $[b_1,+\infty)$ and ϕ_1 is a uniform integrability test function.

8. Integration to the limit for positive integrands

Theorem (Fatou). *Let $f_\bullet$ be a sequence of measurable functions from a measure space $(S,\mathbf{S},\lambda)$ into $\mathbf{R}^+$, and define $f = \liminf f_\bullet$. Then*

$$\int_S f\,d\lambda \leq \liminf \int_S f_\bullet\,d\lambda. \tag{8.1}$$

If f and each function f_n are integrable, and if

$$\int_S f\,d\lambda = \lim \int_S f_\bullet\,d\lambda, \tag{8.2}$$

then

(8.3) $$\lim \int_S |f-f_\bullet|\, d\lambda = 0.$$

If the sequence $f_\bullet$ converges in measure to a function g, f can be replaced by g in these assertions.

As of 1991 the oldest generation upholds tradition by still calling this theorem "Fatou's lemma."

Proof. Apply Beppo-Levi's theorem to yield

(8.4) $$\int_S f\, d\lambda = \int_S \sup_{k>1} \inf_{m>k} f_m\, d\lambda = \lim_{k\to\infty} \int_S \inf_{m>k} f_m\, d\lambda$$
$$= \liminf_{k\to\infty} \int_S \inf_{m>k} f_m\, d\lambda \le \liminf \int_S f_\bullet\, d\lambda,$$

as was to be proved.

To prove (8.3) when (8.2) is true and all the functions involved are integrable, majorize the integral in (8.3):

(8.5) $$\int_S |f-f_n|\, d\lambda \le -\int_S (f-f_n)\, d\lambda + 2\int_S f\, d\lambda - 2\int_S \inf_{k\ge n} f_k\, d\lambda.$$

When $n\to\infty$, the first integral on the right tends to 0 by hypothesis. The third integral on the right tends monotonely to the second by the Beppo-Levi theorem . Thus the right side of (8.5) tends to 0, as asserted.

If the sequence $f_\bullet$ converges in measure to g, choose a subsequence along which the sequence $\lambda[f_\bullet]$ has as limit the inferior limit in (8.1), and then choose a further subsequence along which $f_\bullet$ converges almost everywhere to g. Apply what has been proved to this subsequence to find that (8.1) is valid with f replaced by g. If all the functions involved are integrable and (8.2) is true, then what has already been proved implies that the sequence $\lambda[|f-f_\bullet|]$ has limit 0 along every subsequence of $f_\bullet$ for which $f_\bullet$ converges to f almost everywhere. Thus every subsequence of $\lambda[|f-f_\bullet|]$ has a further subsequence with limit 0, and therefore the sequence itself has limit 0, as asserted.

9. The dominated convergence theorem

Theorem (Lebesgue). *Let $f_\bullet$ be a sequence of measurable functions from a measure space $(S,\mathbf{S},\lambda)$ into $\bar{\mathbf{R}}$, and suppose that* $\sup|f_\bullet|$ *is integrable. If, either in measure or almost everywhere, the sequence converges to a function f, then*

(9.1) $$\int_S f\, d\lambda = \lim \int_S f_\bullet\, d\lambda.$$

This theorem is called the *dominated convergence theorem*, because the condition of integrability of $\sup|f_\bullet|$ is commonly phrased in the uneconomical form *there is a positive integrable function g such that* $|f_n| \le g$ *almost everywhere, for every value of n.*

Proof. According to Fatou's theorem, whether the limit is an in measure or an almost everywhere limit,

$$\int_S \lim\,(g+f_\bullet)\,d\lambda \le \liminf \int_S (g+f_\bullet)\,d\lambda, \tag{9.2}$$

that is,

$$\int_S f\,d\lambda \le \liminf \int_S f_\bullet\,d\lambda. \tag{9.3}$$

Inequality (9.3), together with its application to the sequence $-f_\bullet$, yields (9.1).

The d_λ continuity of the function μ: $A \to \int_A f\,d\lambda$.

If f is integrable, μ is a finite valued signed measure (Section 6). If λ is a finite measure, d_λ is a pseudometric for **S** (Section III.13) and μ *is a continuous function on* $(\mathbf{S},d_\lambda)$; equivalently, if $A_\bullet$ is a sequence of measurable sets whose measures tend to 0, then the sequence $\mu(A_\bullet)$ has limit 0. This conclusion follows from Theorem 9, because the sequence $f\mathbf{1}_{A_\bullet}$ converges to 0 in measure, and the absolute value of each integrand is majorized by $|f|$.

Bounded convergence theorem. This name is given to the dominated convergence theorem's special case in which λ is a finite measure and there is a constant c for which $\sup|f_\bullet| \le c$ almost everywhere. The classical corresponding theorem, that a uniformly convergent sequence of continuous functions on a compact interval of **R** can be integrated (Riemann integral) to the limit, is a special case of the bounded convergence theorem, because each function of the uniformly convergent sequence is bounded, and that fact together with the uniform convergence implies that the sequence has an overall bound. (The fact that Riemann integration is a special case of Lebesgue integration will be proved in Section 20.)

10. Integration over product measures

In this section, let $(S_1,\mathbf{S}_1,\lambda_1)$ and $(S_2,\mathbf{S}_2,\lambda_2)$ be σ finite measure spaces. The measures may or may not be complete. Provide the product space $S = S_1 \times S_2$ with the product σ algebra $\mathbf{S} = \sigma(\mathbf{S}_1 \times \mathbf{S}_2)$ and product measure $\lambda = \lambda_1 \times \lambda_2$. If A is a subset of S, for each point s_1 of S_1 define $A_1(s_1) = \{s_2\colon (s_1,s_2) \in A\}$. It is convenient to write a function f on S as $f(\bullet,\bullet)$ to exhibit it as a function of its arguments in S_1 and S_2. Recall from Sections **II**.5 and **V**.2, that if A is in $\mathbf{S}$, then for each point s_1 of S_1, the set $A_1(s_1)$ is in $\mathbf{S}_2$, and if f is measurable from $(S,\mathbf{S})$ into $\bar{\mathbf{R}}$, then $f(s_1,\bullet)$ is measurable from $(S_2,\mathbf{S}_2)$ into $\bar{\mathbf{R}}$.

Theorem (Fubini, Tonelli). (a) *If f is measurable from $(S,\mathbf{S})$ into $\bar{\mathbf{R}}^+$, the integral $\int_{S_2} f(\bullet,s_2)\,\lambda_2(ds_2)$ is a measurable function from $(S_1,\mathbf{S}_1)$ into $\bar{\mathbf{R}}^+$, and*

$$\int_S f\,d\lambda = \int_{S_1} \lambda_1(ds_1) \int_{S_2} f(s_1,s_2)\,\lambda_2(ds_2). \tag{10.1}$$

(b) *If f is measurable from $(S,\mathbf{S})$ into $\bar{\mathbf{R}}$ and is λ integrable, then $f(s_1,\bullet)$ is λ_2 integrable for λ_1 almost every point s_1, the inner integral in* (10.1) *defines a measurable and λ_1 integrable function from $(S_1,\mathbf{S}_1)$ into $\bar{\mathbf{R}}$, and* (10.1) *is true.*

Observe that under the hypotheses of the theorem, the iterated integral can be evaluated in either order.

Proof of (a) **for finite measures.** If A is in the class $\mathbf{S}_1\times\mathbf{S}_2$ of product sets, it is trivial that

$$\lambda(A) = \int_{S_1} \lambda_2\big(A_1(s)\big)\,\lambda_1(ds). \tag{10.2}$$

It follows that the class Γ of sets A in $\mathbf{S}$ for which (10.2) is true contains the algebra $\sigma_0(\mathbf{S}_1\times\mathbf{S}_2)$ of disjunct unions of product sets. Moreover (bounded convergence theorem) Γ is a monotone class, and therefore (Theorem **II**.6) contains $\sigma(\mathbf{S}_1\times\mathbf{S}_2)=\mathbf{S}$. It is important to note that (10.2) implies that A is λ null if and only if $A_1(s)$ is λ_2 null, for λ_1 almost every point s.

If f is a measurable function from $(S,\mathbf{S},\lambda)$ into $\mathbf{R}^+$, the function

$$\int_{S_2} f(\bullet,s_2)\,\lambda_2(ds_2)$$

is a measurable function from $(S_1,\mathbf{S}_1)$ into $\bar{\mathbf{R}}^+$ and (10.1) is true. In fact (10.1) reduces to (10.2) when f is the indicator function of a set in $\mathbf{S}$, and (10.1) is therefore true (along with the measurability of the inner integral) when f is a positive step function. For general positive f, **V**(1.1) exhibits f as the limit of an increasing sequence of positive step functions. An application of the Beppo-Levi theorem to this increasing sequence yields (10.1).

Proof of (b) **for finite measures.** Apply (a) to $f\vee 0$ and to $-(f\wedge 0)$.

Proof of the theorem. Since the theorem is true for finite measure spaces, it is true if S in the theorem is replaced by a product set $A_1\times A_2$ with A_i in $\mathbf{S}_i$, of finite λ_i measure, for $i = 1, 2$. Since S is a countable disjunct union of such product sets, the theorem is true as stated.

Adaptation to complete measures. In the preceding proof, completeness of the measures λ_1 and λ_2, or its absence, is irrelevant, and λ was not completed. Suppose that λ_1^*, λ_2^*, and λ^* are respectively the completions of λ_1, λ_2, and λ. Since (Section **IV**.1) every null set of a completed measure space is a subset of a null set of the original measure space, (10.2) implies that a λ^* measurable set is λ^* null if and only if $A_1(s_1)$ is λ_2^* null, for λ_1^* almost every s_1, and then that, if A if λ^* measurable, $A_1(s_1)$ is λ_2^* measurable for λ_1^* almost every s_1, and (10.1) is true. Since (Section V.2(c)) every λ^* measurable function on S is λ^* almost everywhere equal to a λ measurable function, it follows that if f is a λ^* measurable function from S into $\bar{\mathbf{R}}^+$, then $f(s_1,\bullet)$ is λ_2^* measurable for λ_1^* almost every s_1, that the inner integral in (10.1) defines a λ_1^* measurable function of s_1, and that (10.1) is true. Apply this result to $f\vee 0$ and $-(f\wedge 0)$ to deduce the obvious version of (10.2) for integrable not necessarily positive functions that are measurable with respect to the σ algebra of λ^* measurable sets. This discussion is not much simpler if λ_1 and λ_2 in the theorem are already complete measures, because even then the measure λ in the theorem is not necessarily complete.

Note on the construction of product measures. Product measures were constructed before integration, in Section **IV**.11. Alternatively, the construction can be deferred until integration is available, when, in the notation of Theorem 10, the product measure λ can be defined by (10.2).

Application to the volume under a graph. Let $(S,\mathbb{S},\lambda)$ be a finite measure space and define $S^+ = S\times\mathbf{R}^+$, $\mathbb{S}^+ = \sigma(\mathbb{S}\times\mathbf{B}(\mathbf{R}^+))$. The space $(S^+,\mathbb{S}^+)$ is a measurable space. Let μ be the measure on $\mathbb{S}^+$ that is the product of the measures λ on $\mathbb{S}$ and Lebesgue measure on $\mathbf{R}$. If f is a function from S into $\mathbf{R}^+$ the subsets

$$\{(s,t):t=f(s)\} \qquad \{(s,t): 0\le t\le f(s)\}$$

of S are respectively the *graph* of f and the *ordinate set* $O(f)$ of f. If f is measurable, let $f_\bullet$ be the increasing sequence of step functions, with limit f, defined by V(1.1), and define $g_n=f_n+2^{-n}$. The ordinate sets $O(f_n)$ and $O(g_n)$ are each a union of products of measurable subsets of S and compact intervals of $\mathbf{R}$. These ordinate sets are therefore measurable subsets of S^+. Define $D_n=O(g_n)-O(f_n)$. The set D_n contains the part of the graph of f over the set $\{s: f(s) < 2^n\}$. Moreover $\mu(D_n)\le 2^{-n}\lambda(S)$, and $\lim D_\bullet$ is the graph. Thus the graph is in $\mathbb{S}^+$ and is μ null. The sequence $O(f_\bullet)$ is an increasing sequence of measurable subsets of S^+, with limit $O(f)$ less the graph of f, and therefore the ordinate set of f is in $\mathbb{S}^+$. According to Theorem 10,

$$\mu\big(O(f)\big)=\int_S f\,d\lambda. \tag{10.3}$$

If S is a compact interval of $\mathbf{R}$ and λ is Lebesgue measure on this interval, (10.3) is the usual formula for the area under the graph of f.

The preceding discussion has been under the hypothesis that f is a measurable function, and under this hypothesis, the measurability of $O(f)$ was proved. Conversely, $O(f)$ measurability implies that f is a measurable function, because $O(f)$ measurability implies (Theorem 10) that the integrand in (10.3) is a measurable function on S and that (10.3) is true.

11. Jensen's inequality

Theorem (Jensen). *Let ϕ be a convex function from an interval I of $\mathbf{R}$ into $\mathbf{R}$, and define ϕ at an endpoint of I not in I as the limit of ϕ at the point. Let f be a measurable function from a probability space into I. If f and $\phi(f)$ are integrable, then*

$$\phi[E\{f\}] \le E\{\phi(f)\}. \tag{11.1}$$

If f and $\phi(f)$ are not supposed integrable, but if ϕ and f are lower bounded, (11.1) *remains true.*

Proof. The graph of ϕ has the property that if $(\xi_0,\phi(\xi_0))$ is a point of the graph, no point of the graph lies strictly below a suitably chosen line L through the point. Hence, if the equation for L is $\eta = a(\xi-\xi_0) + \phi(\xi_0)$, it follows that $\phi(f) \ge a(f-\xi_0) + \phi(\xi_0)$. In particular, if f and $\phi(f)$ are integrable, choose $\xi_0 = E\{f\}$, and integrate this inequality to obtain Jensen's inequality. More generally, if f and ϕ are lower bounded, in which case I can be chosen with a finite left-hand endpoint, $E\{f\}$ and $E\{\phi(f)\}$ are well defined, possibly $+\infty$. To prove the inequality in this case, apply the case of Jensen's inequality already proved to $f \wedge c$ and let c tend to $\sup_S f$.

Application to L^p for a probability space. According to Jensen's inequality, if $1 \le p < +\infty$,

$$E^p\{|f|\} \le E\{|f|^p\}, \tag{11.2}$$

which can be rewritten in the form $|f|_1 \le |f|_p$, if $f \in \mathrm{L}^p$. More generally, if $1 \le p < p' < +\infty$, then $E^{p'/p}\{|f|^p\} \le E\{|f|^{p'}\}$, that is, if $f \in \mathrm{L}^{p'}$,

$$|f|_p \le |f|_{p'}. \tag{11.3}$$

Moreover this inequality is true for $p' = +\infty$ in the sense that direct substitution in the integral over the space S for the L^p norm yields the inequality

$|f|_p \le$ ess $\sup_S |f|$, and this essential supremum, if finite, is $|f|_\infty$. Finally, $\lim_{p\to\infty} |f|_p =$ ess $\sup_S |f|$. In fact *this limit relation is true not only for f on a probability space, but for f on an arbitrary finite measure space.* To prove this, trivial if $\lambda(S) = 0$, suppose that $\lambda(S) > 0$, denote the essential supremum by c and observe that if $c' < c$,

$$c'[\lambda\{f \ge c'\}]^{1/p} \le \Big(\int_S |f|^p \, d\lambda\Big)^{1/p} \le c\,[\lambda(S)]^{1/p}, \tag{11.4}$$

and observe that when $p \to +\infty$, the first and third terms have limits c' and c respectively.

12. Conjugate spaces and Hölder's inequality

If $1 < p < +\infty$ and

$$1/p + 1/q = 1, \tag{12.1}$$

then $q > 1$; q is the *index conjugate to p*, and L^q is *conjugate* to L^p. The relation is symmetric; if p is conjugate to q, then q is conjugate to p. The only self-conjugate index is 2.

Theorem (Hölder's inequality). (p, q, *conjugate indices, measure space* $(S, \mathbf{S}, \lambda)$) *If* $f \in L^p$ *and* $g \in L^q$, *then fg is integrable, and*

$$\Big|\int_S fg \, d\lambda\Big| \le |f|_p \, |g|_q. \tag{12.2}$$

Proof. It is sufficient to prove this inequality when f and g are positive, because replacing f and g in (12.2) by their absolute values does not change the right side and can only increase the left side. The inequality is trivial if either norm on the right vanishes, and it can therefore be assumed that each is strictly positive, and even that each is 1, because multiplying the functions by constants does not affect the inequality. Under these hypotheses, (6.1), with f replaced by g^q, defines a probability measure μ, and an application of Jensen's inequality (11.1), with $\phi(s) = s^p$ on $\mathbf{R}^+$, yields

$$\Big(\int_S fg \, d\lambda\Big)^p = \Big(\int_{\{g>0\}} fg^{1-q}(g^q d\lambda)\Big)^p \le \int_{\{g>0\}} f^p g^{p-pq}(g^q \, d\lambda) \le 1, \tag{12.3}$$

as was to be proved.

Hölder's inequality for the case $p = q = 2$. In this case, (12.2) becomes

$$\Big(\int_S fg \, d\lambda\Big)^2 \le \int_S f^2 \, d\lambda \int_S g^2 \, d\lambda, \tag{12.4}$$

which is trivial when g vanishes almost everywhere, and otherwise is most easily proved by defining $c = \lambda[fg]|g|_2^2$ and noting that $(f - cg)^2$ has a positive integral. This inequality is variously named, in honor of Bunyakovsky, Cauchy, and Schwarz who fortunately are not available to express their appreciation at the dubious honor of being so closely identified with this (now) rather simple inequality.

Hölder's inequality in the discrete context. Suppose that p and q are conjugate indices, $a_\bullet$ and $b_\bullet$ are sequences of numbers, and $\alpha_\bullet$ is a sequence of positive numbers. *If the series* $\Sigma\,|a_\bullet|^p\alpha_\bullet$ *and* $\Sigma\,|b_\bullet|^q\alpha_\bullet$ *converge, then* $\Sigma\,|a_\bullet b_\bullet|\alpha_\bullet$ *converges, and*

$$|\Sigma\, a_\bullet b_\bullet \alpha_\bullet| \le \left(\Sigma|a_\bullet|^p\alpha_\bullet\right)^{1/p}\left(\Sigma\,|b_\bullet|^q\alpha_\bullet\right)^{1/q}. \tag{12.2'}$$

This is the special case of Hölder's inequality in which the space S is countable and $\alpha_\bullet$ is the sequence of measures of its singletons.

13. Minkowski's inequality

This inequality makes an L^p pseudometric possible (see Section 14).

Theorem (Minkowski's inequality). (*Measure space* $(S,\mathbf{S},\lambda)$) *If* $p \ge 1$ *and* f *and* g *are in* L^p, *then* $f+g$ *is in* L^p *and*

$$|f+g\,|_p \le |f\,|_p + |g\,|_p. \tag{13.1}$$

Proof. Inequality (13.1) is trivial when p is 1 or $+\infty$; in the following suppose that $1 < p < +\infty$. It was already noted in Section 4 that linear combinations of functions in L^p are in L^p. It is sufficient to prove (13.1) when f and g are positive. Apply Hölder's inequality to derive

$$\begin{aligned} |f+g\,|_p^p &= \int_S (f+g)^{p-1} f\,d\lambda + \int_S (f+g)^{p-1} g\,d\lambda \\ &\le |f+g\,|_p^{p/q}\left[|f|_p + |g\,|_p\right], \end{aligned} \tag{13.2}$$

which yields (13.1).

Minkowski's inequality in the discrete context. Suppose that $p \ge 1$, $a_\bullet$ and $b_\bullet$ are sequences of numbers, and $\alpha_\bullet$ is a sequence of positive numbers. Then

$$\left[\Sigma\,(|a_\bullet+b_\bullet|^p)\alpha_\bullet\right]^{1/p} \le \left[\Sigma\,|a_\bullet|^p\alpha_\bullet\right]^{1/p} + \left[\Sigma\,|b_\bullet|^p\alpha_\bullet\right]^{1/p}. \tag{13.1'}$$

This is the special case of Minkowski's inequality in which the space S is countable and $\alpha_\bullet$ is the sequence of measures of its singletons.

14. The L^p spaces as normed linear spaces ($1 \leq p < +\infty$, measure space $(S,\mathbb{S},\lambda)$)

Define the distance between the functions f and g in L^p as $|f-g|_p$. Minkowski's inequality yields the triangle inequality for this distance function:

$$(14.1) \qquad |f-h|_p \leq |f-g|_p + |g-h|_p.$$

Thus this definition of distance makes L^p a pseudometric space. A function f is at L^p distance 0 from a function g if and only if the two functions are equal almost everywhere.

Observation. If functions equal almost everywhere are identified with each other, the space of equivalence classes obtained in this way can be made into a metric space in the usual way (Section **0**.13). This procedure involves defining summation and other operations on equivalence classes, operations already defined on functions, and it is not clear what is gained thereby. A comon style makes L^p the metric space of equivalence classes but uses the word "function" and the phrase "except for a set of measure 0" anyway, thus attaining the best of all nonlogical worlds.

L^p **convergence**. Convergence of a sequence of functions in the L^p pseudometric, called L^p *convergence*, or *convergence in the mean of order p*, implies convergence in measure, in view of the inequality

$$(14.2) \qquad \int_S |f-g|^p \, d\lambda \geq \varepsilon^p \lambda\{|f-g| \geq \varepsilon\}.$$

The converse is false; a sequence in L^p may converge in measure to a function that is not even integrable.

Observations on L^p convergence. If a sequence $f_\bullet$ converges in the mean of order p to f then the sequence $|f_\bullet|$ also converges in the mean of order p to $|f|$, because $\big||f|-|f_n|\big| \leq |f-f_n|$. If $f_\bullet$ converges to f in the mean of order p, then the sequence $|f|^p$ converges to $|f|^p$ in the mean of order 1. This fact, already noted for $p=1$, is implied for $p > 1$ by the following inequality, obtained by combining the Hölder and Minkowski inequalities with the elementary inequality $|a^p-b^p| \leq p|a-b|\,(a+b)^{p-1}$ for positive numbers a and b:

$$(14.3) \quad \big|\,|f|^p - |f_n|^p\,\big|_1 \leq p|f-f_n|_p \,\big|\,|f|+|f_n|\,\big|_p^{p/q} \leq p|f-f_n|_p \big(|f_n-f|_p + |2f|_p\big)^{p/q}.$$

The following fact will be used in Chapter **XI**. If a sequence $f_\bullet$ converges to f in the mean of order p and a sequence $g_\bullet$ converges to g in the mean of the conjugate order q then the sequence $f_\bullet g_\bullet$ of products converges to fg in the mean of order 1. This convergence assertion follows from the inequalities

(14.4)
$$|fg-f_ng_n|_1 \leq |(f-f_n)g|_1 + |(g-g_n)f_n|_1$$
$$\leq |f-f_n|_p\, |g|_q + |g-g_n|_q\Big(|f_n-f|_p + |f|_p\Big).$$

Completeness of L^p. *The space* L^p *is complete.* In fact if $f_\bullet$ is an L^p Cauchy sequence, (14.2) implies that the sequence is a Cauchy sequence for convergence in measure, and therefore (Theorem V.12) the sequence converges in measure to some function f. Apply Fatou's theorem to derive the inequality

(14.5)
$$|f-f_n|_p \leq \liminf |f_\bullet - f_n|_p.$$

The L^p Cauchy condition implies that the right side of this inequality tends to 0 as $n \to +\infty$, and therefore L^p is complete.

15. Approximation of L^p functions

It is sometimes useful to approximate an L^p, function, in the sense of L^p distance, by functions in various special classes.

Approximation of L^p functions by step functions $(1 \leq p < +\infty$, measure space $(S,\mathbf{S},\lambda)$, L^p distance$)$. *The class of* L^p *step functions is dense in* L^p. In fact, if f is in L^p and is positive, let $f_\bullet$ be the monotone increasing step function sequence with limit f defined by V(1.1). The sequence $|f-f_\bullet|_p$ has limit 0 according to the dominated convergence theorem. If f is not necessarily positive, its positive and negative parts can be approximated separately.

A more delicate approximation result is sometimes needed. Let $\mathbf{S}_0$ be a subalgebra of $\mathbf{S}$, generating $\mathbf{S}$, that is, $\sigma(\mathbf{S}_0) = \mathbf{S}$. It will now be shown that the preceding result remains true when λ is a finite measure and the step functions are to be linear combinations of indicator functions of sets in $\mathbf{S}_0$. All that needs proof is that if $A_1,\dots,A_n$ are sets of finite measure, in $\mathbf{S}$, and if $c_1,\dots,c_n$ are strictly positive constants, then the function

$$f = \sum_1^n c_j \mathbf{1}_{A_j}$$

can be approximated arbitrarily in the L^p distance sense by functions of the same type, except that the sets involved are in $\mathbf{S}_0$. According to Theorem IV.3(b), there is a set A_j' in $\mathbf{S}_0$ with the property that $\lambda(A_j \Delta A_j') < \big(\varepsilon/(c_j n)\big)^p$. Define

$$g = \sum_1^n c_j \mathbf{1}_{A_j'}.$$

Then

(15.1) $$|c_j \mathbf{1}_{A_j} - c_j \mathbf{1}_{A'_j}|_p \leq c_j \lambda^{1/p}(A_j \Delta A'_j) < \varepsilon/n, \quad (j \leq n).$$

Add these inequalities to find that $|f\text{-}g|_p \leq \varepsilon$.

According to this result if $S = \mathbf{R}^N$ and the measure is finite, approximating step functions can be chosen to be linear combinations of indicator functions of right semiclosed intervals, in view of Theorem **IV**.8(a) when $N = 1$ and Section **IV**.10 when $N > 1$.

Separability of $\mathrm{L}^p(S,\mathbf{S},\lambda)$. *If the measure space is* σ *finite,* $1 \leq p < +\infty$, *and the* σ *algebra* $\mathbf{S}$ *is generated up to null sets by a countable subclass of* $\mathbf{S}$, *then* L^p *is complete and separable*. Completeness was shown in Section 14. To prove separability, observe first that it is sufficient to prove separability when λ is finite valued, because the class of L^p functions vanishing off a set of finite measure is dense in L^p. For λ finite valued, the space $(\mathbf{S},d_\lambda)$ is separable according to Theorem **IV**.3(d); equivalently the subspace of L^p consisting of indicator functions of sets is separable. It follows that the subspace of rational valued L^p step functions and therefore the class of all L^p step functions is separable. The latter class is dense in L^p according to the first assertion in this section, and it follows that L^p is separable.

Topology of L^p **for functions on a separable metric space** $(1 \leq p < +\infty)$. It is supposed that the functions are defined on a separable metric space S, the measure space is $(S,\mathbf{B}(S),\lambda)$ and the measure is σ finite.

(a) Since the hypotheses of the preceding paragraphs are satisfied, *the space* L^p *is a complete separable pseudometric space.*

(b) *The class of continuous functions in* L^p *is dense in* L^p. It is sufficient to show that every step function in L^p, or even every indicator function of a measurable set of finite measure, or even every indicator function of a closed set of finite measure is an L^p limit of a sequence of continuous functions in L^p. If A is a closed set of finite measure, it has an open neighborhood G of finite measure according to Theorem **IV**.5. If d is the metric of S, define

(15.2) $$f_n(s) = \Big[\exp[-nd(s,A)] - \exp[-nd(s,S-G]\Big] \vee 0.$$

The sequence $f_\bullet$ is a uniformly bounded sequence of positive continuous functions with limit $\mathbf{1}_A$. Each function vanishes outside G. It follows that $\lim |\mathbf{1}_A - f_\bullet|_p = 0$.

Example. Let S be the circle in $\mathbf{R}^2$ of radius 1 and center the origin, and let λ be the completion of a finite measure on $\mathbf{B}(S)$. It is an important fact that the class of complex valued trigonometric polynomials, that is, the linear combinations with complex coefficients, of the functions of the sequence $f_\bullet$ defined by $f_n(z) = z^n = \exp in\theta$, $0 \le \theta \le 2\pi$, in which n runs through the positive and negative integers, is dense in L^p for $1 \le p < +\infty$. In view of the results in this section, to prove this, it is sufficient to prove that an arbitrary continuous function f on S can be approximated arbitrarily closely in L^p by trigonometric polynomials. A sequence of trigonometric polynomials will be defined that converges to f uniformly, and therefore also in L^p distance. It will be convenient to write f as a function F of the central angle θ, on the interval $[0,2\pi]$. Furthermore it will be convenient to extend F to $\mathbf{R}$, making the function continuous with period 2π. Consider the sequence $g_\bullet$ of trigonometric polynomials, defined by

(15.3) $$g_n(\theta) = \sum_{k=-n}^{n} a_k e^{ik\theta}, \quad a_k = \frac{1}{2\pi}\int_0^{2\pi} F(\alpha)e^{-ik\alpha}\,\lambda(d\alpha) \qquad (n \ge 0),$$

where λ is Lebesgue measure. Fourier series theory (to be discussed in Section VII.7) suggests that under appropriate conditions on F, $g_\bullet$ is a sequence with limit F. In the present context it is simpler to apply *Césaro averages.* Define

(15.4) $$g_n^*(\theta) = [g_0(\theta)+\cdots+g_{n-1}(\theta)]/n = \int_0^{2\pi} K_n(\alpha)F(\theta+\alpha)\lambda(d\alpha),$$

where

(15.5) $$K_n(\alpha) = \frac{1}{2\pi}\,\frac{\sin^2 n\alpha/2}{n\sin^2\alpha/2}.$$

Choosing $F = 1$ makes $g_n^* = 1$, and it follows that

$$\int_0^{2\pi} K_n(\alpha)\,\lambda(d\alpha) = 1.$$

Finally,

(15.6) $$g_n^*(\theta)-F(\theta) = \int_{-\pi}^{\pi} K_n(\alpha)[F(\theta+\alpha)-F(\theta)]\,\lambda(d\alpha)$$

Choose $\varepsilon > 0$, and choose δ so small that F has oscillation at most ε in every interval of length 2δ. It follows, separating out the integration in the interval $(-\delta,\delta)$, that

(15.7) $$|g_n^*(\theta)-F(\theta)| \le \varepsilon + 2\left(\sup|F|\right)\left(\sin^{-2}(\delta/2)\right)/n \le 2\varepsilon$$

for sufficiently large n. It follows that the sequence $g_\bullet^*$ converges uniformly to F, as was to be proved.

16. Uniform integrability

Let $\{f_i,\ i \in I\}$ be a family of integrable functions on a finite measure space $(S,\mathbf{S},\lambda)$. Then (dominated convergence theorem) for each value of i,

$$\lim_{\alpha\to+\infty} \int_{\{|f_i|\geq\alpha\}} |f_i|\,d\lambda = 0. \tag{16.1}$$

Definition. A family of integrable functions on a finite measure space is *uniformly integrable* if (16.1) is true uniformly for i in I.

This definition implies that finitely many integrable functions form a uniformly integrable family and, slightly more generally, that the functions in a finite number of uniformly integrable families are uniformly integrable. The family of linear combinations of members of a uniformly integrable family is uniformly integrable if there is a uniform bound on the coefficient absolute values.

Theorem. *A family $\{f_i,\ i \in I\}$ of integrable functions on a finite measure space $(S,\mathbf{S},\lambda)$ is uniformly integrable if and only if*

(a) *the family is* L^1 *bounded, that is,* $\sup \|f_\bullet\|_1 < +\infty$, *and*

(b) *if* μ_i *is the measure defined by* $\mu_i(A) = \int_A |f_i|\,d\lambda$,
then $\lim_{\lambda(A)\to 0} \mu_i(A) = 0$, *uniformly for i in I.*

The condition (b) is described in Section IX.4 as *uniform absolute continuity* of the family of measures $\mu_\bullet$.

Proof. In the following proof, it will be supposed that the functions are all positive, since only the absolute values of the functions are involved. Observe that

$$\mu_i(A) \leq \int_{A\cap\{f_i<\alpha\}} f_i\,d\lambda + \int_{\{f_i\geq\alpha\}} f_i\,d\lambda \leq \alpha\lambda(A) + \int_{\{f_i\geq\alpha\}} f_i\,d\lambda. \tag{16.2}$$

If the sequence $f_\bullet$ is uniformly integrable, the last integral can be made uniformly small by choosing α large. The inequality with $A = S$ therefore implies the L^1 boundedness of the sequence $f_\bullet$. If $\varepsilon > 0$, the last integral can be made $< \varepsilon$, uniformly for all i, by choosing α sufficiently large, and then the preceding term can be made uniformly small by choosing $\lambda(A)$ small. Thus (b) is satisfied.

Conversely, if conditions (a) and (b) are satisfied, $\|f_i\|_1=\lambda[f_i] \geq \alpha\lambda\{f_i \geq \alpha\}$. Hence $\lambda\{f_i \geq \alpha\}$ is uniformly small when α is large, and, according to (b), this implies that (16.1) is true, uniformly as i varies.

17. Uniform integrability in terms of uniform integrability test functions

Theorem. *Let $\{f_i,\ i \in I\}$ be a family of measurable functions on a finite measure space $(S,\mathbf{S},\lambda)$. If the family is uniformly integrable, there is a convex uniform integrability test function ϕ for which*

$$\sup \int_S \phi(|f_\bullet|)\, d\lambda < +\infty. \tag{17.1}$$

Conversely, if there is a uniformly integrable test function ϕ for which (17.1) *is true, the family is uniformly integrable.*

Proof. To economize on absolute value signs, it will be assumed in the proof (without loss of generality) that the functions are positive. If there is uniform integrability, there is a positive monotone decreasing function ψ_1 on $\mathbf{R}^+$, with limit 0 at $+\infty$, for which

$$\int_{\{f_i \geq \alpha\}} f_i\, d\lambda \leq \psi_1(\alpha) \tag{17.2}$$

for all i. The very special case of this converse in Section 7 was treated in such a way that the discussion is applicable here; the function ψ_1 can be majorized by a continuous monotone function ψ satisfying (17.2), and the desired convex uniform integrability test function is defined in terms of ψ by the procedure of Section 7. Conversely, if there is a uniform integrability test function ϕ satisfying (17.1), then

$$\int_S \phi(f_i)\, d\lambda \geq \frac{\phi(\alpha)}{\alpha} \int_{\{f_i \geq \alpha\}} f_i\, d\lambda. \tag{17.3}$$

When α becomes infinite the left side is bounded uniformly as i varies, the fraction on the right becomes infinite, and therefore there is uniform integrability.

18. L^1 convergence and uniform integrability

Theorem. *Let $f_\bullet$ be a sequence of integrable functions on a finite measure space $(S,\mathbf{S},\lambda)$, converging in measure to a function f. Then the sequence is uniformly integrable if and only if there is L^1 convergence.*

Proof. If there is uniform integrability, the sequence $f_\bullet$ is L^1 bounded, and therefore (Fatou's theorem) f is integrable. Then the sequence $f_\bullet - f$ is uniformly integrable. Given $\varepsilon > 0$,

$$(18.1)\qquad \int_{\{|f-f_n|<\varepsilon\}} |f-f_n|\, d\lambda \le \varepsilon\lambda(S).$$

On the other hand, $\lim \lambda\{|f-f_\bullet| \ge \varepsilon\}=0$ because there is convergence in measure, and therefore (Theorem 16)

$$(18.2)\qquad \lim_{n\to\infty} \int_{\{|f-f_n|\ge\varepsilon\}} |f-f_n|\, d\lambda = 0.$$

Relations (18.1) and (18.2) combine to yield L^1 convergence. Conversely, if there is L^1 convergence, uniform integrability of $f_\bullet$ will follow from that of $g_\bullet = |f-f_\bullet|$. The L^1 convergence implies L^1 boundedness of the sequence $g_\bullet$. There remains the proof that if $\lambda(A)$ is small the integral $\lambda[g_n 1_A]$ is uniformly small as n varies. Choose $\varepsilon > 0$, and choose k so large that $\lambda[g_n] < \varepsilon$ when $n > k$. Then, if $\lambda(A)$ is so small that $\max_{n\le k} \lambda[g_n 1_A] < \varepsilon$, it follows that $\lambda[g_n 1_A] < \varepsilon$ for all n.

Remark on integrating term by term to the limit for a convergent sequence of measurable functions. Aside from the rather specialized Beppo-Levi theorem, there are only two general criteria for integrating a convergent function sequence term by term to the limit: Lebesgue dominated convergence and uniform integrability. If the measure space is a finite measure space, dominated convergence implies uniform integrability. For such a measure space the uniform integrability criterion is more general, but the dominated convergence criterion is more natural in many contexts. In other contexts, for example, in martingale theory (Sections **XI**.14-16), uniform integrability is more natural – and even necessary – because the dominated convergence criterion is insufficiently general.

19. The coordinate space context

Let f be a real valued random variable on a probability space $(S,\mathbf{S},P)$, and define F: $F(\alpha) = P\{f \le \alpha\}$, the distribution function of f. One of the special aspects of probability theory is that many questions about random variables are formulated in terms of distributions and do not otherwise involve the probability space on which the random variables are defined. In the present context, the distribution function F determines the Lebesgue-Stieltjes probability measure λ_F on $\mathbf{R}$, and the coordinate variable x on the probability space $(\mathbf{R},\mathbf{B}(\mathbf{R}),\lambda_F)$ then has the distribution function F. In any question involving the distribution function of f or of a Borel measurable function of f, the probability space $(\mathbf{R},\mathbf{B}(\mathbf{R}),\lambda_F)$ can serve as well as the original probability space. For example, if ϕ is a Borel measurable function from $\mathbf{R}$ into $\mathbf{R}$, then $\phi(f)$ is integrable on $(S,\mathbf{S},P)$ if and only if $\phi(x)$ is integrable on $(\mathbf{R},\mathbf{B}(\mathbf{R}),\lambda_F)$, and

$$(19.1)\qquad E\{\phi(f)\} = \int_{\mathbf{R}} \phi\, d\lambda_F = \int_{-\infty}^{+\infty} \phi(\alpha)\, dF(\alpha).$$

The expectation is an integral over the space on which f is defined, the second expression is an integral over $\mathbf{R}$, and the third is an alternative form of the second, but can also be interpreted in terms of Riemann-Stieltjes integration if ϕ is bounded and λ_F almost everywhere continuous (see Section 20). The point is that the measure relations involved in defining these integrals are identical. In particular, if f is integrable,

$$(19.2)\qquad E\{f\} = \int_{\mathbf{R}} \alpha\lambda_F(d\alpha) = \int_{-\infty}^{+\infty} \alpha dF(\alpha).$$

Application to independent random variables. If $(S_1,\mathbf{S}_1,\lambda_1)$ and $(S_2,\mathbf{S}_2,\lambda_2)$ are σ finite measure spaces, and λ is the product measure $\lambda_1\times\lambda_2$, let f_1 be an integrable function on the first space and f_2 on the second. Then (Theorem 10) f_1f_2 is integrable on the product space, and

$$(19.3)\qquad \int_{S_1\times S_2} f_1f_2\, d\lambda = \int_{S_1} f_1\, d\lambda_1 \int_{S_2} f_2\, d\lambda_2.$$

Conversely if f_1f_2 is λ integrable and neither factor vanishes almost everywhere on its space, then each factor is integrable on its space. The probability version of this is phrased somewhat differently and it is not quite obvious that the probability version is a special case. The probability version deals with two mutually independent real valued random variables f_1,f_2 on a probability space. The standard theorem is that if they are integrable, then f_1f_2 is integrable, and $E\{f_1f_2\} = E\{f_1\}E\{f_2\}$. Conversely, if neither random variable vanishes almost surely, their individual integrability is implied by that of their product. To put this probabilistic context into the product space context, all that has to be done is to observe that the joint distribution of the two random variables is the product measure of their separate distributions, so that f_1 and f_2 can be replaced by the coordinate functions of $\mathbf{R}^2$ on which a product measure is defined, the product of the distributions of f_1 and f_2.

More generally, if f_1 and f_2 are random variables on some probability space (with measure denoted by P) and have the joint distribution function F, that is,

$$F(\alpha_1,\alpha_2) = P\{f_1\le\alpha_1, f_2\le\alpha_2\},$$

then the Lebesgue-Stieltjes measure λ_F is a probability measure on $\mathbf{R}^2$, and, if ϕ is a Borel measurable function from $\mathbf{R}^2$ into $\mathbf{R}$, for which $\phi(f_1,f_2)$ is integrable,

$$(19.4)\qquad E\{\phi(f_1,f_2)\} = \int_{\mathbf{R}^2} \phi(x_1,x_2)\, d\lambda_F = \int_{\mathbf{R}^2} \phi(\alpha_1,\alpha_2)\, dF(\alpha_1,\alpha_2).$$

Here x_1 and x_2 are the coordinate variables of $(\mathbf{R}^2, \mathbf{B}(\mathbf{R}^2), \lambda_F)$. If f_1 and f_2 are mutually independent, and if ϕ is the indicator function of the set $\{(\alpha_1, \alpha_2) \in \mathbf{R}^2: \alpha_1 + \alpha_2 \leq \alpha\}$ then (19.4) yields the distribution function G of the random variable $f_1 + f_2$, defined on S, the same distribution function as that of $x_1 + x_2$, defined on $\mathbf{R}^2$, in terms of the distribution functions F_1 of f_1 and F_2 of f_2. Theorem 10 yields G as the evaluation of an iterated integral:

$$(19.5) \qquad G(\alpha) = \int_{\mathbf{R}} F_1(\alpha - \beta)\, dF_2(\beta).$$

Thus the distribution function of the sum of two mutually independent random variables is the *convolution* of the distribution functions of the summands.

Coordinate functions can be used to replace any finite number of random variables for many purposes. The corresponding replacement for infinitely many random variables is provided by a probability measure on infinite dimensional coordinate space (see Kolmogorov's theorem V.6).

20. The Riemann integral

In this section, the concept of Riemann integration is placed in a general context. The application to the simplest context, in which the domain of definition of the integrand is a compact interval of $\mathbf{R}$, is made at the end of the discussion. Let S be a compact metric space, and λ be the completion of a finite measure on $\mathbf{B}(S)$. Throughout this section, all measure concepts refer to λ, unless otherwise specified.

Each point of S has arbitrarily small neighborhoods with null boundaries, because the boundary of a ball centered at the point is null, except perhaps for at most countably many values of the radius. The space S can be expressed in many ways as a finite disjunct union of sets with null boundaries. For example, S is the union of finitely many balls with null boundaries, and the union can be made into a disjunct union of subsets of the balls. A *partition of S* is defined as the cells (summands) of a finite disjunct sequence of subsets of S with null boundaries and union S. A cell of such a partition is the union of its open interior and part of its null boundary and is therefore measurable. If π and π' are partitions, and if each set of the π' partition is a subset of a set in the π partition, π' is a *refinement* of π. If π and π' are partitions, they can be combined into a refinement of both in which the cells are the intersections of the cells of the two partitions.

Let f be a function from S into $\mathbf{R}$. When S is a compact interval of $\mathbf{R}$, the Riemann integration procedure, which defines an integral with respect to f under rather strong restrictions on f, with λ on intervals defined as interval length, is what is given in the traditional introduction to integration on compact intervals of $\mathbf{R}$. The principles involved may perhaps be more readily understood in the present more general context. If S and λ are as described in the preceding paragraph, suppose that the *Darboux sums*

(20.1) $$\Sigma_n (\sup \overline{S_n} f)\lambda(S_n) \quad \text{and} \quad \Sigma_n (\inf \overline{S_n} f)\, \lambda(S_n),$$

corresponding to a partition with cells $S_\bullet$ having closures $\overline{S_\bullet}$, are well defined and finite. If so, the union of those cell closures $\overline{S_n}$ that are not λ null must contain the compact support of λ. The first sum, the *upper Darboux sum* for the partition, is at least as large as the second sum, the *lower Darboux sum*. These sums are finite if and only if f is bounded on each nonnull partition cell closure, and if so, f is bounded on the support of λ. If this condition is satisfied for one partition, it is satisfied for all the refinements of that partition. Moreover, considering only finite sums, in going from a partition to a refinement, an upper sum can only change by decreasing, because each partition cell is replaced by one or more partition refinement cells whose sum is at most the original partition cell. Similarly, in going from a partition to a refinement, a lower sum can only change by increasing. The upper sum for a partition π is at least as large as the lower sum for a partition π', because the upper sum for π is at least as large as the upper sum for the combined partition, which in turn is at least as large as the lower sum for the combined partition, which in turn is at least as large as the lower sum for π'. The function f is *Riemann integrable* on S for a given measure λ, if the Darboux upper and lower sums are finite for some partition, and if the infimum of the class of upper sums is equal to the supremum of the class of lower sums. The value assigned to the Riemann integral is then this common extreme value, which is necessarily finite. In the following, the integral defined in the preceding chapters is identified as the "Lebesgue integral", to distinguish it from the Riemann integral.

Theorem (*S compact metric, λ the completion of a finite measure of Borel sets*). *A function from S into* $\mathbf{R}$ *is Riemann integrable for λ if and only if the function is bounded on the compact support of λ and is continuous at λ almost every point of this support. If Riemann integrable, the function is measurable and Lebesgue integrable, with Lebesgue integral equal to the Riemann integral.*

Proof. It will be convenient to evaluate the Darboux sums as L integrals. Define $\bar{f}_\pi$ and $\underline{f}_\pi$ on each cell of a partition π as the supremum and infimum of f on the closure of the cell. Under these definitions the upper and lower sums, supposed finite, are equal respectively to the Lebesgue integrals on S of $\bar{f}_\pi$ and $\underline{f}_\pi$. If π' is a refinement of π, $\underline{f}_\pi \le \underline{f}_{\pi'} \le f \le \bar{f}_{\pi'} \le \bar{f}_\pi$. Suppose that there is at least one partition with well-defined and finite upper and lower sums. Let C_2 be the infimum of the upper sums, the limit of a sequence of upper sums for a sequence $\pi_\bullet$ of partitions. Replace each partition π_n by the combination of $\pi_1,\dots,\pi_n$, if necessary, to ensure that π_{n+1} is a refinement of π_n. Then $\bar{f}_{\pi_\bullet}$ is a decreasing sequence of measurable functions. If $\bar{f}$ is the sequence limit, then (bounded convergence theorem)

$$C_2 = \lim \int_S \bar{f}_{\pi_\bullet}\, d\lambda = \int_S \bar{f}\, d\lambda. \tag{20.2}$$

Since f_{π_n} is continuous at each point of the interior of each cell of π_n and therefore continuous at almost every point of S, the function $\bar{f}$ is upper semicontinuous at almost every point of S. Similarly, the supremum C_1 of the lower sums is the integral of the limit $\underline{f}$ of an increasing sequence of measurable almost everywhere continuous functions, and $\underline{f}$ is lower semicontinuous at almost every point of S. Moreover $\underline{f} \leq f \leq \bar{f}$, and

$$C_1 = \int_S \underline{f}\, d\lambda \leq \int_S \bar{f}\, d\lambda = C_2. \tag{20.3}$$

There is Riemann integrability, that is, $C_1 = C_2$, if and only if $\underline{f}=f=\bar{f}$ almost everywhere on S. In this case, f is measurable, continuous at almost every point of S, bounded on the compact support of λ, and its Riemann integral is equal to its Lebesgue integral.

Conversely, suppose that f is bounded on the compact support of λ and is continuous at λ almost every point of this compact support, that is, at almost every point of S. Let A be a compact non λ null subset of S with the property that f is continuous at every point of A. The function f is uniformly continuous on A in the sense that if $\eta > 0$, there is a strictly positive δ so small that if s is a point of A and s' is a point of S within distance δ of s, then $|f(s)-f(s')| < \eta$. (Note that this statement is stronger than the statement that the restriction to A of f is uniformly continuous on A, but that the proof of this stronger statement differs only trivially from that of the weaker one.) Therefore if $\varepsilon > 0$ there is a partition of S with the property that the oscillation of f on each cell containing points of A is at most ε. It follows that there are partition sequences for which $\bar{f} - \underline{f} \leq \varepsilon$ on A. Let $A_\bullet$ be an increasing sequence of compact subsets of the set of points of continuity of λ with the property that $\lim \lambda(A_\bullet)$ is the measure of the compact support of λ. There is a sequence of partitions, each a refinement of its predecessor, for which the nth partition cells containing points of A_n are so small that the oscillation of f on them is at most $1/n$, and that the upper and lower Darboux sums for the nth partition have as respective limits the extreme values C_2 and C_1 of such sums. The corresponding limit functions $\underline{f}$ and $\bar{f}$ are then equal almost everywhere on the support of λ and therefore f is Riemann integrable.

The case when S is a compact interval of R. In this case, the proof that the stated condition in Theorem 20 is sufficient shows that partitions whose cells are intervals suffice in the analysis, that is, yield the common extreme limiting value to the upper and lower Darboux sums. The Riemann integral in this context is also called the *Riemann-Stieltjes* integral when the measure λ is not Lebesgue measure, that is, when the measure of an interval is not the positive difference between endpoint coordinates. When Riemann integrals are used over the whole line **R**, they are *improper* Riemann integrals, whose values are defined as the limit of values of the integrals over an increasing sequence of compact intervals with union **R**.

21. Measure theory vs. premeasure theory analysis

Before the advent of measure theory, classical analysis dealt for the most part with smooth functions, considering others as pathological, useful – if at all – only as a source of counterexamples designed to show the scope of definitions and theorems. If a sequence of functions converged, but not uniformly, the convergence was difficult to exploit, for example, in term-by-term integration to the limit, unless the domain of the functions could be divided into finitely many subdomains of uniform convergence, and small remainder sets. In many contexts, measure theory widened the class of admissible domains and functions to the classes of measurable sets and measurable functions, and in so doing made it possible to apply the usual limiting procedures without leaving admissible classes. What was unexpected was that, in a reasonable sense, most of the old concepts were very nearly still present. Egoroff's theorem showed that uniform convergence was nearly present whenever there was convergence. Lusin's theorem showed that the new measurable functions were nearly continuous. On the other hand, measure theory could be applied in abstract contexts where topology was inappropriate. In fact probabilists had been doing this for centuries, except that they called measure "probability" and integrals "expectations", and of course they lacked the refinements and rigor of modern analysis and were scorned because no one was sure what mathematical probability was, and it certainly was not part of respectable analysis.

At the present time, one of the more profound differences between probability and other aspects of measure theory is that if a mathematician explicitly evaluates an integral of a function defined on a space other than a subset of $\mathbf{R}^N$, the mathematician is almost surely a probabilist, even more surely if "a.s." rather than "a.e." appears in the analysis. But probabilists, like other mathematicians, are not above the use of classical methods, based essentially on antiderivatives, to evaluate integrals of functions on $\mathbf{R}^N$.

VII
Hilbert Space

1. Analysis of L^2

The L^p space of functions on a measure space $(S,\mathbf{S},\lambda)$ was discussed in Section VI.14. The special case $p = 2$ plays a strong role in analysis and will now be discussed in more detail in an abstract form. It is useful to allow complex valued measurable functions in this study. Complex conjugates are indicated by upper bars. Recall that a complex valued function f is measurable if and only if its real part $\Re f$ and imaginary part $\Im f$ are.

The following are basic properties of L^2, some obtained by setting $p = 2$ in Section VI.14.

(a) The space is a linear space, that is, linear combinations (with complex coefficients) of members of the space are in the space.

(b) Each pair f, g of members has an *inner product*

$$(f,g) = \int_S f\bar{g}\, d\lambda \tag{1.1}$$

that satisfies the following conditions.

(i) Hermitian symmetry: $(f,g) = \overline{(g,f)}$.

(ii) Linearity in the first argument: for complex constants a_1, a_2:

$$(a_1f_1+a_2f_2,g) = a_1(f_1,g)+a_2(f_2,g).$$

(iii) $(f,f) \geq 0$, with equality if and only if $f = 0$ a.e.

The number $|f| = (f,f)^{1/2}$ is the *norm* of f. (The subscript 2 will be omitted from the norm notation throughout this discussion of L^2.) A glance at the proof in Section VI.12 of the Hölder inequality for the case $p = 2$ shows that the properties under (b) imply the Schwarz et al. inequality

$$|(f,g)| \leq |f|\,|g| \tag{1.2}$$

without the necessity of referring back to the definition of the inner product as an integral. More explicitly,

$$(1.3)\qquad |g|^2\left[|f|^2|g|^2-|(f,g)|^2\right]=\left\|\,|g|^2f-(f,g)g\,\right\|^2\geq 0.$$

Inequality (1.2) implies Minkowski's inequality in the present context:

$$(1.4)\qquad |f+g|\leq|f|+|g|,$$

and this in turn yields the triangle inequality for L^2,

$$(1.5)\qquad |f-h|\leq|f-g|+|g-h|.$$

(c) The pseudometric space L^2 is complete. It is convenient to write in a special way that a sequence $f_\bullet$ of functions converges to f in the L^2 pseudometric, that is, that there is a limit in the mean of order 2:

$$(1.6)\qquad \text{l.i.m.} f_\bullet=f;$$

"of order 2" will be omitted below.

2. Hilbert space

The properties listed in Section 1 suggest a postulate system that makes it possible to ignore the measure background. That is, instead of functions $f, g, \ldots$ postulate that there is an abstract space $\mathfrak{H}$, *Hilbert space*, consisting of points, written in boldface: **f**, **g**,... with the properties listed in Section 1(a), 1(b)(i) and (ii). The inner product and norm notation used for L^2 will also be used for $\mathfrak{H}$. Properties (a), (b) (i), and (ii), with the points of $\mathfrak{H}$ in boldface instead of L^2 functions in italics, need no changes in wording. The product of the complex constant 0 and a member of $\mathfrak{H}$ is a member **0** of $\mathfrak{H}$, the identity element of the space considered as a group with group operation addition. Property (b)(iii) can now be restated without reference to a measure space: $|\mathbf{f}|=0$ if and only if $\mathbf{f}=\mathbf{0}$, and (c) becomes: $\mathfrak{H}$ is a complete metric space under the definition that the distance between **f** and **g** is $|\mathbf{f}-\mathbf{g}|$. One way of looking at this discussion is that as so defined $\mathfrak{H}$ satisfies properties (a), (b), and (c) if these properties are restated in the language of equivalence classes, in which functions are identified if they are equal almost everywhere.

The axioms given above are for complex Hilbert space. Real Hilbert space has the same axioms except that the inner product is real valued, and in (b)(ii) only real coefficients a_1 and a_2 are allowed. These axioms correspond to the properties of real valued functions (or better, equivalence classes of real valued functions) in an L^2 space. In this chapter the Hilbert space will always be complex unless the contrary is stated.

The preceding paragraphs outline the axioms of Hilbert space. Readers who prefer measure theory to axiomatics can think of the space as an L^2 space partitioned into equivalence classes, but they may lose the flavor of the geometry of the space. To stress this flavor, points of $\mathfrak{H}$ will be called "vectors". The intuitive simplified picture is that these are analogs of vectors from the origin of a finite dimensional Euclidean space; the Euclidean distance between the endpoints of two such vectors corresponds to the distance between two Hilbert space vectors. In fact the Euclidean space picture is a special (real) Hilbert space according to an example below.

Continuity of the inner product. The inequality

$$(2.1)\qquad \left|(\mathbf{f},\mathbf{g}) - (\mathbf{f}_1,\mathbf{g}_1)\right| = \left|(\mathbf{f}-\mathbf{f}_1,\mathbf{g}_1) + (\mathbf{f}-\mathbf{f}_1,\mathbf{g}-\mathbf{g}_1) + (\mathbf{f}_1,\mathbf{g}-\mathbf{g}_1)\right|$$
$$\le |\mathbf{f}-\mathbf{f}_1|\,|\mathbf{g}_1| + |\mathbf{f}-\mathbf{f}_1|\,|\mathbf{g}-\mathbf{g}_1| + |\mathbf{f}_1|\,|\mathbf{g}-\mathbf{g}_1|$$

implies continuity of the inner product at the pair $\mathbf{f}_1, \mathbf{g}_1$. It follows that the norm is a continuous function of its argument.

Dimensionality. The *dimensionality* of a Hilbert space is *infinite* if there are arbitrarily large finite sets of linearly independent vectors; otherwise it is the largest cardinality of a set of linearly independent vectors.

Hilbert space isomorphisms and unitary operators. Two Hilbert spaces are *isomorphic* if there is a 1-1 linear correspondence between them preserving inner products. A *unitary transformation* on a Hilbert space is a linear transformation $\mathbf{T}$ of the space onto itself which preserves inner products. Observe that preservation of norms implies preservation of the inner product. In fact if $\mathbf{T}$ is a norm preserving linear transformation of a Hilbert space onto itself, or even only into itself, then the two equations

$$|\mathbf{f}+\mathbf{g}| = |\mathbf{T}(\mathbf{f}+\mathbf{g})|, \quad |\mathbf{f}+i\mathbf{g}| = |\mathbf{T}(\mathbf{f}+i\mathbf{g})|$$

imply, when the equalities are squared and written out in detail, that $(\mathbf{f},\mathbf{g}) = (\mathbf{Tf},\mathbf{Tg})$.

Example. The axioms of Hilbert space are modeled on the properties of complex function L^2 space, which in turn is an obvious example if the functions are replaced by equivalence classes, identifying two functions if they are equal almost everywhere. In particular, if the measure space is the space of strictly positive integers (or of those $\le N$), if all subsets of the space are measurable, and if each singleton has measure 1, L^2 becomes the space of infinite (or length N) sequences $a_\bullet$ of complex numbers, with

$$(2.2)\qquad \|a_\bullet\| = \left(\Sigma\, |a_\bullet|^2\right)^{1/2} < +\infty, \quad \left(a_\bullet, b_\bullet\right) = \Sigma\, a_\bullet \bar{b}_\bullet.$$

This special Hilbert space is denoted by l^2. In this special L^2 context, null sets are empty, and therefore "almost everywhere" never appears.

Orthogonality. Vectors **f** and **g** in $\mathfrak{H}$ are *orthogonal to each other* if $(\mathbf{f},\mathbf{g})=0$. The vector **f** is *orthogonal to a set of vectors* if it is orthogonal to each. If **f** is orthogonal to **g**, then $|\mathbf{f}+\mathbf{g}|^2 = |\mathbf{f}|^2 + |\mathbf{g}|^2$, the Hilbert space version of the Pythagorean theorem. Without orthogonality,

$$|\mathbf{f}+\mathbf{g}|^2 = |\mathbf{f}|^2 + |\mathbf{g}|^2 + 2\Re(\mathbf{f},\mathbf{g}). \tag{2.3}$$

A set $\mathbf{f}_\bullet$ of vectors is an *orthonormal set* if the vectors are pairwise orthogonal and each has norm 1. An orthonormal set is *complete* if there is no vector other than **0** orthogonal to every vector in the set. If the set is not complete there is a vector of norm 1 orthogonal to every vector in the set. In l^2, the sequence $(1,0,0,\ldots)$, $(0,1,0,\ldots)$,... is an orthonormal sequence, which is complete because orthogonality of $(a_1,a_2,\ldots)$ to the nth member of this sequence implies that $a_n = 0$.

Subspaces. A subset of $\mathfrak{H}$ is a *subspace* (or a *closed linear manifold*) if it is a closed set that is linear, that is, that contains the linear combinations of its members. Observe that the closure of a linear subset of $\mathfrak{H}$ is also a linear set and therefore is a subspace. The vectors orthogonal to a given set of vectors form a subspace. In particular the set of vectors orthogonal to a subspace $\mathfrak{M}$ is the *orthogonal subspace*, denoted by $\mathfrak{M}^\perp$. The geometric notion of a subspace is a hyperplane through the origin.

3. The distance from a subspace

If A is a subset of a Hilbert space, let $d(\mathbf{f},A)$ be the distance from a vector **f** to A.

Lemma. *If $\mathfrak{M}$ is a subspace of a Hilbert space, there is a unique vector in $\mathfrak{M}$ at distance $d(\mathbf{f},\mathfrak{M})$ from* **f**.

Proof. If **f** is in $\mathfrak{M}$, $d(\mathbf{f},\mathfrak{M}) = 0$ and the lemma is trivial. If **f** is not in $\mathfrak{M}$, let $\mathbf{g}_\bullet$ be a sequence of points of $\mathfrak{M}$ for which $\lim |\mathbf{f}-\mathbf{g}_\bullet| = d(\mathbf{f},\mathfrak{M})$. It will be shown that the sequence $\mathbf{g}_\bullet$ is a Cauchy sequence, converging to the point of $\mathfrak{M}$ closest to **f**. Note the equation

$$|\mathbf{g}_m-\mathbf{g}_n|^2 = |\mathbf{f}-\mathbf{g}_m|^2 + |\mathbf{f}-\mathbf{g}_n|^2 - 2\Re(\mathbf{f}-\mathbf{g}_m,\mathbf{f}-\mathbf{g}_n). \tag{3.1}$$

The inequality

$$|\mathbf{f}-(\mathbf{g}_m+\mathbf{g}_n)/2| \le |\mathbf{f}-\mathbf{g}_m|/2 + |\mathbf{f}-\mathbf{g}_n|/2 \tag{3.2}$$

implies that the left side of (3.2) tends to $d(\mathbf{f},\mathfrak{M})$ when m and n become infinite. The limit equation

$$(3.3)\qquad d^2(\mathbf{f},\mathfrak{M}) = \lim_{m,n\to\infty} |\mathbf{f}-(\mathbf{g}_m+\mathbf{g}_n)/2|^2$$

$$= \lim_{m,n\to\infty} \Big[|\mathbf{f}-\mathbf{g}_m|^2+|\mathbf{f}-\mathbf{g}_n|^2+2\Re\big(\mathbf{f}-\mathbf{g}_m,\mathbf{f}-\mathbf{g}_n\big)\Big]/4$$

implies

$$(3.4)\qquad \lim_{m,n\to\infty} \Re\big(\mathbf{f}-\mathbf{g}_m,\mathbf{f}-\mathbf{g}_n\big) = d^2(\mathbf{f},\mathfrak{M}).$$

This limit relation combines with (3.1) to yield the fact that $\mathbf{g}_\bullet$ is a Cauchy sequence: $\lim_{m,n\to\infty} |\mathbf{g}_m-\mathbf{g}_n|= 0$. Thus the sequence $\mathbf{g}_\bullet$ has limit $\mathbf{g}$, at distance $d(\mathbf{f},\mathfrak{M})$ from $\mathbf{f}$, by continuity of the distance function. There is only one vector of $\mathfrak{M}$ at distance $d(\mathbf{f},\mathfrak{M})$ from $\mathbf{f}$, because if $\mathbf{g}'$ and $\mathbf{g}''$ are vectors in $\mathfrak{M}$ at distance $d(\mathbf{f},\mathfrak{M})$ from $\mathbf{f}$ then $\mathbf{g}'=\mathbf{g}''$ because what was just proved implies that the sequence $\mathbf{g}',\mathbf{g}'',\mathbf{g}',\ldots$, in which the two vectors alternate, is a Cauchy sequence.

4. Projections

The *projection* of a Hilbert space $\mathfrak{H}$ onto a subspace $\mathfrak{M}$ is the transformation taking each point $\mathbf{f}$ of $\mathfrak{H}$ into the closest point to $\mathbf{f}$ on $\mathfrak{M}$.

Theorem. *The projection* T *of* $\mathfrak{H}$ *onto* $\mathfrak{M}$ *has the following properties.*

(a) **T** *is idempotent*: $\mathbf{T}^2 = \mathbf{T}$.

(b) $\mathbf{Tf}=\mathbf{f}$ *for* $\mathbf{f}$ *in* $\mathfrak{M}$; $\mathbf{Tf}=0$ *for* $\mathbf{f}$ *in* $\mathfrak{M}^\perp$.

(c) **T** *is Hermitian symmetric*: $(\mathbf{Tf},\mathbf{g}) = (\mathbf{f},\mathbf{Tg})$.

(d) *For every vector* $\mathbf{f}$, $\mathbf{f}-\mathbf{Tf}$ is orthogonal to $\mathfrak{M}$. *Thus the equation* $\mathbf{f} = \mathbf{Tf} + (\mathbf{f}-\mathbf{Tf})$ *is a representation of* $\mathbf{f}$ *as the sum of a vector in* $\mathfrak{M}$ *and one in* $\mathfrak{M}^\perp$. *Such a representation is unique, and the second summand is the projection of* $\mathbf{f}$ *on* $\mathfrak{M}$. *Moreover* $(\mathfrak{M}^\perp)^\perp = \mathfrak{M}$.

(e) **T** *is linear*: $\mathbf{T}(a\mathbf{f}+b\mathbf{g}) = a\mathbf{Tf}+b\mathbf{Tg}$, *for all complex constants* a,b, *and all vectors* $\mathbf{f}$, $\mathbf{g}$.

Conversely, if **T** *is a transformation from* $\mathfrak{H}$ *into a subspace* $\mathfrak{M}$, *and if* $\mathbf{f}-\mathbf{Tf}$ *is orthogonal to* $\mathfrak{M}$ *for all* $\mathbf{f}$, *then* **T** *is the projection on the subspace.*

Proof of (a). This property follows trivially from the projection definition.

Proof of (b), (c), and (d). If $\mathbf{g}$ is a point of $\mathfrak{M}$, the inequality

$$(4.1)\qquad |f-Tf|^2 \le |f-(Tf+g)|^2 = |f-Tf|^2 + |g|^2 - 2\Re(f-Tf,g)$$

implies that

$$(4.2)\qquad |g|^2 \ge 2\Re(f-Tf,g).$$

The vector **g** can be replaced here by c**g**, with an arbitrary complex constant c, and then if c tends to 0, (4.2) yields $(f-Tf,g) = 0$, that is, **f**−**Tf** is orthogonal to $\mathfrak{M}$.

Thus $f = Tf + (f-Tf)$ is a representation of **f** as a sum $f = f_1+f_2$ of a vector in $\mathfrak{M}$ and one in $\mathfrak{M}^\perp$. If $f = f_1'+f_2'$ is a second such representation, $f_1-f_1' = f_2'-f_2$ is a vector in both $\mathfrak{M}$ and $\mathfrak{M}^\perp$ and is therefore **0**. Thus this representation of **f** as a sum of orthogonal components is unique. Since the projection of $\mathfrak{H}$ on $\mathfrak{M}^\perp$ provides another such representation, it follows that **f**−**Tf** is the projection of $\mathfrak{H}$ on $\mathfrak{M}^\perp$ and that $\mathfrak{M} = (\mathfrak{M}^\perp)^\perp$. Direct computation yields, in terms of the representations of **f** and **g** as sums of components in $\mathfrak{M}$ and $\mathfrak{M}^\perp$,

$$(4.3)\qquad f = f_1+f_2,\quad g = g_1+g_2,\quad (Tf,g) = (f_1,g_1) = (f,Tg).$$

Proof of (e). In the notation of (4.3), $af+bg = (af_1+bg_1) + (af_2+bf_2)$ represents the left side as the sum of vectors in $\mathfrak{M}$ and $\mathfrak{M}^\perp$ from which the linearity property (e) can be read off.

Proof of the converse. Under the hypotheses of the converse, $f = Tf + (f-Tf)$ expresses **f** as the sum of vectors in $\mathfrak{M}$ and $\mathfrak{M}^\perp$, and **T** is therefore the projection on $\mathfrak{M}$.

5. Bounded linear functionals on $\mathfrak{H}$

A *bounded linear functional* on $\mathfrak{H}$ is a function $\mathbb{L}$ from $\mathfrak{H}$ to the complex plane which is linear and bounded, that is, for all complex constants a, b, and all vectors **f**, **g**

$$(5.1)\qquad \mathbb{L}(af+bg) = a\mathbb{L}(f) + b\mathbb{L}(g),\quad |\mathbb{L}(f)| \le c\,|f|,$$

for some positive constant c. The minimum value of c satisfying (5.1) is the *norm* of $\mathbb{L}$. Observe that $\mathbb{L}$ is continuous, because $|\mathbb{L}(f-g)| \le c|f-g|$.

Example. If **g** is a vector, the function $f \to (f,g)$ is a bounded linear functional. In fact, linearity of the functional is an inner product axiom, and the boundedness follows from the inequality $|(f,g)| \le |g||f|$, according to which the norm of this linear functional is at most $|g|$. The norm of the functional is actually $|g|$ because there is equality in this inequality when $f = g$. Different choices of $\mathbf{g}$ yield different linear functionals, because $(\mathbf{f},\mathbf{g}_1) = (\mathbf{f},\mathbf{g}_2)$ for all $\mathbf{f}$ implies that $\mathbf{g}_1-\mathbf{g}_2$ is orthogonal to every vector, in particular is orthogonal to itself, and is therefore $\mathbf{0}$. According to the following theorem, every bounded linear functional has the form of this example. Thus, for L^2 on a measure space, this theorem sets up a linear, norm and order preserving correspondence between bounded linear functionals on L^2 and the members of the dual space, which is also L^2.

Theorem. *If* $\mathbb{L}$ *is a bounded linear functional, there is a vector* $\mathbf{h}$ *for which* $\mathbb{L}(\bullet) = (\bullet,\mathbf{h})$.

Proof. The set of vectors on which $\mathbb{L}$ vanishes is a subspace $\mathfrak{M}$ of $\mathfrak{H}$. If the theorem is true, $\mathbf{h}$ must be orthogonal to $\mathfrak{M}$. Let $\mathbf{T}$ be the projection of $\mathfrak{H}$ on $\mathfrak{M}^\perp$. Then for every vector $\mathbf{f}$, $\mathbf{f}-\mathbf{Tf}$ is in $\mathfrak{M}$, $\mathbb{L}(\mathbf{f}) = \mathbb{L}(\mathbf{Tf})$, and for every vector $\mathbf{g}$,

$$\mathbb{L}\left[\mathbb{L}(\mathbf{f})\mathbf{Tg}-\mathbb{L}(\mathbf{g})\mathbf{Tf}\right] = 0, \tag{5.2}$$

It follows that for all vectors $\mathbf{f}$ and $\mathbf{g}$, $\mathbb{L}(\mathbf{f})\mathbf{Tg}-\mathbb{L}(\mathbf{g})\mathbf{Tf}$ is a vector in $\mathfrak{M}$ and $\mathfrak{M}^\perp$ and is therefore $\mathbf{0}$. Thus $\mathfrak{M}^\perp$ is the space of multiples of some vector $\mathbf{h}_0$ in $\mathfrak{M}^\perp$. If $\mathbf{h}_0=\mathbf{0}$, the theorem is true with $\mathbf{h}=\mathbf{0}$ and $\mathbb{L}$ vanishes identically. If $\mathbf{h}_0\neq\mathbf{0}$, it can be supposed that $|\mathbf{h}_0| = 1$. Then $\mathbf{f} = a\mathbf{h}_0 + \mathbf{f}_1$ with $\mathbf{f}_1$ in $\mathfrak{M}$, for some constant a depending on $\mathbf{f}$, and in fact $(\mathbf{f},\mathbf{h}_0) = a$. Finally, $\mathbb{L}(\mathbf{f}) = (\mathbf{f},\mathbf{h}_0)\mathbb{L}(\mathbf{h}_0)$, that is, the theorem is true with $\mathbf{h} = \overline{\mathbb{L}(\mathbf{h}_0)}\mathbf{h}_0$.

6. Fourier series

If $\mathbf{f}_\bullet$ is an orthonormal sequence and $\mathbf{f}$ is an arbitrary vector, the number $(\mathbf{f},\mathbf{f}_n)$ is the nth *Fourier coefficient of* $\mathbf{f}$ (relative to $\mathbf{f}_\bullet$) and $\Sigma(\mathbf{f},\mathbf{f}_\bullet)\mathbf{f}_\bullet$ is the corresponding *Fourier series of* $\mathbf{f}$. The notation $\mathbf{f} \sim \Sigma\, a_\bullet\mathbf{f}_\bullet$ will mean that the series is the Fourier series for $\mathbf{f}$, that is, that the coefficients are the Fourier coefficients. The key convergence properties of orthogonal series are given by the following theorem.

Theorem. *Let* $\mathbf{f}_\bullet$ *be an orthonormal sequence in a Hilbert space* H .

(a) *The series* $\Sigma a_\bullet\mathbf{f}_\bullet$ *converges if and only if* $\Sigma\, |a_\bullet|^2 < +\infty$.

(b) *If* $\mathbf{f} = \Sigma a_\bullet\mathbf{f}_\bullet$ *and* $\mathbf{g} = \Sigma b_\bullet\mathbf{g}_\bullet$ *are convergent series, then* $(\mathbf{f},\mathbf{g}) = \Sigma\, a_\bullet\bar{b}_\bullet$.

(c) **(Bessel's inequality)** *If* $\mathbf{f} \sim \Sigma a_\bullet\mathbf{f}_\bullet$ *,then* $|\mathbf{f}|^2 \geq \Sigma\, |a_\bullet|^2$.

Proof of (a). The series $\Sigma a_\bullet\mathbf{f}_\bullet$ converges if and only if the Cauchy condition

$$\lim_{j,k\to\infty} \Big|\sum_{n=j}^{k} a_n\mathbf{f}_n\Big|^2 = \lim_{j,k\to\infty} \sum_{n=j}^{k} |a_n|^2 = 0, \tag{6.1}$$

and therefore if and only if $\Sigma\,|a_\bullet|^2 < +\infty$.

Proof of (b). By direct calculation, if **f** and **g** are given as the indicated sums,

$$(6.2)\qquad \Big(\mathbf{f}-\sum_{n=1}^{m} a_n\mathbf{f}_n, \mathbf{g}-\sum_{n=1}^{m} b_n\mathbf{g}_n\Big) = \big(\mathbf{f},\mathbf{g}\big) - \sum_{n=1}^{m} a_n\bar{b}_n.$$

Apply continuity of the inner product to obtain (b) when $m \to +\infty$.

Proof of (c). It is sufficient to prove Bessel's inequality for a finite orthonormal sequence. In this case, if $\mathbf{f} \sim \Sigma a_\bullet\mathbf{f}_\bullet$, there is no convergence problem, and (special case of (6.2))

$$(6.3)\qquad |\mathbf{f}-\Sigma a_\bullet\mathbf{f}_\bullet|^2 = |\mathbf{f}|^2 - \Sigma\,|a_\bullet|^2 \geq 0.$$

7. Fourier series properties

The following theorem lists the Fourier series properties with special emphasis on their geometric significance.

Theorem. *Let $\mathbf{f}_\bullet$ be an orthonormal sequence in a Hilbert space $\mathfrak{H}$.*

(a) *The class $\mathfrak{H}'$ of the sums $\Sigma a_\bullet\mathbf{f}_\bullet$ with $\Sigma\,|a_\bullet|^2 < +\infty$ is a subspace of $\mathfrak{H}$.*

(b) *If $\mathbf{f} \sim \Sigma a_\bullet\mathbf{f}_\bullet$ then the series converges; denote the series sum by $\mathbf{f}'$. Then $\mathfrak{H}'$ is the class of vector sums $\mathbf{f}'$.*

(c) *$\mathbf{f}'$ is the projection of $\mathbf{f}$ on $\mathfrak{H}'$, and $\mathbf{f}-\mathbf{f}'$ is the projection of $\mathbf{f}$ on $\mathfrak{H}'^{\perp}$; $\mathfrak{H}' = \mathfrak{H}$ if and only if $\mathbf{f}_\bullet$ is a complete orthonormal sequence. In this case, $\mathbf{f} \sim \Sigma a_\bullet\mathbf{f}_\bullet$ implies that $\mathbf{f} = \Sigma a_\bullet\mathbf{f}_\bullet$.*

(d) (**Parseval identity**) *If $\mathbf{f}_\bullet$ is complete and $\mathbf{f} = \Sigma a_\bullet\mathbf{f}_\bullet$ and $\mathbf{g} = \Sigma b_\bullet\mathbf{g}_\bullet$, then*

$$(7.1)\qquad |\mathbf{f}|^2 = \Sigma\,|a_\bullet|^2 \qquad \big(\mathbf{f},\mathbf{g}\big) = \Sigma a_\bullet\bar{b}_\bullet.$$

Equivalence of the Parseval identity and equality in Bessel's inequality. If the Parseval identity holds for all pairs **f**, **g**, then it is trivial that there is equality in Bessel's inequality; one need only take **f** = **g**. Conversely, if equality holds in Bessel's inequality for all **f**, then Parseval's identity holds for all pairs **f**, **g**. In fact if equality in the Bessel inequality is written out for **f**, **g**, **f**+**g**, and *i***f**+**g**, the equations imply the Parseval identity.

Proof of (a). Rather than proving directly that $\mathfrak{H}'$ is linear and closed, it is more instructive to note that by its very definition $\mathfrak{H}'$ is in 1-1 correspondence with the Hilbert space of sequences (Fourier coefficients) $a_\bullet$ satisfying $\Sigma\ |a_\bullet|^2 < +\infty$, described in the example in Section 2. The correspondence preserves inner products, and the two spaces must therefore be isomorphic Hilbert spaces.

Proof of (b). According to Theorem 6, the series converges. The second assertion of (b) states that if $\mathbf{f}' = \Sigma\, a_\bullet \mathbf{f}_\bullet$ then the coefficients are the Fourier coefficients. Apply the continuity of the inner product:

$$(7.2)\qquad \left(\mathbf{f}',\mathbf{f}_k\right) = \lim_{m\to\infty}\left(\sum_{n=1}^{m} a_n\mathbf{f}_n,\mathbf{f}_k\right) = a_k.$$

Proof of (c). The fact that the vectors $\mathbf{f}$ and $\mathbf{f}'$ have the same Fourier coefficients implies that the difference $\mathbf{f}-\mathbf{f}'$ is orthogonal to every $\mathbf{f}_n$ and therefore to every linear combination of these vectors, and therefore (continuity of the inner product) to the closure $\mathfrak{H}'$ of these linear combinations. Thus $\mathbf{f} = \mathbf{f}'+(\mathbf{f}-\mathbf{f}')$ is the representation of $\mathbf{f}$ as the sum of a vector in $\mathfrak{H}'$ and one in $\mathfrak{H}'^{\perp}$ that is, as the sum of its projections on these mutually orthogonal subspaces. Completeness of $\mathbf{f}_\bullet$ is equivalent to $\mathfrak{H}'^{\perp} = \mathbf{0}$.

Proof of (d). This is now simply an application of Theorem 6(b).

8. Orthogonalization (Erhardt Schmidt procedure)

If $\mathbf{g}_1,\ldots,\mathbf{g}_N$ are linearly independent vectors of a Hilbert space, the following procedure yields an orthonormal set $\mathbf{f}_1,\ldots,\mathbf{f}_N$ of linear combinations of these vectors, and the given vectors are in turn linear combinations of those in the orthonormal set. Define

$$\mathbf{f}_1 = \mathbf{g}_1/|\mathbf{g}_1|,$$

$$(8.1)\qquad \mathbf{f}_2' = \mathbf{g}_2 - \left(\mathbf{g}_2,\mathbf{f}_1\right)\mathbf{f}_1, \qquad \mathbf{f}_2 = \mathbf{f}_2'/|\mathbf{f}_2'|,$$

$$\mathbf{f}_3' = \mathbf{g}_3 - \left(\mathbf{g}_3,\mathbf{f}_1\right)\mathbf{f}_1 - \left(\mathbf{g}_3,\mathbf{f}_2\right)\mathbf{f}_2, \qquad \mathbf{f}_3 = \mathbf{f}_3'/|\mathbf{f}_3'|,$$

$$\cdots$$

The idea guiding this procedure is that when $j > 1$, the vector $\mathbf{f}_j'$ is $\mathbf{g}_j$ less the sum of its Fourier series for the orthonormal set $\mathbf{f}_1,\ldots,\mathbf{f}_{j-1}$

Dimensionality of a Hilbert space and complete orthonormal sets. If a Hilbert space has finite dimensionality N, the result just proved shows that there is an orthonormal set of N vectors. This set is obviously complete.

Suppose that a Hilbert space $\mathfrak{H}$ has infinite dimensionality, but that the space is separable, that is, there is a dense sequence $\mathbf{g}_\bullet$ of vectors. Delete from this sequence any vector $\mathbf{g}_k$ that is a linear combination of $\mathbf{g}_1,\ldots,\mathbf{g}_{k-1}$. The remaining vectors have the property that their linear combinations are dense in the space and that every finite subset is a linearly independent set. Apply the Schmidt procedure to these remaining vectors to find an orthonormal sequence $\mathbf{f}_\bullet$. The linear combinations of the vectors $\mathbf{f}_\bullet$ are dense in the Hilbert space. The closure of the linear combinations is a subspace of $\mathfrak{H}$ according to Theorem 6, and therefore is $\mathfrak{H}$. Thus the sequence $\mathbf{f}_\bullet$ is a complete orthonormal set, an infinite sequence, because $\mathfrak{H}$ was supposed infinite dimensional.

Conversely, if a Hilbert space has a countable complete orthonormal set the space is a separable metric space. In fact if the sequence $\mathbf{f}_\bullet$ is a complete orthonormal set, the class of all finite linear combinations of these vectors with rational real and imaginary part coefficients is a countable dense set.

Separable Hilbert spaces and l^2. If a Hilbert space $\mathfrak{H}$ is separable and if $\mathbf{f}_\bullet$ is a finite or infinite complete orthonormal sequence, as determined by the dimensionality of $\mathfrak{H}$, every Hilbert space vector $\mathbf{f}$ is the sum of its Fourier series, $\mathbf{f} = \Sigma\, a_\bullet \mathbf{f}_\bullet$, and according to (7.1),

$$|\mathbf{f}|^2 = \Sigma\, |a_\bullet|^2 = |a_\bullet|^2,$$

where the first norm is that for $\mathfrak{H}$ and the second is that for the Hilbert space l^2. Thus there is a one-to-one correspondence between the vectors of the Hilbert spaces $\mathfrak{H}$ and l^2, which preserves norms, and therefore, as remarked in Section 2, preserves inner products. That is, these two Hilbert spaces are isomorphic. This argument shows that up to isomorphisms, for a given dimensionality, l^2 is the most general Hilbert space, at least if the Hilbert space is supposed separable. The conclusion is actually valid for the most general Hilbert space, if uncountable infinite series are allowed, under the usual summation conventions.

9. Fourier trigonometric series

Let S be the circle in $\mathbf{R}^2$ of radius 1 with center the origin, equivalently the interval $[-\pi,\pi]$ of $\mathbf{R}$, for $-\pi$ identified with π. The measure involved is one-dimensional Lebesgue measure. The sequence $f_\bullet$ defined on $[-\pi,\pi]$ by $f_n(s) = (2\pi)^{-1/2}e^{ins}$, $n = \cdots,-1,0,1,\cdots$ is an orthonormal sequence. According to Section **VI.14**, the set of linear combinations of these functions is dense in the pseudometric space L^2. It follows that every function in L^2 is equal to the sum of its Fourier series, in the sense that the partial sums converge to the function in the mean. Actually it was proved by Carleson that there is almost everywhere convergence in this particular context.

10. Two trigonometric integrals

The following integrals, in which ds refers to Lebesgue measure and only Riemann integration is needed because the integrands are continuous, will be used in discussing Fourier transforms:

$$(10.1)\qquad \int_{-\infty}^{\infty} \frac{\sin \beta s}{s}\, ds \begin{cases} =\pi \\ =0, \\ =-\pi \end{cases} \qquad \int_{-\infty}^{\infty} \frac{1-\cos \beta s}{s^2}\, ds \begin{cases} =\beta\pi & >0 \\ =0 & \text{if } \beta=0 \\ =\beta\pi & <0. \end{cases}$$

The first integral, which is not absolutely convergent, is defined by

$$(10.2)\qquad \lim_{\alpha\to\infty} \int_{-\alpha}^{\alpha} \frac{\sin s}{s}\, ds$$

and is evaluated by residues. Replace $\sin s$ by $\sin ts$ in the first integrand and integrate with respect to t over the interval $[0,\beta]$ to evaluate the second integral.

11. Heuristic approach to the Fourier transform via Fourier series

Throughout this section the measure involved will be Lebesgue measure on **R** or on a subinterval of **R**.

The Fourier transform. If f is a Lebesgue measurable function from **R** into **R**, the *Fourier transform* Uf of f is defined by

$$(11.1)\qquad (\mathrm{U}f)(t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} f(s)e^{its}\, ds,$$

under suitable hypotheses on f, and the integral is the *Fourier integral.* (In the following, the left side of (11.1) will be abbreviated to U$f(t)$.) If i is replaced by $-i$ in the integrand, the integral is the *inverse Fourier integral,* defining the *inverse Fourier transform* U*. The Fourier integral is well defined when $f \in \mathrm{L}^1(\mathbf{R})$, but it is sometimes necessary to define the Fourier integral in a way that requires less of f. There is a general principle that possibly with generalizations of the definitions of the integrals involved, U and U* are inverse operations, and that these operations preserve inner products, that is, on suitably restricted domains, UU* and U*U are the identity and

$$(f,g) = (\mathrm{U}f,\mathrm{U}g) = (\mathrm{U}^*f,\mathrm{U}^*g).$$

This rather vague remark will be justified by the Fourier–Plancherel theorem in the next section, and by Lévy's theorem on the representation of a distribution function in terms of its characteristic function in Section X.14.

The following unrigorous argument leads from trigonometric Fourier series (which make correspond to a function in $L^2([-\pi,\pi])$ (Lebesgue measure) the sequence in l^2 of its Fourier coefficients) to Fourier integrals (which, in the context of the Fourier-Plancherel theorem, make correspond to a function in $L^2(\mathbf{R})$ its generalized Fourier transform, also a function in $L^2(\mathbf{R})$). Fourier series based on a complete orthornormal sequence define a transformation from one Hilbert space onto a second; Fourier transforms, in the context of the Fourier-Plancherel theorem, define a transformation from a Hilbert space onto itself. Both transformations preserve inner products.

The relations to be applied between the Fourier trigonometric series of a function f in $L^2([-\pi,\pi])$ and its coefficients, are

(11.2)
$$f(s) = (2\pi)^{-1/2}\Sigma_{-\infty}^{\infty}\, b_n e^{ins}, \quad b_n = (2\pi)^{-1/2}\int_{-\pi}^{\pi} f(s)\, e^{-ins} ds,$$
$$\int_{-\pi}^{\pi} |f(s)|^2\, ds = \Sigma_{-\infty}^{\infty} |b_n|^2.$$

Here the first sum is convergent in the mean. These relations can be rewritten for functions defined on the interval $[-l\pi,l\pi]$, in the form

(11.3)
$$f(s) = (2\pi)^{-1/2}\Sigma_{-\infty}^{\infty}\, b(n,l)\, e^{isn/l}/l, \quad b(n,l) = (2\pi)^{-1/2}\int_{-l\pi}^{l\pi} f(s)\, e^{-isn/l}\, ds,$$
$$\int_{-l\pi}^{l\pi} |f(s)|^2\, ds = \Sigma_{-\infty}^{\infty} |b(n,l)|^2/l,$$

in which $b(n,l)$ is not normalized like a Fourier coefficient. When l becomes infinite, the sums are approximating sums for Riemann integration and thereby suggest, with a change of notation and the help of a grain of salt, the relations

$$f(s) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \mathrm{U}^*f(t)\, e^{ist}\, dt = \mathrm{U}\mathrm{U}^*f, \qquad \mathrm{U}^*f(t) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} f(s)\, e^{-ist}\, ds,$$

(11.4)

$$\int_{-\infty}^{\infty} |f(s)|^2\, ds = \int_{-\infty}^{\infty} |\mathrm{U}^*f(t)|^2\, dt.$$

These manipulations suggest that if f is in L^2 and Uf and U*f are suitably defined, if necessary by loosening the definition of an integral, then Uf and U*f are in L^2, UU* and U*U are the identity, and U and U* preserve L^2 norms; as has already been noted, U and U* then also preserve inner products.

12. The Fourier-Plancherel theorem

This theorem is the most elegant version of (11.2) in the Fourier transform context. In the following, the measure is Lebesgue measure on **R**. Observe that if f is in $L^2(\mathbf{R})$ it does not follow that f is in $L^1(\mathbf{R})$, and therefore the Fourier integral of f cannot be defined as an ordinary integral. The Fourier-Plancherel theorem solves this problem by using an L^2 definition of the Fourier integral.

The clearest way to deal with the Fourier-Plancherel theorem is to use both L^2 and Hilbert space terminology. To make the reasoning both perspicuous and correct a careful distinction will be made between L^2 functions and their equivalence classes. If f is a function, the corresponding equivalence class, consisting of all functions equal to f almost everywhere, will be denoted by **f** and identified as a point of the Hilbert space $\mathfrak{H}$ of equivalence classes. (The only exception is the notation for an indicator function, already in boldface, but it will be clear from the context whether an indicator function or an equivalence class is meant.) Thus, for example, convergence in the mean of a sequence $f_\bullet$ is equivalent to Hilbert space convergence of $\mathbf{f}_\bullet$. In the following theorem Uf is defined as a function, a certain limit in the mean. Such a limit is any function in an equivalence class of functions in L^2. This equivalence class is unaffected by a change of f on a null set. Thus U determines a transformation of equivalence classes, that is, a transformation **U** from $\mathfrak{H}$ into itself.

Theorem. *If $f \in L^2(\mathbf{R})$, the mean limits*

$$\text{(12.1)} \qquad \mathrm{U}f(t) = \text{l.i.m.}_{\alpha\to\infty} (2\pi)^{-1/2} \int_{-\alpha}^{\alpha} f(s)\, e^{ist}\, ds,$$

$$\mathrm{U}^*f(t) = \text{l.i.m.}_{\alpha\to\infty} (2\pi)^{-1/2} \int_{-\alpha}^{\alpha} f(s)\, e^{-ist}\, ds$$

exist. The transformations **U** and **U*** *are unitary operators on* $\mathfrak{H}$*, and each is the inverse of the other.*

Observe that if f is in both L^2 and L^1, the mean limits in (12.1), once they are known to exist, can be chosen as the ordinary integrals.

Proof. The Fourier transforms of the indicator functions of compact intervals can be calculated as ordinary integrals, as can – using (10.1) – the inner products of the Fourier transforms:

(12.2)

$$\mathrm{U}1_{[a,b]}(t) = \frac{e^{itb} - e^{ita}}{it\sqrt{2\pi}}, \qquad \mathrm{U}1_{[c,d]}(t) = \frac{e^{itd} - e^{itc}}{it\sqrt{2\pi}},$$

$$\left(\mathrm{U}1_{[a,b]}, \mathrm{U}1_{[c,d]}\right) = \int_{-\infty}^{\infty} \frac{\cos t(b-d) + \cos t(a-c) - \cos t(b-c) - \cos t(a-d)}{2\pi t^2}\, dt$$

$$= \left(1_{[a,b]}, 1_{[c,d]}\right).$$

Observe that Uf in (12.2) is in $L^2(\mathbf{R})$ but not in $L^1(\mathbf{R})$. It follows from (12.2) that the Fourier transform U defines a linear transformation, which preserves inner products, on the class of step functions constant on intervals. The corresponding Hilbert space transformation **U** is thereby defined on a certain linear set, dense in $\mathfrak{H}$ according to Section **VI**.15.

Now (Section **0**.12(b)) a uniformly continuous function from a dense subset of a complete metric space into a complete metric space has a unique continuous extension to the whole space. In the present context, the extension of **U** is a linear map of $\mathfrak{H}$ into itself, with a corresponding extension of U. Moreover (continuity of the Hilbert space inner product), **U** preserves inner products. If $f \in L^2(\mathbf{R})$ and I is a finite interval, the restriction to I of f is in $L^2(I)$ and therefore (Schwarz et al. inequality) also in $L^1(I)$, that is, $f1_I$ is in both $L^2(\mathbf{R})$ and $L^1(\mathbf{R})$. If $f_\bullet$ is a sequence of step functions, vanishing outside I, convergent in the mean to $f1_I$, the Fourier transform sequence U$f_\bullet$ has as limit in the mean the Fourier transform of $f1_I$, and therefore Uf, as defined by the extension theorem, is the Fourier transform of $f1_I$. That is, one version of U($f1_I$) is given by an ordinary integral:

(12.3)

$$\mathrm{U}(f1_I)(t) = (2\pi)^{-1/2} \int_I f(s)\, e^{ist}\, ds,$$

$$\int_{-\infty}^{\infty} |\mathrm{U}(f1_I)(t)|^2\, dt = \int_I |f(s)|^2\, ds.$$

It follows that if $I_\bullet$ is an increasing sequence of finite intervals with center the origin and limit **R**, the sequence of integrals

$$(2\pi)^{-1/2} \int_{I_\bullet} f(s)\, e^{ist}\, ds$$

is a Cauchy sequence in L^2. Thus Uf is defined by (12.1) and preserves inner products. Similarly, U^* is defined by (12.1) and preserves inner products. Apply U^* to $U1_{[a,b]}$, evaluated in (12.2):

$$(12.4)\qquad U^*(U1_{[a,b]})(s) = \text{l.i.m.}_{\alpha\to\infty}\int_{-\alpha}^{\alpha} e^{-ist}\frac{e^{itb}-e^{ita}}{2\pi i t}\,dt$$

$$= \text{l.i.m.}_{\alpha\to\infty}\int_{-\alpha}^{\alpha}\frac{\sin t(b-s)-\sin t(a-s)}{2\pi t}\,dt.$$

According to (10.1), the pointwise limit of the second integral is $1_{[a,b]}$, neglecting the points a and b, and therefore this indicator function is also the limit in the mean. Thus U^*U, neglecting null sets, is the identity on the class of indicator functions of intervals, therefore also on the class of linear combinations of these indicator functions, and therefore (continuity of these distance-preserving operators) U^*U is the identity on L^2, that is, U^*U is the identity on $\mathfrak{H}$. The set $U\mathfrak{H}$ is a subspace of $\mathfrak{H}$. If an element of $\mathfrak{H}$ is orthogonal to this subspace, U^*U takes this element into one orthogonal to $\mathfrak{H}$, and it follows that the element is the zero element, that is, U is a unitary operator taking $\mathfrak{H}$ onto itself, and U^*U is the identity. Similarly, U^* is a unitary operator taking $\mathfrak{H}$ onto itself, and UU^* is the identity.

13. Ergodic theorems

If $\mathbf{U}$ is a unitary transformation of a Hilbert space onto itself, the invariant vectors, that is, those vectors $\mathbf{f}$ satisfying the equation $\mathbf{Uf} = \mathbf{f}$, form a subspace. The iterates of unitary transformations on a Hilbert space arise in many applications, and it is an important theorem that the average of the iterates converges to the projection on this invariant subspace. Applications and examples will be given after the proof of the theorem.

Theorem (von Neumann's Hilbert space ergodic theorem). *Let* $\mathbf{U}$ *be a unitary transformation on a Hilbert space* $\mathfrak{H}$, *and let* $\mathbf{P}$ *be the projection of* $\mathfrak{H}$ *onto the invariant subspace* $\mathfrak{M}$ *for* $\mathbf{U}$. *Then for every vector* $\mathbf{f}$,

$$(13.1)\qquad \lim_{n\to\infty}\frac{1}{n+1}\sum_{m=0}^{n}\mathbf{U}^m\mathbf{f} = \mathbf{Pf}.$$

To prove the theorem, following F. Riesz, it is sufficient to prove that (13.1) is true both for $\mathbf{f}$ in $\mathfrak{M}$ and $\mathbf{f}$ in $\mathfrak{M}^{\perp}$, because each average in (13.1) is a linear transformation of $\mathbf{f}$, and $\mathbf{f}$ can be written as the sum of a vector in $\mathfrak{M}$ and one in $\mathfrak{M}^{\perp}$. It is trivial that (13.1) is true for $\mathbf{f}$ in $\mathfrak{M}$. There remains the identification of $\mathfrak{M}^{\perp}$, and the proof that the limit on the left in (13.1) is $\mathbf{0}$ for $\mathbf{f}$ in that subspace. Denote by $\mathfrak{N}_0$ and $\mathfrak{N}$, respectively, the linear class of vectors of the form $\mathbf{g}-\mathbf{Ug}$ and the closure of $\mathfrak{N}_0$, a subspace. It will be shown that $\mathfrak{M}^{\perp} = \mathfrak{N}$ and that the limit in (13.1) is $\mathbf{0}$ when $\mathbf{f}$ is in $\mathfrak{N}$. In the equality

$$(13.2) \qquad (\mathbf{h},\mathbf{g}-\mathbf{Ug}) = (\mathbf{h},\mathbf{g}) - (\mathbf{h},\mathbf{Ug}) = (\mathbf{Uh},\mathbf{Ug}) - (\mathbf{h},\mathbf{Ug})$$

$$= (\mathbf{Uh}-\mathbf{h},\mathbf{Ug}),$$

if $\mathbf{h}$ is orthogonal to $\mathfrak{N}$, the first term on the left vanishes for every $\mathbf{g}$, and therefore the last term vanishes for every $\mathbf{g}$. Hence $\mathbf{Uh}-\mathbf{h}$ is orthogonal to $\mathbf{U}\mathfrak{H}$, that is, to $\mathfrak{H}$, and it follows that $\mathbf{Uh}-\mathbf{h} = \mathbf{0}$, that is, $\mathbf{h}$ is in $\mathfrak{M}$. Conversely, if $\mathbf{h}$ is in $\mathfrak{M}$, the last term in (13.2) vanishes for every g, and the vanishing of the first term means that $\mathbf{h}$ is orthogonal to $\mathfrak{N}$. Thus the subspaces $\mathfrak{M}$, $\mathfrak{N}$ are complementary orthogonal subspaces. If $\varepsilon > 0$, $\mathbf{f}$ is in $\mathfrak{N}$, and $\mathbf{g}-\mathbf{Ug}$ is a vector in $\mathfrak{N}_0$ at distance at most ε from $\mathbf{f}$, then

$$(13.3) \quad \left|\frac{1}{n+1}\sum_{m=0}^{n} \mathbf{U}^m\mathbf{f}\right| \le \left|\frac{1}{n+1}\sum_{m=0}^{n} \mathbf{U}^m[\mathbf{f}-(\mathbf{g}-\mathbf{Ug})]\right| + |(\mathbf{g}-\mathbf{U}^{n+1}\mathbf{g})|/(n+1).$$

The first term on the right is majorized by ε, the second is majorized by $2|\mathbf{g}|/(n+1)$, and therefore the limit of the term on the left is 0 when n becomes infinite, as was to be proved.

Example (a) (**Koopman**). Let $(S,\mathbf{S},\lambda)$ be a finite measure space, let T be a one-to- one transformation of S onto itself, and suppose that T is measure preserving, in the sense that T takes measurable sets into measurable sets of the same measure. If f is a measurable function from S into $\mathbf{R}$, define a transformation U of functions by $(\mathrm{U}f)(s) = f(Ts)$. Then U is a linear transformation, takes measurable functions into measurable functions, and Uf has the same distribution as f. Hence, for every $p \ge 1$, U on L^p is a linear norm-preserving transformation of L^p onto itself. In particular, if $p = 2$, U defines a unitary transformation $\mathbf{U}$ of the Hilbert space of equivalence classes of L^2. A function f is *invariant* if $f(s) = f(Ts)$ for almost all s, that is, if its equivalence class is invariant under $\mathbf{U}$. A set is *invariant* if its indicator function is. For example, the almost everywhere constant functions are invariant functions, and the null sets and their complements are invariant sets. If the latter are the only invariant sets, then an invariant function is necessarily equal almost everywhere to a constant, because if f is an invariant function, every set of the form $\{f \in A\}$, with A in $\mathbf{B}(\mathbf{R})$, is an invariant set.

The study of the iterates of a unitary transformation of Hilbert space gives information on the iterates of measure-preserving transformations. Measure-preserving transformations of the phase space of mechanical systems are of fundamental importance in physics, and the question of possible long-time averages of transforms of a function, that is, of possible limits of the average of the first n transforms of f, is crucial. In the slang of physics, the function f has a space average $\lambda[f] = \lambda[f]/\lambda(S)$, and one important question is whether the long-time average of the transformed functions,

$$\text{(13.4)} \qquad \lim_{n\to\infty}\frac{1}{n+1}\sum_{m=0}^{n} f\big(T^n(s)\big) = \lim_{n\to\infty}\frac{1}{n+1}\sum_{m=0}^{n} \mathrm{U}^n f(s),$$

exists in some sense, and if so whether this limit average is constant and equal to the space average. The following translation of Theorem 13 into measure theory gives one answer to this question.

L^2 Ergodic theorem. *Let $(S,\mathbf{S},\lambda)$ be a finite measure space, and let T be a one-to- one measure-preserving transformation of S onto itself. If $f \in \mathrm{L}^2$ then*

$$\text{(13.5)} \qquad \text{l.i.m.}_{n\to\infty}\frac{1}{n+1}\sum_{m=0}^{n} f\big(T^n(s)\big)$$

exists. The limit is almost everywhere constant $\big($necessarily the constant $\lambda[f])/\lambda(S)\big)$ for every choice of f if and only if the only invariant functions are the almost everywhere constant functions, equivalently if and only if the only invariant sets are the null sets and their complements.

Generalization of the L^2 ergodic theorem. In the preceding discussion, define $f_n = \mathrm{U}^n f$, for $n \geq 0$. The sequence $f_\bullet$ is *stationary*, in the sense that for $k \geq 1$, the k-dimensional distribution of $f_n, f_{n+1}, \ldots, f_{n+k-1}$ is the same for every integer n. That is, the distributions involved are invariant under index set translation.

Suppose, more generally, that $f_\bullet'$ is an arbitrary sequence, indexed by the positive integers, of measurable square integrable functions from a finite measure space $(S',\mathbf{S}',\lambda')$ into $\mathbf{R}$, with the stationarity property that for $k \geq 1$, the k-dimensional distribution of $f_n', f_{n+1}', \ldots, f_{n+k-1}'$ is the same for every positive integer n. As an application of the ideas in Section V.6 on measures in coordinate spaces, it will now be shown that the study of such a stationary sequence can be reduced to the context of the L^2 ergodic theorem. Let the index set I be the set of all integers, and let S be the space $\mathbf{R}^I$ of all sequences (indexed by I) of real numbers. Let f_n be the nth coordinate function of S, and let $\mathbf{S} = \sigma(f_\bullet)$ be the σ algebra of subsets of S generated by the class of sets of the form $\{f_n \in A\}$, where A is a Borel subset of $\mathbf{R}$. In other words, $\mathbf{S}$ is the smallest σ algebra making all the coordinate functions measurable. For each n in I and each $k \geq 0$, assign to the coordinate functions $f_n, \ldots, f_{n+k}$ of S the joint distribution of $f_0', \ldots, f_k'$. These distributions are mutually consistent in the sense of Section V.6, because the joint distributions of the given functions are mutually consistent, and these distributions therefore determine a measure λ on $\mathbf{S}$, according to Theorem V.6. By definition of λ, the joint distribution of the coordinate functions $f_0,\ldots,f_k$ is the same as that of the given functions $f_0',\ldots,f_k'$, and therefore in any question regarding limits of averages of the given functions, there is no loss of generality in replacing $(S'',\mathbf{S}',\lambda')$ by $(S',\mathbf{S},\lambda)$ and replacing $f_\bullet'$ by $\{f_n,\ n \ge 0\}$. The space S has the advantage that if T is the translation taking each point of S into the point with index value increased by 1, then T is a one to one measure preserving transformation of S onto itself and $f_0(T^n s) = f_n(s)$. Thus the problem of the possible limits of successive averages of members of the original stationary sequence $f_\bullet'$ has been reduced to the context of measure preserving transformations. The fact that the mean limit in (13.5) exists implies that the corresponding mean limit for the original primed functions exists.

Example (b). Consider the N-dimensional Hilbert space l^2 defined in Section 2. Let $\mathbf{f}_\bullet$ be a complete orthonormal sequence in the space, let $\alpha_\bullet$ be N numbers of modulus 1, and define

$$\mathbf{U}(\Sigma a_\bullet \mathbf{f}_\bullet) = \Sigma \alpha_\bullet a_\bullet \mathbf{f}_\bullet.$$

The transformation $\mathbf{U}$ is unitary, and $\mathbf{U}^m(\Sigma a_\bullet f_\bullet) = \Sigma \alpha^m a_m \mathbf{f}_m$. Since the sum $1+\alpha+\cdots+\alpha^{n-1}$ is n when $\alpha = 1$, and otherwise is $(1-\alpha^n)/(1-\alpha)$. It follows that

$$\text{(13.6)} \qquad \lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} \mathbf{U}^m \mathbf{f} = \Sigma_{m:\alpha_m=1}\, a_m \mathbf{f}_m = \mathbf{P}\mathbf{f},$$

where $\mathbf{P}$ is the projection of $\mathfrak{H}$ onto the subspace generated by those vectors $\mathbf{f}_m$ with $\alpha_m = 1$, that is, $\mathbf{P}$ is the projection on the invariant subspace for $\mathbf{U}$, as it should be according to the Hilbert space ergodic theorem. This simple example is actually not far from the general case. To formulate a hint of the general case, let $\beta_\bullet$ be the distinct values of $\alpha_\bullet$ and write $\mathbf{U}$ in the form $\mathbf{U} = \Sigma\, \beta_\bullet \mathbf{P}_\bullet$, where $\mathbf{P}_m$ is the projection of the Hilbert space on the subspace of linear combinations of $\mathbf{f}_k$ for those values of k with $\alpha_k = \beta_m$. Thus these projections are on mutually orthogonal subspaces. The general unitary transformation of a separable Hilbert space has nearly the same form: the finite sum is replaced by a continuous sum (a form of Stieltjes integral of a family of projections).

Example (c). Let S be a circle of radius 1, $\mathbf{S} = \mathbf{B}(S)$, and λ be Lebesgue measure on the circle. Let α be an irrational number and T be the rotation of S through α radians about the center. Then T is one to one and measure preserving. Moreover, only the almost everywhere constant functions are invariant under T. To prove this assertion, suppose that T is a square integrable invariant function. Then the Fourier coefficients of f and the transformed function must be equal, and a trivial calculation yields, for the sequence $a_\bullet$ of Fourier coefficients of f,

$$a_n = a_n e^{ni\alpha}, \quad (n=0,\pm1,\pm2,\ldots). \tag{13.7}$$

Then $a_n = 0$ except possibly when $n = 0$, and therefore the Fourier series for f makes f an almost everywhere constant function. The L^2 ergodic theorem yields the limit equation

$$\text{l.i.m.}_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} f(s+m\alpha) = \frac{1}{2\pi}\int_0^{2\pi} f(s)\, ds, \tag{13.8}$$

where ds refers to Lebesgue measure.

The law of large numbers. A theorem that states that the sequence of successive averages of a sequence of functions on a probability space is convergent in some sense, is a *law of large numbers*. If the theorem states that the convergence is almost everywhere convergence, it is a *strong law of large numbers*. Thus the L^2 ergodic theorem, and the corresponding limit equation for a stationary sequence are laws of large numbers when the measure space is a probability space. This sort of romantic nomenclature adds romance (and mystery to nonprobabilists) to measure theory whenever the measure of the space is 1!

According to the *Birkhoff ergodic theorem*, whose proof will be omitted, the limit equation (13.5), and the corresponding limit equation for a stationary sequence, are true in the sense of almost everywhere convergence, even if f is supposed to be only in L^1. When the measure space is a probability space, the Birkhoff ergodic theorem is thus the *strong law of large numbers for stationary sequences of random variables.*

The law of large numbers in the independence case. Let $f_\bullet$ be a sequence of mutually independent random variables, with a common distribution, on a probability space $(S,\mathbf{S},P)$. Recall that if the common distribution is the measure λ of Borel subsets of $\mathbf{R}$, one representation of such a sequence is the sequence of coordinate functions of infinite dimensional Euclidean space $\mathbf{R}\times\mathbf{R}\times\cdots$ with the product measure $\lambda\times\lambda\times\cdots$. The sequence $f_\bullet$ is stationary, and therefore the L^2 ergodic theorem is applicable if f_1 is in L^2. In this case the limit is almost surely constant, according to the 0-1 law (Theorem V.9) and the L^2 law of large numbers takes the form

$$\text{l.i.m.}_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^{n} f_n = E\{f_1\}. \tag{13.9}$$

As already noted, this limit is valid as an almost everywhere limit, even when f_1 is in L^1, and when so stated (13.9) becomes *the strong law of large numbers for identically distributed independent random variables.* This form of the strong law of large numbers will be proved (Theorem **XI**.19) as an application of a martingale convergence theorem.

VIII
Convergence of Measure Sequences

1. Definition of convergence of a measure sequence

This chapter discusses the kinds of convergence of a measure sequence most frequently met in classical analysis. The typical context to be considered is the following. A class **M** of measures on a measurable space $(S,\mathbf{S})$ is given, together with a class Γ of functions from S into **R**. The problem is to find a definition of convergence of the sequence $\lambda_\bullet$ in **M** to a measure λ in **M**, which implies that $\lim\lambda_\bullet[f] = \lambda[f]$ for every function f in Γ. This problem has an easy solution, a solution by definition: *the sequence* $\lambda_\bullet$ *is* Γ *convergent to* λ *if*

$$\lim\lambda_\bullet[f] = \lambda[f] \qquad (f \in \Gamma). \tag{1.1}$$

This chapter is based on this easy solution, but there is an old proverb to take into account: *there is no free theorem*. The easy solution is indeed easy, but is only a first step in finding and characterizing a *useful* solution, that is, one in which S,**S**, Γ, and **M** are chosen in such a way that the choice can be shown to be applicable to the needs of analysis.

Example. The Vitali-Hahn-Saks theorem. Let $(S,\mathbf{S})$ be an arbitrary measurable space, Γ be the class of bounded measurable functions from S into **R**, and **M** be the class of finite measures. Under these definitions, the indicator functions of measurable sets are in Γ, and therefore Γ convergence of a sequence of measures implies setwise convergence. Conversely, the Vitali-Hahn-Saks theorem states that if there is setwise convergence of $\lambda_\bullet$ to a finite valued set function λ, then λ is a measure. Under setwise convergence, the limit relation in (1.1) is true when f is the indicator function of a measurable set, and therefore successively when f is a step function and when f is in Γ. Thus Γ convergence is the same as setwise convergence.

The Vitali-Hahn-Saks theorem is a valuable tool, but setwise convergence of measure sequences is a strong kind of convergence that is rare in classical analysis. What is common is the following context:

(a) the space S is a metric space, either compact or at least separable and locally compact, and $\mathbb{S}$ is the class of Borel subsets of S;

(b) Γ is the class of bounded continuous functions from S into $\mathbf{R}$;

(c) $\mathbb{M}$ is the class of *Radon* measures, those measures finite valued on compact sets.

This will be the context in the rest of this chapter.

Punctured compact metric spaces. If S is a not compact metric space, a function f from S into a metric space is said *to have limit* α *at infinity* if the inverse image of each neighborhood of α is contained in the complement of a compact subset of S. The following argument shows that this terminology can be interpreted literally, under appropriate conventions.

Let S' be a compact metric space and s' be a not isolated point of the space. The space $S = S'-\{s'\}$ is then a locally compact but not compact separable metric space and will be called a *punctured compact metric space*. Conversely, if S is a locally compact but not compact separable metric space, it will now be shown that under a change of metric that does not change the class of open sets, S becomes a punctured compact metric space. To see this let δ_1 be the distance function of S, $t_\bullet$ be a sequence dense in S, and $A_\bullet$ be an increasing sequence of compact subsets of S, with union S. For each value of n, the function $\delta_1(\bullet,t_n)\wedge 1$ is a positive continuous function on S, vanishing only at t_n. The function $\sum \delta_1(\bullet,\tilde{A}_n)\wedge 2^{-n} = g$ is a strictly positive continuous function on S, at most 2^{-n} on $\tilde{A}_n$. If "going to infinity" means proceeding through the sequence $\tilde{A}_\bullet$, this function has limit 0 at infinity. This idea is made precise as follows. Define

$$f_n = \left[\delta_1(\bullet,t_n)\wedge g\right]\wedge 2^{-n}.$$

If s and t are points of S at which $f_n(s) = f_n(t)$ for all n, then $s = t$. Define a new distance function δ for S by

$$(1.1) \qquad \delta(s,t) = \sum |f_\bullet(s) - f_\bullet(t)|.$$

The space S under the δ metric has the same topology as under the δ_1 metric because a sequence $s_\bullet$ has limit s in S under one of the two metrics if and only if it has s as limit in the other. Adjoin a point "∞" to S to obtain a space $S' = S\cup\{\infty\}$, define $f_n(\infty) = g(\infty) = 0$, and define distance on S' by (1.1). Under this definition, f_n and g are continuous on S'. If $s_\bullet$ is a sequence in S, either some subsequence is δ_1 convergent and therefore δ convergent to a point of S, or only a finite number of members of the sequence are in any one set of the sequence $A_\bullet$. In the latter case, each function f_n has limit 0 along the sequence, and therefore the sequence has the point ∞ as limit in the δ metric. Thus S' is a compact metric space, exhibiting S as a punctured compact metric space. "Going to infinity" in S means approaching the point ∞ of S' in the δ metric of

S'. In particular, an open not compact subset of a compact metric space can be remetrized to be a punctured compact metric space. The effect of the new metric is to replace the original boundary by a single point, ∞.

If S is a punctured compact metric space and λ is a Radon measure on **B(S)**, the extension of λ to a measure on $\mathbf{B}(S')$ obtained by assigning a finite measure to the singleton $\{\infty\}$ will be denoted by λ'.

The class M(S). If S is a metric space, the space of finite measures on $\mathbf{B}(S)$ will be denoted by $\mathbf{M}(S)$, and $\{\lambda \in \mathbf{M}(S): \lambda(S) \le c\}$ will be denoted by $\mathbf{M}_c(S)$.

The class C(S) and its norms. If S is a metric space, denote by $\mathbf{C}(S)$ the class of bounded continuous functions from S into **R**, and define the *sup norm* of a function f in $\mathbf{C}(S)$:

$$|f| = \sup_S |f|. \tag{1.2}$$

The space $\mathbf{C}(S)$ is a complete metric space under the *sup norm metric*, in which the distance between two functions f and g is $|f-g|$.

If S is a locally compact but not compact separable metric space and $A_\bullet$ is a sequence of compact subsets of S, with union S define a *local sup norm* of a function f in $\mathbf{C}(S)$:

$$|f|_{\text{loc}} = \sum_n \sup_{A_n}(|f-g| \wedge 2^{-n}). \tag{1.3}$$

The space $\mathbf{C}(S)$ is a metric space under the *local sup norm metric*, in which the distance between two functions f and g is $|f-g|_{\text{loc}}$.

When S is not compact, the sup norm metric of $\mathbf{C}(S)$, for which convergence of a sequence of functions is uniform convergence, is sometimes less useful than the local sup norm metric, for which convergence of a sequence of functions is locally uniform convergence, that is, uniform convergence on every compact subset of S. In the local sup norm metric, the set $\{f \in \mathbf{C}(S): |f| \le c\}$ is a closed subset of $\mathbf{C}(S)$ for every positive c.

Separability of C(S). *The class* $\mathbf{C}(S)$ *on a compact metric space* S *is separable under the sup norm.* In fact, if $\delta(\bullet,\bullet)$ is the metric for S and $s_\bullet$ is a sequence in S, dense in S, then the class of rational coefficient polynomials in finitely many variables, with arguments $\{\delta(\bullet,s_n), n\ge 1\}$ form a countable algebra of continuous functions on S, separating S and containing the constant functions. Hence (Stone-Weierstrass theorem) this class is dense in $\mathbf{C}(S)$.

The class $\mathbf{C}_{00}(S)$. A continuous function f from a metric space S into **R** has *compact support* if the open subset $\{f \neq 0\}$ of S has compact closure. Denote by $\mathbf{C}_{00}(S)$ the class of these functions. Then $\mathbf{C}_{00}(S) \subset \mathbf{C}(S)$, and there is equality when S is compact.

The class $\mathbf{C}_0(S)$ and its separability. If S is a locally compact but not compact separable metric space, $\mathbf{C}_0(S)$ *is the class of continuous functions from* S *into* $\mathbf{R}$ *with limit* 0 *at infinity*. Then $\mathbf{C}_{00}(S) \subset \mathbf{C}_0(S) \subset \mathbf{C}(S)$. If S is a punctured compact metric space, with $S' = S\cup\{\infty\}$, extend each function in $\mathbf{C}_0(S)$ is to S' by setting the function equal to 0 at ∞. The resulting class is a closed subset of the separable metric space $\mathbf{C}(S')$ under the sup norm. Hence $\mathbf{C}_0(S)$ is separable in the sup norm metric, and this assertion must be true whenever S is a locally compact separable metric space.

Relations among $\mathbb{C}_{00}(S)$, $\mathbb{C}_0(S)$, and $\mathbb{C}(S)$. If S is a locally compact but not compact separable metric space, $\mathbf{C}_0(S)$ is a closed subset of $\mathbf{C}(S)$ in the sup norm metric. Furthermore, in this metric, $\mathbf{C}_{00}(S)$ *is dense in* $\mathbf{C}_0(S)$. To show this, it will be shown that if ε is strictly positive and $f \in \mathbf{C}_0(S)$, there is a function f_ε in $\mathbf{C}_{00}(S)$ for which $|f-f_\varepsilon| < \varepsilon$. Since $f=f\vee 0-[(-f)\vee 0)]$ it is sufficient to find f_ε for f positive. If f is positive, define $A = \{f\leq \varepsilon/2\}$, let $\delta(\bullet,A)$ be the distance of a point from A, and choose c to satisfy the inequality $c\delta(\bullet,A) \geq \max_S f$ on the compact set $\{f\geq \varepsilon\}$. Define $f_\varepsilon = f \wedge[c\delta(\bullet,A)]$, a function in $\mathbb{C}_{00}(S)$, and observe that the desired inequality $|f-f_\varepsilon| < \varepsilon$ is satisfied.

Sequential convergence of Radon measures. If S is a metric space, a sequence $\lambda_\bullet$ in $\mathbf{M}(S)$ is $\mathbb{C}_{00}(S)$ *convergent to a Radon measure* λ (also called *vaguely convergent* to λ) if $\lim\lambda_\bullet[f] = \lambda[f]$ for f in $\mathbb{C}_{00}(S)$. If S is compact, then $\mathbb{C}_{00}(S) = \mathbf{C}(S)$, and in this context $\mathbb{C}_{00}(S)$ convergence will be called $\mathbf{C}(S)$ *convergence.* If S is a not compact open subset of a compact space, a sequence $\lambda_\bullet$ in $\mathbf{M}(S)$ is $\mathbf{C}_0(S)$ *convergent* to a Radon measure λ if $\lim \lambda_\bullet[f] = \lambda[f]$ for f in $\mathbf{C}_0(S)$. In each context the limit measure λ is unique if a Radon measure λ is uniquely determined by $\lambda[f]$, as will be shown to be true in the contexts studied in this chapter.

2. Linear functionals on subsets of $\mathbb{C}(S)$

Let S be a topological space and $\mathbf{C}_1(S)$ be a linear subset of $\mathbf{C}(S)$, that is, a subset of $\mathbf{C}(S)$ containing the linear combinations of its members. The sup norm metric of $\mathbf{C}(S)$ will be used when boundedness of a linear functional is discussed.

Functionals on $\mathbf{C}_1(S)$. A function $\mathbf{L}$ from $\mathbf{C}_1(S)$ into $\mathbf{R}$ is honored by the name *functional*. A functional $\mathbf{L}$ is

positive if $f \geq 0$ implies that $\mathbf{L}(f) \geq 0$,
bounded if $|\mathbf{L}(f)| \leq const.|f|$,
linear if $\mathbf{L}(af+bg) = a\mathbf{L}(f) + b\mathbf{L}(g)$ for f, g in $\mathbf{C}_1(S)$ and a,b in $\mathbf{R}$

The following properties (a)-(g) of a functional $\mathbf{L}$ will be needed.

(a) *Positivity and linearity of* L *imply that* $L(f) \le L(g)$ *when* $f \le g$.

(b) L *is continuous if positive, linear, and bounded,* because then

$$|L(f) - L(g)| = |L(f - g)| \le L(|f - g|) \le const.\ |f - g|. \tag{2.1}$$

Conversely, L *is bounded if positive and continuous,* because if $|L(g)| \le 1$ when $|g| \le \eta$, it follows that for f not identically 0,

$$|L(f)| = (1/\eta)|f|)\ L(\eta f/|f|) \le (1/\eta)|f|. \tag{2.2}$$

(c) *If* $C_1(S)$ *includes the constant functions,* L *is bounded if positive and linear* because then

$$|L(f)| \le L(|f|) \le L(1)|f|. \tag{2.3}$$

(d) *If S is a locally compact but not compact separable metric space and* $C_1(S) = C_0(S)$, *then* L *is bounded if positive and linear.* To prove this, define

$$M = \sup\{L(f): f \in C_0(S), 0 \le f \le 1\}.$$

Unless this supremum is finite, to each strictly positive integer n there corresponds a function f_n in $C_0(S)$ for which $0 \le f_n \le 1$ and $L(f_n) \ge n^2$. Define

$$f = \sum_1^\infty n^{-2} f_n.$$

Since the series converges uniformly, $f \in C_0(S)$. Apply the positivity and linearity of L to obtain the inequality

$$L(f) \ge L\Big(\sum_1^k n^{-2} f_n\Big) \ge k$$

for all k. Since this inequality contradicts the finiteness of $L(f)$, it follows that M must be finite. Hence L is bounded, because for g in $C_0(S)$, and not vanishing identically,

$$|L(g)| \le L(|g|/|g|)\ |g| \le M\,|g|. \tag{2.4}$$

(e) The following example exhibits a positive linear unbounded functional on $C_{00}(S)$.

Example. Let S' be the interval $[0,1]$ on **R**, a compact metric space under the Euclidean metric, and delete the point 1 to obtain S. If λ is a Radon measure, the functional $L: f \to \lambda[f]$ on $C_{00}(S)$ is a positive linear functional, but is not bounded unless $\lambda(S)$ is finite.

(*f*) If S is a topological space, **L** is a positive linear functional on $\mathbf{C}_{00}(S)$, and S_0 is an open subset of S with compact closure, each function f in the class $\mathbf{C}_0(S_0)$ has an extension to a function in $\mathbf{C}_{00}(S)$, obtained by defining f as 0 on $S - S_0$. The restriction of **L** to these extensions defines a positive linear functional on $\mathbf{C}_0(S_0)$.

(g) *If S is a locally compact but not compact separable metric space, a bounded positive linear functional on* $\mathbf{C}_{00}(S)$ *can be extended uniquely to one on* $\mathbf{C}_0(S)$. To prove this, recall that $\mathbf{C}_{00}(S)$ is a dense (in the sup norm metric) subset of the closed subset $\mathbf{C}_0(S)$ of $\mathbf{C}(S)$, and therefore a bounded positive linear functional **L** on $\mathbf{C}_{00}(S)$ is a uniformly continuous function from a subset of a metric space into a subset of a complete metric space. Hence (Section 0.12(b)) **L** can be extended uniquely to be a positive linear functional on $\mathbf{C}_0(S)$.

3. Generation of positive linear functionals by measures (S compact metric)

If $\lambda \in \mathbf{M}(S)$, the functional **L** defined on $\mathbf{C}(S)$ by

$$\mathbf{L}(f) = \lambda[f] \tag{3.1}$$

is a positive linear functional, bounded according to Section 2 property (c), but here, more specifically, $|\mathbf{L}(f)| \le \lambda(S)\|f\|$. The following theorem asserts that all positive linear functionals on $\mathbf{C}(S)$ can be generated in this way. Thus measures on compact metric spaces can be defined indirectly, as positive linear functionals. (Actually this can be done on suitably restricted non- metric spaces, but doing so is beyond the scope of this book.)

Theorem (*S compact metric*). *If* **L** *is a positive linear functional on* $\mathbf{C}(S)$, *there is a unique measure* λ *in* $\mathbf{M}(S)$ *for which* (3.1) *is true.*

Proof. (a) There can be only one measure λ satisfying (3.1) because, under (3.1), if $f_\bullet$ is a decreasing sequence in $\mathbf{C}(S)$, with limit the indicator function of a compact subset F of S, then $\lim \mathbf{L}(f_\bullet) = \lambda(F)$. Thus two measures satisfying (3.1) are equal on the class of compact sets. Since the class of sets on which the measures are equal is a monotone class containing the compact (that is, closed, in the present context) sets, the measures are equal on $\mathbf{B}(S)$. (A sequence $f_\bullet$ with the properties used here is exhibited in part (c) of this proof.)

(b) Define set functions λ and λ^* by

$$\lambda(F) = \inf\{\mathbf{L}(f): f \in \mathbf{C}(S), f \ge \mathbf{1}_F\} \qquad (F \text{ compact}), \tag{3.2}$$

(3.3) $$\lambda(G) = \sup\{\lambda(F): F \text{ compact}, F \subset G\} \qquad (G \text{ open}),$$

(3.4) $$\lambda^*(A) = \inf\{\lambda(G): G \text{ open}, G \supset A\} \qquad (A \in 2^S).$$

If F is a set that is both open and compact, the value of $\lambda(F)$ from (3.3) is the same as that from (3.2). In particular, $\lambda(S) = \mathbf{L}(1)$. The set functions λ and λ^* are monotone increasing on their domains of definition. The set function λ^* is the outer measure generated by the class of open sets together with the set function λ on this class, and therefore the definition $d_{\lambda^*}(A,B) = \lambda^*(A\Delta B)$ yields a distance function on the class 2^S. The d_{λ^*} closure $\mathbf{A}$ of the class of compact sets is closed under finite unions and intersections, because the class of compact sets is closed under these operations. It will be shown that $\mathbf{A}$ is a σ algebra including $\mathbf{B}(S)$ and that the restriction of λ^* to $\mathbf{A}$ is a measure, satisfying (3.1) and equal to λ on the closed and open sets.

(c) *If F is a compact set, there are monotone decreasing sequences of continuous functions with limit $\mathbf{1}_F$, and if $f_\bullet$ is such a sequence,* $\lim \mathbf{L}(f_\cdot) = \lambda(F)$. For example, the sequence $\{\exp[(-n\delta(\bullet,F)], n \ge 1\}$, where $\delta(s,F)$ is the distance from the point s to the set F, is a monotone decreasing sequence of continuous functions with limit $\mathbf{1}_F$. Moreover, if $f_\bullet$ is any such sequence and f is a continuous function majorizing $\mathbf{1}_F$, then (Dini's theorem) the sequence $f_\bullet \vee f$ converges uniformly to f, that is, converges in the $\mathbf{C}(S)$ metric, and therefore

$$\lambda(F) \le \lim \mathbf{L}(f_\cdot) \le \lim \mathbf{L}(f_\cdot \vee f) = \mathbf{L}(f).$$

Then (c) is true because f can be chosen to make $\mathbf{L}(f)$ arbitrarily close to $\lambda(F)$.

(d) It is trivial that $\lambda = \lambda^*$ on the class of open sets. To see that *these two set functions are also equal on the class of compact sets*, let F be a compact set and $f_\bullet$ be a monotone decreasing sequence of continuous functions, with limit $\mathbf{1}_F$. If $g_n = f_n + 1/n$, the sequence $g_\bullet$ has these same properties, and the set $G_n = \{g_n > 1\}$ is an open superset of F. Furthermore $\lambda(G_n) \le \mathbf{L}(g_n)$ because g_n majorizes the indicator function of every compact subset of G_n. Hence

$$\lambda(F) \le \lambda^*(F) \le \lambda(G_n) \le \mathbf{L}(g_n) \downarrow \lambda(F),$$

and it follows that $\lambda(F) = \lambda^*(F)$.

(e) *The set function λ is finitely additive on the class of compact sets*. Let A and B be compact sets and $f_\bullet$ and $g_\bullet$ be decreasing sequences of continuous functions with respective limits $\mathbf{1}_A$ and $\mathbf{1}_B$. The equation

(3.5) $$\mathbf{L}(f_n + g_n) = \mathbf{L}(f_n \vee g_n) + \mathbf{L}(f_n \wedge g_n) = \mathbf{L}(f_n) + \mathbf{L}(g_n)$$

yields, when $n \to \infty$, the equation

$$\lambda(A \cup B) + \lambda(A \cap B) = \lambda(A) + \lambda(B), \tag{3.6}$$

which implies that λ is additive on the class of compact sets.

(f) *The class* **A** *contains the open sets.* It is sufficient to prove that if F is a compact subset of an open set G, then

$$\lambda(G) \geq \lambda(F) + \lambda(G - F), \tag{3.7}$$

because if (3.7) is applied to an increasing sequence $F_\bullet$ of compact subsets of G for which $\lim \lambda(F_\bullet) = \lambda(G)$, (3.7) implies that $\lim d_{\lambda^*}(G, F_n) = \lim \lambda(G - F_n) = 0$. To prove (3.7) observe that if F' is a compact subset of the open set $G - F$, then $\lambda(G) \geq \lambda(F \cup F') = \lambda(F) + \lambda(F')$, and this inequality implies (3.7).

(g) **A** *is an algebra.* All that remains to be proved is that $\tilde{\mathbf{A}} \subset \mathbf{A}$. Since the class **A** contains the open sets, **A** is the class of d_{λ^*} limits of the class of all sets that are either compact or open, a class closed under complementation. Therefore **A** is also closed under complementation.

(h) *The outer measure* λ^* *is finitely additive on* **A** because the additivity equation (3.6) is true for compact sets.

(i) *The outer measure* λ^* *is countably additive on* **A**, and **A** is a σ algebra, because on the one hand λ^* is countably subadditive and on the other hand, if A is a disjunct countable union $\cup A_\bullet$ of sets in **A**, $\lambda^*(A) \geq \Sigma\, \lambda^*(A_\bullet)$ because the inequality is true for the partial sums. This countable additivity shows that **A** is closed under countable unions, and is therefore a σ algebra including $\mathbf{B}(S)$.

(j) *Equation* (3.1) *is correct with* λ *extended to* $\mathbf{B}(S)$ *by* λ^*. Since $\lambda(S) = \mathbf{L}(1)$, it is sufficient to prove that (3.1) is true for f increased by a constant function, and therefore it is sufficient to prove (3.1) for strictly positive f, and it is even sufficient to prove that $\mathbf{L}(f) \leq \lambda[f]$ for strictly positive f, because this inequality can then be applied to the function $c\mathbf{1} - f$ for $c > \max_S f$. Fix f, supposed strictly positive, choose a to satisfy the inequality $0 < a < \min_S f$, choose $\varepsilon > 0$, and define the compact set $A_j = \{a + j\varepsilon \leq f \leq a + (j+1)\varepsilon\}$ for $j = 0, \ldots, k$, with k large enough to satisfy the inequality $a + (k+1)\varepsilon > \max f$. Suppose further that a has been chosen to make each set $\{f = a + j\varepsilon\}$ λ null; only countably many points of $\mathbf{R}$ have to be avoided to make this choice. Let f_j be a continuous function majorizing $\mathbf{1}_{A_j}$ and define $g = \Sigma_j \left[a + (j+1)\varepsilon\right] f_j$. Then

$$\mathbf{L}(f) \leq \mathbf{L}(g) = \Sigma_j \left[a + (j+1)\varepsilon\right] \mathbf{L}(f_j), \tag{3.8}$$

$$\lambda\Big[\Sigma_j \left[(a + (j+1)\varepsilon\right] \mathbf{1}_{A_j}\Big] \leq \lambda[f] + \varepsilon\lambda(S).$$

Since each function f_j can be chosen to make $\mathbf{L}(f_j)$ arbitrarily close to $\lambda(A_j)$, (3.8) implies

$$\mathbf{L}(f) \le \lambda[f] + \varepsilon\lambda(S), \tag{3.9}$$

and therefore $\mathbf{L}(f) \le \lambda[f]$ if f is strictly positive, as was to be proved.

4. $\mathbb{C}(S)$ convergence of sequences in $\mathbb{M}(S)$ (S compact metric)

$\mathbb{C}(S)$ sequential convergence of Radon measures was defined in Section 1. It will be seen below that there is a metric on $\mathbf{M}(S)$, consistent with $\mathbb{C}(S)$ sequential convergence, making $\mathbf{M}(S)$ complete and $\mathbf{M}_c(S)$ compact.

Example. Let S be the compact interval $[0,1]$ on $\mathbf{R}$, and define λ_n as the probability measure in $\mathbf{M}(S)$ supported by the singleton $\{1/n\}$. The sequence $\lambda_\bullet$ is $\mathbb{C}(S)$ convergent to the probability measure supported by the singleton $\{0\}$. Observe that each measure λ_n assigns the measure 1 to the open interval $(0,1)$, and 0 to the singleton $\{0\}$, but the limit measure assigns the value 0 to that interval and 1 to that singleton. Thus $\mathbb{C}(S)$ convergence on S does not imply $\mathbb{C}(S)$ convergence on the compact subset $\{0\}$ and does not imply setwise convergence of the sequence of measures, the convergence prescribed in the Vitali-Hahn-Saks theorem. According to Theorem 6, the latter convergence implies $\mathbb{C}(S)$ convergence when S is compact metric.

Theorem (S *compact metric*). *If* $\lambda_\bullet$ *is a sequence in* $\mathbb{M}(S)$ *for which the sequence* $\lambda_\bullet[f]$ *has a finite limit for f in a dense subset of* $\mathbb{C}(S)$, *then the sequence* $\lambda_\bullet$ *is* $\mathbb{C}(S)$ *convergent.*

Proof. Let $\mathbf{C}$ be a dense subset of $\mathbb{C}(S)$ on which the sequence $\lambda_\bullet$ has a finite limit. There is a function g in $\mathbf{C}'$ with $g > 1$. If $f \in \mathbb{C}(S)$ and if $\varepsilon > 0$, there is a function f_ε in $\mathbf{C}'$ with $|f - f_\varepsilon| < \varepsilon$. Then $f < f_\varepsilon + \varepsilon g$, and therefore

$$\limsup \lambda_\bullet[f] \le \lim \lambda_\bullet[f_\varepsilon] + \varepsilon \lim \lambda_\bullet[g]. \tag{4.1}$$

Apply this inequality to $-f$ to find that the difference between the limit superior in (4.1) and the limit inferior of the same sequence is at most $2\varepsilon \lim \lambda_\bullet[g]$ for every ε, and is therefore 0. Thus the sequence $\lambda_\bullet[f]$ has a finite limit for f in $\mathbb{C}(S)$. This limit defines a positive linear functional on $\mathbb{C}(S)$, necessarily of the form $f \to \lambda[f]$ for some measure λ, that is, there is $\mathbb{C}(S)$ convergence to λ.

5. Metrization of $\mathbb{M}(S)$ to match $\mathbb{C}(S)$ convergence; compactness of $\mathbb{M}_C(S)$ (S compact metric)

A metric for $\mathbf{R}^\infty$. Let $\mathbf{d}$ be a metric on the space $\mathbf{R}^\infty$ of infinite sequences of real numbers under which $\mathbf{R}^\infty$ is complete and under which convergence means coordinatewise convergence. For example, if ξ: $\xi_\bullet$ and η: $\eta_\bullet$ are points of $\mathbf{R}^\infty$, define

$$\mathbf{d}(\xi,\eta) = \sum_1^\infty 2^{-n} \wedge |\xi_n - \eta_n|. \tag{5.1}$$

A metric for $\mathbb{M}(S)$. Although $\mathbb{C}(S)$ sequential convergence was defined in Section 1, no corresponding metric was exhibited. A metric on $\mathbb{M}(S)$ consistent with the $\mathbb{C}(S)$ sequential convergence definition will now be defined. Let $h_\bullet$ be a dense subset of $\mathbb{C}(S)$. For each measure λ in $\mathbb{M}(S)$, the sequence $\lambda[h_\bullet]$ is a point of $\mathbf{R}^\infty$. If μ is a second measure in $\mathbb{M}(S)$, define the distance between the two measures by

$$d_{\mathbb{M}}(\lambda,\mu) = \mathbf{d}\big(\lambda[h_\bullet],\mu[h_\bullet]\big). \tag{5.2}$$

This definition satisfies the axioms for a metric.

Theorem (*S compact metric*).

(a) *A sequence in* $\mathbb{M}(S)$ *is* $d_{\mathbb{M}}$ *convergent if and only if the sequence is* $\mathbb{C}(S)$ *convergent.*

(b) $\mathbb{M}(S)$ *in the* $d_{\mathbb{M}}$ *metric is a complete metric space.*

(c) ($\mathbb{C}(S)$ *in the sup norm metric,* $\mathbb{M}(S)$ *in the* $d_{\mathbb{M}}$ *metric). The function* $(f,\lambda)\to\lambda[f]$ *from* $\mathbb{C}(S)\times\mathbb{M}(S)$ *into* $\mathbf{R}$ *is continuous.*

(d) *For each positive constant c,* $\mathbb{M}_C(S)$ *is a compact subset of* $\mathbb{M}(S)$, *as is the subset* $\{\lambda \in \mathbb{M}(S): \lambda(S) = c\}$.

In the following proof, $h_\bullet$ is the sequence in the definition of $d_{\mathbb{M}}$. The $d_{\mathbb{M}}$ metric depends on the choice of $h_\bullet$, but in view of this theorem, the $\mathbb{C}(S)$ topology, that is, the class of open sets, is independent of this choice.

Proof of (a). Under the $d_{\mathbb{M}}$ metric, a sequence $\lambda_\bullet$ in $\mathbb{M}(S)$ has limit λ in $\mathbb{M}(S)$ if and only if $\lim \lambda_\bullet[f] = \lambda[f]$ is satisfied for f in $h_\bullet$. Hence (Theorem 4), there is $\mathbb{C}(S)$ convergence if and only if there is convergence in the $d_{\mathbb{M}}$ metric.

Proof of (b). If $\lambda_\bullet$ is a $d_{\mathbb{M}}$ Cauchy sequence in $\mathbb{M}(S)$, the conditions of Theorem 4 are satisfied, with $h_\bullet$ the dense set of functions, and therefore the sequence $\lambda_\bullet$ is $\mathbb{C}(S)$ convergent.

Proof of (c). If $f_\bullet$ in $\mathbf{C}(S)$ is convergent in the sup norm sequence and $\lambda_\bullet$ in $\mathbf{M}(S)$ is a $d_\mathbf{M}$ convergent sequence, the following inequality makes (c) obvious:

(5.3) $|\lambda[f] - \lambda_n[f_n]| \le \lambda_n[|f-f_n|] + |\lambda[f] - \lambda_n[f]| \le |f-f_n|\lambda_n[1] + |\lambda[f] - \lambda_n[f]|.$

Proof of (d). If $\lambda_\bullet$ is a sequence in $\mathbf{M}_C(S)$, the sequence $\lambda_\bullet[f]$ is bounded for each f in $\mathbf{C}(S)$, and (Bolzano-Weierstrass theorem and the diagonal procedure) there is a subsequence of $\lambda_\bullet$ along which this sequence of integrals converges to a finite limit for every function h_n, and therefore this subsequence is $\mathbf{C}(S)$ convergent. The limit measure λ is in $\mathbf{M}_C(S)$, because

$$\lim \lambda_\bullet[1] = \lim \lambda_\bullet(S) = \lambda[1] = \lambda(S).$$

More generally, this argument shows that if A is a compact subset of $\mathbf{R}^+$, the set of measures $\{\lambda \in \mathbf{M}(S): \lambda(S) \in A\}$ is compact in the $d_\mathbf{M}$ metric.

6. Properties of the function $\mu \to \mu[f]$ from $\mathbf{M}(S)$, in the $d_\mathbf{M}$ metric, into $\mathbf{R}$ (S compact metric)

If f is in $\mathbf{C}(S)$, the function $\mu \to \mu[f]$ from $\mathbf{M}(S)$ into $\mathbf{R}$ is continuous under the $d_\mathbf{M}$ metric of $\mathbf{M}(S)$, because convergence in the $d_\mathbf{M}$ metric implies $\mathbf{C}(S)$ convergence. The next theorem treats this function for other choices of f. The following observation is made to clarify the hypotheses of the theorem.

Observation on semicontinuity. If f is a function from S into $\mathbf{R}$, denote by $f^\wedge$ the upper limit function of f: $f^\wedge(s) = f(s) \vee \limsup_{t\to s} f(t)$. This function is upper semicontinuous, majorizes f, and there is equality at a point if and only if f is upper semicontinuous at the point. Thus if f is upper semicontinuous λ almost everywhere, f coincides λ almost everywhere with the upper semicontinuous function $f^\wedge$. In the other direction, a function coinciding λ almost everywhere with an upper semicontinuous function need not be upper semicontinuous at any point. For example, if λ vanishes on singletons and f is defined as 0 except at the points of a countable dense set, at which f is defined as 1, f coincides with the continuous function $\mathbf{0}$ at λ almost every point although $f^\wedge$ is identically 1.

Theorem (*S compact metric, $d_\mathbf{M}$ metric of* $\mathbf{M}(S)$, $\lambda \in \mathbf{M}(S)$, *$f$ bounded and Borel measurable from S into* $\mathbf{R}$, $A \in \mathbf{B}(S)$).

(a) *If f is upper [lower] semicontinuous λ almost everywhere on S, the function $\mu \to \mu[f]$ from $\mathbf{M}(S)$ into $\mathbf{R}$ is upper [lower]semicontinuous at λ. If f is continuous λ almost everywhere on S, the function $\mu \to \mu[f]$ is continuous at λ.*

(b) *In particular, if $\lambda_\bullet$ is an infinite sequence in* $\mathbf{M}(S)$ *with* $\mathbf{C}(S)$ *limit (that is,*

$d_{\mathbf{M}}$ *limit)* λ:

(i) *if* $\lambda(\tilde{A}\cap\partial A) = 0$, *for example, if* A *is closed, then* $\limsup \lambda_\bullet(A) \le \lambda(A)$;

(ii) *if* $\lambda(A\cap\partial A) = 0$, *for example, if* A *is open, then* $\liminf \lambda_\bullet(A) \ge \lambda(A)$;

(iii) *if the conditions on* A *in* (i) *and* (ii) *are both satisfied, that is, if* $\lambda(\partial A) = 0$, *then* $\lim \lambda_\bullet(A) = \lambda(A)$.

(c) *Conversely, if* $\lambda \in \mathbf{M}(S)$, *if* $\lambda_\bullet$ *is an infinite sequence in* $\mathbf{M}(S)$ *and* $\lim \lambda_\bullet(A) = \lambda(A)$ *whenever* $\lambda(\partial A) = 0$, *then the sequence* $\lambda_\bullet$ *has* $\mathbf{C}(S)$ *limit* λ.

Proof of (a). Let $f^\wedge$ be the upper limit function of f. Then $f^\wedge$ is a bounded upper semicontinuous function and as such is the limit of a decreasing sequence $f_\bullet$ in $\mathbf{C}(S)$. The function $\mu \to \mu[f^\wedge]$ is the limit of the decreasing sequence $\{\mu \to \mu[f_n],\ n \ge 1\}$ of continuous functions from $\mathbf{M}(S)$ into $\mathbf{R}$, and is therefore upper semicontinuous. If f is upper semicontinuous at λ almost every point of S, then $f = f^\wedge$ at λ almost every point of S, and therefore

$$\lambda[f] = \lambda[f^\wedge] \ge \limsup_{\mu\to\lambda} \mu[f^\wedge] \ge \limsup_{\mu\to\lambda} \mu[f]. \tag{6.1}$$

This inequality is the condition that the function $\mu \to \mu[f]$ be upper semicontinuous at λ. Apply this result to $-f$ to obtain the corresponding result in the lower semicontinuous context, and combine these two results to obtain the last assertion in (a).

Proof of (b). Recall that $\lambda[\mathbf{1}_A]$ is defined as $\lambda(A)$. Assertions (i) and (ii) are applications of (a) and the fact that the indicator function of a closed [open] set is upper [lower] semicontinuous. Assertion (b)(iii) follows from (b)(i) and (b)(ii), or from (a).

Proof of (c). Choose $\varepsilon > 0$, f in $\mathbf{C}(S)$, and a with $a < \min_S f$ and define the compact set $A_j = \{a+j\varepsilon \le f \le a+(j+1)\varepsilon\}$ for $j = 0,\dots,k$, with k so large that $a+(k+1)\varepsilon > \max f$. The number a can be chosen in such a way that the sets $\{f = a+j\varepsilon\}$ are λ and λ_n null for all n and j, because this condition excludes only countably many values of a. Thus each boundary set ∂A_j is λ and λ_n null for all n. The integral of f over S with respect to either λ or λ_n is the sum of the integrals of f over $A_0, A_1,\dots$, and therefore

$$\left|\lambda[f] - \Sigma_j (a+j\varepsilon)\lambda(A_j)\right| \le \varepsilon\lambda(S), \quad \left|\lambda_n[f] - \Sigma_j (a+j\varepsilon)\lambda_n(A_j)\right| \le \varepsilon\lambda_n(S). \tag{6.2}$$

Under condition (b)(iii), when $n \to +\infty$ the right side of the second inequality tends to the right side of the first, and the sum in the second inequality tends to the sum in the first. It follows that

$$\left|\lambda[f] - \lambda_n[f]\right| \le 3\varepsilon\lambda(S) \tag{6.3}$$

for large n, and therefore the sequence $\lambda_\bullet$ has $\mathbf{C}(S)$ limit λ.

Observation. Theorem 6 implies that if $\lambda_\bullet$ is a $\mathbf{C}(S)$ convergent sequence of measures in $\mathbf{M}(S)$, with limit measure λ, and if S_0 is a compact subset of S with λ null boundary, then the sequence of restrictions of $\lambda_\bullet$ to $\mathbf{B}(S_0)$ is a sequence in $\mathbf{M}(S_0)$ with $\mathbf{C}(S_0)$ limit the restriction of λ to $\mathbf{B}(S_0)$.

7. Generation of positive linear functionals on $\mathbb{C}_0(S)$ by measures (S a locally compact but not compact separable metric space)

The following theorem is the adaptation of Theorem 3 to linear functionals on $\mathbb{C}_0(S)$ and $\mathbb{C}_{00}(S)$.

Theorem (*S a locally compact but not compact separable metric space, $\mathbb{C}_{00}(S)$ in the sup norm metric*). *If* $\mathbf{L}$ *is a positive linear functional on* $\mathbb{C}_0(S)$ [*a positive bounded linear functional on* $\mathbb{C}_{00}(S)$], *there is a unique measure* λ *in* $\mathbf{M}(S)$ *for which* $\mathbf{L}(f) = \lambda[f]$.

Proof. It can be supposed that S is a punctured compact metric space. Since a positive linear functional on $\mathbb{C}_0(S)$ (under the sup norm) is bounded (Section 2 property (d)), and since (Section 2 property (g)), a positive bounded linear functional on $\mathbb{C}_{00}(S)$ can be extended uniquely into a positive linear functional on $\mathbb{C}_0(S)$, it is sufficient to prove the part of the theorem for functionals on $\mathbb{C}_0(S)$. If g' is a function on $S' = S \cup \{\infty\}$, denote by g'_S the restriction of g' to S. Define a functional $\mathbf{L}'$ on $\mathbb{C}(S')$ by

$$(7.1) \qquad \mathbf{L}'(f') = \mathbf{L}\big([f'-f'(\infty)]_S\big) + f'(\infty)M,$$

where M is at least as large as the constant M in (2.4). This functional $\mathbf{L}'$ is obviously linear, and is positive because, under posiivity of f',

$$\mathbf{L}\big([f'-f'(\infty)]_S\big) \geq -\mathbf{L}\big([f'(\infty)-f']\vee 0]_S\big) \geq -f'(\infty)M.$$

According to Theorem 3 there is a unique measure λ' in $\mathbf{M}(S')$ for which $\mathbf{L}'(f') = \lambda'[f']$, and therefore if λ is the restriction of λ' to $\mathbf{B}(S)$ and $f \in \mathbb{C}_0(S)$, it follows that $\mathbf{L}(f) = \lambda[f]$ and that λ is a uniquely defined measure generating $\mathbf{L}$.

Unbounded positive linear functionals on $\mathbb{C}_{00}(S)$. Suppose that $\mathbf{L}$ is a positive and linear but not bounded functional on $\mathbb{C}_{00}(S)$. Let S_0 be an open subset of S, with compact closure. As pointed out in Section 2 under property (f), $\mathbf{L}$ defines a positive linear functional $\mathbf{L}(S_0,\bullet)$ on $\mathbb{C}_0(S_0)$, with $\mathbf{L}(S_0,f_0)$ equal to the value of $\mathbf{L}$ for the extension of f_0 with value 0 on $S-S_0$. It follows that there is a measure $\lambda(S_0,\bullet)$ in $\mathbf{M}(S_0)$ generating $\mathbf{L}(S_0,\bullet)$: $\mathbf{L}(S_0,f_0) = \lambda[S_0,f_0]$.

Moreover if S_1 is an open subset of S_0, then $\mathbf{L}(S_1,\bullet) = \mathbf{L}(S_0,\bullet)$ on $\mathbf{C}_0(S_1)$ in the sense that if $f_1 \in \mathbf{C}_0(S_1)$ then $\mathbf{L}(S_1,f)$ is $\mathbf{L}(S_0,\bullet)$ evaluated at the extension of f_1 by 0 to S_0. Hence $\lambda[S_0,\bullet] = \lambda[S_1,\bullet]$ on subsets of S_1. Thus there is a Radon measure λ on $\mathbf{B}(S)$ for which $\mathbf{L}(f) = \lambda[f]$ for f in $\mathbf{C}_{00}(S)$. Since $\mathbf{L}$ is not bounded, $\lambda(S) = +\infty$.

8. $\mathbb{C}_0(S)$ and $\mathbb{C}_{00}(S)$ convergence of sequences in $\mathbb{M}(S)$ (S a locally compact but not compact separable metric space)

As the following example shows, if $\lambda_\bullet$ is a $\mathbf{C}(S_{00})$ convergent sequence in $\mathbf{M}(S)$, it is not necessarily true that $\sup \lambda_\bullet(S)$ is finite.

Example. Let S be the interval $(0,1)$ and λ_n be the measure supported by the singleton $\{1/n\}$, with $\lambda_n(\{1/n\}) = n$. Then the sequence $\lambda_\bullet$ has $\mathbf{C}(S_{00})$ limit the identically vanishing measure, but the sequence is not $\mathbf{C}(S_0)$ convergent, and in fact such a sequence cannot be $\mathbf{C}(S_0)$ convergent according to the next theorem.

Theorem *(S a locally compact but not compact separable metric space, with metric δ, $\mathbf{C}_0(S)$ and $\mathbf{C}_{00}(S)$ in the sup norm metric, and $\lambda_\bullet$ a sequence in $\mathbf{M}(S)$).*

(a) *If* $\lim \lambda_\bullet[f]$ *exists and is finite for f in $\mathbf{C}_0(S)$ then* $\sup \lambda_\bullet(S) < +\infty$, *and the sequence $\lambda_\bullet$ is $\mathbf{C}_0(S)$ and $\mathbf{C}_{00}(S)$ convergent to a measure λ in $\mathbf{M}(S)$.*

(b) *If* $\sup \lambda_\bullet(S) < +\infty$, *and* $\lim \lambda_\bullet[f]$ *exists (finite) for f in a dense subset of $\mathbf{C}_0(S)$ (for example, for f in $\mathbf{C}_{00}(S)$), then the sequence $\lambda_\bullet$ is $\mathbf{C}_0(S)$ convergent.*

(c) *If S is a punctured compact metric space, with $S' = S \cup \{\infty\}$, let μ' be the extension of a measure μ in $\mathbf{M}(S)$ to a measure in $\mathbf{M}(S')$ obtained by assigning a value to $\mu'(\{\infty\})$. Then $\lambda_\bullet$ is $\mathbf{C}_0(S)$ convergent to λ if and only if there are choices of $\lambda_\bullet'(\{\infty\})$ and $\lambda'(\{\infty\})$ for which $\lambda_\bullet'$ is $\mathbf{C}(S')$ convergent to λ'.*

Proof of (a), It is trivial that $\mathbf{C}_0(S)$ convergence implies $\mathbf{C}_{00}(S)$ convergence. The positive linear functional $\mathbf{L}$ defined on $\mathbf{C}_0$ by $\mathbf{L}(f) = \lim \lambda_\bullet[f]$ is generated by a measure in $\mathbf{M}(S)$, according to Theorem 7, and this measure is the $\mathbf{C}_0(S)$ limit of $\lambda_\bullet$. If the sequence $\lambda_\bullet$ is $\mathbf{C}_0(S)$ convergent to λ, but if the sequence $\lambda_\bullet(S)$ is not bounded, there is a subsequence $\lambda_{a_\bullet}$ of $\lambda_\bullet$ for which $\lambda_{a_n}(S) \geq n^2$. The functional

$$f \to \sum_{n=1}^{\infty} n^{-2} \lambda_{a_n}[f]$$

is positive and linear on $\mathbf{C}_0(S)$ and as such is generated by some finite measure μ, that is,

$$(8.1)\qquad \mu[f] = \sum_{n=1}^{\infty} n^{-2}\lambda_{a_n}[f] \qquad (f \in \mathbf{C}_0(S)).$$

The function $s \to f_k(s) = 1 - \exp[-k\delta(s,\infty)]$ on S is in $\mathbf{C}_0(S)$. The fact that the sequence $f_\bullet$ is a monotone increasing positive sequence on S, with limit the function **1**, yields the impossible inequality

$$(8.2)\qquad \mu(S) = \lim \mu[f_\bullet] = \lim \sum_{n=1}^{\infty} n^{-2}\lambda_{a_n}[f_\bullet] \ge \sum_{n=1}^{\infty} n^{-2}\lambda_{a_n}(S) = +\infty,$$

and therefore the sequence $\lambda_\bullet(S)$ is bounded.

Proof of (b). The proof is parallel to that of Theorem 4. Let $\mathbf{C}_0'(S)$ be a dense subset of $\mathbf{C}_0(S)$ for which the sequence $\lambda_\bullet$ has a finite limit. If $f \in \mathbf{C}_0(S)$ and $\varepsilon > 0$, there is a function f_ε in $\mathbf{C}_0'(S)$ at distance $< \varepsilon$ from f. Hence

$$(8.3)\qquad \limsup \lambda_\bullet[f] \le \lim \lambda_\bullet[f_\varepsilon] + \varepsilon \sup \lambda_\bullet(S).$$

The rest of the proof of (b) is the same as the corresponding part of the proof of Theorem 4.

Proof of (c). If the sequence $\lambda_\bullet$ has $\mathbf{C}_0(S)$ limit λ, and $\sup\lambda_\bullet(S) = \alpha$, extend λ_n to a measure λ_n' in $\mathbf{M}(S)$ by defining $\lambda_n'(\{\infty\}) = \alpha - \lambda_n(S)$, so that $\lambda_n'(S') = \alpha$. Let f' be a function in $\mathbf{C}(S')$, and let $[f'-f'(\infty)]_S$ be the restriction of $f'-f'(\infty)$ to S. This restriction is in $\mathbf{C}_0(S)$ and

$$(8.4)\qquad \lim \lambda_\bullet'[\mathbf{f}'] = \lim \lambda_\bullet\big[[f'-f'(\infty)]_S\big] + f'(\infty)\alpha = \lambda\big[[f'-f'(\infty)]_S\big] + f'(\infty)\alpha.$$

It follows that the sequence $\lambda_\bullet'$ is $\mathbf{C}(S')$ convergent with limit λ', where $\lambda'(\{\infty\}) = \alpha - \lambda(S)$. Conversely, if there are extensions $\lambda_\bullet'$ of $\lambda_\bullet$ and λ' of λ for which the sequence $\lambda_\bullet'$ is $\mathbf{C}(S')$ convergent to λ' in $\mathbf{M}(S')$, and if f is in $\mathbf{C}_0(S)$, define f' on S' as f on S and as 0 at ∞. Then the sequence $\lambda_\bullet'[f'] = \lambda_\bullet(f)$ has limit $\lambda'[f'] = \lambda[f]$. Hence $\lambda_\bullet$ is $\mathbf{C}_0(S)$ convergent.

9. Metrization of $\mathbb{M}(S)$ to match $\mathbb{C}_0(S)$ convergence; compactness of $\mathbb{M}_c(S)$ (S a locally compact but not compact separable metric space, c a strictly positive number)

If $h_{0\bullet}$ is a sequence dense in $\mathbf{C}_0(S)$ under the sup norm metric, define the $\mathbf{C}_0(S)$ distance between measures λ and μ in $\mathbf{M}(S)$ by

$$(9.1)\qquad d_{0\mathbf{M}_c}(\lambda,\mu) = \mathbf{d}\big(\lambda(h_{0\bullet}),\mu(h_{0\bullet})\big),$$

as suggested by the distance definition (5.2) between measures on a compact metric space. In the present context, Theorem 5 takes the following form.

Theorem *(S a locally compact but not compact separable metric space, c a strictly positive number).*

(a) *A sequence in* $\mathbf{M}_c(S)$ *is* $d_{0\mathbf{M}}$ *convergent if and only if the sequence is* $\mathbf{C}_0(S)$ *convergent.*
(b) $\mathbf{M}_c(S)$ *in the* $d_{0\mathbf{M}}$ *metric is a compact metric space.*
(c) *($\mathbf{C}(S)$ in the sup norm metric, $\mathbf{M}_c(S)$ in the $d_{0\mathbf{M}_c}$ metric) The function* $(f,\lambda)\to\lambda[f]$ *from* $\mathbf{C}(S)\times\mathbf{M}_c(S)$ *into* $\mathbf{R}$ *is continuous.*

The proof of this theorem is a mild modification of the proof of Theorem 5 and is left to the reader. Although $\mathbf{M}_c(S)$ is compact in this metric, the following example shows that the set $\{\lambda \in \mathbf{M}(S)\colon \lambda(S) = c\}$ is not compact. The point is (see the next section), that when $\lambda_\bullet$ is a $\mathbf{C}_0(S)$ convergent sequence in $\mathbf{M}_c(S)$, with limit λ, then $\lambda(S) \le \liminf \lambda_\bullet(S)$ and there may be strict inequality.

Example. If S is the interval $(0,1)$ on $\mathbf{R}$ and λ_n is the probability measure supported by the singleton $\{1/n\}$, the sequence $\lambda_\bullet$ is $\mathbf{C}_0(S)$ convergent to the identically 0 measure.

10. Properties of the function $\mu\to\mu[f]$, from $\mathbf{M}(S)$ in the $d_{0\mathbf{M}}$ metric into $\mathbf{R}$ (S a locally compact but not compact separable metric space)

The following theorem is the adaptation of Theorem 6 to the present context.

Theorem *(S a locally compact but not compact separable metric space, $d_{0\mathbf{M}}$ metric of $\mathbf{M}_c(S)$, $\lambda \in \mathbf{M}_c(S)$, f bounded and Borel measurable from S into $\mathbf{R}$, $A \in \mathbf{B}(S)$).*

(a) *If f is upper [lower] semicontinuous λ almost everywhere on S, with limit superior ≤ 0 [limit inferior ≥ 0] at infinity, the function $\mu\to\mu[f]$ from $\mathbf{M}_c(S)$ into $\mathbf{R}$ is upper [lower] semicontinuous at λ. If f is continuous λ almost everywhere on S, with limit 0 at infinity, the function $\mu\to\mu[f]$ is continuous at λ.*

(b) *In particular, if $\lambda_\bullet$ is an infinite sequence in $\mathbf{M}_c(S)$, with $\mathbf{C}_0(S)$ limit λ,*
(i) *if $\lambda(\bar{A}\cap\partial A) = 0$ and A has compact closure, for example, if A is compact, then* $\limsup \lambda_\bullet(A) \le \lambda(A)$,
(ii) *if $\lambda(A\cap\partial A) = 0$, for example, if A is open, then* $\liminf \lambda_\bullet(A) \ge \lambda(A)$,
(iii) *if the conditions on A in* (i) *and* (ii) *are both satisfied, that is, if A has compact closure and $\lambda(\partial A) = 0$, then* $\lim \lambda_\bullet(A) = \lambda(A)$.

(c) *Conversely, if* $\lambda \in \mathbf{M}(S)$, *and if* $\lambda_\bullet$ *is a sequence in* $\mathbf{M}(S)$, *for which* $\sup \lambda_\bullet(S) < +\infty$ *and* $\lim \lambda_\bullet(A) = \lambda(A)$ *whenever* $\lambda(\partial A) = 0$ *and A has compact closure, then the sequence* $\lambda_\bullet$ *has* $\mathbf{C}_0(S)$ *limit* λ.

The proof follows that of Theorem 6. Alternatively, the theorem can be reduced to Theorem 6.

Observation. Theorem 10 implies that if $\lambda_\bullet$ is a $\mathbf{C}_0(S)$ convergent sequence of measures in $\mathbf{M}(S)$ with limit measure λ and S_0 is a compact subset of S with λ null boundary, then the sequence of restrictions of $\lambda_\bullet$ to $\mathbf{B}(S_0)$ is a sequence in $\mathbf{M}(S_0)$ with $\mathbf{C}(S_0)$ limit the restriction of λ to $\mathbf{B}(S_0)$.

11. Stable $\mathbb{C}_0(S)$ convergence of sequences in $\mathbb{M}(S)$ (S a locally compact but not compact separable metric space)

Let $\lambda_\bullet$ be a sequence in $\mathbf{M}(S)$ with $\mathbf{C}_0(S)$ limit λ in $\mathbf{M}(S)$. The sequence is *stably* $\mathbf{C}_0(S)$ convergent to λ if $\lim \lambda_\bullet(S) = \lambda(S)$. The point of this strengthening of $\mathbf{C}_0(S)$ convergence is that the sequence of measures is not allowed to unload measure at infinity.

By definition, $\lambda_\bullet$ is stably $\mathbf{C}_0(S)$ convergent to λ if and only if $\lim\lambda_\bullet[f] = \lambda[f]$, whenever f is in $\mathbf{C}(S)$ and is either identically constant or has limit 0 at infinity; in other words if and only if $\lim\lambda_\bullet[f] = \lambda[f]$ whenever f is continous and has a finite limit at infinity. If S is a punctured compact metric space, this condition has an elegant formulation: $\lim\lambda_\bullet[f] = \lambda[f]$ whenever f can be defined at the point ∞ of S' to become a member of $\mathbf{C}(S')$. A trivial computation shows that $\lambda_\bullet$ converges to λ in this sense if and only if, when λ_n and λ are extended to measures λ_n' and λ' in $\mathbf{M}(S')$, by assigning the measure value 0 to the singleton $\{\infty\}$, it follows that the sequence $\lambda_\bullet'$ is $\mathbf{C}(S')$ convergent to λ'. Instead of the measure value 0, an arbitrary positive measure value can be used.

12. Metrization of $\mathbb{M}(S)$ to match stable $\mathbb{C}_0(S)$ convergence (S a locally compact but not compact separable metric space)

It can be assumed that S is a punctured compact metric space, with $S' = S\cup\{+\infty\}$. The space S' is a compact metric space in which the point ∞ plays a special role. Define a distance on $\mathbf{M}(S')$ as in Section 5 in terms of a dense sequence in $\mathbf{C}(S')$ but adjoin the function $\mathbf{1}_{\{\infty\}}$ to this dense sequence. In other words a sequence $\lambda_\bullet'$ of measures in $\mathbf{M}(S')$ is convergent to λ' in the sense of this distance if and only if there is $\mathbf{C}(S')$ convergence and $\lim \lambda_\bullet'(\{\infty\}) =$

$\lambda(\{\infty\})$. Under this metric the space $\mathbf{M}(S')$ of measures is complete, and the subset of these measures for which $\{\infty\}$ is null is a closed subset. If $\lambda \in \mathbf{M}(S)$, extend λ to λ' in $\mathbf{M}(S')$ by defining $\lambda'(\{\infty\}) = 0$ and define the distance $d_{\mathbf{M}}'$ between two measures in $\mathbf{M}(S)$ as the distance just defined between the corresponding primed measures in $\mathbf{M}(S')$. Under this metric, $\mathbf{M}(S)$ is a complete metric space for which convergence is stable $\mathbf{C}_0(S)$ convergence. As the Example in Section 9 showed, in general the $d_{\mathbf{M}}'$ metric does not make the space $\mathbf{M}_c(S)$ compact. The following theorem restates the fact that the $d_{\mathbf{M}}'$ metric is adapted to stable $\mathbf{C}_0(S)$ convergence and includes other properties of this type of convergence.

Theorem (*S a locally compact but not compact separable metric space, $\lambda_\bullet$ a sequence in* $\mathbf{M}(S)$, *c a positive constant*).

(a) *If the sequence $\lambda_\bullet$ is $\mathbf{C}_0(S)$ convergent to a measure in $\mathbf{M}(S)$, this convergence is stable $\mathbf{C}_0(S)$ if and only if to every strictly positive ε there corresponds a compact subset A_ε of S, with the property that* $\sup \lambda_\bullet(\tilde{A}_\varepsilon) < \varepsilon$.

(b) *A sequence in $\mathbf{M}(S)$ is $d_{\mathbf{M}}'$ convergent if and only if the sequence is stably $\mathbf{C}_0(S)$ convergent; $\mathbf{M}(S)$ in the $d_{\mathbf{M}}'$ metric is a complete metric space.*

(c) *The function $(f,\lambda) \rightarrow \lambda[f]$ from $\mathbf{C}(S)\times\mathbf{M}(S)$ ($\mathbf{C}(S)$ in the local sup norm metric, $\mathbf{M}(S)$ in the $d_{\mathbf{M}}'$ metric) into $\mathbf{R}$ is continuous on the set $\{|f| \le c\}\times\mathbf{M}(S)$. In particular, $\lim \lambda_\bullet[f] = \lambda[f]$ when $\lambda_\bullet$ is stably $\mathbf{C}_0(S)$ convergent to λ and f is in $\mathbf{C}(S)$.*

Proof of (a). Without loss of generality, it can be assumed that S is a punctured compact metric space. Extend $\lambda_\bullet$ and λ to measures $\lambda_\bullet'$ and λ' on $\mathbf{B}(S')$ by defining the extended measures to be 0 on the singleton $\{\infty\}$. It was noted above that the sequence $\lambda_\bullet'$ is $\mathbf{C}(S')$ convergent to λ' if and only if $\lambda_\bullet$ is stably $\mathbf{C}_0(S)$ convergent to λ. If there is $\mathbf{C}(S')$ convergence to λ', and if ε is strictly positive, choose a ball B in S', with a λ null boundary, with center the point ∞, radius so small that $\lambda(B) < \varepsilon$. Only countably many radius values are exceptional for the λ null property of the ball boundary. Then $\lim \lambda_\bullet(B) = \lambda(B)$. Hence $\lambda_n(B) < \varepsilon$ for all but finitely many values of n, and the radius of B can be decreased further, if necessary to make this inequality valid for all values of n. The trace on S of the ball obtained in this way is the complement of the desired set A_ε. Conversely, if to every strictly positive ε there is a compact subset A_ε of S with the stated properties, it can be assumed, increasing the set, if necessary, that $\lambda(\tilde{A}_\varepsilon) < \varepsilon$ and ∂A_ε $(=\partial\tilde{A}_\varepsilon)$ is λ null. (For example, the open set $\tilde{A}_\varepsilon$ can be taken as the trace on S of a sufficiently small ball in S', with center the point ∞.) Then $\lim \lambda_\bullet(S) = \lambda(S)$ because

$$\limsup |\lambda(S)-\lambda_\bullet(S)| \le \lambda(\tilde{A}_\varepsilon) + \limsup |\lambda(A_\varepsilon)-\lambda_\bullet(A_\varepsilon)| \tag{12.1}$$

$$+ \limsup \lambda_\bullet(\tilde{A}_\varepsilon) < 2\varepsilon.$$

Proof of (b). The proof was given at the beginning of this section.

Proof of (c). Suppose that $f_\bullet$ is a uniformly bounded sequence of functions in $\mathbf{C}(S)$, convergent to f in the local sup norm metric, and that $\lambda_\bullet$ is a stably $\mathbf{C}_0(S)$ convergent sequence in $\mathbf{M}(S)$, with limit λ. Choose A_ε with the properties stated in (a), increasing this set, if necessary to satisfy the conditions $\lambda[f1_{\tilde{A}_\varepsilon}] < \varepsilon$, $\lambda(\partial A_\varepsilon) = 0$. Define $b = \sup|f_\bullet|$. According to Theorems 10 and 6, the sequence of restrictions of $\lambda_\bullet$ to $\mathbf{B}(A_\varepsilon)$ is $\mathbf{C}(A_\varepsilon)$ convergent to the restriction of λ to $\mathbf{B}(A_\varepsilon)$. Furthermore, the sequence of restrictions to A_ε of $f_\bullet$ is convergent in the sup norm topology of $\mathbf{C}(A_\varepsilon)$ to the restriction of f to A_ε. It follows, according to Theorem 6, that $\lim \lambda_\bullet[f_\bullet 1_{A_\varepsilon}] = \lambda[f1_{A_\varepsilon}]$, and therefore

$$\limsup |\lambda(f) - \lambda_\bullet(f_\bullet)| \le \limsup \lambda_n[f_n 1_{\tilde{A}_\varepsilon}] + \lambda[f1_{\tilde{A}_\varepsilon}] \le (c+1)\varepsilon.$$

Hence (c) is true.

13. Properties of the function $\mu \rightarrow \mu[f]$, from $\mathbf{M}(S)$ in the $d_{\mathbf{M}}'$ metric into $\mathbf{R}$ (S a locally compact but not compact separable metric space)

The following theorem is the adaptation of Theorem 10 to the present context.

Theorem (*S a locally compact but not compact separable metric space, $d_{\mathbf{M}}'$ metric of $\mathbf{M}(S)$, $\lambda \in \mathbf{M}(S)$, f bounded and Borel measurable from S into $\mathbf{R}$, $A \in \mathbf{B}(S)$*).

(a) *If f is upper [lower] semicontinuous λ almost everywhere on S, the function $\mu \rightarrow \mu(f)$ from $\mathbf{M}(S)$ into $\mathbf{R}$ is upper [lower] semicontinuous at λ. If f is continuous λ almost everywhere on S, then the function $\mu \rightarrow \mu[f]$ is continuous at λ.*

(b) *In particular, if $\lambda_\bullet$ is an infinite sequence in $\mathbf{M}(S)$, with stable $\mathbf{C}_0(S)$ limit λ,*
  (i) *if $\lambda(\tilde{A} \cap \partial A) = 0$, for example, if A is closed, then* $\limsup \lambda_\bullet(A) \le \lambda(A)$,
  (ii) *if $\lambda(A \cap \partial A) = 0$, for example, if A is open, then* $\liminf \lambda_\bullet(A) \ge \lambda(A)$,
  (iii) *if the conditions on A in* (i) *and* (ii) *are both satisfied, that is, if $\lambda(\partial A) = 0$, then* $\lim \lambda_\bullet(A) = \lambda(A)$.

(c) *Conversely, if $\lambda \in \mathbf{M}(S)$, if $\lambda_\bullet$ is a sequence in $\mathbf{M}(S)$ for which* $\lim \lambda_\bullet(A) = \lambda(A)$ *whenever $\lambda(\partial A) = 0$, then the sequence $\lambda_\bullet$ has stable $\mathbf{C}_0(S)$ limit λ.*

The proof follows that of Theorem 6; alternatively this theorem can be deduced from Theorem 6.

14. Application to analytic and harmonic functions

A function defined on an open plane set is *harmonic* if it has continuous second partial derivatives and satisfies Laplace's equation. The real part of an analytic function is harmonic, and conversely a harmonic function on a simply connected domain is necessarily the real part of an analytic function. As an example of the application of sequential convergence of measures in classical analysis the Riesz-Herglotz representation of a positive harmonic function on a disk in terms of a measure on the disk boundary will be derived.

All disks considered have the origin as center. Let B_α be the open disk of radius α, choose $\beta > 0$, and let u be a harmonic function with domain B_β. Then $u = \Re f$ for some function f analytic on B_β, given by a power series

$$f(z) = \Sigma_0^\infty a_n z^n = \Sigma_0^\infty a_n r^n e^{ni\theta} \qquad (z = re^{i\theta}, r < \beta). \tag{14.1}$$

If $0 < \alpha < \beta$, there is uniform convergence when $r = \alpha$, and therefore

$$\int_0^{2\pi} f(\alpha e^{is})\, e^{-nis}\, l(ds) \begin{cases} = 2\pi a_n \alpha^{n+1} & (n \geq 0), \\ = 0 & (n < 0). \end{cases} \tag{14.2}$$

Here $l(ds)$ refers to length (Lebesgue measure) on the interval $[0,2\pi]$, and the integral can be evaluated as a Riemann integral. Aside from a multiplicative constant, the evaluations in (14.2) are the evaluations of the (trigonometric) Fourier coefficients of the restriction of f to ∂B_α. The evaluations (14.1) and (14.2) yield, for $z = re^{i\theta}$ with $r < \alpha < \beta$,

$$\begin{aligned} f(z) &= \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \int_0^{2\pi} f(\alpha e^{is})(r/\alpha)^{|n|} e^{ni(s-\theta)}\, l(ds) \\ &= \frac{1}{2\pi\alpha} \int_{\partial B_a} f(\zeta) \frac{\alpha^2 - |z|^2}{|\zeta - z|^2}\, l_\alpha(d\zeta), \end{aligned} \tag{14.3}$$

where the measure l_α is length on ∂B_α. This expression for f can also be obtained using the Cauchy integral formula. The harmonic function $u = \Re f$ is therefore given by the corresponding integral

$$u(z) = \frac{1}{2\pi\alpha} \int_{\partial B_a} u(\zeta) \frac{\alpha^2 - |z|^2}{|\zeta - z|^2}\, l_\alpha(d\zeta) \qquad (|z| < \alpha < \beta). \tag{14.4}$$

This representation of u in B_α in terms of the values of u on ∂B_α is the *Poisson integral representation* of u.

Suppose now that u is positive and harmonic in B_β. The Riesz-Herglotz representation of u generalizes (14.4) by providing a *Poisson-Stieltjes representation* of u in B_β in terms of a measure on ∂B_β even though u is not defined on the boundary. To derive this representation, put (14.4) in a slightly different form:

$$u(z) = \int_{\partial B_\alpha} \frac{\alpha^2-|z|^2}{|\zeta-z|^2}\,\lambda_\alpha(d\zeta) \qquad (|z| < \alpha < \beta) \tag{14.5}$$

where λ_α is the measure of Borel subsets of ∂B_α defined by

$$\lambda_\alpha(A) = \frac{1}{2\pi\alpha}\int_A u(\zeta)\, l(_\alpha(d\zeta). \tag{14.6}$$

Then $\lambda_\alpha(\partial B_\alpha) = u(0)$. The measure λ_α can be thought of as a measure on $\bar{B}_\beta$ carried by ∂B_α. When α tends to β along an increasing sequence, the corresponding sequence $\lambda_{\alpha\bullet}$ of measures is a bounded sequence of measures on the compact space $\bar{B}_\beta$, and therefore there is a $\mathbf{C}(\bar{B}_\beta)$ convergent subsequence with limit measure λ carried by ∂B_β. In view of Theorem 5(c),

$$u(z) = \int_{\partial B_\beta} \frac{\beta^2-|z|^2}{|\zeta-z|^2}\,\lambda(d\zeta) \qquad (|z| < \beta). \tag{14.7}$$

This is the Riesz-Herglotz representation of u in terms of a measure on ∂B_β. In a more thorough discussion it is shown that λ is uniquely determined by u and that λ_α is $\mathbf{C}(S)$ convergent to λ when α tends to β.

IX
Signed Measures

1. Range of values of a signed measure

Signed measures, defined in Section III.1, have values either in $(-\infty,+\infty]$ or $[-\infty,+\infty)$, to avoid the possibility of adding $+\infty$ to $-\infty$. It will be shown in Section 2 that a signed measure is actually bounded on the side where it is finite. For a signed measure space $(S,\mathbf{S},\lambda)$, the signed measure λ has its values in $(-\infty,+\infty]$ if and only if $\lambda(S) > -\infty$, its values in $[-\infty,+\infty)$ if and only if $\lambda(S) < +\infty$, and λ is finite valued if and only if $\lambda(S)$ is finite.

2. Positive and negative components of a signed measure

If $(S,\mathbf{S},\lambda)$ is a signed measure space and A is a measurable set, define

(2.1) $\quad \lambda^+(A) = \sup_{B\subset A} \lambda(B), \quad \lambda^-(A) = -\inf_{B\subset A} \lambda(B), \quad |\lambda| = \lambda^+ + \lambda^-.$

The set functions λ^+, $-\lambda^-$ and $|\lambda|$ are respectively the *positive, negative,* and *total variations* of λ. It will be shown that all three set functions are countably additive. The signed measure is *finite*, or σ *finite,* when the total variation is. A measurable set A is a *positivity set* of λ if $\lambda^-(A) = 0$, a *negativity set* if $\lambda^+(A) = 0$. A positivity [negativity] set A is *maximal* if every positivity [negativity] set is a subset of A, neglecting $|\lambda|$ null sets. If S is the union of a positivity and a negativity set, the summands are obviously maximal.

Theorem. *Let* $(S,\mathbf{S},\lambda)$ *be a signed measure space. Then:*

(a) λ^+, λ^-, *and* $|\lambda|$ *are measures*; $\lambda^+(S)$ *is finite if* $\lambda < +\infty$, $\lambda^-(S)$ *is finite if* $\lambda > -\infty$.

(b) **(Jordan decomposition)** $\lambda = \lambda^+ - \lambda^-$. *Moreover, if* $\lambda = \lambda_1 - \lambda_2$ *is any representation of* λ *as the difference between two measures, then* $\lambda^+ \le \lambda_1$ and $\lambda^- \le \lambda_2$.

(c) **(Hahn decomposition)** *S is the disjunct union of a positivity set S^+ for λ and a negativity set S^- for λ, each maximal, and unique up to $|\lambda|$ null sets.*

(d) *For every measurable set A,* $\lambda^+(A) = \lambda(A \cap S^+)$, $\lambda^-(A) = -\lambda(A \cap S^-)$.

Proof. Suppose, for definiteness, that $\lambda < +\infty$, and choose a sequence $A_\bullet$ of measurable sets for which $\lim \lambda(A_\bullet) = \lambda^+(S)$. Let B_n be the union of those cells of the partition of S generated by $A_1,..,A_n$ (see Section II.3), at which λ is positive. Then $\lambda(B_n) \geq \lambda(A_n)$. As n increases, the partition becomes finer, $B_n \cup B_{n+1} \cup \cdots$ is B_n augmented perhaps by cells of B_{n+1} and so on. Thus

$$\lambda^+(S) \geq \lambda\left(\bigcup_n^\infty B_\bullet\right) \geq \lambda(B_n) \geq \lambda(A_n). \tag{2.2}$$

Define $S^+ = \limsup B_\bullet$ and $S^- = S - S^+$. Then $\lambda(S^+) = \lambda^+(S)$, and therefore $\lambda^+(S) < +\infty$. Moreover, in view of the maximal property of S^+, $\lambda^-(S^+) = \lambda^+(S^-) = 0$. Thus S^+ and S^- are the sets of a Hahn decomposition and the Jordan decomposition can be written in the form $\lambda(A) = \lambda(A \cap S^+) + \lambda(A \cap S^-)$. Only the minimal character of λ^+ and λ^-, stated in (b) is still to be proved, and this character follows at once from the definitions of these measures.

Example. Let $(S,\underline{S},\lambda)$ be a measure space, let f be a measurable function from the space into $\overline{\mathbf{R}}$, define $f^+ = f \vee 0$, $f^- = (-f) \vee 0$, and suppose that f^+ is λ integrable. Then the set function μ: $A \rightarrow \mu(A) = \lambda[f 1_A]$ is a signed measure, a finite valued signed measure if and only if f is integrable. Obviously

$$\mu^+(A) = \lambda[f^+ 1_A], \quad \mu^-(A) = \lambda[f^- 1_A], \quad |\mu|(A) = \lambda[|f| 1_A],$$

and the sets S^+ and S^- of a Hahn decomposition for μ can be chosen respectively as the sets $\{f \geq 0\}$, $\{f < 0\}$.

3. Lattice property of the class of signed measures

Theorem. *If λ and λ' are signed measures on a measurable space, there is a signed measure $\lambda \vee \lambda'$ majorizing λ and λ' and majorized by every other signed measure majorant of λ and λ'.*

The theorem implies that in addition to the smallest measure majorant $\lambda \vee \lambda'$ of λ and λ' there is also a largest measure minorant, $-(-\lambda) \vee (-\lambda')$. This minorant is denoted by $\lambda \wedge \lambda'$. Obviously $\lambda^+ = \lambda \vee \lambda_0$, and $\lambda^- = -(\lambda \wedge \lambda_0)$, with λ_0 the identically vanishing measure.

Proof. If $\lambda - \lambda'$ is a well-defined signed measure, that is, if $\lambda(S)$ and $\lambda'(S)$ are not both $+\infty$ or both $-\infty$, let S^+ be a maximal positivity set and let S^- be a maximal negativity set, for $\lambda - \lambda'$. Define

$$(\lambda\vee\lambda')(A) = \lambda(A\cap S^+) + \lambda'(A\cap S^-).$$

This sum defines a measure with the required properties. There remains the case when $\lambda(S)$ and $\lambda'(S)$ are both $+\infty$ or both $-\infty$. Assume the first possibility. If $\lambda(A) = \lambda'(A) = +\infty$, define $(\lambda\vee\lambda')(A) = +\infty$. If either $\lambda(A)$ or $\lambda'(A)$ is finite, let λ_A and λ'_A be, respectively, the restrictions of λ and λ' to the measurable subsets of A, and let A^+ and A^- be, respectively, maximal positivity and negativity sets for $\lambda_A - \lambda'_A$. If A_0 is a measurable subset of A, define

$$(\lambda\vee\lambda')(A_0) = \lambda(A_0\cap A^+) + \lambda'(A_0\cap A^-).$$

This value is independent of the choice of the superset A of A_0 and yields the required measure.

Example. If $(S,\mathbf{S},\lambda)$ is a measure space and if f and f' are measurable functions from the space into $\bar{\mathbf{R}}$, for which either $f\vee 0$ or $(-f)\vee 0$ is λ integrable and either $f'\vee 0$ or $(-f')\vee 0$ is λ integrable, define $\mu(A)=\lambda[f\mathbf{1}_A]$ and $\mu'(A)=\lambda[f'\mathbf{1}_A]$. Then μ and μ' are signed measures, $(\mu\vee\mu')(A) = \lambda[(f\vee f')\mathbf{1}_A]$, and $(\mu\wedge\mu')(A) = \lambda[(f\wedge f')\mathbf{1}_A]$.

4. Absolute continuity and singularity of a signed measure

Let λ be a measure and μ be a signed measure on a measurable space $(S,\mathbf{S})$. The signed measure μ is *absolutely continuous relative to* λ, or λ *absolutely continuous*, if μ vanishes on λ null sets, equivalently if $|\mu|$ vanishes on λ null sets. At the other extreme, μ is *singular relative to* λ, or λ *singular*, if there is a λ null set whose complement is $|\mu|$ null. Only the identically vanishing signed measure is both λ absolutely continuous and λ singular. If λ and μ are measures, μ is λ singular if and only if λ is μ singular, that is, each of the measures is carried by a null set of the other.

Theorem *(Measure space $(S,\mathbf{S},\lambda)$). If a signed measure μ on $\mathbf{S}$ is λ absolutely continuous, or singular, μ^+ and μ^- have this same character.*

The theorem is clear from the definitions of μ^+ and μ^-.

Alternative approach to the absolute continuity definition. (Notation as above, but λ and μ are both measures.) It is trivial that the limit equation

(4.1) $$\lim_{\lambda(A)\to 0}\mu(A) = 0$$

implies that μ is λ absolutely continuous. Conversely, if μ is a finite measure that is λ absolutely continuous, then (4.1) is true. To prove this, it is sufficient to prove that if $A_\bullet$ is a set sequence for which $\lim \lambda(A_\bullet) = 0$ then every subsequence of $A_\bullet$ has a further subsequence along which $\mu(A_\bullet)$ tends to 0. Given a subsequence of $A_\bullet$, choose a further subsequence $B_\bullet$ for which $\Sigma\lambda(B_\bullet)$ converges, and define $C_n = \cup_n^\infty B_\bullet$. Then the sequence $C_\bullet$ is monotone decreasing with limit a λ null set, a set which is necessarily also a μ null set, and therefore the sequence $\mu(C_\bullet)$, which majorizes the sequence $\mu(B_\bullet)$, is monotone decreasing with limit 0.

Uniform absolute continuity. If $\{\mu_i, i \in I\}$ is a family of finite measures on a measure space $(S,\mathbf{S},\lambda)$ and $\lim_{\lambda(A)\to 0} \mu_i(A) = 0$, uniformly as i varies in I, the family of measures is *uniformly absolutely continuous relative to* λ. If λ is a finite measure and d_λ is the pseudometric on $\mathbf{S}$ determined by λ, that is, $d_\lambda(A,B) = \lambda(A\Delta B)$, then *the condition of uniform absolute continuity of the family* $\mu_\bullet$ *of measures is equivalent to the uniform* d_λ *continuity of the family* $\mu_\bullet$ *considered as a family of functions from* $(\mathbf{S},d_\lambda)$ into $\mathbf{R}$. In fact uniform d_λ continuity at the empty set is precisely the definition of uniform absolute continuity, and uniform d_λ continuity at the empty set implies uniform d_λ continuity because $d_\lambda(A,B) = d_\lambda(A\Delta B,\varnothing)$.

5. Decomposition of a signed measure relative to a measure

Theorem (Lebesgue decomposition). (*Measure space* $(S,\mathbf{S},\lambda)$). *If* μ *is a signed measure on* $\mathbf{S}$, μ *is the sum,* $\mu = \mu_{ac} + \mu_s$, *of uniquely determined* λ *absolutely continuous and* λ *singular signed measures.*

Proof. If there is a decomposition, $\mu = \mu_{ac}+\mu_s$, into absolutely continuous and singular measures, it is unique, because if there were two such decompositions, the difference between the singular components would be the negative of the difference between the absolutely continuous components. These differences would be both λ absolutely continuous and λ singular and would therefore vanish identically. In view of Theorem 4 it will be sufficient to derive the Lebesgue decomposition when μ is a measure. If $c = \sup\{\mu(A): \lambda(A) = 0\}$, then $c = 0$ if and only if μ is λ absolutely continuous, in which case there is nothing to prove. If $c > 0$ and $A_\bullet$ is a sequence of sets for which $\lambda(A_n)=0$ and $\lim \mu(A_\bullet) = c$, define $B = \cup A_\bullet$, and then define μ_{ac} by $\mu_{ac}(A) = \mu(A\cap \tilde{B})$, μ_s by $\mu_s(A) = \mu(A\cap B)$.

Examples. (a) Let f be a measurable function from a measurable space $(S,\mathbf{S})$ into $\bar{\mathbf{R}}^+$ and define a measure μ by $\mu(A) = \lambda[f1_A]$. Then μ is λ absolutely continuous. In particular, f may be identically $+\infty$, in which case μ is $+\infty$ on all non λ null sets.

(b) Let S be $\mathbf{R}$ and λ be Lebesgue measure. If $A = r_\bullet$ is a countable subset of $\mathbf{R}$ and $p_\bullet$ is a sequence in $\bar{\mathbf{R}}^+$, let μ be the measure, carried by A, with $\mu(\{r_n\}) = p_n$. This measure is λ singular. It is finite if $\Sigma p_\bullet$ converges, σ finite if the series diverges but each summand is finite.

(c) **The Cantor set on R and a corresponding measure, singular relative to Lebesgue measure, with no nonnull singletons.** Let S be the interval $[0,1]$ on $\mathbf{R}$ and λ be Lebesgue measure on that interval. A monotone increasing continuous function F will now be defined on S, with $F(0) = 0$ and $F(1) = 1$, with the property that the measure λ_F generated by F is λ singular and that the derivative F' exists and vanishes λ almost everywhere on S. Let S be on the axis of abscissas of coordinate axes in a plane and choose a sequence $s_\bullet$ on the ordinate axis, as follows.

Step 0: choose an arbitrary point s_1 of the interval $(0,1)$ of the ordinate axis.
Step 1: the point s_1 divides $(0,1)$ into two open intervals; choose points s_2 in the upper interval, s_3 in the lower.
Step n: 2^n points having been chosen, dividing $(0,1)$ into 2^n open intervals, choose a point in each, going down from the top. These choices should be made in such a way that the sequence $s_\bullet$ is dense in S.

Next, delete from S, on the axis of abscissas, a sequence $I_\bullet$ of open subintervals, as follows.

Step 0': delete from S an open interval I_1 with closure in $(0,1)$.
Step 1': the deletion of I_1 leaves a right and a left interval in S; corresponding to the choices of s_2 and s_3, delete I_2, an open interval with closure in the interior of the right interval, and delete I_3, an open interval with closure in the interior of the left interval.
Step n': 2^n intervals having been left in S; delete from the interior of each of these an open interval whose closure is in that interior, ordering the sequence $I_\bullet$ of deletions from right to left. Choose these intervals to make $\Sigma\, \lambda(I_\bullet) = 1$.

Having ordered $s_\bullet$ and $I_\bullet$ in this way, define F on I_n to be identically s_n. As so defined, F is monotone increasing, defined λ almost everywhere on S, and its range of values is dense on $[0,1]$. Define F elsewhere on S by continuity to obtain a function with the desired properties. The set $S - \bigcup I_\bullet$ is a perfect nowhere dense λ null set, and the measure λ_F generated by F is λ singular. This set was devised by Cantor, who chose each point s_n in the middle of its interval and chose each interval I_n as the middle third of its interval. With these choices the set is the *Cantor set.* Lebesgue used Cantor's set to obtain the continuous monotone function F, generating a measure singular with respect to his measure.

6. A basic preparatory result on singularity

Lemma. *Let* λ *and* μ *be finite measures on a measurable space* $(S,\mathbf{S})$, *and let* Γ *be the class of positive measurable functions* g *satisfying the inequality*

$$\int_A g\,d\lambda \le \mu(A) \qquad (A \in \mathbf{S}). \tag{6.1}$$

Then, $g_1 \vee g_2$ *is in* Γ *if* g_1 *and* g_2 *are, and, unless* μ *is* λ *singular,* Γ *contains a function not vanishing* λ *almost everywhere,*

Proof. If g_1 and g_2 are in Γ, their maximum on $A \cap \{g_1 > g_2\}$ is g_1 and therefore satisfies (6.1) on this set. Similarly their maximum satisfies (6.1) on $A \cap \{g_2 \ge g_1\}$. Hence their maximum satisfies (6.1) and accordingly is in the class Γ. For $n \ge 1$, let A_n be a maximal positivity set of the signed measure μ-λ/n, so that $\mu(\tilde{A}_n) \le \lambda(\tilde{A}_n)/n \le \lambda(S)/n$, and $A_\bullet$ is an increasing sequence, neglecting λ null sets. If A_n is λ null for every value of n, then $\mu(\cap \tilde{A}_\bullet) = 0$, that is, μ is carried by $\cup A_\bullet$ and must be λ singular. On the other hand, if some set A_n is not λ null, then the function $\mathbf{1}_{A_n}/n$ is in Γ and does not vanish almost everywhere.

7. Integral representation of an absolutely continuous measure

Theorem (Radon-Nikodym). *Let* μ *be a finite signed measure and* λ *be a* σ *finite measure on a measurable space* $(S,\mathbf{S})$. *There is then a* λ *integrable function* f, *uniquely determined up to* λ *null sets, satisfying*

$$\mu_{ac}(A) = \int_A f\,d\lambda \qquad (A \in \mathbf{S}). \tag{7.1}$$

For c *a constant, the inequality* $\mu_{ac} \ge c\lambda$ $[\mu_{ac} \le c\lambda]$ *implies the* λ *almost everywhere inequality* $f \ge c$ $[f \le c]$.

The function f in (7.1), the *Radon-Nikodym derivative of* μ_{ac} *with respect to* λ, or *density with respect to* λ *of the measure* μ_{ac}, is sometimes denoted by $d\mu_{ac}/d\lambda$.

Proof. In view of Theorem 4, in proving the existence of f it will be sufficient to assume that μ is a measure. Moreover it will be sufficient to consider only the case of finite λ, because the result in that case can be applied individually to each of a disjunct sequence of measurable sets on which λ is finite valued. Under these hypotheses of positivity and finiteness it will now be shown that any order supremum of the class Γ in Lemma 6, can be taken as the function f in (7.1). If Γ contains only functions that vanish λ almost everywhere, μ is singular, and the theorem is true with f identically 0. Otherwise, there is a sequence $f_\bullet$ of members of Γ whose pointwise supremum is an essential order supremum of Γ (Theorem V.18). Since $g_1 \vee g_2$ is in Γ whenever g_1 and g_2 are, f_n can be replaced by $f_1 \vee \cdots \vee f_n$, to make the sequence $f_\bullet$ monotone increasing. An application of the Beppo-Levi theorem now shows that the limit g of the sequence, an essential order supremum of Γ, is in Γ. The measure

$$\mu' = \mu - \int_\bullet g \, d\lambda$$

must be λ singular, or an application of Lemma 6 to μ' would yield a not λ almost everywhere vanishing integrand whose sum with g is in Γ, contrary to the maximality of g . Thus $\mu - \mu' = \mu_{ac}$, and g is the desired Radon-Nikodym derivative of μ_{ac}.

Going back to the general case of signed measures, to show that the Radon-Nikodym derivative is unique up to λ null sets, suppose that f and another integrable function f' have the same integral over every measurable set A. The fact that

$$\int_A (f - f') \, d\lambda = 0 \text{ when } A = \{f \geq f'\} \text{ i}$$

implies that $f \leq f'$, λ almost everywhere; by symmetry, $f' \leq f$, λ almost everywhere, and therefore finally $f = f'$, λ almost everywhere.

To prove the last statement of the theorem, observe that if $\mu_{ac} \geq c\lambda$ then $\mu_{ac} - c\lambda$ is a positive measure, and its Radon-Nikodym derivative with respect to λ must be positive and be $(d\mu_{ac} d\lambda) - c$ up to λ null sets. The inequality in the other direction is treated similarly.

8. Bounded linear functionals on $\mathbf{L}^1$

Let $(S, \mathbf{S}, \lambda)$ be a finite measure space. A *bounded linear functional on* $\mathbf{L}^1$ is a function $\mathbf{L}$ from $\mathbf{L}^1$ into $\mathbf{R}$ satisfying the following conditions:

$\mathbf{L}(af + bg) = a\mathbf{L}(f) + b\mathbf{L}(g)$ for f, g in $\mathbf{L}^1$ and constants a, b in $\mathbf{R}$;
$|\mathbf{L}(g)| \leq \text{const.} |g|_1$ for g in $\mathbf{L}^1$;
$\mathbf{L}(g) = 0$ if g vanishes λ almost everywhere.

The first and third conditions imply that $\mathbf{L}$ defines a unique function of equivalence classes of members of $\mathbf{L}^1$, if two members are put in the same equivalence class if and only if they are equal almost everywhere. In the following, however, functions rather than equivalence classes will be treated. The *norm* $|\mathbf{L}|$ of $\mathbf{L}$ is the smallest constant for which the second condition is valid. The functional $\mathbf{L}$ is *positive* when $f \geq 0$ implies that $\mathbf{L}(f) \geq 0$.

If f is a bounded measurable function, with essential supremum $|f|_\infty$, the function on $\mathbf{L}^1$ defined by

$$(8.1) \qquad L(g) = \int_S gf\,d\lambda$$

is a bounded linear functional, positive if f is almost everywhere positive, and $|L(f)| \le |f|_\infty |f|_1$. Moreover, this inequality for f in L^1 is not satisfied with any constant smaller than $|f|_\infty$, because this inequality with a constant α implies that $|f| \le \alpha$ almost everywhere, according to the last assertion in Theorem 7. The norm of L is therefore the essential supremum of f. A similar argument shows that L is a positive functional if and only if $f \ge 0$ almost everywhere. The following theorem states that every bounded linear functional on L^1 has this form.

Theorem (σ *finite measure space* $(S,\mathbf{S},\lambda)$). *If* L *is a bounded linear functional on* L^1, *there is a unique (neglecting null sets) bounded measurable function* f *for which* L *is given by* (8.1). *Moreover* $\|L\| = |f|_\infty$, *and* L *is positive if and only if* $f \ge 0$ *almost everywhere.*

This theorem sets up a linear, norm and order preserving correspondence between bounded linear functionals on L^1 and the members of the dual space L^∞.

Proof. Define a set function μ on $\mathbf{S}$ by $\mu(A) = L(1_A)$. The properties of L imply that μ is a λ absolutely continuous finite signed measure. Hence there is a λ integrable function f, the Radon-Nikodym derivative of μ with respect to λ, for which, for A in $\mathbf{S}$,

$$|\mu(A)| = |L(1_A)| = |\lambda[f1_A]| \le \|L\| \, |1_A|_1 = \|L\| \, \lambda(A),$$

and therefore (Theorem 7) $|f| \le \|L\|$ almost everywhere. Thus (8.1) is true for g the indicator function of a measurable set, and therefore (8.1) is true when g is a linear combination of such functions. Since every function in L^1 can be approximated arbitrarily closely in the $L^1(\lambda)$ distance sense by such step functions (see Section VI.15), this evaluation of L is valid for g in L^1.

9. Sequences of signed measures

In this and the next two sections, sequences of signed measures on a measurable space are treated, and a method based on a property of complete pseudometric spaces is applied, namely, the property (Section 0.12(c)) that if a complete pseudometric space is a countable union of closed sets, then at least one summand has an inner point.

If λ is a signed measure on a measurable space $(S,\mathbf{S})$, the inequality

$$(9.1) \qquad |\lambda|(S) \le 2\sup\{|\lambda(A)| : A \in \mathbf{S}\}$$

implies that, for a finite set $\lambda_\bullet$ of signed measures, $\sup |\lambda_\bullet|(S) < +\infty$ whenever

$\sup |\lambda_\bullet(A)| < +\infty$ for every set A in $\mathbf{S}$, but it is not at all obvious how this conclusion must be weakened for a sequence of signed measures. According to the following theorem, no weakening is necessary.

Theorem. *Let $\lambda_\bullet$ be a sequence of signed measures on a measurable space $(S,\mathbf{S})$. Then the boundedness condition*

$$(9.2) \qquad \sup|\lambda_\bullet(A)| < +\infty \textit{ for every measurable set } A$$

implies the uniform boundedness condition

$$(9.3) \qquad \sup\left\{ |\lambda_n(A)|: n \ge 1, A \in \mathbf{S} \right\} < +\infty,$$

that is, $\sup \|\lambda_\bullet\|(S) < +\infty$.

Inequality (9.2), with $A = S$, implies that each signed measure λ_n in this theorem is finite valued. The theorem is trivial if the signed measures are measures, because then the conclusion follows from (9.2) with this choice of A.

The condition (9.2) is equivalent to the pair of conditions

(a) *the signed measures $\lambda_\bullet$ are all finite valued,*
(b) $\limsup |\lambda_\bullet(A)| < +\infty$ *for every measurable set A,*

and (a) is equivalent to the condition

(a') the values $\lambda_\bullet(S)$ are all finite.

Under (b) without (a), $\lambda_n(S)$ is finite valued when n is sufficiently large, say $n \ge n_0$, and therefore Theorem 9 is applicable to the sequence $\{\lambda_n, n \ge n_0\}$.

Proof of the theorem. Choose a finite measure λ on $\mathbf{S}$, relative to which every set function λ_n is absolutely continuous; for example, define

$$\lambda = \Sigma_1^\infty n^{-2}\lambda_n/\|\lambda_n\|(S).$$

It will be proved first that there are strictly positive constants c, δ for which $\lambda(A) < \delta$ implies that $\sup \|\lambda_\bullet\|(A) \le c$. Under the pseudometric d_λ: $d_\lambda(A,B) = \lambda(A\Delta B)$, $\mathbf{S}$ is a complete pseudometric space (Section III.14), and each function λ_n is a continuous function from $\mathbf{S}$ into $\mathbf{R}$. The set

$$(9.4) \qquad \bigcap_{n=1}^{\infty} \{A \in \mathbf{S}: |\lambda_n(A)| \le k\}) = \{A \in \mathbf{S}: \sup|\lambda_\bullet(A)| \le k\}$$

is a d_λ closed subset of $\mathbf{S}$ that increases as k increases, tending to $\mathbf{S}$. It follows (pseudometric version of Section 0.12(c)) that there is a value of k for which the set in (9.4) contains a ball, say one with center B_0 and radius δ. That is, there are numbers c_1 and δ for which $|\lambda_n(C)| \le c_1$ for all values of n, whenever $d_\lambda(B_0,C) < \delta$. Now suppose that A is a set with $\lambda(A) < \delta$. Then $d_\lambda(B_0 \cup A, B_0) < \delta$, $d_\lambda(B_0 - A, B_0) < \delta$ and therefore

$$(9.5)\quad |\lambda_n(A)| = |\lambda_n(B_0 \cup A) - \lambda_n(B_0 - A)| \le d_\lambda(B_0 \cup A, B_0) + d_\lambda(B_0 - A, B_0) \le 2c_1.$$

This is the desired inequality, with $c = 2c_1$, and the proof of the theorem is now complete if S can be written as the union of finitely many sets of arbitrarily small strictly positive λ measure. Many measure spaces, for example, a finite interval of **R** with λ defined as Lebesgue measure, have this property. On the other hand, a probability space does not have this property if the class of measurable sets is the whole space and the empty set. The following completion of the proof of the theorem shows that the latter example, in which there is a nonnull set whose subsets have measure either 0 or that of the set, exhibits the only context that must be taken into account. Let α be the supremum of the class Γ of constants η for which there is a measurable set A with $\lambda(A) \ge \eta$ and $\sup|\lambda_\bullet|(A) < +\infty$. The proof of the theorem will be completed by proving that α is in Γ and that $\alpha = \lambda(S)$.

Proof that $\alpha \in \Gamma$. For $n \ge 1$, choose a measurable set A_n with $\lambda(A_n) > \alpha - 1/n$, and with the property that $\sup|\lambda_\bullet|(A_n) < +\infty$. Define $A^k = \cup_1^k A_\bullet$. Then $\alpha - 1/n \le \lambda(A^k) \le \alpha$. If some set A^k has λ measure α, then α is in Γ, as desired. If $\lambda(A^k) < \alpha$ for all n, the union of these sets is a set A of λ measure α. Choose n so large that $\lambda(A^k) > \alpha - \delta$. Then

$$\sup|\lambda_\bullet|(A) \le \sup|\lambda_\bullet|(A^k)| + c.$$

Thus α is in Γ.

Proof that $\alpha = \lambda(S)$. Let B be a measurable set, of measure α, for which $\sup|\lambda_\bullet|(B) < +\infty$. If $S-B$ is λ null there is nothing more to prove, and the following argument shows that the hypothesis that $S-B$ is not λ null leads to a contradiction. If $S-B$ is not λ null, this set has the following two properties:

(i) $S-B$ has no measurable subset C for which $0 < \lambda(C) < \delta$, because such a set could be adjoined to B, thereby contradicting the maximal character of α.

(ii) $S-B$ has no measurable subset C, of strictly positive λ measure, all of whose subsets have measure either 0 or $\lambda(C)$, because such a set could be adjoined to B, thereby contradicting the maximal character of α

Hence there is a measurable subset C of $S-B$, of λ measure strictly between 0 and $\lambda(S-B)$. Either C or $(S-B)-C$ has λ measure at most $\lambda(S-B)/2$ Choose the one, say C_1, with this property, and continue, replacing $S-B$ in the reasoning by $S-(B \cup C_1)$ and so on, obtaining a sequence $C_\bullet$ of subsets of $S-B$, with strictly positive λ measures tending to 0. Since such a sequence contradicts property (i), the set $S-B$ must be λ null, as was to be proved.

10. Vitali-Hahn-Saks theorem (continued)

A part of the Vitali-Hahn-Saks theorem was proved in Section **III**.10. That part is repeated as part (a) of the following more complete version.

Theorem (Vitali-Hahn-Saks). *Let $\lambda_\bullet$ be a sequence of finite measures on the measurable space $(S,\mathbf{S})$.*

(a) *If $\lambda_\bullet$ converges setwise to a finite valued set function λ, then λ is a measure.*

(b) *In* (a), *if there is a finite measure μ with respect to which each measure λ_n is absolutely continuous then λ is also μ absolutely continuous, and the sequence $\lambda_\bullet$ is equiuniformly μ absolutely continuous.*

As already noted, part (a) of the theorem is simply a restatement of Theorem **III**.10.

Proof of (b). In view of the setwise convergence of $\lambda_\bullet$, the limit measure λ vanishes on μ null sets and therefore is μ absolutely continuous. If the sequence $\lambda_\bullet$ is not uniformly μ absolutely continuous, there are a subsequence $\lambda_{a_\bullet}$ and a sequence $B_\bullet$ of measurable sets, with the property that $\lim \mu(B_\bullet) = 0$ but $\inf \lambda_{a_\bullet}(B_\bullet) > 0$, contradicting the **Observation** in Section **III**.10.

Relation to Fatou's theorem. Let $f_\bullet$ be a sequence of positive measurable integrable functions on a measure space $(S,\mathbf{S},\mu)$, with a measurable integrable function f as an almost everywhere or in measure limit and define $\lambda_n(A) = \mu[f_n \mathbf{1}_A]$ and $\lambda(A) = \mu[f\mathbf{1}_A]$. If $\lim \mu[f_\bullet] = \mu[f]$, then (Fatou's theorem) the sequence $f_\bullet$ has limit f in the L^1 sense. Moreover (Theorem **VI**.18) in that case if μ is a finite measure, the sequence $f_\bullet$ is μ uniformly integrable; equivalently $\lambda_\bullet$ is d_μ equicontinuous. The uniform integrability of $f_\bullet$ implies the uniform absolute continuity of $\lambda_\bullet$, and the L^1 convergence of $f_\bullet$ implies that the sequence $\lambda_\bullet$ is setwise convergent to λ and in fact that the sequence of variations $|\lambda_\bullet - \lambda|$ has limit 0, a stronger form of convergence than setwise convergence. On the other hand, the sequence $f_\bullet$ in Theorem 10(b) converges to f only in the weak sense that there is setwise convergence of $\lambda_\bullet$ to λ.

11. Theorem 10 for signed measures

Theorem. *If $\lambda_\bullet$ is a sequence of finite signed measures on a measurable space $(S,\mathbf{S})$, setwise convergent to a finite valued set function λ, then λ is a signed measure.*

Proof. The set function λ is obviously finitely additive. Choose a finite measure μ on $\mathbf{S}$, relative to which every measure λ_n is absolutely continuous, say

$\mu = \Sigma_1^\infty n^{-2}|\lambda_n|/|\lambda_n|(S)$. Then $\lambda_\bullet$ is a sequence of continuous functions from the complete pseudometric space $(\mathbf{S}, d_\mu)$ into $\mathbf{R}$, and the sequence has limit λ. According to Theorem 9, λ is bounded. There remains the proof of countable additivity. According to a standard metric space theorem (pseudometric version of (Section 0.12(f)), the function λ must have a d_λ continuity point, say the set A, at which the sequence is uniformly convergent. If $B \in \mathbf{S}$,

$$\lambda(B) = \lambda(A \cup B) - \lambda(A - B). \tag{11.1}$$

Now when B tends to $\emptyset$ in the d_μ metric, that is, when $\mu(B)$ tends to 0, both $A \cup B$, and $A-B$ tend to A in this metric, and therefore by d_λ continuity of λ at A, (11.1) implies that $\lambda(B)$ tends to 0. Thus λ is d_μ continuous at $\emptyset$. It follows that if a monotone decreasing sequence of measurable sets has limit $\emptyset$, the corresponding sequence of λ values has limit 0, that is, λ is countably additive.

X
Measures and Functions of Bounded Variation on R

1. Introduction

This chapter is devoted to the context of set functions and corresponding point functions on **R**, a context so important that it deserves a treatment beyond the discussion of Lebesgue-Stieltjes measures in Chapter IV. The first topic will be the derivation of monotone functions, for which a covering lemma will be needed.

2. Covering lemma

In the following lemma, the fact will be used that *if $I_\bullet$ is an arbitrary collection of intervals of* **R** *(which may or may not include one or both endpoints), and if A is the union of their interiors, then the difference $B = \cup I_\bullet - A$ is countable.* To see this, for each point of B choose a corresponding interval in $I_\bullet$ containing the point as an endpoint. The intervals chosen with the point of B as their left endpoint have pairwise disjoint interiors and are therefore countable in number, as are the intervals chosen with the point of B as their right endpoint. Hence B is countable. Incidentally this countability implies that B and $\cup I_\bullet$ are Borel sets.

Lemma (Aldaz). *If λ is a Lebesgue-Stieltjes measure on* **R**, *$I_\bullet$ is a collection of intervals of* **R** *(which may or may not include one or both endpoints), and $c<1/2$, there is a finite disjunct sequence $J_\bullet$ of members of $I_\bullet$ with $\Sigma\, \lambda(J_\bullet)>c\lambda(\cup I_\bullet)$.*

Proof. In the notation used at the beginning of this section, the set B is countable, and its points are endpoints of a countable subcollection of $I_\bullet$. By Lindelöf's theorem, the union of the $I_\bullet$ interval interiors is the same as some countable subunion of the interiors. Thus it can be assumed that the given collection $I_\bullet$ is a finite or infinite sequence. Choose n so large that

$$\sum_{j=1}^{n} \lambda(I_j) > 2c\lambda(\cup I_\bullet),$$

and for convenience in referencing define $J_i = I_i$ for $i \le n$. It is no restriction to assume that each interval J_i contains a point s_i not in any other of these n intervals and that these intervals are ordered in such a way that $s_\bullet$ is monotone increasing. Then $J_i \subset (-\infty, s_j)$ and $J_k \subset (s_j, \infty)$ when $i < j < k$. The intervals $J_\bullet$ with even indices are pairwise disjoint, as are the intervals with odd indices. Hence either the union of the even indexed intervals or of the odd indexed intervals has measure $> c\lambda(\cup J_\bullet)$ and provides the intervals demanded by the lemma.

3. Vitali covering of a set

A class of intervals of $\mathbf{R}^N$ *covers a set in the sense of Vitali* if each point of the set is a point of an arbitrarily small interval in the class.

Theorem. *Let λ be a Lebesgue-Stieltjes measure on $\mathbf{R}$. If a class of closed intervals of $\mathbf{R}$ covers a subset A of $\mathbf{R}$ in the sense of Vitali, there is a countable disjunct sequence of those intervals with union including λ almost every point of A.*

Proof. Let λ^* be the outer measure generated by $\mathbf{B}(\mathbf{R})$ and λ. If A is λ^* null there is nothing to prove. Since it is sufficient to prove the theorem for a bounded set A, it can be supposed that $0 < \lambda^*(A) < +\infty$. Let G be an open superset of A with $\lambda(G) < (10/9)\lambda^*(A)$ and let $I_\bullet$ be the collection of the intervals of the Vitali cover of A that are subsets of G. According to Lemma 2, there are finitely many pairwise disjoint intervals $J_\bullet$ in $I_\bullet$ with union of measure at least $\lambda^*(A)/3$ and therefore leaving a subset of A of outer measure at most $(7/9)\lambda^*(A)$ uncovered. Replace A in this argument by $A - \cup J_\bullet$, replace the given Vitali cover by its members not meeting $A - \cup J_\bullet$, and repeat the argument to find pairwise disjoint members of this cover, not meeting $J_\bullet$, and together with $J_\bullet$ leaving a subset of A of outer measure at most $(7/9)^2\lambda^*(A)$ uncovered. And so on, through the powers of 7/9.

4. Derivation of Lebesgue-Stieltjes measures and of monotone functions

In this section, λ and μ are Lebesgue-Stieltjes measures on $\mathbf{R}$, and λ is complete. The functions F_λ and F_μ are corresponding monotone increasing right continuous functions that generate these measures, and λ^*, μ^* are respectively the outer measures generated by λ, μ. Let s be a point of $\mathbf{R}$ and I be a closed interval containing s. The upper [lower] derivate at s of F_μ with respect to F_λ is defined as the limit superior [inferior], of $\mu(I)/\lambda(I)$, when I shrinks to s except that these derivates are left undefined when $\lambda(I)=0$ for some choice of I. In the latter case, s is in a maximal interval of such points, and the class of such points is a λ null countable disjunct union of such intervals. If the upper and lower derivates exist and are equal at s, that is, if there is a limit at s, the limit is the *derivative at s of* μ *with respect to* λ, or of F_μ *with respect to* F_λ, and will be denoted by $(d\mu/d\lambda)(s)$, or by $(dF_\mu/dF_\lambda)(s)$. Observe that if s is a discontinuity point of F_λ, that is, if $\lambda(\{s\}) > 0$, then the derivative exists at s and is $\mu(\{s\})/\lambda(\{s\})$. If λ is Lebesgue measure, the derivative will be denoted, as usual, by F_μ'.

Theorem. *Let* λ *and* μ *be Lebesgue-Stieltjes measures on* **R**. *Then*

(a) *the derivative* $d\mu/d\lambda$ *exists and is finite* λ *almost everywhere on* **R**;

(b) *this derivative (extended arbitrarily to all of* **R**) *is* λ *measurable and integrable, and is a version of the Radon–Nikodym derivative with respect to* λ *of the absolutely continuous component of* μ *relative to* λ.

Proof of (a). It is sufficient to prove the theorem for the derivative at the points of a finite open interval J. Choose numbers a, b, with $a < b$, and let A be the set of points s in J at which the upper and lower derivates exist and are respectively $> b$ and $< a$. The closed subintervals I of J with the property that I contains a point of A and that $\mu(I)/\lambda(I) > b$, cover A in the sense of Vitali, and Theorem 3 therefore implies that there is a disjunct countable union B of closed intervals covering $\lambda+\mu$ almost every point of A and satisfying the inequality $\mu(B) > b\lambda(B)$. Then $\mu(B) > b\lambda^*(A)$. Now if $\gamma > 0$, all the above can be done with the additional condition on the Vitali covering intervals: they are to be subsets of an open set J' covering A, chosen in such a way that $\mu(J') \le \lambda^*(A)+\gamma$. It follows that $\mu^*(A) \ge b\lambda^*(A)$. A parallel proof shows that $\mu^*(A) \le a\lambda^*(A)$, and therefore that $\lambda^*(A) = 0 \;(= \lambda(A))$. The set where the derivative does not exist for s in J is the union of all sets defined like A, as a and b range through the rational numbers. It follows that this limit exists λ almost everywhere on J.

The following argument to prove λ almost everywhere finiteness of the derivative should be compared with the application of martingale theory to derivation in Section **XI**.17. Choose numbers α and β, with $\beta > \alpha$. It will be shown that the derivative is finite λ almost everywhere by showing that the derivative is λ measurable and satisfies the inequality

$$\int_{\alpha+}^{\beta} (d\mu/d\lambda)\, d\lambda \le \mu((\alpha,\beta]), \tag{4.1}$$

where here and below it is to be understood that the notation for the integration limits means that the integration is over the right semiclosed interval $(\alpha,\beta]$. Define right semiclosed subintervals of $(\alpha,\beta]$ by

$$I_{n,m} = \big(\alpha+m(\beta-\alpha)2^{-n}, \alpha+(m+1)(\beta-\alpha)2^{-n}\big] \qquad (n \ge 1,\ m = 0,\dots,2^n-1),$$

and define a function x_m on $(\alpha,\beta]$, constant on each interval $I_{n,m}$, equal to 0 on the interval when the interval is λ null and otherwise equal to $\mu(I_{n,m})/\lambda(I_{n,m})$ on the interval. The function x_n is Borel measurable, its integral with respect to λ over the interval $(\alpha,\beta]$ is $\mu\big((\alpha,\beta]\big)$, and $\lim x_\bullet = d\mu/d\lambda$ at each point at which the derivative exists. It follows that the derivative is λ measurable and Fatou's theorem yields (4.1). Since the measure $A \rightarrow \lambda\big[(d\mu/d\lambda)1_A\big]$ is majorized by μ on right semiclosed intervals, this majorization is valid on Borel sets and therefore on the domain of λ; that is, for every λ measurable set A,

$$\int_A d\mu/d\lambda\, d\lambda \le \mu(A). \tag{4.2}$$

Proof of (b) **when μ is λ absolutely continuous.** Let f be (a version of) the Radon-Nikodym derivative of μ with respect to λ. Then (4.1) becomes

$$\int_{\alpha+}^{\beta} (d\mu/d\lambda)\, d\lambda \le \mu\big((\alpha,\beta]\big) = \int_{\alpha+}^{\beta} f\, d\lambda. \tag{4.3}$$

If f is bounded, the sequence $x_\bullet$ is uniformly bounded, and the Lebesgue dominated convergence theorem can be applied above, instead of Fatou's theorem. Hence in this case there is equality in (4.3) and therefore in (4.2). If μ_n is the measure with Radon-Nikodym derivative $f \wedge n$, the derivates of μ are at least as large as those of μ_n, and therefore

$$\int_{\alpha+}^{\beta} d\mu/d\lambda\, d\lambda \ge \int_{\alpha+}^{\beta} d\mu_n/d\lambda\, d\lambda = \int_{\alpha+}^{\beta} f\wedge n\, d\lambda. \tag{4.4}$$

Let $n \rightarrow +\infty$ in (4.4) to obtain the reverse of the inequality in (4.3). Thus there is equality in (4.3), therefore also in (4.2), and by uniqueness of the Radon-Nikodym derivative it follows that $f = d\mu/d\lambda$ up to λ null sets.

Proof of (b) **when μ is λ singular.** If μ is λ singular, choose A in (4.2) as a μ null set that carries λ to find that $d\mu/d\lambda$ vanishes λ almost everywhere.

In view of the Lebesgue decomposition of μ relative to λ, the fact that (b) is true both when μ is λ absolutely continuous and when μ is λ singular implies that (b) is true for arbitrary μ.

5. Functions of bounded variation

Let F be a function from a compact interval I: $[a,b]$ of $\mathbf{R}$ into $\mathbf{R}$, choose a finite increasing sequence $t_1 = a < \cdots < t_n = b$, and consider the sums

$$V^+(F,t_\bullet) = \sum_{j=1}^{n-1} \left[\left(F(t_{j+1})-F(t_j)\right)\vee 0\right], \quad V^-(F,t_\bullet) = \sum_{j=1}^{n-1} \left[\left(F(t_j)-F(t_{j+1})\right)\vee 0\right],$$

(5.1)

$$V(F,t_\bullet) = \sum_{j=1}^{n-1} \left|F(t_{j+1})-F(t_j)\right|.$$

The values $V^+(F,t_\bullet)$, $-V^-(F,t_\bullet)$, and $V(F,t_\bullet)$ are, respectively the *positive, negative,* and *total variations of F on the sequence* $t_\bullet$. They satisfy the equations

(5.2) $$V(F,t_\bullet) = V^+(F,t_\bullet) + V^-(F,t_\bullet), \qquad F(b)-F(a) = V^+(F,t_\bullet) - V^-(F,t_\bullet).$$

The suprema of $V^+(F,t_\bullet)$, $V^-(F,t_\bullet)$, and $V(F,t_\bullet)$ over all sequences $t_\bullet$ are, respectively, the *positive*, minus the *negative*, and the *total variations of F on I*, denoted by $V^+(F,I)$, $V^-(F,I)$, and $V(F,I)$. If the total variation on I is finite, the other two variations are also finite, and F is *of bounded variation on I.* Obviously $|F(b)-F(a)| \le V(F,I)$. If F is of bounded variation on I, define *the total variation function* $\mathbf{|F|}$ on I by

$$\mathbf{|F|}(a) = 0; \quad \mathbf{|F|}(t) = V(F,[a,t]) \qquad (t > a),$$

and define the positive and negative variation functions F^+ and $-F^-$ in the corresponding way.

The following properties are readily checked, taking advantage of the fact that if more points are added to the sequence $t_\bullet$, each of the three sums in (5.1) can only increase.

(a) Maximizing successions of finite sequences for the three sums in (5.1) can be combined to yield one maximizing succession for all three, and thereby, if F is of bounded variation, to derive the equalities (5.2) for the variations on I. Apply this with I replaced by a subinterval with left endpoint a to find that if F is of bounded variation,

(5.3) $$\mathbf{|F|} = F^+ + F^-, \quad F - F(a) = F^+ - F^-.$$

(b) If F and G are functions from I into **R**, then $V(F+G,I) \le V(F,I)+V(G,I)$; V^+ and V^- satisfy the same inequality.

(c) The set function $V(F,\bullet)$ is additive in the sense that if I_1 and I_2 are compact subintervals of I, disjoint except for a common endpoint, then

(5.4) $$V(F,I_1 \cup I_2) = V(F, I_1) + V(F,I_2),$$

and the other two variations are additive in the same sense. It follows that the functions $|F|$, F^+, and F^- are monotone increasing on I; and in fact if $a \le s < t \le b$,

$$V^-(F,[s,t]) = F^-(t) - F^-(s), \quad V^+(F,[s,t]) = F^+(t) - F^+(s),$$

(5.5)

$$V(F,[s,t]) = |F|(t) - |F|(s).$$

Theorem. *Let F be a function from a compact interval I into* **R**. *Then the following holds.*

(a) (**Jordan decomposition**) *F is of bounded variation if and only if F is the difference between two monotone increasing functions, for example, as exhibited by* (5.3).

(b) *If F is of bounded variation, then*
 (i) *F has a right and left limit at every point, and*
 (ii) *F is right [left] continuous at a point if and only if $|F|$ is; in more detail,*

$$F^+(s+) - F^+(s) = [F(s+) - F(s)] \vee 0, \qquad F^+(s) - F^+(s-) = [F(s) - F(s-)] \vee 0.$$

(5.6)

$$-[F^-(s+) - F^-(s)] = [F(s+) - F(s)] \wedge 0, \quad -[F^-(s) - F^-(s-)] = [F(s) - F(s-)] \wedge 0.$$

In particular, if F is of bounded variation, and right continuous at the interior points of I, then F^+ and F^- have this same continuity property.

(c) *F^+ and F^- are minimal in the sense that if F is of bounded variation and $F = F_2 - F_1$, with each function F_i monotone increasing, then*

(5.7) $F^+(t) - F^+(s) \le F_2(t) - F_1(s), \quad F^-(t) - F^-(s) \le F_1(t) - F_1(s) \qquad (a \le s < t \le b).$

The point of the first equality in (5.6) is that F^+ has the same right jump at s as F if F has a positive right jump there, and F^+ is right continuous at s if F is right continuous there or has a negative jump there. The other inequalities have corresponding significance. Observe that (b) implies

(5.8) $\quad |F|(s+) - |F|(s) = |F(s+) - F(s)|, \; |F|(s) - |F|(s-) = |F(s) - F(s-)|.$

The notation of the preceding paragraphs will be used in the following proof.

Proof of (a). It is trivial that if the function F is monotone increasing, then its total variation on I is $F(b) - F(a)$, and it follows that the difference between two monotone increasing functions on I is of bounded variation on I. Conversely, if

F is of bounded variation, (5.3) exhibits F as the difference between two monotone increasing functions.

Proof of (b). A function of bounded variation has right and left limits at all points because monotone functions have this property. Now (notation of (5.1))

$$\left[F(t_{j+1})-F(t_j)\right]\vee 0 \le F^{+}(t_{j+1})-F^{+}(t_j);$$

moreover the sum over the values of j of the differences between the two sides of this inequality is the difference $V^+(F,I)-V^+(F,t_\bullet)$. Choose $\varepsilon > 0$ and choose $t_\bullet$ in such a way that this difference is at most ε. Then, all the more,

$$(5.9)\qquad 0 \le F^{+}(t_2)-F^{+}(a) - \left[F(t_2)-F(a)\right]\vee 0 \le \varepsilon.$$

Add a point to $t_\bullet$ if necessary — adding a point does not invalidate (5.9) — to make t_2-a so small that $|F(t_2)-F(a+)| \le \varepsilon$ and $F^{+}(t_2)-F^{+}(a+) \le \varepsilon$. With this choice of t_2, (5.9) yields

$$(5.10)\qquad 0 \le F^{+}(a+)-F^{+}(a) - \left[F(a+)-F(a)\right]\vee 0 \le 3\varepsilon,$$

and therefore the first equality in (5.6) is true when $s = a$, in fact, for every value of s. The corresponding arguments yield the other equalities.

Proof of (c). It is sufficient to prove (c) for $s = a$ and $t = b$. Under the hypotheses of (c) (notation of (5.1)),

$$(5.11)\qquad \sum_{j=1}^{n-1}\left[\left(F(t_{j+1})-F(t_j)\right)\vee 0\right] \le \sum_{j=1}^{n-1}\left[\left(F_2(t_{j+1})-F_2(t_j)\right)\vee 0\right] = F_2(b)-F_2(a),$$

and the supremum of the left side for all sequences $t_\bullet$ is $F^{+}(b)-F^{+}(a)$. Thus the first inequality in (5.7) is true, and the corresponding argument yields the second inequality.

6. Functions of bounded variation vs. signed measures

Let F be a function of bounded variation on the compact interval I. Then $F = F_2-F_1$, where F_1 and F_2 are monotone increasing. Moreover, if F is right continuous except possibly at a, F_1 and F_2 can be chosen with this same continuity property. (Apply Theorem 5, or simply replace the original choices of F_1 and F_2 by their right limit functions at the points of I other than a.) Under this right continuity condition define (notation of Section IV.8) the signed measure λ_F by $\lambda_F = \lambda_{F_2}-\lambda_{F_1}$. Observe that this definition is independent of the choice of F_1 and F_2 (satisfying the stated right continuity condition), that the λ_F measure of the interval $[a,t]$ is $F(t)-F(a)$, and that this evaluation of λ_F on each interval $[a,t]$ uniquely determines λ_F. Thus to a function of bounded variation on

I, right continuous except possibly at a, corresponds a finite signed measure on $\mathbf{B}(I)$. Conversely if λ is a finite signed measure on $\mathbf{B}(I)$, define F_λ on I by $F_\lambda(a) = 0$ and $F_\lambda(t) = \lambda([a,t])$ for $t > a$. The function F_λ is right continuous except possibly at a and is monotone increasing if and only if λ is a measure. In view of the Jordan decomposition of a signed measure, F_λ is the difference between two monotone increasing functions and is therefore of bounded variation. Moreover λ is the signed measure generated by F_λ by the above defined procedure, because this procedure assigns the same measure as λ to each interval $[a,t]$. Finally it is trivial to check that F is, up to an additive constant, the function of bounded variation determined by the signed measure λ_F. The minimal properties of F^+ and F^- are in exact correspondence with the minimal properties of positive and negative variations of a signed measure on $\mathbf{B}(I)$, as applied to intervals, and it follows that if F is of bounded variation, then $(\lambda_F)^+ = \lambda_{F^+}$ and $(\lambda_F)^- = \lambda_{F^-}$.

7. Absolute continuity and singularity of a function of bounded variation

Let F be a monotone increasing function on the compact interval I:$[a,b]$, right continuous except perhaps at a, and let G be a function defined and of bounded variation on I, right continuous except perhaps at a. The function F *is absolutely continuous* [*singular*] *relative to* G if λ_F is absolutely continuous [singular] relative to λ_G. The function F is simply described as *absolutely continuous* [*singular*] if F is absolutely continuous [singular] relative to Lebesgue meaasure. The Lebesgue decomposition of a signed measure in the present context states that a function of bounded variation relative to a function is the sum of an absolutely continuous and a singular function (relative to the specified monotone function). Here all functions are supposed right continuous except possibly at a.

For example, suppose that F is a monotone increasing function on I, right continuous except possibly at a. Suppose that, if $\varepsilon > 0$, there is a corresponding $\delta > 0$ with the property that if $I_\bullet$ is a finite set of pairwise disjoint subintervals of I, open relative to I, and if $\lambda_G(\cup I_\bullet) \le \delta$ then $\lambda_F(\cup I_\bullet) \le \varepsilon$. The same assertion will then be true for countable interval unions and is equivalent to the condition that λ_F vanishes on λ_G null sets. Thus this condition is necessary and sufficient for absolute continuity. The condition can be phrased trivially without reference to measure because the values of λ_F and λ_G on intervals have simple expressions without intervention of measure theory. Similarly, singularity of a function of bounded variation relative to a monotone increasing function can be expressed without the help of measure theory.

8. The convergence set of a sequence of monotone functions

Let $F_\bullet$ be a sequence of monotone increasing functions from **R** into **R**. (The changes to be made below if the domain of definition of the functions is an interval of **R** will be obvious.) Suppose that the set S of points of convergence of the sequence (to a finite value) is dense in **R**, and define F on S as the limit of $F_\bullet$. Then F is a monotone increasing function from S into **R**, and the left and right limit functions of F, defined on **R**, are respectively left and right continuous. Moreover at each point s of **R**,

$$(8.1)\qquad \begin{aligned} &\limsup F_\bullet(s) \le \lim F_\bullet(s') = F(s') && (s < s' \in S),\\ &\liminf F_\bullet(s) \ge \lim F_\bullet(s') = F(s') && (s > s' \in S), \end{aligned}$$

so that

$$(8.2)\qquad F(s-) \le \liminf F_\bullet(s) \le \limsup F_\bullet(s) \le F(s+).$$

Thus the sequence $F_\bullet$ converges at every point s at which $F(s-) = F(s+)$. It follows that $\mathbf{R}-S$ is countable. *The function $s \to F(s+)$ is a right continuous function on* **R**, *and the sequence $F_\bullet$ converges to this function at its continuity points.*

9. Helly's compactness theorem for sequences of monotone functions

Roughly speaking, Helly's theorem states that for each positive constant c, the class

$$\{F\colon F \text{ is monotone increasing from } \mathbf{R} \text{ into } \mathbf{R} \text{ and } |F| \le c\}$$

of functions is compact. In later sections, this assertion will be put in precise topological contexts, in terms of the measures generated by monotone increasing functions.

Theorem (Helly). *If $F_\bullet$ is a uniformly bounded sequence of monotone increasing functions from* **R** *into* **R**, *there is a subsequence converging on* **R**.

The theorem is true for monotone functions on an arbitrary interval, with or without its endpoints, and the proof requires only trivial modifications.

Proof. According to the Bolzano-Weierstrass theorem, if s is a point of $\mathbf{R}$, some subsequence of $F_\bullet$ converges to a limit at s. If S' is a finite subset of $\mathbf{R}$, the preceding remark, applied repeatedly, yields a subsequence of $F_\bullet$ converging at every point of S'. Finally, if S' is countably infinite, the diagonal procedure yields a subsequence of $F_\bullet$ converging at every point of S'. Suppose now that S' has been chosen to be countable and dense in $\mathbf{R}$. The monotonicity of the functions has not yet been used, but according to Section 8, the fact that the functions of the given sequence are monotone increasing implies that the convergence set of $F_\bullet$ is $\mathbf{R}$ less an at most countable set. Hence a further subsequence of $F_\bullet$ can be chosen to obtain convergence everywhere on $\mathbf{R}$.

10. Intervals of uniform convergence of a convergent sequence of monotone functions

Theorem. *A sequence $F_\bullet$ of monotone increasing functions from* $\mathbf{R}$ *into* $\mathbf{R}$, *converging to a finite valued monotone increasing function F at the set of points of continuity of F, converges uniformly on each compact interval of continuity of F.*

Proof. If the sequence $F_\bullet$ does not converge uniformly on a compact continuity interval I of F, there is a subsequence $G_\bullet$ of $F_\bullet$, a sequence $s_\bullet$ of points of I, and a strictly positive ε, with the property that

$$|F(s_n)-G_n(s_n)| \geq \varepsilon \qquad (n \geq 1). \tag{10.1}$$

It can be supposed, going to further subsequences if necessary, that the sequence $s_\bullet$ is convergent, with limit s in I. If $s' < s < s''$ and if s' and s'' are points of continuity of F, then

$$F(s') \leq \liminf G_\bullet(s_\bullet) \leq \limsup G_\bullet(s_\bullet) \leq F(s''). \tag{10.2}$$

Since the first and last terms of this inequality have limit $F(s)$ when s' and s'' tend to s, it follows that the sequences $F(s_\bullet)$ and $G_\bullet(s_\bullet)$ have the same limit $F(s)$. This conclusion contradicts (10.1) and therefore there is the stated uniform convergence.

11. $\mathbb{C}(I)$ convergence of measure sequences on a compact interval I

If $I = [a,b]$ is a compact interval of $\mathbf{R}$, then (Section IV.9) a measure λ in $\mathbf{M}(I)$, the class of finite measures on the Borel subsets of I, generates a unique bounded monotone increasing function F_λ, the distribution function corresponding to λ, defined on I, right continuous on (a,b), and vanishing at a. Conversely, a monotone increasing function F on I, right continuous on (a,b) and vanishing at a, generates a unique measure λ_F in $\mathbf{M}(I)$. The monotone function generated by λ_F is F and the measure generated by F_λ is λ. The following theorem relates the convergence of a sequence of finite measures on $\mathbf{B}(I)$ to the convergence of the corresponding sequence of monotone functions.

Theorem. *Let I: $[a,b]$ be a compact subinterval of* **R**.

(a) *If a sequence $\lambda_\bullet$ in $\mathbf{M}(I)$ is $\mathbf{C}(I)$ convergent to λ, then the sequence $F_{\lambda_\bullet}$ converges to F_λ at the continuity points of F_λ and at b.*

(b) *Let $F_\bullet$ be a sequence of monotone increasing functions on $[a,b]$, right continuous on (a,b) and vanishing at a, and let F be a function with these same properties. If $F_\bullet$ converges to F at the continuity points of F, and at b, then the sequence $\lambda_{F_\bullet}$ is $\mathbf{C}(I)$ convergent to λ_F.*

Proof of (a). If $\lambda_\bullet$ is $\mathbf{C}(I)$ convergent to λ, and if $a < s \le b$, with $\lambda(\{s\}) = 0$ when $s < b$, that is, if s is a continuity point of F_λ when $s < b$, then $[a,s]$ is a set with a λ null boundary relative to I, and it follows from Theorem 8.6(b)(iii) that $\lim \lambda_\bullet([a,s]) = \lambda([a,s])$, that is, $\lim F_\bullet(s) = F_\lambda(s)$.

Proof of (b). If the sequence $F_\bullet$ converges to F at the continuity points of F, and at b, the sequence $\lambda_{F_\bullet}(I)$ is bounded, and therefore a subsequence of $\lambda_{F_\bullet}$ is $\mathbf{C}(I)$ convergent to some measure λ, according to Section **VIII**.5(c). According to part (a) of the present theorem, the corresponding subsequence of $F_\bullet$ is convergent to F_λ, at the continuity points of F_λ and at b, and therefore $F_\lambda = F$, that is, $\lambda = \lambda_F$. But then the $\mathbf{C}(I)$ convergent subsequences of $\lambda_\bullet$ must all have the same limit measure λ_F, and therefore the sequence $\lambda_\bullet$ is $\mathbf{C}(I)$ convergent to λ_F.

12. $\mathbf{C}_0(\mathbf{R})$ convergence of a measure sequence

In many applications one is dealing with monotone functions and measures on open subintervals of **R**. Since in fact the most common application is to **R** itself, this section deals with that case; the generalization to an arbitrary open subinterval of **R** is trivial. The monotone functions considered will be bounded, the measures will be finite, and **R** will be treated (see Section **VIII**.1) as a punctured compact space, the space obtained by removing the point at infinity of a one-point metric compactification of **R**. If λ is a finite measure on $\mathbf{B}(\mathbf{R})$, F_λ is the right continuous monotone function defined on **R** by setting $F_\lambda(s) = \lambda((-\infty,s])$. Then $F_\lambda(-\infty+) = 0$. In the other direction, if F is a bounded right continuous monotone increasing function on **R**, with $F(-\infty+) = 0$, the measure λ_F is the corresponding Lebesgue-Stieltjes measure generated by setting $\lambda_F\big((-\infty,b]\big) = F(b)$. The monotone function generated by λ_F is F, and the measure generated by F_λ is λ.

Theorem. (a) *If a sequence* $\lambda_\bullet$ *in* $\mathbf{M}(\mathbf{R})$ *is* $\mathbf{C}_0(\mathbf{R})$ *convergent to a measure* λ, *then the sequence* $F_{\lambda_\bullet}$ *is bounded, and*

$$\lim[F_{\lambda_\bullet}(b)-F_{\lambda_\bullet}(a)] = F_\lambda(b)-F_\lambda(a) = \lambda\big((a,b]\big) \tag{12.1}$$

whenever a and b are continuity points of F_λ.

(b) *Let* $F_\bullet$ *be a bounded sequence of monotone increasing right continuous functions on* $\mathbf{R}$ *with* $F_\bullet(-\infty+) = 0$. *If* $\lim[F_\bullet(b)-F_\bullet(a)]$ *exists for a and b in a dense subset of* $\mathbf{R}$, *then the sequence* $\lambda_{F_\bullet}$ *is* $\mathbf{C}_0(\mathbf{R})$ *convergent to a measure* λ.

Proof of (a). If the sequence $\lambda_\bullet$ has $\mathbf{C}_0(\mathbf{R})$ limit λ, boundedness of the sequence $F_{\lambda_\bullet}$ follows from the boundedness of $\lambda_\bullet(S)$. The limit equation (12.1) is true because, according to Theorem VIII.10(b)(iii), $\lim \lambda_\bullet(A) = \lambda(A)$ when A has compact closure and λ null boundary.

Proof of (b). Under the hypotheses of (b), the sequence $\lambda_{F_\bullet}(S)$ is bounded, and therefore a subsequence of $\lambda_{F_\bullet}$ is $\mathbf{C}_0(\mathbf{R})$ convergent to some measure, λ, according to Section VIII.9. According to (a) the corresponding subsequence of $[F_\bullet(b)-F_\bullet(a)]$ converges to $F_\lambda(b)-F_\lambda(a) = \lambda((a,b])$ whenever $\{a\}$ and $\{b\}$ are λ null, and therefore whenever neither a nor b is in a certain countable set. But then two $\mathbf{C}_0(\mathbf{R})$ convergent subsequences of $\lambda_\bullet$ must converge to the same values on intervals $(a,b]$ whenever neither a nor b is in an exceptional countable set, and therefore the two limit measures must agree on intervals whose endpoints are in a dense set. Thus the $\mathbf{C}_0(\mathbf{R})$ convergent subsequences of $\lambda_\bullet$ all have the same limit measure λ, and therefore the sequence $\lambda_{F_\bullet}$ is convergent to λ.

Example. Let F be a probability distribution function on $\mathbf{R}$, that is, F is a right continuous monotone increasing function with limits 0 and 1 at $-\infty$, and $+\infty$, respectively. Define $F_n(s) = F(s+n)$. When either $n\to-\infty$ or $n\to+\infty$, the corresponding sequence of measures has $\mathbf{C}_0(\mathbf{R})$ limit the identically vanishing measure. The sequence $F_1,F_{-1},F_2,F_{-2},\ldots$ does not converge at any point, but the corresponding sequence of measures again has the identically vanishing measure as $\mathbf{C}_0(\mathbf{R})$ limit. This example shows that there may not be $F_{\lambda_\bullet}$ convergence when there is $\mathbf{C}_0(\mathbf{R})$ convergence of $\lambda_\bullet$. Of course, even in this example the difference sequence limit in (12.1) exists, and is identically 0, as it should be according to Theorem 12.

13. Stable $\mathbf{C}_0(\mathbf{R})$ convergence of a measure sequence

The added condition that $\mathbf{C}_0(\mathbf{R})$ convergence of a sequence $\lambda_\bullet$ to λ be stable is the condition $\lim \lambda_\bullet(\mathbf{R}) = \lambda(\mathbf{R})$. According to Theorem **VIII**.13, a bounded sequence $\lambda_\bullet$ of measures on **R** is stably $\mathbf{C}_0(\mathbf{R})$ convergent with limit λ if and only if $\lim \lambda_\bullet(A) = \lambda(A)$ for every set A with a λ null boundary. Moreover it is sufficient if A in the stated condition is an interval with λ null boundary, because then there is $\mathbf{C}_0(\mathbf{R})$ convergence according to Theorem 12, and in addition $\lim \lambda_\bullet(\mathbf{R}) = \lambda(\mathbf{R})$ because $\partial\mathbf{R} = \emptyset$, Thus $\lambda_\bullet$ has stable $\mathbf{C}_0(\mathbf{R})$ limit λ if and only if not only $\lim F_{\lambda_\bullet} = F_\lambda$ at the continuity points of F_λ, but also $\lim F_{\lambda_\bullet}(+\infty-) = F_\lambda(+\infty-)$.

According to Theorem **VIII**.13, if f is a bounded continuous function on **R**, then under stable $\mathbf{C}_0(\mathbf{R})$ convergence of $\lambda_\bullet$ to λ,

$$\lim \int_{\mathbf{R}} f(s)\, dF_{\lambda_\bullet}(s) = \int_{\mathbf{R}} f(s)\, dF_\lambda(s). \tag{13.1}$$

Conversely the validity of (13.1) for every bounded continuous function f with a finite limit at $+\infty$ and the same limit at $-\infty$ is essentially the definition of stable $\mathbf{C}_0(\mathbf{R})$ convergence.

Example. If $x_\bullet$ is a sequence of random variables on a probability space, and if the sequence converges in measure, then the sequence $\lambda_\bullet$ of distributions of the random variables is stably $\mathbf{C}_0(\mathbf{R})$ convergent, because $\lim E\{f(x_\bullet)\} = \lambda_\bullet[f]$ is a convergent sequence whenever ϕ is a bounded continuous function from **R** into **R**.

The Lévy metric for stable $\mathbf{C}_0(\mathbf{R})$ convergence. According to Section **VIII**.12, there is a stable $\mathbf{C}_0(\mathbf{R})$ metric for the class $\mathbf{M}(\mathbf{R})$ of finite measures on **R**, that is, a metric under which this class is a complete metric space, with convergence in the metric the same as stable $\mathbf{C}_0(\mathbf{R})$ convergence. An equivalent metric adapted to the class of corresponding bounded monotone increasing functions (right continuous and with $F(-\infty+) = 0$) was devised by P. Lévy. If F is such a function, fill in its graph at each discontinuity point s with the vertical line segment from $((s,F(s-))$ to $(s,F(s))$. If F and G are two such functions, each line of slope -1, in the plane of the filled graphs, determines a line segment with endpoints on the graphs, and Lévy defined the distance from F to G as the maximum length of all these line segments between the two graphs. The verification that this distance definition has the stated properties is left to the reader.

14. The characteristic function of a measure

Let λ be a finite measure on **R** with corresponding distribution function F_λ. The *characteristic function of* λ, also called the *characteristic function of* F_λ, is the function Φ from **R** into the complex numbers defined by

$$(14.1)\qquad \Phi(t) = \int_{\mathbf{R}} e^{ist}\lambda(ds) = \int_{\mathbf{R}} e^{ist}\,dF_\lambda(s).$$

The characteristic function of a real valued measurable function is defined as the characteristic function of its distribution. Thus, in a probability context, the characteristic function of a random variable x is the function $t \mapsto E\{e^{ixt}\}$. The characteristic function of a measure λ is bounded in absolute value by $\lambda(\mathbf{R})$ and is continuous, with value $\lambda(\mathbf{R}) = F_\lambda(\infty-)$ at the origin. If λ is absolutely continuous with respect to Lebesgue measure, $\Phi/\sqrt{2\pi}$ is the Fourier transform of F_λ' and therefore, according to the general remarks on Fourier transforms in Section **VII.**11, F_λ' would be expected in some sense to be the inverse Fourier transform of $\Phi/\sqrt{2\pi}$,

$$(14.2)\qquad F_\lambda'(s) = \frac{1}{2\pi}\int_{\mathbf{R}} e^{-its}\,\Phi(t)\,dt,$$

and thereby to be determined by its characteristic function. These remarks are not rigorous mathematics but in fact even without a hypothesis of absolute continuity, a version of (14.2) suggested by formal integration is true, according to the following theorem.

Theorem (Lévy). *If* λ *is a finite measure on* **R**, *with characteristic function* Φ, *then*

$$(14.3)\qquad \frac{F_\lambda(b)+F_\lambda(b-)}{2} - \frac{F_\lambda(a)+F_\lambda(a-)}{2} = \lim_{\alpha\to\infty}\int_{-\alpha}^{\alpha}\frac{e^{-itb} - e^{-ita}}{-2\pi i t}\,\Phi(t)\,dt.$$

Proof. The integral on the right becomes, after the formula for Φ is inserted and the integration order is reversed,

$$(14.4)\qquad \int_{\mathbf{R}}\lambda(ds)\int_{-\alpha}^{\alpha}\frac{e^{-it(b-s)} - e^{-it(a-s)}}{-2\pi i t}\,dt$$

$$= \int_{\mathbf{R}}\lambda(ds)\int_{-\alpha}^{\alpha}\frac{\sin t(b-s) - \sin t(a-s)}{2\pi t}\,dt.$$

According to Section **VII.**10, when $\alpha\to+\infty$ the inner integral on the right has limit 1 for s in (a,b), limit 1/2 at a and b, and limit 0 elsewhere; since this integral is bounded uniformly as s varies, the Lebesgue dominated convergence theorem is applicable and yields (14.3).

15. Stable $\mathbb{C}_0(\mathbf{R})$ convergence of a sequence of probability distributions

In the following discussion, the issue is the stable $\mathbf{C}_0(\mathbf{R})$ convergence of a sequence of finite measures on **R**. It is trivial that under this convergence (13.1) is true for every bounded continuous function f from **R** into the complex numbers as well as into the real numbers. The most important case is that of a sequence of probability measures on **R** converging to a probability measure on **R**, and the discussion is restricted to that case, but the results are easily translated to cover the general case.

Theorem (**Lévy**). *Let* $\lambda_\bullet$ *be a sequence of probability measures on* **R**, *with characteristic function sequence* $\Phi_\bullet$.

If $\lambda_\bullet$ *is stably* $\mathbf{C}_0(\mathbf{R})$ *convergent to a probability measure* λ, *equivalently if the sequence* $F\lambda_\bullet$ *converges to* $F\lambda$ *at the continuity points of* $F\lambda$, *then the sequence* $\Phi_\bullet$ *converges to the characteristic function of* λ, *uniformly on every finite interval. Conversely, if the sequence* $\Phi_\bullet$ *converges, and the convergence is uniform on some neighborhood of the origin, then the limit function is the characteristic function of a probability measure* λ *and the sequence* $\lambda_\bullet$ *converges stably* $\mathbf{C}_0(\mathbf{R})$ *to* λ.

Proof. If the sequence $\lambda_\bullet$ converges stably $\mathbf{C}_0(\mathbf{R})$ to λ, Theorem VIII.13 implies that the sequence $\Phi_\bullet$ of characteristic functions converges to the characteristic function Φ of λ, and that the convergence is uniform on finite intervals. In fact, according to that theorem, if $t_\bullet$ is a sequence in **R**, with limit t, then $\lim \Phi_\bullet(t_\bullet) = \Phi(t)$.

Conversely, supppose that the sequence $\Phi_\bullet$ of characteristic functions converges to some function Φ. The sequence $\lambda_\bullet$ is a bounded sequence of measures, and therefore there is a $\mathbf{C}_0(\mathbf{R})$ convergent subsequence, with some limit measure λ. According to Theorem 14, λ is uniquely determined by Φ, and therefore the sequence $\lambda_\bullet$ is $\mathbf{C}_0(\mathbf{R})$ convergent, because every $\mathbf{C}_0(\mathbf{R})$ convergent subsequence has the same limit measure λ. Since the sequence $\Phi_\bullet$ has limit Φ,

$$\lim \int_{\mathbf{R}} (1 - \cos ts)\, \lambda_\bullet(ds) = 1 - \Re\Phi(t). \tag{15.1}$$

The integrals on the left define a bounded sequence of functions of t, and therefore Lebesgue's dominated convergence theorem is applicable to yield

$$\begin{aligned}\lim \frac{1}{\delta} \int_0^\delta dt \int_{\mathbf{R}} (1 - \cos ts)\, \lambda_\bullet(ds) \\ = \lim \int_{\mathbf{R}} \left(1 - \frac{\sin \delta s}{\delta s}\right) \lambda_\bullet(ds) = 1 - \frac{1}{\delta} \int_0^\delta \Re\Phi(t)\, dt.\end{aligned} \tag{15.2}$$

Since there is $\mathbf{C}_0(\mathbf{R})$ convergence of $\lambda_\bullet$ to λ,

$$(15.3)\qquad \lim \int_{\mathbf{R}} \frac{\sin \delta s}{\delta s}\, \lambda_\bullet(ds) = \int_{\mathbf{R}} \frac{\sin \delta s}{\delta s}\, \lambda(ds) = \frac{1}{\delta}\int_0^\delta \Re\Phi(t)\, dt.$$

When δ tends to 0, the Lebesgue dominated convergence theorem is applicable to the second integral in (15.3), yielding the limit $\lambda(\mathbf{R})$. Now by hypothesis, Φ is the limit of a sequence of characteristic functions, converging uniformly in some neighborhood of the origin, and therefore $\Phi(0) = 1$ and Φ is continuous in a neighborhood of the origin. Thus the third integral in (15.3) has limit 1 when δ tends to 0. Hence λ is a probability measure, the sequence $\lambda_\bullet$ is stably $\mathbf{C}_0(\mathbf{R})$ convergent to λ, and, according to the first part of the theorem, Φ is the characteristic function of λ.

16. Application to a stable $\mathbb{C}_0(\mathbf{R})$ metrization of $\mathbb{M}(\mathbf{R})$

Theorem 15 suggests that one simple stable $\mathbf{C}_0(S)$ metrization of the class of probability measures on $\mathbf{R}$ is by way of characteristic functions. For example, if λ_1 and λ_2 are probability measures on $\mathbf{R}$, with respective characteristic functions Φ_1 and Φ_2, define the distance between the probability measures as

$$(16.1)\qquad \sum_{n=1}^{\infty} 2^{-n} \sup\left\{ |\Phi_1(s) - \Phi_2(s)| : |s| \le n \right\}.$$

Under this metric, a sequence $\lambda_\bullet$ of probability measures converges stably $\mathbf{C}_0(\mathbf{R})$ to a probability measure λ if and only if the corresponding characteristic function sequence converges to the characteristic function of λ locally uniformly on $\mathbf{R}$, and the Cauchy condition for convergence implies the existence of a stable $\mathbf{C}_0(\mathbf{R})$ limit. Thus the metric is a metric for $\mathbf{C}_0(\mathbf{R})$ convergence, making the class of probability measures a complete metric space. It is left to the reader to check that this metric also makes $\mathbf{M}(\mathbf{R})$ a complete metric space.

17. General approach to derivation

Let $(S,\mathbf{S})$ be a measurable space, λ be a finite measure on $\mathbf{S}$, and μ be a finite signed measure on $\mathbf{S}$. In many contexts, each point s of S is associated with a system S_s of subsets of S in such a way that, at λ almost every point s, the denominator in the following relation does not vanish, and that the limit

$$(17.1)\qquad \lim \frac{\mu(A)}{\lambda(A)} = f(s)$$

exists and is finite, when A runs through S_s in some prescribed order or partial order. For example, according to Theorem X.4, a Lebesgue-Stieltjes signed measure on **R** has derivative lim $\mu(A)/\lambda(A)$ at λ almost every point s, when A is a closed interval containing s that shrinks to s. A second example, in an abstract context, will be given in Section **XI**.17: a sequence of finer and finer partitions of S is chosen, $A_n(s)$ is the cell of the nth partition containing s, and then for λ almost every s, lim $\mu(A_\bullet(s))/\lambda(A_\bullet(s))$ exists.

In such a context, the difference between the limit superior and limit inferior of $\mu(A)/\lambda(A))$ at s when A runs through S_s in the prescribed way will be denoted by $\bar{D}_s[\mu,\lambda]$, and the limit, when it exists, will be denoted by $D_s[\mu,\lambda]$. If, for every pair λ, μ, $D_\bullet[\mu,\lambda]$ exists λ almost everywhere, then D will be called a *derivation operator* for the measurable space, a *Radon-Nikodym derivation operator* if $D_\bullet[\mu,\lambda]$ is a version of the Radon-Nikodym derivative of the absolutely continuous component of μ relative to λ. The exact context is irrelevant in the following. Recall the notation $|\bullet|$ for the absolute variation of a signed measure. The inequality

$$(17.2)\qquad \left|\frac{\mu(A(s))}{\lambda(A(s))}-a\right| = \left|\frac{\mu(A(s))-\lambda(A(s))}{\lambda(A(s))}\right| \le \frac{|\mu-a\lambda|(A(s))}{\lambda(A(s))}$$

suggests a useful stronger condition on the derivation operator than $D_s[\lambda,\mu] = a$, the condition

$$(17.3)\qquad D_s[|\mu-a\lambda|,\lambda] = 0,$$

which has the advantage that the ratios leading to derivation are ratios of positive quantities, and the limit to be verified is 0. A point s at which (17.3) is true is a *Lebesgue point* for μ relative to λ, and of course relative to the prescribed derivation procedure.

Theorem. *Let* $(S,\mathbf{S})$ *be a measurable space,* D *be a Radon-Nikodym derivation operator for* $(S,\mathbf{S})$, μ *be a finite signed measure on* $\mathbf{S}$, *and* λ *be a finite measure on* $\mathbf{S}$. *Then* λ *almost every point of* S *is a Lebesgue point for* μ *relative to* λ.

In view of the Lebesgue decomposition of μ relative to λ, it is sufficient to prove the theorem when μ is λ singular and when μ is λ absolutely continuous.

Proof when μ is λ singular. When μ is λ singular, that is, by definition, when $|\mu|$ is λ singular, $D_s[|\mu|,\lambda] = D_s[\mu,\lambda] = 0$ for λ almost every s, because D is a Radon-Nikodym derivation operator.

Proof when μ is λ absolutely continuous. Let f be the Radon-Nikodym derivative of μ relative to λ. If r is a real number, then $|\mu-r\lambda|$ is a finite measure absolutely continuous relative to λ, with Radon-Nikodym derivative $|f-r|$:

$$(17.4)\qquad D_s[|\mu-r\lambda|,\lambda] = |f(s)-r|,$$

for λ almost every s. Hence there is a λ null set B with the property that (17.4) is true for every rational value of r, when s is not in B. Moreover, if r is an arbitrary real number and if r' is rational,

$$\text{(17.5)} \qquad \frac{1}{\lambda(A)}\Big|\,|\mu-r\lambda|(A)-(|\mu-r'\lambda|(A)\Big|$$

$$=\frac{1}{\lambda(A)}\Big|\int_A [|f-r|-|f-r'|]\,d\lambda\Big| \le \frac{1}{\lambda(A)}\int_A |r-r'|\,d\lambda = |r-r'|.$$

It follows that when s is not in B, $\bar{D}_s[|\mu-r\lambda|,\lambda] \le 2|r-r'|$, and therefore the derivative on the left in (17.4) exists, and (17.4) is true. Hence, for s not in B, (17.4) is true with r replaced by $f(s)$, that is, every point of $S-B$ is a Lebesgue point for μ relative to λ.

Example. Let the measure space be **R** under Lebesgue measure, let ds refer to Lebesgue measure, let derivation refer to the usual difference quotient limit procedure, let f be a Lebesgue measurable function on **R**, integrable over every finite interval, and define $\mu(A) = \lambda[f1_A]$, for bounded Lebesgue measurable sets A. The fact that almost every point is a Lebesgue point of μ relative to Lebesgue measure is the fact that for (Lebesgue measure) almost every point t,

$$\text{(17.6)} \qquad \lim_{\delta\to 0} \frac{1}{\delta}\int_t^{t+\delta} |f(t)-f(s)|\,ds = 0.$$

This result is stronger than the result that the integrand f is almost everywhere the derivative of its integral.

18. A ratio limit lemma

In many applications of analysis one deals with sequences of integrals of the form

$$\text{(18.1)} \qquad \lambda[K_n] = \int_0^b K_n(s)\,\lambda(ds),$$

where $K_\bullet$ is a sequence of positive functions, λ is a measure on the integration interval $[0,b]$, and the sequence $K_\bullet$ has limit 0 on $(0,b]$ but converges to 0 in such a way that as n increases, the integration outside an interval containing 0 becomes negligible, compared to the value of the integral inside the interval. Thus, roughly, $\lambda[K_n] \sim K_n(0)\lambda(I_n)$ for I some interval with left-hand endpoint 0, whose choice is the same for every choice of λ. If this guess is indeed true, then if μ is a second measure on $[0,b]$, the ratio $\mu[K_n]/\lambda[K_n]$ would be near

$\mu(I_n)/\lambda(I_n)$ for large n, and therefore would tend to $d\mu/d\lambda$ at 0, if this derivative exists. The following lemma gives a simple set of conditions on $K_\bullet$ that will be used in identifying a context in which this reasoning can be made precise. It will be convenient to write integrals with respect to a measure in terms of the monotone increasing function generating the measure.

Lemma. *Let $K_\bullet$ be a sequence of strictly positive continuous functions on the interval* $[0,b]$. *Suppose that*

(a) *each derivative K_n' exists, is continuous, and $K_n' \le 0$;*

(b) *if $0 < a < b$, the function* $\liminf K_\bullet/K_\bullet(a)$ *has limit $+\infty$ at* 0, *and has Lebesgue integral $+\infty$ over every neighborhood of* 0.

Let F and G be monotone increasing functions on $[0,b]$, *with $F(0) = G(0) = 0$, right continuous on $(0,b)$, and satisfying one of the following conditions:*

(c') $F(0+) = 0,\ G(0+) > 0$;

(c'') $G(0+) = 0,\ \lim_{s\to 0} F(s)/G(s) = 0$, *and* $\limsup_{s\to 0} s/G(s) < +\infty$.

Then

$$\lim \frac{\int_0^b K_\bullet(s)\, dF(s)}{\int_0^b K_\bullet(s)\, dG(s)} = 0. \tag{18.2}$$

Observe that the hypotheses on $K_\bullet$ imply that $\lim K_\bullet(a)/K_\bullet(0) = 0$.

Proof. In the following, ds will refer to one-dimensional Lebesgue measure. Integration by parts of Riemann-Stieltjes integrals will be used.

Under (c'), if $0 < a < b$,

$$\int_0^a K_n\, dF \le K_n(a)F(a) - F(a)\int_0^a K_n'\, ds = F(a)K_n(0), \tag{18.3}$$

$$\int_a^b K_n\, dF \le F(b)K_n(a),$$

and therefore the ratio in (18.2) is at most

$$\frac{F(a)K_n(0) + F(b)K_n(a)}{G(0+)K_n(0)}. \tag{18.4}$$

When n increases, the ratio in (18.4) tends to $F(a)/G(0+)$, which can be made arbitrarily small by choosing a small. Hence (18.2) is true under (c').

Under (c"), choose ε strictly positive but small enough to satisfy the inequality $\limsup_{s\to 0} s/G(s) < 1/\varepsilon$, and then choose a strictly positive and so small that $F \le \varepsilon G$ and $G \ge \varepsilon s$ on $[0,a]$. Then

$$\text{(18.5)}\qquad \int_0^a K_n\,dF = F(a)K_n(a) - \int_0^a FK_n'\,ds \le F(a)K_n(a) - \varepsilon\int_0^a GK_n'\,ds$$

$$= K_n(a)[F(a)-\varepsilon G(a)] + \varepsilon\int_0^a K_n\,dG \le \varepsilon\int_0^a K_n\,dG.$$

Similarly,

$$\text{(18.6)}\qquad \int_0^a K_n\,dG = G(a)K_n(a) - \int_0^a GK_n'\,ds \ge G(a)K_n(a) - \varepsilon\int_0^a sK_n'\,ds$$

$$= K_n(a)[G(a)-\varepsilon a] + \varepsilon\int_0^a K_n\,ds \ge \varepsilon\int_0^a K_n\,ds.$$

Hence

$$\text{(18.7)}\qquad \limsup \frac{\int_0^b K_\bullet(s)\,dF(s)}{\int_0^b K_\bullet(s)\,dG(s)} \le \varepsilon + \limsup \frac{\int_a^b K_\bullet dG}{\int_0^a K_\bullet\,ds}.$$

Since the numerator on the right is majorized by $K_n(a)G(b)$, the superior limit on the right-hand side is majorized by

$$\text{(18.8)}\qquad G(b)\left[\liminf_{n\to\infty}\int_0^a [K_n/K_n(a)]\,ds\right]^{-1}$$

$$\le G(b)\left[\int_0^a \liminf_{n\to\infty}[K_n/K_n(a)]\,ds\right]^{-1} = 0,$$

and therefore the lemma is true.

19. Application to the boundary limits of harmonic functions

Recall the Riesz-Herglotz representation in Section **VIII.14** of a positive harmonic function on a disk B_β of radius β, center the origin:

$$\text{(19.1)}\qquad u(z) = \int_{\partial B_\beta} \frac{\beta^2 - r^2}{|\zeta - re^{i\theta}|^2}\,\lambda(d\zeta) \qquad (z = re^{i\theta}).$$

In particular, when λ is one dimensional Lebesgue measure divided by $2\pi\beta$ the function u is identically 1. Observe that when ϑ is fixed and z tends to a boundary point $\beta e^{i\vartheta}$, the integrand tends to 0 at all other boundary points. Thus this context is precisely that envisaged in Section 18.

Theorem (Fatou-Doob). *If u and h are positive harmonic functions on a disk, determined respectively by the Riesz-Herglotz measures λ_u and λ_h, then, on radial approach to λ_h almost every boundary point of the disk, u/h has limit the Radon-Nikodym derivative of the absolutely continuous component of λ_u relative to λ_h.*

The Radon-Nikodym derivative in question is uniquely determined only up to a λ_h null set, and this lack of uniqueness matches the fact that the radial limit exists aside from a λ_h null boundary set. In other words, the radial limit function (augmented say by 0 where the limit does not exist) is one version of this Radon-Nikodym derivative. According to Theorem 4, one version of this Radon-Nikodym derivative at a boundary point can be obtained as the pointwise derivative defined at the boundary point s as the limit, when this limit exists, of $\mu_u(I)/\mu_h(I)$ when the closed interval I containing s shrinks to s, and it is this derivative that will be considered from now on. It will be proved that the radial limit of u/h exists at every boundary point that (i) is a Lebesgue point for λ_u relative to λ_h and (ii) is a point at which the derivative of Lebesgue measure on the boundary relative to λ_h exists and is finite. The set of such boundary points is the complement of a λ_h null set. It is no restriction to consider only the boundary point $z = \beta$. Write the Riesz-Herglotz representations of u and h at the point at distance r from the origin on the positive real axis in the form

$$(19.2)\qquad u(r) = \int_0^\pi \frac{\beta^2 - r^2}{|\beta e^{is} - r|^2}\, dF(s), \qquad\qquad h(r) = \int_0^\pi \frac{\beta^2 - r^2}{|\beta e^{is} - r|^2}\, dG(s),$$

where F and G are positive monotone increasing functions on $[0,\pi]$ vanishing at 0. To prove the assertion on the existence of a radial limit, it is sufficient to prove that $\lim_{r\to\beta} u(r)/h(r) = 0$ under the combined assumptions that dF/dG vanishes at the origin, and that $\lim\sup_{s\to 0} s/G(s)$ is finite. Lemma 18 will be applied to obtain this result.

Let $\beta_\bullet$ be an increasing sequence of strictly positive numbers with limit β. Take b in the lemma as π, and define $K_n(s)$ as the value of the integrand in (19.2) with $r = \beta_n$. Condition (a) of the lemma is obviously satisfied. The limit inferior in condition (b) is an actual limit that majorizes $const.s^{-2}$. Thus the lemma is applicable and yields the desired zero limit.

With a little more care, this method shows that the limit in question exists not only as a radial limit but as a nontangential limit.

XI
Conditional Expectation; Martingale Theory

1. Stochastic processes

In probability theory, a family $\{x_t, t \in I\}$ of random variables on a probability space $(S,\mathbf{S}',P)$, that is, a family of measurable functions from S into $\bar{\mathbf{R}}$, is sometimes glorified by the name *stochastic process* or, simply, *process*. At the point s of S, the random variable x_t has value $x_t(s)$, a point of the state space of the process, and the function $x_\bullet(s)$, from I into the state space, is a *sample function* of the process. The set I is the *parameter set* of the family. There is no definition of the term *stochastic process* more specific than that just given, but the term is usually reserved for families with some interesting property. To loosely paraphrase a judge discussing a somewhat different concept, probabilists cannot define stochastic processes, but they recognize one when they see it! Typical examples of the application of the term are to martingales and to Markov processes, both defined in this chapter.

An attentive reader will observe, and perhaps resent, that in other chapters a function is f or g, and so on, whereas in this chapter a function is more likely to be x or y, and so on, at the other end of the alphabet. This difference is traditional, and is one of the principal features that distinguishes probability from the rest of measure theory.

2. Conditional probability and expectation

Let $(S,\mathbf{S}',P)$ be a probability space, x be an integrable random variable, and A be a measurable set. In Section III.6, in the context of a discrete probability space, the conditional expectation $E\{x|A\}$ was defined by

$$E\{x|A\} = E\{x1_A\}/P\{A\} \tag{2.1}$$

when A is not null. With this definition, meaningful whenever x is an integrable random variable on a probability space, if $S_1,\dots,S_n$ is a partition of S, that is, if these sets are pairwise disjoint measurable sets with union S, and if $E\{x|S_j\}$ is defined arbitrarily when S_j is null,

$$\Sigma E\{x|S_\bullet\}P\{S_\bullet\} = E\{x\}, \tag{2.2}$$

and more generally, if A is a union of sets of the partition,

$$\sum_{S_j \subset A} E\{x|S_j\}P\{S_j\} = \int_A x dP. \tag{2.3}$$

Let **S** be the algebra of unions of partition sets. A function from S into **R** is measurable from (S,**S**) into **R**, in short, is **S** measurable, if and only if the function is constant on each partition cell. If $E\{x|\mathbf{S}\}$ is defined as the random variable with value $E\{x|S_j\}$ on each nonnull cell S_j and is any constant on each null partition cell, (2.3) can be written in the elegant form

$$\int_A E\{x|\mathbf{S}\}\, dP = \int_A x dP \qquad (A \in \mathbf{S}). \tag{2.4}$$

The integrand on the left is uniquely determined on the nonnull partition cells by (2.4) and **S** measurability. This discussion leads to the following Kolmogorov definition of conditional expectation.

Definition of conditional expectation. Let (S,**S**',P) be a probability space, **S** be a σ algebra of measurable sets, and x be an integrable random variable. Then $E\{x|\mathbf{S}\}$, *the conditional expectation of x given* **S**, is a random variable satisfying the following conditions:

(a) $E\{x|\mathbf{S}\}$ is **S** measurable and integrable;

(b) $E\{x|\mathbf{S}\}$ satisfies (2.4).

This definition is not vacuous, because the right side of (2.4) defines a function of A, a measure on **S**, which is absolutely continuous relative to the restriction $P_\mathbf{S}$ of P to **S**, and therefore (Radon-Nikodym theorem) there is a function satisfying (a) and (b), uniquely determined up to $P_\mathbf{S}$ null sets. The symbol $E\{x|\mathbf{S}\}$ refers to a function, any function, satisfying (a) and (b), a function, not an equivalence class of functions. This convention means that in almost every discussion involving conditional expectations, *almost everywhere* or *almost surely*, or their respective abbreviations *a.e.*, *a.s.*, will be sure to appear almost everywhere.

In the extreme case when **S** is the σ algebra consisting of the whole space and the empty set, $E\{x|\mathbf{S}\} = E\{x\}$; *almost everywhere* is unnecessary in this case. At the other extreme, when $\mathbf{S} = \mathbf{S}'$, $E\{x|\mathbf{S}\} = x$ almost everywhere. The new conditional expectation definition is consistent with the old, in the sense that if $S_\bullet$ is a partition of S, as discussed at the beginning of this section, and if **S** is the σ algebra of unions of partition cells, $E\{x|\mathbf{S}\}$ is constant on each partition cell S_j; the constant is arbitrary if S_j is null, and is the jth conditional expectation in (2.2) if S_j is not null.

Definition of conditional probability. If A is measurable, $P\{A|\mathbf{S}\}$ is defined as $E\{1_A|\mathbf{S}\}$; more precisely, the first function is to be taken as any one of the possible conditional expectation functions. (This last proviso will be omitted from now on in similar contexts.)

Conditioning by a random variable family. If $x_\bullet$ is a family of random variables, $E\{x|x_\bullet\}$ is defined as $E\{x|\sigma(x_\bullet)\}$, where (see Section V.1) $\sigma(x_\bullet)$ is the σ algebra of sets defined by measurable conditions on the random variables $x_\bullet$, that is, the smallest σ algebra $\mathbf{S}$ making every random variable x_t measurable from $(S,\mathbf{S})$ into $\bar{\mathbf{R}}$.

Example (a), **Densities on $\mathbf{R}^2$**. Let P be a probability distribution on the plane, given by a Radon-Nikodym (Lebesgue measurable) density $p(\bullet,\bullet)$ with respect to two-dimensional Lebesgue measure, that is, the measure assigned to a Lebesgue measurable subset of $\mathbf{R}^2$ is the integral of p (with respect to Lebesgue measure on $\mathbf{R}^2$) over the set. To avoid a few *almost everywheres*, it is supposed in the following, without loss of generality, that p is Borel measurable. Let x_1 and x_2 be the coordinate functions on the plane. Then the distributions of x_1 and x_2 are the marginal distributions given by the respective densities (with respect to one- dimensional Lebesgue measure),

$$(2.5) \qquad p_1 = \int_{-\infty}^{+\infty} p(\bullet,\alpha)\,d\alpha, \quad \text{and} \quad p_2 = \int_{-\infty}^{+\infty} p(\alpha,\bullet)\,d\alpha.$$

Here $d\alpha$ refers to Lebesgue measure on $\mathbf{R}$. The random variables x_1 and x_2 are mutually independent if and only if p defines a product measure, a measure necessarily given (P almost everywhere on $\mathbf{R}^2$) by the density that is the product of the marginal densities relative to one-dimensional Lebesgue measure: $p(\alpha,\beta) = p_1(\alpha)p_2(\beta)$.

Without the hypothesis of independence, one version of the conditional distribution of x_2 for given x_1 is given by

$$(2.6) \qquad P\{x_2 \in A_2 | x_1\} = \int_{A_2} p(x_1,\alpha)\,d\alpha / p_1(x_1),$$

determined by the conditional density (relative to Lebesgue measure on $\mathbf{R}$) $p(x_1,\bullet)/p_1(x_1)$. In fact the function of x_1 in (2.6), when integrated with respect to P measure over the $\mathbf{R}^2$ set $\{x_1 \in A_1\}$ yields $P\{x_1 \in A_1, x_2 \in A_2\}$. Observe that the random variable $p_1(x_1)$ vanishes P almost nowhere. The fact that p is Borel measurable implies that every integrand here is Borel measurable on whatever space is relevant.

Example (b). Let P be a probability distribution on $\mathbf{R}^2$, and again let x_1 and x_2 be the coordinate functions on $\mathbf{R}^2$. Suppose that P is carried by the line through the origin of slope 1. Then $x_1=x_2$ almost surely. In this case $P\{x_2 \in A_2 | x_1\} = 1_{A_2}(x_1)$ almost surely.

Example (c). Transition functions and the processes they generate. Let $(S,\mathbb{S})$ be a measurable space. A *transition function* on the space is a function $(s,A)\to p(s,A)$ from $S\times\mathbb{S}$ into **R** with the following properties:

(c1) For each point s of S, the function $p(s,\bullet)$ is a probability measure on $\mathbb{S}$.

(c2) For each set A in $\mathbb{S}$, the function $p(\bullet,A)$ is $\mathbb{S}$ measurable.

Transition functions generalize the stochastic matrices discussed in Section III.7. In intuitive language, in a probabilistic context, $p(s,A)$ is the probability of a transition from the point s into a point of the set A. For example, let $(S,\mathbb{S})$ be a measurable space, p be a probability measure on $\mathbb{S}$, and $p(\bullet,\bullet)$ be a transition function on $(S,\mathbb{S})$. Define a measure P on the measurable space $\big(S\times S,\sigma(\mathbb{S}\times\mathbb{S})\big)$ for which (generalization of III(7.3))

(2.7)
$$P\{A_1\times A_2\} = \int_{A_1} p(s_1,A_2)p(ds_1) \qquad (A_i \in \mathbb{S}),$$

by the definition (integral over $S\times S$)

(2.8)
$$P(A) = \int p(ds_1)\int 1_A\, p(s_1,ds_2) \qquad \big(A \in \sigma(\mathbb{S}\times\mathbb{S})\big).$$

(It is left to the reader to check that the inner integral actually defines an $\mathbb{S}$ measurable function.) For this measure, if x_1 and x_2 are the coordinate functions of $S\times S$,

(2.9)
$$P\{x_1 \in A_1\} = p\{A_1\}, \quad P\{x_2 \in A_2 \| x_1\} = p(x_1,A_2) \quad \text{a.s.} \quad (A_i \in \mathbb{S}),$$

and if f is a Borel measurable function from **R** into **R** for which $f(x_2)$ is integrable then

(2.10)
$$E\{f(x_2) \| x_1\} = \int_S f(\zeta)\, p(x_1,d\zeta) \qquad \text{a.s.}$$

One can go on, following Section III.7, but no new ideas beyond measure niceties are involved. These measure niceties are so unnice that the probability measure P defined on $\sigma(\mathbb{S}\times\mathbb{S})$ by an initial distribution together with a transition function does not quite provide the most general measure on $\sigma(\mathbb{S}\times\mathbb{S})$ unless restrictions are imposed on the measurable space $(S,\mathbb{S})$.

In particular, a common construction of a probability measure P on the n-fold product $(S^n, s(\mathbb{S}^n))$ applies n transition functions $p^{(\cdot)}$ on $\mathbb{S}$ together with an initial point s_0 by the definition (integral over S^n)

$$(2.11) \qquad P(A) = \int p^{(1)}(s_0, ds_1) \int p^{(2)}(s_1, ds_2) \ldots \int 1_A \, p^{(n)}(s_{n-1}, ds_n) \qquad (A \in \sigma(\mathbb{S}^n)).$$

If $x_1, \ldots, x_n$ are the coordinate functions of S^n, these functions are random variables for which

$$(2.12) \quad P\{x_{m+1} \in A_{m+1} | x_0, \ldots, x_m\} = P\{x_{m+1} \in A_{m+1} | x_m\} = p^{(m+1)}(x_m, A_{m+1}) \qquad \text{a.s. } (0 \le m < n)$$

and more generally, if f is a Borel measurable function from $\mathbf{R}$ into $\mathbf{R}$ for which $f(x_{m+1})$ is integrable then (2.12) implies

$$(2.13) \quad E\{f(x_{m+1}) | x_0, \ldots, x_m\} = E\{f(x_{m+1}) | x_m\} = \int_S f(\zeta) p^{(m+1)}(x_m, d\zeta) \qquad \text{a.s. } (0 \le m < n).$$

Observe that the conditional probabilities and expectations in (2.12) and (2.13) depend only on the last conditioning variable x_m. This property will be defined as the Markov property in the general definition of this property in Section 4.

3. Conditional expectation properties

In the following list of properties of conditional expectations, since all σ algebras are to be σ algebras of measurable sets and all random variables are to be integrable, these hypotheses are omitted in the listing except in (d), (e), (f), and (h), in which the omission might lead to misunderstanding. Some of the properties can be stated more generally: $E\{x | \mathbb{S}\}$ can be defined for x positive, not necessarily integrable, but somewhat restricted, and (d) and (e) can be correspondingly extended, but the restrictions make the extensions not very useful. Proofs are given at the end of the list.

(a) *If $x=y$ a.e. then $E\{x | \mathbb{S}\} = E\{y | \mathbb{S}\}$ a.e.*

(b) *If a and b are constants, $E\{ax+by | \mathbb{S}\} = aE\{x | \mathbb{S}\} + bE\{y | \mathbb{S}\}$ a.e.*

(c) *If $x \le y$ a.e. then $E\{x | \mathbb{S}\} \le E\{y | \mathbb{S}\}$ a.e. In particular, $|E\{x | \mathbb{S}\}| \le E\{|x| | \mathbb{S}\}$ a.e.*

(d) **Beppo-Levi's theorem for conditional expectations.** *If $x_\bullet$ is an increasing sequence of positive random variables with integrable limit x, then*

$$(3.1) \qquad \lim E\{x_\bullet | \mathbb{S}\} = E\{x | \mathbb{S}\} \text{ a.e.}$$

In particular, if $A_\bullet$ is a disjunct sequence of measurable sets, with union A, then

$$\sum P\{A_\bullet|\mathbf{S}\} = P\{A|\mathbf{S}\} \text{ a.e.} \tag{3.2}$$

(e) **Fatou's theorem for conditional expectations.** *Let $x_\bullet$ be a sequence of positive integrable random variables, and define $x = \liminf x_\bullet$. If x is integrable, then*

$$E\{x|\mathbf{S}\} \le \liminf E\{x_\bullet|\mathbf{S}\} \quad \text{a.e.} \tag{3.3}$$

Moreover $\lim E\{|x-x_\bullet||\mathbf{S}\} = 0$ *almost everywhere where*

$$E\{x|\mathbf{S}\} = \lim E\{x_\bullet|\mathbf{S}\}. \tag{3.4}$$

(f) **Lebesgue's dominated convergence theorem for conditional expectations.** *Let $x_\bullet$ be a sequence of random variables, and suppose that* $\sup|x_\bullet|$ *is integrable. If the sequence $x_\bullet$ converges almost everywhere [in measure] to a random variable x, then* $\lim E\{x_\bullet|\mathbf{S}\} = E\{x|\mathbf{S}\}$ *almost everywhere [in measure].*

(g) *If x is a constant function, or more generally, if x is $\mathbf{S}$ measurable and integrable, then $E\{x|\mathbf{S}\} = x$ a.e.*

(h) *If y is a bounded $\mathbf{S}$ measurable function and x is integrable, then*

$$E\{yx|\mathbf{S}\} = yE\{x|\mathbf{S}\} \quad \text{a.e.} \tag{3,5}$$

(i) *If $\mathbf{S} \subset \mathbf{T}$ then $E\Big\{E\{x|\mathbf{S}\}\Big|\mathbf{T}\Big\} = E\Big\{E\{x|\mathbf{T}\}\Big|\mathbf{S}\Big\} = E\{x|\mathbf{S}\}$ a.e.*

(j) *If $\mathbf{S}$ and the random variable x are mutually independent, then $E\{x|\mathbf{S}\} = E\{x\}$ a.e.*

(k) *If $\mathbf{T}$ is independent of the pair x, $\mathbf{S}$, then $E\{x|\mathbf{S},\mathbf{T}\} = E\{x|\mathbf{S}\}$ a.e.* (Property (j) is a slightly concealed special case of (k).)

(l) **Jensen's inequality for conditional expectations.** *Let ϕ be a convex function from an interval I of $\mathbf{R}$ into $\mathbf{R}$ and x be a random variable with range in I. If x and $\phi(x)$ are integrable, then $\phi[E\{x|\mathbf{S}\}] \le E\{\phi(x)|\mathbf{S}\}$ a.e.*

Observe that if $p \ge 1$, this inequality implies that

$$E\{|E\{x|\mathbf{S}\}|^p\} \le E\{E\{|x|^p|\mathbf{S}\}\} = E\{|x|^p\}, \tag{3.6}$$

and therefore, to show that the conditional expectation of a random variable can be approximated in the sense of L^p distance by conditional expectations of other random variables, it is sufficient to show that L^p distance approximation is possible for the unconditioned random variables.

(m) *The class of random variables*

$$\left\{ E\{x|\mathbf{S}\} : \mathbf{S} \text{ is a } \sigma \text{ algebra of measurable sets} \right\}$$

is uniformly integrable.

(n) **Hölder's inequality for conditional expectations.** *If* $x \in L^p$ *and* $y \in L^q$, *where* $1/p + 1/q = 1$, *then* xy *is integrable and*

$$|E\{xy|\mathbf{S}\}| \le E^{1/p}\{|x|^p|\mathbf{S}\}\, E^{1/q}\{|y|^q|\mathbf{S}\} \quad a.e. \tag{3.7}$$

(o) **Minkowski's inequality for conditional expectations.** *If* $p \ge 1$ *and* x *and* y *are in* L^p, *then* $x+y$ *is in* L^p *and*

$$E^{1/p}\{|x+y|^p|\mathbf{S}\} \le E^{1/p}\{|x|^p|\mathbf{S}\} + E^{1/p}\{|y|^p|\mathbf{S}\} \quad a.e. \tag{3.8}$$

(p) *If* $x \in L^2$ *then* $E\{x|\mathbf{S}\} \in L^2$, *and* $x-E\{x|\mathbf{S}\}$ *is the unique up to null sets random variable, in* L^2, *which is orthogonal to every* $\mathbf{S}$ *measurable random variable in* L^2.

Properties (a)–(g) follow at once from the defining properties of conditional expectations and the corresponding properties of expectations, or from translations of the expectation proofs into conditional expectation proofs. The easiest way to prove (f) when there is convergence in measure is to remark that according to the dominated convergence theorem, convergence in measure of $x_\bullet$ to x, when $\sup|x_\bullet|$ is integrable, implies that $x_\bullet$ converges to x in the L^1 metric. It follows that the sequence $E\{x_\bullet|\mathbf{S}\}$ converges to $E\{x|\mathbf{S}\}$ in this metric, and therefore converges in measure. The choice $\mathbf{S} = \mathbf{S}'$ shows that under the hypothesis of convergence in measure, there need not be almost everywhere convergence of the sequence of conditional expectations.

Proof of (h). It is sufficient to prove (h) when x and y are positive functions. If y is the indicator function of a set in $\mathbf{S}$, then (h) becomes a trivial consequence of the conditional expectation definition. It follows that (h) is true for y a positive linear combination of indicator functions of sets in $\mathbf{S}$, and therefore for y a limit of a bounded increasing sequence of such functions; the class of these limits is the class of positive bounded $\mathbf{S}$ measurable functions.

Proof of (i). The equality between first and third terms is a special case of (g), because the inner conditional expectation in the first term is a $\mathbf{T}$ measurable function. To prove equality between the second and third term it must be verified that $E\{x|\mathbf{T}\}$ has the same integral as x over a set in $\mathbf{S}$, and this is true because the two functions have the same integral over every set in $\mathbb{T}$, a larger σ algebra than $\mathbf{S}$.

Proof of (j). If $A \in \mathbf{S}$, the random variables x and $\mathbf{1}_A$ are mutually independent, and therefore (Section VI.19) $E\{x\mathbf{1}_A\} = E\{x\}P\{A\}$, that is, the function identically equal to $E\{x\}$ satisfies the defining equality for $E\{x|\mathbf{S}\}$.

Proof of (k). It must be shown, generalizing the preceding proof slightly, that $E\{x|\mathbf{S}\}$ has the same integral as x over every set in $\sigma(\mathbf{S},\mathbf{T})$. Now $\sigma(\mathbf{S},\mathbf{T})$ is the σ algebra generated by the class of disjunct finite unions of sets of the form $A\cap B$, with A in $\mathbf{S}$ and B in $\mathbf{T}$. In view of the fact that this algebra is dense in $\sigma(\mathbf{S},\mathbf{T})$ under the set distance d_P, it is sufficient to prove that $E\{x|\mathbf{S}\}$ has the same integral as x over every set $A\cap B$. The independence assumption implies the desired equality by way of the fact that the expectation of the product of mutually independent integrable random variables is the product of their expectations:

$$(3.9) \qquad E\{x\mathbf{1}_{A\cap B}\} = E\{x\mathbf{1}_A \mathbf{1}_B\} = E\{x\mathbf{1}_A\}E\{\mathbf{1}_B\}$$

$$= E\big\{E\{x|\mathbf{S}\}\mathbf{1}_A\big\}E\{\mathbf{1}_B\} = E\big\{E\{x|\mathbf{S}\}\mathbf{1}_A\mathbf{1}_B\big\} = E\big\{E\{x|\mathbf{S}\}\mathbf{1}_{A\cap B}\big\}.$$

Proof of (l). If $\phi(\xi) \ge a\xi+b$ for all ξ in I, then

$$(3.10) \qquad E\{\phi(x)|\mathbf{S}\} \ge E\{ax+b|\mathbf{S}\} \quad \text{a.e.}$$

The convex function ϕ is the supremum of all the linear functions it majorizes and, neglecting null sets, is therefore (Section V.18) the supremum of a countable subset of these linear functions, the limit of an increasing sequence $f_\bullet$ of maxima of finitely many of these linear functions. To obtain (l), apply the Beppo-Levi theorem for conditional expectations to (3.10), with $ax+b$ replaced by $f_\bullet(x)$.

Proof of (m). According to Theorem VI.7, integrability of $|x|$ implies that there is a convex uniform integrability test function ϕ for which $\phi(|x|)$ is integrable. Apply Jensen's inequality for conditional expectations,

$$(3.11) \qquad E\Big\{\phi[E\{|x||\mathbf{S}\}]\Big\} \le E\Big\{E\{\phi(|x|)|\mathbf{S}\}\Big\} = E\{\phi(|x|)\},$$

to find that the class of functions

$$(3.12) \qquad \Big\{E\{\phi(|x|)|\mathbf{S}\} : \mathbf{S} \textit{ is a } \sigma \textit{ algebra of measurable sets}\Big\}$$

is L^1 bounded. This fact implies the stated uniform integrability.

Proof of (n). If x and y are step functions, written as linear combinations of indicator functions of pairwise disjoint measurable sets, the conditional expectations relative to **S** of their rth powers have the form $\Sigma\, a_\bullet^r P\{A_\bullet|\mathbf{S}\}$, and the discrete context Hölder inequality yields (n) for this special case. In the general case, xy is integrable, according to Hölder's inequality, and it need only be shown that $E\{xy|\mathbf{S}\}$, $E\{|x|^p|\mathbf{S}\}$ and $E\{|y|^q|\mathbf{S}\}$ can be approximated arbitrarily closely in the L^1 sense by the corresponding expressions with step functions. As pointed out in a remark after the statement of Jensen's inequality for conditional expectations, it is sufficient to show that these approximations are possible for the unconditioned random variables, and the possibility of the latter approximations was shown in Sections VI.14-15.

Proof of (o). The Minkowski inequality for conditional expectations can be reduced to the Hölder inequality for conditional expectations just as in the unconditioned context, treated in Section VI.13. Alternatively the Minkowski inequality for conditional expectations follows from the unconditioned Minkowski inequality when the random variables are step functions, and an approximation procedure yields the general case.

Proof of (p). The square integrablity of $E\{x|\mathbf{S}\}$ follows from Jensen's (or Schwarz et al.'s) inequality for conditional expectations. Denote by $L^2(\mathbf{S})$ the class of square integrable **S** measurable functions. The defining equation for a conditional expectation implies that the random variable $y = x - E\{x|\mathbf{S}\}$ is orthogonal to the indicator function of a set in **S**, and therefore orthogonal to the class of linear combinations of such functions, that is, to the class of **S** measurable step functions. Hence y is orthogonal to the L^2 closure $L^2(\mathbf{S})$ of this class. Conversely, if z is in $L^2(\mathbf{S})$ and has the property that $x-z$ is orthogonal to $L^2(\mathbf{S})$ then z satisfies the conditions for $E\{x|\mathbf{S}\}$. A more elegant formulation of this characterization of $E\{x|\mathbf{S}\}$ is given in the following paragraph.

Hilbert space description of conditional expectations. Denote by $\mathfrak{H}$ the Hilbert space of equivalence classes of square integrable random variables, and denote by $\mathfrak{M}$ the space of equivalence classes of square integrable **S** measurable functions, in each case identifying two random variables when they are equal almost everywhere. According to (p), $E\{\cdot|\mathbf{S}\}$ acting on $\mathfrak{H}$ is the Hilbert space projection onto $\mathfrak{M}$.

4. Filtrations and adapted families of functions

Let $(S,\mathbf{S})$ be a measurable space and I be an ordered set, with order relation symbol $\leq$. A *filtration* of the space is a map $i \to \mathbf{S}_i$ from I into the class of sub σ algebras of **S**, increasing in the sense that $\mathbf{S}_i \subset \mathbf{S}_j$ when $i \leq j$. The triple $(S,\mathbf{S},\mathbf{S}_\bullet)$ is a *filtered measurable space*. If P is a probability measure defined on **S**, $(S,\mathbf{S},\mathbf{S}_\bullet,P)$ is a *filtered probability space*. If $x_\bullet$ is a family of measurable functions, indexed by I, from $(S,\mathbf{S})$ into some measurable state space, and if, for every index point i, x_i is not only measurable from $(S,\mathbf{S})$ into the state space, but even measurable from $(S,\mathbf{S}_i)$ into the state space, the family $x_\bullet$ is *adapted* to $\mathbf{S}_\bullet$ and $(S,\mathbf{S},\mathbf{S}_\bullet,x_\bullet,P)$ is a *filtered stochastic process*. These concepts are fundamental in modern probability theory and are stated here in a general form, but when applied in this book the parameter set I will always be a set of integers in their natural order, sometimes with $+\infty$ [$-\infty$] adjoined when the set of integers is unbounded above [below], and the state space will be a subset of $\mathbf{R}$ unless specified otherwise. In discussing filtered stochastic processes, the full notation $(S,\mathbf{S},\mathbf{S}_\bullet,x_\bullet,P)$ will be abbreviated to show only those items needed to avoid ambiguity.

Example. If $x_\bullet$ is a finite or infinite sequence of random variables on a probability space and $\mathbf{S}_n = \sigma(x_1,\ldots,x_n)$, the σ algebra of sets determined by measurable conditions on $x_1,\ldots,x_n$ (see Section V.1), then $\mathbf{S}_\bullet$ is a filtration, the one with the minimal σ algebras to which the sequence $x_\bullet$ is adapted.

Transition functions and Markov processes. In Section III.7, the Markov property was defined in a very special context. The general definition is the following. If $(S,\mathbf{S},\mathbf{S}_\bullet,x_\bullet,P)$ is an adapted process with ordered parameter set I, the process is a *Markov process* , that is, it has the *Markov property*, if whenever i and j are parameter values with $i < j$ and A is a measurable subset of the state space then

$$(4.1) \qquad P\{x_j \in A \,|\, \mathbf{S}_i\} = P\{x_j \in A \,|\, x_i\} \quad \text{a.e.}$$

This property means, roughly, that conditional probabilities of future events, given the whole past up through time i, actually depend on the past only by way of x_i. In Section 2 it was shown how to construct a finite sequence of random variables with the Markov property, based on an initial point and transition functions. This construction will be applied in Section 21.

The idea of a Markov process is that certain conditional probabilities (and necessarily corresponding expectations when the state space is $\mathbf{R}$) depend only on the last conditioning variable. Martingale theory is based in part on the same idea - dependence only on the last conditioning variable - but imposes stringent hypotheses on the values of certain conditional expectations rather than on conditional probabilities.

5. Martingale theory definitions

Set σ algebras are basic in measure theory, because they are the natural domains of definition of measures. In martingale theory, the effect of varying these set σ algebras is studied systematically. A natural way to initiate such a study is to see the effect on a conditional expectation of varying the conditioning σ algebra. It turns out that an analysis of this effect leads to applications in many parts of mathematics.

Let $(S,\mathbf{S},\mathbf{S}_\bullet,x_\bullet,P)$ be a filtered stochastic process. The process is a *martingale* if, for every index point i, x_i is integrable, and for $i<j$,

$$x_i = E\{x_j|\mathbf{S}_i\} \quad \text{a.e.} \tag{5.1}$$

Submartingales and *supermartingales* are defined like martingales except that, in (5.1) equality is replaced by inequality, "≤" for submartingales, "≥" for supermartingales. Thus the negative of a submartingale is a supermartingale, and a process that is both a submartingale and a supermartingale is a martingale. These processes can of course be defined by integral equalities and inequalities without the explicit use of conditional expectations. For example, the integrated form of the defining martingale equation (5.1) is

$$\int_A x_i\,dP = \int_A x_j\,dP \qquad (A \in \mathbf{S}_i;\, i<j). \tag{5.1'}$$

To obtain the integrated versions of the submartingale and supermartingale inequalities replace "=" in (5.1') by "≤" and "≥", respectively.

Observe that, if $\mathbf{S}_j' = \sigma(x_i\colon i \le j)$, then $\mathbf{S}_j' \subset \mathbf{S}_j$, and $\{x_\bullet,\mathbf{S}_\bullet'\}$ is a martingale, or submartingale, or supermartingale if $\{x_\bullet,\mathbf{S}_\bullet\}$ is. In fact, say in the martingale case, if both sides of (5.1) are operated on by $E\{\bullet|\mathbf{S}_i'\}$, the left side becomes x_i almost everywhere according to Section 3(g), and the right side becomes $E\{x_j|\mathbf{S}_i'\}$ almost everywhere according to Section 3(i).

When there is no question about the filtration, the notation $x_\bullet$ is used instead of $(x_\bullet,\mathbf{S}_\bullet)$. A subset of a martingale, submartingale, or supermartingale is, respectively, a martingale, submartingale, or supermartingale relative to the corresponding subset of the given.filtration

Nonmathematical interpretation. If the parameter set is thought of as representing the flow of time, if $\mathbf{S}_i$ is thought of as a representation of all past events through time i in some context, and if x_i represents some present evaluation in this context, then the martingale, submartingale, and supermartingale conditions state, respectively, that given the whole past, what one expects to get in the future is what one has already, something more, or something less. This rather vague interpretation should not be scorned, because it has led to basic theorems in martingale theory. See Section 10 for an application of this interpretation to suggest a mathematical theorem.

6. Martingale examples

(a) Let $(S,\mathbf{S},\mathbf{S}_\bullet,P)$ be a filtered probability space and x be an integrable random variable. For each parameter point i, define $x_i = E\{x|\mathbf{S}_i\}$. An application of Section 3(i) shows that $(x_\bullet,\mathbf{S}_\bullet)$ is a martingale. This martingale is uniformly integrable, according to Section 3(m). Moreover if another index point α is adjoined to I, following all the others in the ordering, and if x_α is defined as x, $\mathbf{S}_\alpha$ as $\mathbf{S}$, the extended family is still a martingale. Every martingale whose parameter set has a last element is of this type. The following is a particular case. On a probability space, let x be an integrable random variable and $y_\bullet$ be a finite or infinite sequence of random variables. Define $\mathbf{S}_n = \sigma(y_1,\dots,y_n)$ and $x_n = E\{x|\mathbf{S}_n\}$. Then $(x_\bullet,\mathbf{S}_\bullet)$ is a martingale, and if the pair $(x,\mathbf{S})$ is adjoined at the end, the new process, with last random variable x, is a martingale.

(b) Let $y_\bullet$ be a sequence of mutually independent integrable random variables on a probability space, and define $x_n = y_1+\dots+y_n$, $\mathbf{S}_n = \sigma(y_1,\dots,y_n) = \sigma(x_1,\dots,x_n)$. Then, if every y_j has zero expectation, $(x_\bullet,\mathbf{S}_\bullet)$ is a martingale; if every y_j has positive [negative] expectation $(x_\bullet,\mathbf{S}_\bullet)$ is a submartingale [supermartingale]. To prove the submartingale assertion, for example, apply Sections 3(g) and 3(j): if $m < n$,

$$(6.1)\quad E\{x_n|\mathbf{S}_m\} = E\{x_m|\mathbf{S}_m\} + E\{x_n - x_m|\mathbf{S}_m\} = x_m + E\{x_n - x_m\} \geq x_m \quad \text{a.e.}$$

7. Elementary properties of (sub- super-) martingales.

(a) *The function* $i \to E\{x_i\}$ *is a constant function if* $x_\bullet$ *is a martingale, monotone increasing [decreasing] if* $x_\bullet$ *is a submartingale [supermartingale].*

(b) *If the parameter set is a set of successive integers, it is sufficient for the martingale equality* (5.1) *that every almost everywhere one-step equality* $x_n = E\{x_{n+1}|\mathbf{S}_n\}$ *be valid,* because then

$$x_n = E\Big\{E\{x_{n+2}|\mathbf{S}_{n+1}\}\Big|\ \mathbf{S}_n\Big\} = E\{x_{n+2}|\mathbf{S}_n\} \quad \text{a.e.},$$

and so on. Similarly, *the one-step submartingale and supermartingale inequalities suffice in the context of the stated parameter set.*

(c) *If* $(x_\bullet,\mathbf{S}_\bullet)$ *is a martingale [submartingale] and* ϕ *is a convex [convex monotone increasing] function on* $\mathbf{R}$, *with* $\phi(x_i)$ *integrable for all index points* i, *then* $(\phi(x_\bullet),\mathbf{S}_\bullet)$ *is a submartingale.* To prove this assertion for $x_\bullet$ a submartingale, first apply the monotonicity of ϕ, then Jensen's inequality for conditional expectations, to obtain

$$(7.1)\qquad \phi(x_i) \leq \phi\big[E\{x_j|\mathbf{S}_i\}\big] \leq E\{\phi(x_j)|\mathbf{S}_i\} \quad \text{a.e.} \qquad (i < j),$$

which is the submartingale inequality for $\phi(x_\bullet)$. If $x_\bullet$ is a martingale, the first inequality is replaced by an equality, and monotonicity of ϕ is not needed.

According to (c), if $x_\bullet$ is a submartingale the process $x_\bullet \vee c$ is a submartingale for every constant c, and the process $e^{x_\bullet}$ is also a submartingale if its random variables are integrable. If $x_\bullet$ is a martingale and $p \geq 1$, the process $|x_\bullet|^p$ is a submartingale, if its random variables are integrable.

(d) *If* $(x_\bullet, \mathbf{S}_\bullet)$ and $(y_\bullet, \mathbf{S}_\bullet)$ *are submartingales, then the process* $(x_\bullet \vee y_\bullet, \mathbf{S}_\bullet)$ *is a submartingale* because

$$x_i \le E\{x_j \mid \mathbf{S}_i\} \le E\{x_j \vee y_j \mid \mathbf{S}_i\} \quad \text{a.e.} \qquad (i < j),$$

and the corresponding inequality is true for $y_\bullet$ and therefore for $x_\bullet \vee y_\bullet$.

8. Optional times

In Section 5, nonmathematical interpretations of martingale theory concepts were suggested. More specifically, in the context of a gambler betting on each of a sequence of successive plays, if $(x_\bullet, \mathbf{S}_\bullet)$ is a martingale, x_i can be interpreted as the gambler's money at time i, and $\mathbf{S}_i$ can be interpreted as the relevant past: the gambler's fortune at all past times, together perhaps with other past items, up through time i. Under this interpretation, the martingale equality states that the gambler expects, given the past through play i, to have, after a later play, the same fortune as at time i. This is a somewhat crude version of the idea that the gambling context is fair; the submartingale case corresponds to a favorable context, the supermartingale case to an unfavorable context. If the gambler's fortune is checked at random times, chosen by what has been going on, in other words at times determined by the past up to and including the present, it is reasonable to suppose that the context seems fair, favorable, or unfavorable when checked at those times if it seems so when checked at the end of each play. That is, if a game looks fair, favorable, or unfavorable, it will look the same way at random times chosen without foreknowledge. The problem suggested by this reasoning is to devise a mathematical formulation of the concept of a random time, and then to show that the martingale, submartingale, and supermartingale inequalities remain valid if parameter values are chosen at random in this sense. The following definition of a random time seems to be the appropriate one for a filtration with a finite or infinite sequence as parameter set.

Optional time definition. Let $\mathbf{S}_\bullet$ (finite or infinite parameter sequence) be a filtration of a measurable space $(S, \mathbf{S})$. A random variable α defined on the space, and taking on only parameter values and $+\infty$, is an *optional time* if $\{\alpha = n\} \in \mathbf{S}_n$ for all finite n, equivalently if

$$\{\alpha \le n\} = \cup_{j=1}^{n} \{\alpha = j\} \in \mathbf{S}_n$$

for all finite n. The σ algebra $\mathbf{S}_\alpha$ is defined as the class of sets in $\mathbf{S}$ for which $A \cap \{\alpha = n\} \in \mathbf{S}_n$ for all finite n, equivalently for which $A \cap \{\alpha \le n\} \in \mathbf{S}_n$ for all finite n. This class is a σ algebra: it obviously contains the countable unions of its members, and it is closed under complementation because $A \in \mathbf{S}_\alpha$ implies that $\tilde{A} \cap \{\alpha = n\} = \{\alpha = n\} \cap \{A \cap \{\alpha = n\}\}^{\sim} \in \mathbf{S}_n$.

A random variable identically equal to a parameter value k is an optional time, and the definition of $\mathbf{S}_k$, formulated for k considered as a constant optional time, yields $\mathbf{S}_k$ as originally defined. The maximum and minimum of two optional times are also optional times.

Example: Entry (hitting) time. Let $(x_\bullet,\mathbf{S}_\bullet)$ be an adapted finite or infinite sequence on a probability space, let A' be a measurable set of the state space, and define the function α as the first time an $x_\bullet$ sample sequence enters A':

$$\alpha = \min\{j: x_j \in A'\},$$

with $\alpha = +\infty$ if no x_j value is a point of A'. Then α is an optional time, because

$$\{\alpha \le n\} = \cup_{i=1}^{n}\{x_i \in A'\} \in \mathbf{S}_n.$$

If the parameter set is $1,\ldots,m$, the finite valued optional time $\alpha \wedge m$ is usually more useful than α as just defined.

9. Optional time properties

In the following list, the optional times are based on a finite or infinite filtration sequence $\mathbf{S}_\bullet$. It is supposed that all optional times have values in the parameter sequence.

(a) *If α is an optional time, α is $\mathbf{S}_\alpha$ measurable,* because $\{\alpha = k\}\cap\{\alpha = n\} \in \mathbf{S}_n$ for all n.

(b) *If α and β are optional times and $\alpha \le \beta$, then $\mathbf{S}_\alpha \subset \mathbf{S}_\beta$,* because if $A \in \mathbf{S}_\alpha$,

$$A\cap\{\beta \le n\} = \big(A\cap\{\alpha \le n\}\big)\cap\{\beta \le n\} \in \mathbf{S}_n$$

for all n.

According to (b), if $\alpha, \beta,\ldots$ is an increasing finite or infinite sequence of optional times, the sequence $\mathbf{S}_\alpha, \mathbf{S}_\beta,\ldots$ is a filtration.

(c) *If $(x_\bullet,\mathbf{S}_\bullet)$ is an adapted process, and α is an optional time, then x_α is $\mathbf{S}_\alpha$ measurable*, because if A' is a measurable subset of the state space,

$$\{x_\alpha \in A'\}\cap\{\alpha = n\} = \{x_n \in A'\}\cap\{\alpha = n\} \in \mathbf{S}_n.$$

According to (c), if $(x_\bullet,\mathbf{S}_\bullet)$ is an adapted process and $\alpha_\bullet$ is an increasing finite or infinite sequence of optional times for $\mathbf{S}_\bullet$, then the sequence $(x_{\alpha_\bullet},\mathbf{S}_{\alpha_\bullet})$ is an adapted process.

10. The optional sampling theorem

The following theorem is the simplest theorem on the invariance of a process type under optional sampling.

Theorem (Doob). Let $(x_\bullet, \mathbf{S}_\bullet)$ *be a martingale with finite or infinite parameter sequence and* α *and* β *be bounded optional times, with* $\alpha \le \beta$. *Then* x_α *and* x_β *are integrable, and*

$$x_\alpha = E\{x_\beta \mid \mathbf{S}_\alpha\} \qquad a.e. \tag{10.1}$$

If the process is a submartingale [supermartingale], equality in (10.1) *is replaced by the inequality* $\le$ $[\ge]$.

The point of the theorem is that if $(x_\bullet, \mathbf{S}_\bullet)$ is a martingale, or is a submartingale, or is a supermartingale, the two element filtered process $(x_\alpha, \mathbf{S}_\alpha), (x_\beta, \mathbf{S}_\beta)$ has this same property.

Proof. If k is the maximum value of β, then $|x_\beta| \le |x_1| + \cdots + |x_k|$, and therefore x_β is integrable, as is x_α according to the same reasoning. Suppose first that the given process is a martingale. The martingale equality (10.1) will be proved first for the very special case $\beta \le \alpha+1$. In this case, if $A \in \mathbf{S}_\alpha$, define $A_i = \{\alpha = i, \beta = i+1\}$. Then $A_i = \{\alpha = i\} \cap \{\beta = i\}^\sim \in \mathbf{S}_i$, and therefore

$$\int_A (x_\beta - x_\alpha)\, dP = \sum_{i \ge 1} \int_{A_i} (x_{i+1} - x_i)\, dP = 0. \tag{10.2}$$

Equation (10.2) is the integrated version of the martingale equality. For general β, consider the increasing sequence α, $(\alpha+1)\wedge\beta, \ldots, (\alpha+k)\wedge\beta = \beta$ of optional times. According to what has just been proved, if the $x_\bullet$ process is evaluated at these optional times, the one-step martingale equality is satisfied, and it was noted above that this fact makes the process a martingale. Hence the first and last members of this martingale form a martingale, as was to be proved. If $x_\bullet$ is a submartingale [supermartingale], the only difference in the discussion is that the integrals on the right in (10.2) are positive [negative], as desired for the proof of the theorem.

According to this theorem, if $\alpha_\bullet$ is an increasing sequence of bounded optional times for $\mathbf{S}_\bullet$, the process $(x_{\alpha_\bullet}, \mathbf{S}_{\alpha_\bullet})$, is a martingale, submartingale, or supermartingale, if the original process is. This is the desired result suggested by the gambling interpretation, at least if the optional times are bounded; fairness or unfairness is preserved under sampling at optional times. This result is not true for unbounded optional times without restrictions on the optional times or on the process or on both.

11. The maximal submartingale inequality

The simplest nontrivial application of this sampling theorem is to the following maximal inequality.

Theorem. *If* $x_1,\ldots,x_n$ *is a submartingale and* $c \in \mathbf{R}^+$*, then*

$$cP\{\max x_\bullet \geq c\} \leq E\{x_n \vee 0\}. \tag{11.1}$$

Proof. Define $A = \{\max x_\bullet \geq c\}$ and $\alpha = \min\{j: x_j \geq c\}$, on A, $\alpha = n$ on $\tilde{A}$, that is, α is the minimum of n and the hitting time of the state space set $[c,+\infty)$. Then α and the function identically n are optional times, and therefore (x_α, x_n) is a submartingale. The integral submartingale inequality $E\{x_\alpha\} \leq E\{x_n\}$ yields

$$E\{x_n\} \geq cP\{A\} + \int_{\tilde{A}} x_n\,dP, \tag{11.2}$$

which implies a stronger inequality than (11.1).

This proof of the maximal inequality is given as an easy application of optional sampling, but in fact a direct proof of the inequality, which amounts to proving the integrated submartingale inequality in this case, is equally easy. The upcrossing inequality in Section 13 shows the power of the optional sampling theorem in a more complicated context.

12. Upcrossings and convergence

Let $\xi_1,\ldots,\xi_n$, a, b be real numbers with $a < b$. Define α_k as follows, with the understanding that if the indicated condition on j is not satisfied for any value of j, α_k is defined as n.

$$\begin{aligned}
\alpha_1 &= \min\{j: \xi_j \leq a\}, \\
\alpha_k &= \min\{j: j > \alpha_{k-1}, \xi_j \geq b\} && (k \text{ even}, \geq 2), \\
&= \min\{j: j > \alpha_{k-1}, \xi_j \leq a\} && (k \text{ odd}, \geq 3).
\end{aligned} \tag{12.1}$$

The *number of upcrossings* U *of* $[a,b]$ *by* $\xi_\bullet$ is the number of times $\xi_\bullet$ proceeds from below a to above b, that is, this number is 0 if $\alpha_2 = \alpha_1$, and otherwise is the maximum value of $k/2$ with k even, $\alpha_k > \alpha_{k-1}$, and $x_{\alpha_k} \geq b$. If $\xi_\bullet$ is an infinite sequence, α_k is defined in the obvious way and is the limit as $n \to +\infty$ of the number of upcrossings of $[a,b]$ by $\xi_1,\ldots,\xi_n$.

An infinite sequence $\xi_\bullet$ has a (not necessarily finite) limit if and only if for every pair $[a,b]$ of numbers with $a<b$, the number of upcrossings of $[a,b]$ by the sequence is finite. It is sufficient if this upcrossing condition is satisfied for rational a and b.

13. The submartingale upcrossing inequality

Theorem (Doob, Snell). *Let $x_1,\dots,x_n$ be a submartingale, a,b be real numbers with $a < b$, and* **U** *be the number of upcrossings of $[a,b]$ by $x_\bullet$. Then*

$$E\{\mathrm{U}\} \le \frac{E\{x_n \vee a - x_1 \vee a\}}{b-a} \le \frac{E\{|x_n - x_1|\}}{b-a}. \tag{13.1}$$

Proof. Define α_k as above for each sample sequence of $x_\bullet$. The random variable α_k is an optional time, because the condition $\alpha_k = m$ is a condition on $x_1,\dots,x_m$. Hence the process $x_{\alpha_1},\dots,x_{\alpha_n}$ is a submartingale, and therefore each of the summands in the following equality has positive expectation:

$$x_n - x_1 = \sum_{j=2}^{n} (x_{\alpha_j} - x_{\alpha_{j-1}}). \tag{13.2}$$

On the set $\{\mathrm{U} = k\}$, each of the first k summands with even j has value $\ge (b-a)$, and the later summands with even j are all 0 except possibly the first one, $x_{\alpha_{2k+2}} - x_{\alpha_{2k+1}}$. If this difference does not vanish, $x_{\alpha_{2k+1}} \le a$, $\alpha_{2k+2} = n$ and $x_n < b$. Thus the difference is at least $(x_n - a)\wedge 0$. Applying these inequalities in (13.2), and ignoring the summands with odd j, which have positive expectations, yields

$$E\{x_n - x_1\} \ge (b-a)E\{\mathrm{U}\} + E\{(x_n - a)\wedge 0\}. \tag{13.3}$$

If $x_\bullet$ is replaced in this inequality by the submartingale $x_\bullet \vee a$, the number of upcrossings is unchanged, and (13.3) yields (13.1).

14. Forward (sub-super-) martingale convergence

The basic martingale theory convergence theorems are due, at various levels of generality, to Doob, Jessen, and Lévy. These convergence theorems are the forward one, for the parameter set of strictly positive integers, and the backward one, for the parameter set $\dots,-1, 0$. The backward theorem is simpler because in that context the parameter set has a last point. The following is the forward convergence theorem.

Theorem (Forward martingale convergence). *Let $(x_n, \mathbb{S}_n,\ n \ge 1)$ be an adapted process, and define $\mathbb{S}_\infty = \sigma(\cup\, \mathbb{S})$.*

(a) *If the process is an* L^1 *bounded submartingale, martingale, or supermartingale, then* $\lim x_\bullet = x_\infty$ *exists almost surely and is integrable.*

(b) *If the process is a lower bounded supermartingale, that is, if* $x_\bullet \ge$ const., *the process is* L^1 *bounded, and the extended process $(x_n, \mathbb{S}_n,\ n \le +\infty)$ is a supermartingale.*

(c) *If the process in* (a) *is uniformly integrable, then the limit* x_∞ *is both an almost sure and an* L^1 *limit, and the extended process* $(x_n, \mathbf{S}_n,\ n \le +\infty)$ *is also respectively a submartingale, martingale or supermartingale. In particular, if* x *is an arbitrary integrable random variable,*

$$\lim E\{x|\mathbf{S}_\bullet\} = E\{x|\mathbf{S}_\infty\} \quad a.e. \tag{14.1}$$

Proof of (a). If $U_n(a,b)$ is the number of upcrossings of $[a,b]$ by $x_1,\dots,x_n$, the sequence $U_\bullet[a,b]$ is an increasing sequence of positive random variables, with limit $U[a,b]$, the number of upcrossings by the infinite sequence. In view of the assumed L^1 boundedness, Theorem 13 implies that the limiting expectation $E\{U[a,b]\} = \lim E\{U_\bullet[a,b]\}$ is finite, and therefore that $U[a,b]$ is almost surely finite. This means that each summand on the right in the following equation is a null set.

$$\begin{aligned}(14.2)\quad &\{\limsup x_\bullet > \liminf x_\bullet\}\\ &= \bigcup \Big\{ \{\limsup x_\bullet > b > a > \liminf x_\bullet\}: a, b \text{ rational} \Big\}.\end{aligned}$$

It follows that there is a (possibly infinite) almost sure limit x_∞ of the sequence $x_\bullet$. Moreover x_∞ is integrable and therefore almost surely finite because, according to Fatou's theorem,

$$E\{|x|\} \le \liminf E\{|x_\bullet|\} < +\infty.$$

Proof of (b). If $x_\bullet$ is a lower bounded. supermartingale, it can be supposed positive for the purposes of the theorem, at the cost of an additive constant. The integral form of the supermartingale inequality on the whole space shows that $E\{x_\bullet\}$ is a decreasing sequence and the supermartingale is therefore L^1 bounded. An application of Fatou's theorem yields, for every parameter value m

$$\int_A x_m\, dP \ge \liminf \int_A x_\bullet\, dP \ge \int_A x_\infty\, dP \qquad (A \in \mathbf{S}_m). \tag{14.3}$$

The inferior limit is actually a limit, because the integral sequence is a decreasing sequence, for subscript values at least m. The inequality between first and third terms is the integrated form of the supermartingale inequality, and therefore the extended process $(x_n, \mathbf{S}_n,\ n \le +\infty)$ is a supermartingale.

Proof of (c). If the sequence $x_\bullet$ is uniformly integrable, convergence implies L^1 convergence (Section **VI**.18). Under the uniform integrability hypothesis, if $A \in \mathbf{S}_m$, and if $x_\bullet$ is a submartingale,

$$\int_A x_m\, dP \le \int_A x_n\, dP \qquad (n > m). \tag{14.4}$$

When $n \to \infty$, the integrand x_n is replaced by x_∞, and the inequality becomes the integrated form of the submartingale inequality $x_m \le E\{x_\infty|\mathbf{S}_m\}$ a.e. Thus (c) is true for submartingales, and the corresponding proof is applicable to martingales and supermartingales. In particular, if $x_n = E\{x|\mathbf{S}_n\}$, there is uniform integrability, and the limit function x_∞ is S_∞ measurable. Moreover, when $m < n$, x has the same integral as x_n over every $\mathbf{S}_m$ set, therefore has the same integral as x_∞ over every $\mathbf{S}_m$ set, that is, x has the same integral as x_∞ over every set in the algebra $\cup\mathbf{S}_\bullet$. Since this algebra is dense in $\mathbf{S}_\infty$ under the dp metric, these two random variables have the same integral over every $\mathbf{S}_\infty$ set, that is $x_\infty = E\{x|\mathbf{S}_\infty\}$ almost surely, as asserted in (c).

15. Backward martingale convergence

It will be convenient to treat backward convergence of a martingale separately from that of submartingales and supermartingales.

Theorem (Backward martingale convergence theorem). *If $(x_n, \mathbf{S}_n, n \le 0)$ is a martingale and $\mathbf{S}_{-\infty} = \cap\mathbf{S}_\bullet$, the martingale is uniformly integrable, and for every parameter value m,*

(15.1) $$\lim_{n\to-\infty} x_n = E\{x_m|\mathbf{S}_{-\infty}\}$$

almost everywhere and in the L^1 *sense.*

Proof. Uniform integrability follows from a conditional expectation property (Section 3(m)) and implies that the sequence $x_\bullet$ is L^1 bounded. The proof method of Theorem 14 is applicable to show that $\lim_{n\to-\infty} x_n = x_{-\infty}$ exists almost everywhere and in the L^1 sense. If m is arbitrary, an application of the L^1 convergence and the integrated form of the martingale equality yields

(15.2) $$\int_A x_{-\infty}\,dP = \lim_{n\to-\infty}\int_A x_n\,dP = \int_A x_m\,dP \qquad (A \in \mathbf{S}_{-\infty}),$$

in which the value of the second integral is the same for all n; (15.2) implies that $x_{-\infty} = E\{x_m|\mathbf{S}_{-\infty}\}$ a.e.

Observe that this limit result is equivalent to asserting that if $(\mathbf{S}_n, n \le 0)$ is a filtration and x is an integrable random variable, then if $\mathbf{S}_{-\infty}$ is defined as $\cap\mathbf{S}_\bullet$,

(15.3) $$E\{x|\mathbf{S}_{-\infty}\} = \lim_{n\to-\infty} E\{x|\mathbf{S}_n\} \text{ a.e.}$$

and the limit equation is also true in the L^1 sense. The process $\{x_n, \mathbf{S}_n, n \ge -\infty\}$ is a martingale.

16. Backward supermartingale convergence

Since the negative of a submartingale is a supermartingale, only one of the two need be treated in a convergence theorem. The following phrasing is for supermartingales, for which the phrasing is slightly more elegant.

Theorem (Backward supermartingale convergence theorem). *Suppose that* $(x_n, \mathbf{S}_n, n \le 0)$ *is a supermartingale, and define* $\mathbf{S}_{-\infty} = \bigcap_\bullet \mathbf{S}_\bullet$. *Then*

(a) $+\infty \ge l \leftarrow \ge \cdots \ge E\{x_{-1}\} \ge E\{x_0\}$;

(b) $\lim x_\bullet = x_{-\infty}$ *exists almost surely*; $-\infty < x_{-\infty} \le +\infty$ *almost surely*; *and* $x_{-\infty} \wedge 0$ *is integrable.*

(c) *If* $l < +\infty$, *then* $x_{-\infty}$ *is almost surely finite, the supermartingale is uniformly integrable, and the convergence is both almost sure and in the* L^1 *sense. Moreover in this case the enlarged process* $\{x_n, \mathbf{S}_n, -\infty \le n\}$ *is a supermartingale.*

Proof of (a). The assertion is trivial, stated only to set the context for (b) and (c). The example $x_n \equiv -n$ for $n \le 0$ shows that the limit l may be $+\infty$ and the sequence $x_\bullet$ may have limit identically $+\infty$.

Proof of (b). The upcrossing argument used in the proof of Theorem 14 shows that the submartingale $-x_\bullet$ is almost everywhere convergent to a limit $-x_{-\infty}$, which may not be finite valued. Apply Fatou's theorem and the submartingale inequality to the submartingale $-(x_\bullet \wedge 0)$ to obtain the inequality

$$E\{x_{-\infty} \wedge 0\} \ge \lim E\{x_\bullet \wedge 0\} \ge E\{x_0 \wedge 0\}, \tag{16.1}$$

which implies that $x_{-\infty} \wedge 0$ is integrable. Thus (b) is true.

Proof of (c). The process $x_\bullet$ can be written as the sum of a uniformly integrable martingale and a positive supermartingale:

$$x_n = E\{x_0 | \mathbf{S}_n\} + [x_n - E\{x_0 | \mathbf{S}_n\}] \qquad \text{a.s.} \tag{16.2}$$

In view of this decomposition of $x_\bullet$ and of Theorem 15, it can be assumed in proving (c) that $x_\bullet$ is a positive supermartingale with finite l. Choose $c > 0$, and observe that the supermartingale $x_\bullet \wedge c$ is bounded, and

$$E\{x_{-\infty} \wedge c\} = \lim E\{x_\bullet \wedge c\} \ge E\{x_n \wedge c\} \qquad (n \le 0). \tag{16.3}$$

It follows that $E\{x_{-\infty}\} \ge E\{x_n\}$, and therefore $E\{x_{-\infty}\} \ge l$. On the other hand, according to Fatou's theorem, $E\{x_{-\infty}\} \le l$. Thus there is equality in Fatou's theorem, and therefore the sequence $x_\bullet$ is uniformly integrable, and (Section VI.18) the convergence is L^1 convergence. The supermartingale inequality

$$\int_A x_m\,dP \geq \int_A x_n\,dP \qquad (m < n), \tag{16.4}$$

for A in $\mathbf{S}_m$ will now be applied for A in the smaller σ algebra $\mathbf{S}_{-\infty}$. With this choice of A, inequality (16.4) is valid whenever $m < n$. In view of the uniform integrability of $x_\bullet$, (16.4) is also valid for $m = -\infty$, and the inequality then becomes the supermartingale inequality for the pair $x_{-\infty}, x_n$. Thus the process $\{x_n, \mathbf{S}_n, -\infty \leq n\}$ is a supermartingale.

17. Application of martingale theory to derivation

Let $(S,\mathbf{S})$ be a measurable space, Q be a finite measure on $\mathbf{S}$, and P be a probability measure on $\mathbf{S}$. In the following discussion of derivation of Q with respect to P, probabilities and probabilistic statements all refer to the probability space $(S,\mathbf{S},P)$. For derivation with respect to a finite not identically vanishing measure P_1 other than a probability measure define $P = P_1/P_1(S)$.

If $\pi{:}S_\bullet$ is a partition of S into a finite number of pairwise disjoint measurable sets, with union S, define

$$x_\pi = \frac{Q(S_j)}{P(S_j)} \text{ on } S_j \tag{17.1}$$

for each nonnull partition cell S_j; if S_j is null, define x_π on S_j as any constant. Then x_π is a random variable, with expectation

$$E\{x_\pi\} = \Sigma_{j\geq 1}\{Q(S_j)\colon P(S_j) > 0\} \leq Q(S), \tag{17.2}$$

and there is equality if every partition cell null for P is also null for Q. Now suppose that $\pi_\bullet$ is a sequence of partitions of S, with each partition a refinement of its predecessor, that is each partition cell of π_{n+1} is a subset of a partition cell of π_n. Define $x_n = x_{\pi_n}$. If $\mathbf{S}_n$ is the σ algebra of unions of the partition cells $S_{n\bullet}$ of π_n, then x_n is $\mathbf{S}_n$ measurable. Furthermore, on each non- null cell S_{nk},

$$\begin{aligned} &E\{x_{n+1}\,|\,\mathbf{S}_n\} \\ &\quad = \Sigma_j\,\{Q(S_{n+1j})\colon P(S_{n+1j} > 0,\ S_{n+1j} \subset S_{nk}\}/P(S_{nk}) \leq x_n \qquad \text{a.e.,} \end{aligned} \tag{17.3}$$

with equality if every partition cell S_{n+1j} null for P is also null for Q. Thus $\{x_\bullet, \mathbf{S}_\bullet\}$ is a positive supermartingale, and in particular is a martingale if Q is P absolutely continuous. (More accurately, as defined above the process random variables are *almost surely* positive; they could have been defined to be positive, but there is no advantage in doing so except to make this sentence superfluous.)

According to Theorem 14, the supermartingale $x_\bullet$ is almost everywhere convergent to an almost surely finite limit function x_∞. This convergence result is more intuitive if formulated as follows. *If s is P almost any point of S, and if $S_n(s)$ is the partition cell of S_n that contains s, then the sequence $Q[S_\bullet(s)]/P[S_\bullet(s)]$ is well defined and has a finite limit.*

The limit function x_∞ will now be analyzed. Let x be the Radon-Nikodym derivative of the absolutely continuous component of Q relative to P, and define $\mathbf{S}' = \sigma(\bigcup \mathbf{S}_\bullet)$, the σ algebra generated by the class of partition cells. It will be proved that $x_\infty = E\{x|\mathbf{S}'\}$ almost surely. It is sufficient to consider the absolutely continuous and singular cases separately.

Absolutely continuous case. Suppose that Q is absolutely continuous relative to P:

$$Q(A) = \int_A x\,dP \qquad (A \in \mathbb{S}). \tag{17.4}$$

In this case, the sequence $x_\bullet$ is a uniformly integrable martingale that converges almost everywhere and in the L^1 sense to $E\{x|\mathbf{S}'\}$. In fact in this case

$$x_n = \Big[\int_{S_{nj}} x\,dP\Big]\Big/P(S_{nj}) = E\{x|\mathbf{S}_n\} \qquad \text{a.s.,} \tag{17.5}$$

on each nonnull partition cell S_{nj}, and Theorem 14 is applicable. In particular, if $\mathbf{S}' = \mathbf{S}$, the limit function is almost surely x.

Singular case. In general, since (Theorem 14(a)) $x_\bullet$ extended by the limit function x_∞ is a supermartingale,

$$\int_A x_\infty\,dP \le \int_A x_m\,dP \le Q(A) \qquad (A \in \mathbb{S}_m) \tag{17.6}$$

for all m. The class of sets A for which the first term is majorized by the third term is a monotone class including the algebra $\bigcup \mathbf{S}_\bullet$ and therefore includes $\mathbf{S}'$. In particular, if Q is singular relative to P, the set A can be chosen so that A is Q null and $\tilde{A}$ is P null. *Hence x_∞ vanishes P almost everywhere, when Q is P singular.*

Example (a). **Derivative of a monotone increasing function.** Let S be an interval of $\mathbf{R}$ of unit length, F be a monotone increasing right continuous function on S, Q be the measure λ_F generated by F, and P be Lebesgue measure on S. Choose pairwise disjoint intervals as partition cells of π_n, choosing them so that the maximum interval length of a partition set of π_n tends to 0 as $n\to\infty$. The σ algebra $\mathbf{S}'$ then contains every interval and is therefore $\mathbf{B}(\mathbf{R})$. The convergence result obtained suggests (but does not imply) that F' exists Lebesgue measure almost everywhere and is the Radon-Nikodym derivative dQ_{ac}/dP. In fact this result was proved in Section X.4. If S is not of unit length, this application is still valid; all that need be done is to make P into Lebesgue measure divided by the Lebesgue measure of S. The method is applicable with P and Q arbitrary Lebesgue-Stieltjes measures on **R**, but the derivation method in Section X.4 yielded much stronger results. The advantage of the partition method of derivation is that it is applicable in very general contexts.

Example (b). A simple singular context. Let S be the interval $[0,1]$ on **R**, **S** be **B**(S), P be Lebesgue measure on **S**, and Q be the probability measure on **S** carried by the singleton $\{0\}$. Then Q is singular relative to P. Let π_n be the partition of S into pairwise disjoint right closed intervals of length 2^{-n}: $[0,2^{-n}],\ldots,(1-2^{-n},1]$. The random variable x_n defined in this section has value 2^n on $[0,2^{-n}]$ and is 0 elsewhere on S. The sequence $x_\bullet$ is a supermartingale with limit 0 except at the origin, thus has limit 0 almost everywhere, as it should. Actually, in this case the supermartingale is even a martingale, because no partition set is null. This example exhibits a positive martingale $x_\bullet$, whose random variables all have expectation 1 but for which $\lim x_\bullet = 0$ almost surely.

18. Application of martingale theory to the 0-1 law

Martingale theory provides an instructive proof of the 0-1 law, Theorem V.9, although the elementary proof given in Section V.9 is easier. In the language of that theorem, if $A \in \bigcap \mathbf{G}_\bullet$, then A is independent of each σ algebra $\mathbb{F}_n$. Hence $P\{A|\mathbb{F}_n\} = P\{A\}$ for all n. On the other hand, since $A \in \mathbb{F}_\infty$, the forward martingale convergence theorem states that $\lim P\{A|\mathbb{F}_\bullet\} = P\{A|\mathbb{F}_\infty\} = \mathbf{1}_A$ almost everywhere. Hence $P\{A\}$ must be either 0 or 1.

19. Application of martingale theory to the strong law of large numbers

The martingale theory convergence theorems make it easy to devise convergence theorems in various contexts. One need only define a filtration and take conditional expectations. Most results obtained in this way are uninteresting, but the following application illustrates the interesting possibilities.

Theorem (Kolmogorov). *Let $x_\bullet$ be a sequence of mutually independent random variables with a common distribution, and suppose that x_1 is integrable. Then*

$$\lim_{n\to\infty} (x_1+\cdots+x_n)/n = E\{x_1\} \quad a.e. \tag{19.1}$$

Proof. Define $s_n = x_1+\cdots+x_n$ and $\mathbf{S}_{-n} = \sigma(s_n, s_{n+1},\ldots)$ for $n \geq 1$. Then $\mathbf{S}_\bullet$ is a filtration, and the process $\{E\{x_1 \| \mathbf{S}_n\}, \mathbf{S}_n, n \leq -1\}$ is a martingale. Each random variable of the process has expectation $E\{x_1\}$. By symmetry,

$$(19.2) \qquad E\{x_1|s_n,s_{n+1},\dots\} = E\{x_j|s_n,s_{n+1},\dots\} \text{ a.e. } (j=1,\dots,n).$$

Averaging the n equalities here, it follows that the nth term of the backward martingale is

$$(19.3) \qquad E\{s_n/n|s_n,s_{n+1},\dots\} = s_n/n \text{ a.e.}$$

According to the backward martingale convergence theorem, the sequence of these conditional expectations converges almost everywhere and in L^1 to a limit random variable $x_{-\infty}$, with $E\{x_{-\infty}\} = E\{x_1\}$. Furthermore, according to Application Section V.10(c) of the 0-1 law, $x_{-\infty}$ must be almost surely constant, and therefore equal almost surely to its expectation, as asserted in (19.1).

Theorem 19 answers the coin-tossing question raised in Section IV.14; it is true that, **in the mathematical context,** the number of heads tossed in n tosses of a balanced coin, divided by n, has almost sure limit 1/2. Whether this is true or not in real life must await an examination of an **experiment,** a nonmathematical concept (although that fact is sometimes not made clear in elementary probability texts), in which a coin is tossed infinitely often. Up to the present time, no one has been able to toss a coin that often, and this is sufficient reason for mathematicians to hand the problem to philosophers and ingenious physicists.

20. Application of martingale theory to the convergence of infinite series

Let $y_\bullet$ be a sequence of square integrable mutually independent random variables, with zero expectations, on some probability space. These random variables are then mutually orthogonal (see Section VI.19), and therefore the convergence of the series $\sum \|y_\bullet\|_2^2$ implies that the series converges in the mean, that is, the series of mutually orthogonal random variables is L^2 convergent. The independence condition is far stronger than orthogonality, and Kolmogorov's theorem, that the series converges almost surely, will now be proved as an application of martingale convergence. According to Section 6, Example (b), the sequence $x_\bullet$ of partial sums of $y_\bullet$ is a martingale, and to prove almost sure convergence it is only necessary to remark that the convergence of $\sum \|y_\bullet\|_2^2$ implies L^2 boundedness of the martingale, which is stronger than the L^1 boundedness condition in Theorem 14, and thereby ensures almost everywhere convergence.

21. Application of martingale theory to the boundary limits of harmonic functions

According to equation VIII(14.4), if u is a function harmonic on an open neighborhood B of the closure of a disk D of radius α, center z', and if z is a point of the disk, then $u(z)$ is a weighted average of its values on the disk boundary ∂D:

$$(21.1)\quad u(z) = \int_{\partial D} u(\zeta)\,\mu_D(z,d\zeta), \quad \mu_D(z,A) = \frac{1}{2\pi\alpha}\int_A \frac{\alpha^2-|z-z'|^2}{|\zeta-z|^2}\,l(d\zeta) \qquad (A \subset \partial D),$$

where l is length on ∂D and $\mu_D(z,\bullet)$ is a measure on the Borel subsets of B, carried by ∂D. (An application of (21.1) to the identically 1 function shows that $\mu_D(z,\bullet)$ is a probability measure.) This measure is called "harmonic measure on ∂D relative to z." It is convenient to extend μ_D by defining $\mu_D(z,\bullet)$ for z in $B-D$ as the probability measure of Borel subsets of B supported by the singleton $\{z\}$. The function $(z,A) \to \mu_D(z,A)$ is a transition function on B, and can serve as the transition probability function of Markov processes as detailed in Sections 2 and 4. One simple Markov process with this transition probability function is the following.

Let B be a disk of radius β, center the origin, and $D_\bullet$ be a sequence of disks with the origin as center, and radii strictly increasing, with limit β. Consider a sequence $\{z_m, m \ge 0\}$ of random variables with values in B, that is, measurable functions from some probability space into B, for which $z_\bullet$ is a Markov process with state space B, with mth transition probability function μ_{D_m} and initial point z_0 in D_1. That is, z_0 is a constant function, with value in D_1, and for $m > 0$ and A a Borel subset of B,

$$(21.2)\qquad P\{z_{m+1} \in A \mid z_0,\dots,z_m\} = \mu_{D_{m+1}}(z_m,A) \quad \text{a.s.}$$

This choice of transition functions means that, successively, z_1 is almost surely on ∂D_1, z_2 is almost surely on ∂D_2, and so on. Thus, for example, if C is a Borel subset of $\partial D_1\times\partial D_2\times\partial D_3$, then (integration over $\partial D_1\times\partial D_2\times\partial D_3$)

$$(21.3)\quad P\{(z_1,z_2,z_3) \in C\} = \int \mu_{D_1}(z_0,d\zeta_1) \int \mu_{D_2}(\zeta_1,d\zeta_2) \int 1_C\, \mu_{D_3}(\zeta_2,d\zeta_3).$$

It follows that if f is a Borel measurable function on B, bounded on ∂D_{n+1}, and $n > 0$, then

$$(21.4)\qquad E\{f(z_{m+1}) \mid z_0,\dots,z_m\} = E\{f(z_{m+1}) \mid z_m\} = \int_{\partial D_{m+1}} f(\zeta)\,\mu_{D_{m+1}}(z_m,d\zeta) \quad \text{a.s.}$$

The fact that there is such a process follows from Kolmogorov's theorem V.6, according to which there is a probability measure on the product space $B \times B \times \cdots$, with the property that the coordinate functions determine a Markov process with initial point z_0 and the prescribed transition functions. Almost all the sample sequences of $z_\bullet$ are sequences with initial point z_0 and with z_m on ∂D_m.

Now suppose that u is a harmonic function on B. An application of (21.4) with $f = u$ yields

$$(21.5) \quad E\{u(z_{m+1}) \mid z_0, \ldots, z_m\} = \int_{\partial D_{m+1}} u(\zeta)\, \mu_{D_{m+1}}(z_m, d\zeta) = u(z_m) \qquad \text{a.s.}$$

Thus the sequence $u(z_\bullet)$ is a martingale relative to the filtration $\mathbf{S}_\bullet$. According to Theorem 14, positivity of u , or, trivially more generally, lower boundedness of u, implies that this martingale converges almost surely, that is, u has a finite limit along almost every sample path of the $z_\bullet$ process to the boundary. The application to the bounded harmonic functions $\Re z$ and $\Im z$ shows that almost every sample sequence of $z_\bullet$ converges, necessarily to a point of ∂B. If u is an arbitrary positive harmonic function, the almost everywhere convergence of $u(z_\bullet)$ means that u has a boundary limit along almost every sample sequence of $z_\bullet$ to ∂B.

More generally, if h is a strictly positive harmonic function on B, the function u/h is a function with an average property determined by *h harmonic measure* μ_D^h: (21.1) yields

$$(21.6) \quad \frac{u(z)}{h(z)} = \int_{\partial D} \frac{u(\zeta)}{h(\zeta)} h(\zeta)\, \mu_D(z, d\zeta)/h(z) = \int_{\partial D} \frac{u(\zeta)}{h(\zeta)} \mu_D^h(z, d\zeta) \quad (z \in D),$$

where $\mu_D^h(z, d\zeta) = \mu_D(z, d\zeta)h(\zeta)/h(z)$. The h harmonic measure μ_D^h can serve as the transition function of a Markov process in the same way as the particular case μD for the special case $h \equiv 1$. (Set $u = h$ to find that $\mu_D^h(z, \bullet)$ is a probability measure carried by ∂D.). For each choice of h, there is a sequence $z_\bullet$ of random variables, with z_m almost surely on ∂D_m, along which the function u/h has an almost sure limit. The measure properties of the sequence $z_\bullet$ depend on the choice of h. It can be shown that, whatever the choice of h, $\lim z_\bullet = z$ exists almost certainly. The random variable z is almost surely on ∂B, but the distribution of z on ∂B depends on the choice of h.

Observe the difference between Theorem X.19 and the present result. Both state that u/h has a finite limit along almost all paths of a certain sort to ∂B but in Theorem X.19 the paths are radii, and the present probabilistic result has paths of a quite different sort. (In a deeper investigation, the martingale method leads to continuous paths (Brownian motion paths conditioned by h) along which u/h has its boundary limits.) The advantage of the martingale method is that it can be adapted to be applicable to harmonic functions on an arbitrary open set and to functions with average properties analogous to those of harmonic functions, for example, to solutions of elliptic and parabolic partial differential equations.

Notation

Index

Graduate Texts in Mathematics

continued from page ii

119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.

图书在版编目（CIP）数据

测度论 = Measure Theory：英文/（美）杜波（Doob，J. L.）著．—北京：世界图书出版公司北京公司，2009.6

ISBN 978-7-5100-0478-0

Ⅰ．测… Ⅱ．杜… Ⅲ．测度论—研究生—教学参考资料—英文 Ⅳ．O174.12

中国版本图书馆 CIP 数据核字（2009）第 084724 号

书　　名：	Measure Theory
作　　者：	J. L. Doob
中 译 名：	测度论
责任编辑：	高蓉　刘慧
出 版 者：	世界图书出版公司北京公司
印 刷 者：	三河国英印务有限公司
发　　行：	世界图书出版公司北京公司（北京朝内大街 137 号　100010）
联系电话：	010-64021602，010-64015659
电子信箱：	kjb@ wpcbj. com. cn
开　　本：	24 开
印　　张：	9.5
版　　次：	2009 年 06 月
版权登记：	图字：01-2009-1096
书　　号：	978-7-5100-0478-0/O · 693　　**定　　价：** 28.00 元