(c) Let $A_r(i_1, \ldots, i_k)$ consist of the numbers in the unit interval in whose base-r expansions the digits $i_1, \ldots, i_k$ nowhere appear consecutively in that order. Show that it is trifling. What does this imply about the monkey that types at random?

1.5. ↑ The *Cantor set* C can be defined as the closure of $A_3(1)$.

(a) Show that C is uncountable but trifling.

(b) From $[0,1]$ remove the open middle third $(\frac{1}{3}, \frac{2}{3})$; from the remainder, a union of two closed intervals, remove the two open middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$. Show that C is what remains when this process is continued ad infinitum.

(c) Show that C is perfect [A15].

1.6. Put $M(t) = \int_0^1 e^{ts_n(\omega)}\, d\omega$, and show by successive differentiations under the integral that

$$M^{(k)}(0) = \int_0^1 s_n^k(\omega)\, d\omega. \tag{1.38}$$

Over each dyadic interval of rank n, $s_n(\omega)$ has a constant value of the form $\pm 1 \pm 1 \pm \cdots \pm 1$, and therefore $M(t) = 2^{-n}\sum \exp t(\pm 1 \pm 1 \pm \cdots \pm 1)$, where the sum extends over all 2^n n-long sequences of $+1$'s and -1's. Thus

$$M(t) = \left(\frac{e^t + e^{-t}}{2}\right)^n = (\cosh t)^n. \tag{1.39}$$

Use this and (1.38) to give new proofs of (1.16), (1.18), and (1.28). (This, the method of moment generating functions, will be investigated systematically in Section 9.)

1.7. ↑ By an argument similar to that leading to (1.39) show that the Rademacher functions satisfy

$$\int_0^1 \exp\left[i \sum_{k=1}^n a_k r_k(\omega)\right] d\omega = \prod_{k=1}^n \frac{e^{ia_k} + e^{-ia_k}}{2} = \prod_{k=1}^n \cos a_k.$$

Take $a_k = t2^{-k}$, and from $\sum_{k=1}^\infty r_k(\omega)2^{-k} = 2\omega - 1$ deduce

$$\frac{\sin t}{t} = \prod_{k=1}^\infty \cos \frac{t}{2^k} \tag{1.40}$$

by letting $n \to \infty$ inside the integral above. Derive Vieta's formula

$$\frac{2}{\pi} = \frac{\sqrt{2}}{2} \frac{\sqrt{2+\sqrt{2}}}{2} \frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2} \cdots.$$

1.8. A number ω is normal in the base 2 if and only if for each positive ϵ there exists an $n_0(\epsilon, \omega)$ such that $|n^{-1}\sum_{i=1}^n d_i(\omega) - \frac{1}{2}| < \epsilon$ for all n exceeding $n_0(\epsilon, \omega)$. Theorem 1.2 concerns the entire dyadic expansion, whereas Theorem 1.1 concerns only the beginning segment. Point up the difference by showing that for $\epsilon < \frac{1}{2}$ the $n_0(\epsilon, \omega)$ above cannot be the same for all ω in N—in other words, $n^{-1}\sum_{i=1}^n d_i(\omega)$ converges to $\frac{1}{2}$ for all ω in N, but not uniformly. But see Problem 13.9.

1.9. 1.3↑ (a) Using the finite form of Theorem 1.3(ii), together with Problem 1.3(b), show that a trifling set is nowhere dense [A15].

(b) Put $B = \bigcup_n (r_n - 2^{-n-2}, r_n + 2^{-n-2}]$, where $r_1, r_2, \ldots$ is an enumeration of the rationals in $(0, 1]$. Show that $(0,1] - B$ is nowhere dense but not trifling or even negligible.

(c) Show that a compact negligible set is trifling.

1.10. ↑ A set of the first category [A15] can be represented as a countable union of nowhere dense sets; this is a topological notion of smallness, just as negligibility is a metric notion of smallness. Neither condition implies the other:

(a) Show that the nonnegligible set N of normal numbers is of the first category by proving that $A_m = \bigcap_{n=m}^\infty [\omega: |n^{-1}s_n(\omega)| < \frac{1}{2}]$ is nowhere dense and $N \subset \bigcup_m A_m$.

(b) According to a famous theorem of Baire, a nonempty interval is *not* of the first category. Use this fact to prove that the negligible set $N^c = (0,1] - N$ is not of the first category.

1.11. Prove:

(a) If x is rational, (1.33) has only finitely many irreducible solutions.

(b) Suppose that $\varphi(q) \geq 1$ and (1.35) holds for infinitely many pairs p, q but only for finitely many relatively prime ones. Then x is rational.

(c) If φ goes to infinity too rapidly, then A_φ is negligible (Theorem 1.6). But however rapidly φ goes to infinity, A_φ is nonempty, even uncountable. *Hint:* Consider $x = \sum_{k=1}^\infty 1/2^{\alpha(k)}$ for integral $\alpha(k)$ increasing very rapidly to infinity.

SECTION 2. PROBABILITY MEASURES

Spaces

Let Ω be an arbitrary space or set of points ω. In probability theory Ω consists of all the possible results or outcomes ω of an experiment or observation. For observing the number of heads in n tosses of a coin the space Ω is $\{0, 1, \ldots, n\}$; for describing the complete history of the n tosses Ω is the space of all 2^n n-long sequences of H's and T's; for an infinite sequence of tosses Ω can be taken as the unit interval as in the preceding section; for the number of α-particles emitted by a substance during a unit interval of time or for the number of telephone calls arriving at an exchange Ω is $\{0, 1, 2, \ldots\}$; for the position of a particle Ω is three-dimensional Euclidean space; for describing the motion of the particle Ω is an appropriate space of functions; and so on. Most Ω's to be considered are interesting from the point of view of geometry and analysis as well as that of probability.

Viewed probabilistically, a subset of Ω is an *event* and an element ω of Ω is a *sample point*.

Assigning Probabilities

In setting up a space Ω as a probabilistic model, it is natural to try and assign probabilities to as many events as possible. Consider again the case $\Omega = (0, 1]$—the unit interval. It is natural to try and go beyond the definition (1.3) and assign probabilities in a systematic way to sets other than finite unions of intervals. Since the set of nonnormal numbers is negligible, for example, one feels it ought to have probability 0. For another probabilistically interesting set that is not a finite union of intervals, consider

$$(2.1) \qquad \bigcup_{n=1}^{\infty} [\omega: -a < s_1(\omega), \ldots, s_{n-1}(\omega) < b, \ s_n(\omega) = -a],$$

where a and b are positive integers. This is the event that the gambler's fortune reaches $-a$ before it reaches $+b$; it represents ruin for a gambler with a dollars playing against an adversary with b dollars, the rule being that they play until one or the other runs out of capital.

The union in (2.1) is countable and disjoint, and for each n the set in the union is itself a union of certain of the intervals (1.9). Thus (2.1) is a countably infinite disjoint union of intervals, and it is natural to take as its probability the sum of the lengths of these constituent intervals. Since the set of normal numbers is not a countable disjoint union of intervals, however, this extension of the definition of probability would still not cover all the interesting sets (events) in $(0, 1]$.

It is, in fact, not fruitful to try to predict just which sets probabilistic analysis will require and then assign probabilities to them in some *ad hoc* way. The successful procedure is to develop a general theory that assigns probabilities at once to the sets of a class so extensive that most of its members never actually arise in probability theory. That being so, why not ask for a theory that goes all the way and applies to *every* set in a space Ω? In the case of the unit interval, should there not exist a well-defined probability that the random point ω lies in A, whatever the set A may be? The answer turns out to be no (see p. 45), and it is necessary to work within subclasses of the class of all subsets of a space Ω. The classes of the appropriate kinds—the fields and σ-fields—are defined and studied in this section. The theory developed here covers the spaces listed above, including the unit interval, and a great variety of others.

Classes of Sets

It is necessary to single out for special treatment classes of subsets of a space Ω, and to be useful, such a class must be closed under various of the

operations of set theory. Once again the unit interval provides an instructive example.

*Example 2.1.** Consider the set N of normal numbers in the form (1.24), where $s_n(\omega)$ is the sum of the first n Rademacher functions. Since a point ω lies in N if and only if $\lim_n n^{-1}s_n(\omega) = 0$, N can be put in the form

$$(2.2) \qquad N = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} [\omega: |n^{-1}s_n(\omega)| < k^{-1}].$$

Indeed, because of the very meaning of union and of intersection, ω lies in the set on the right here if and only if for every k there exists an m such that $|n^{-1}s_n(\omega)| < k^{-1}$ holds for all $n \geq m$, and this is just the definition of convergence to 0—with the usual ϵ replaced by k^{-1} to avoid the formation of an uncountable intersection. Since $s_n(\omega)$ is constant over each dyadic interval of rank n, the set $[\omega: n^{-1}s_n(\omega)| < k^{-1}]$ is a finite disjoint union of intervals. The formula (2.2) shows explicitly how N is constructed in steps from these simpler sets. ■

A systematic treatment of the ideas in Section 1 thus requires a class of sets that contains the intervals and is closed under the formation of countable unions and intersections. Note that a singleton [A1] $\{x\}$ is a countable intersection $\bigcap_n (x - n^{-1}, x]$ of intervals. If a class contains all the singletons and is closed under the formation of *arbitrary* unions, then of course it contains *all* the subsets of Ω. As the theory of this section and the next does not apply to such extensive classes of sets, attention must be restricted to countable set-theoretic operations and in some cases even to finite ones.

Consider now a completely arbitrary nonempty space Ω. A class $\mathscr{F}$ of subsets of Ω is called a *field*† if it contains Ω itself and is closed under the formation of complements and finite unions:

(i) $\Omega \in \mathscr{F}$;
(ii) $A \in \mathscr{F}$ implies $A^c \in \mathscr{F}$;
(iii) $A, B \in \mathscr{F}$ implies $A \cup B \in \mathscr{F}$.

Since Ω and the empty set $\varnothing$ are complementary, (i) is the same in the presence of (ii) as the assumption $\varnothing \in \mathscr{F}$. In fact, (i) simply ensures that $\mathscr{F}$ is nonempty: If $A \in \mathscr{F}$, then $A^c \in \mathscr{F}$ by (ii) and $\Omega = A \cup A^c \in \mathscr{F}$ by (iii).

By DeMorgan's law, $A \cap B = (A^c \cup B^c)^c$ and $A \cup B = (A^c \cap B^c)^c$. If $\mathscr{F}$ is closed under complementation, therefore, it is closed under the formation of finite unions if and only if it is closed under the formation of finite intersec-

*Many of the examples in the book simply illustrate the concepts at hand, but others contain definitions and facts needed subsequently.

†The term *algebra* is often used in place of *field*.

tions. Thus (iii) can be replaced by the requirement

(iii′) $A, B \in \mathscr{F}$ implies $A \cap B \in \mathscr{F}$.

A class $\mathscr{F}$ of subsets of Ω is a *σ-field* if it is a field and if it is also closed under the formation of *countable* unions:

(iv) $A_1, A_2, \ldots \in \mathscr{F}$ implies $A_1 \cup A_2 \cup \cdots \in \mathscr{F}$.

By the infinite form of DeMorgan's law, assuming (iv) is the same thing as assuming

(iv′) $A_1, A_2, \ldots \in \mathscr{F}$ implies $A_1 \cap A_2 \cap \cdots \in \mathscr{F}$.

Note that (iv) implies (iii) because one can take $A_1 = A$ and $A_n = B$ for $n \geq 2$. A field is sometimes called a *finitely additive* field to stress that it need not be a σ-field. A set in a given class $\mathscr{F}$ is said to be *measurable* $\mathscr{F}$ or to be an *$\mathscr{F}$-set*. A field or σ-field of subsets of Ω will sometimes be called a field or σ-field *in* Ω.

Example 2.2. Section 1 began with a consideration of the sets (1.2), the finite disjoint unions of subintervals of $\Omega = (0, 1]$. Augmented by the empty set, this class is a field $\mathscr{B}_0$: Suppose that $A = (a_1, a_1'] \cup \cdots \cup (a_m, a_m']$, where the notation is so chosen that $a_1 \leq \cdots \leq a_m$. If the $(a_i, a_i']$ are disjoint, then A^c is $(0, a_1] \cup (a_1', a_2] \cup \cdots \cup (a_{m-1}', a_m] \cup (a_m', 1]$ and so lies in $\mathscr{B}_0$ (some of these intervals may be empty, as a_i' and a_{i+1} may coincide). If $B = (b_1, b_1'] \cup \cdots \cup (b_n, b_n']$, the $(b_j, b_j']$ again disjoint, then $A \cap B = \bigcup_{i=1}^m \bigcup_{j=1}^n \{(a_i, a_i'] \cap (b_j, b_j']\}$; each intersection here is again an interval or else the empty set, and the union is disjoint, and hence $A \cap B$ is in $\mathscr{B}_0$. Thus $\mathscr{B}_0$ satisfies (i), (ii), and (iii′).

Although $\mathscr{B}_0$ is a field, it is not a σ-field: It does not contain the singletons $\{x\}$, even though each is a countable intersection $\bigcap_n (x - n^{-1}, x]$ of $\mathscr{B}_0$-sets. And $\mathscr{B}_0$ does not contain the set (2.1), a countable union of intervals that cannot be represented as a finite union of intervals. The set (2.2) of normal numbers is also outside $\mathscr{B}_0$. ■

The definitions above involve distinctions perhaps most easily made clear by a pair of artificial examples.

Example 2.3. Let $\mathscr{F}$ consist of the finite and the cofinite sets (A being cofinite if A^c is finite). Then $\mathscr{F}$ is a field. If Ω is finite, then $\mathscr{F}$ contains all the subsets of Ω and hence is a σ-field as well. If Ω is infinite, however, then $\mathscr{F}$ is not a σ-field. Indeed, choose in Ω a set A that is countably infinite and has infinite complement. (For example, choose a sequence $\omega_1, \omega_2, \ldots$ of distinct points in Ω and take $A = \{\omega_2, \omega_4, \ldots\}$.) Then $A \notin \mathscr{F}$, even though A is the union, necessarily countable, of the singletons it contains and each singleton is in $\mathscr{F}$. This shows that the definition of σ-field is indeed more restrictive than that of field. ■

Example 2.4. Let $\mathscr{F}$ consist of the countable and the cocountable sets (A being cocountable if A^c is countable). Then $\mathscr{F}$ is a σ-field. If Ω is uncountable, then it contains a set A such that A and A^c are both uncountable.[†] Such a set is not in $\mathscr{F}$, which shows that even a σ-field may not contain all the subsets of Ω; furthermore, this set is the union (uncountable) of the singletons it contains and each singleton is in $\mathscr{F}$, which shows that a σ-field may not be closed under the formation of arbitrary unions. ■

The largest σ-field in Ω is the *power class* 2^Ω, consisting of *all* the subsets of Ω; the smallest σ-field consists only of the empty set and Ω itself.

The elementary facts about fields and σ-fields are easy to prove: If $\mathscr{F}$ is a field, then $A, B \in \mathscr{F}$ implies $A - B = A \cap B^c \in \mathscr{F}$ and $A \vartriangle B = (A - B) \cup (B - A) \in \mathscr{F}$. Further, it follows by induction on n that $A_1, \ldots, A_n \in \mathscr{F}$ implies $A_1 \cup \cdots \cup A_n \in \mathscr{F}$ and $A_1 \cap \cdots \cap A_n \in \mathscr{F}$.

A field is closed under the finite set-theoretic operations, and a σ-field is closed also under the countable ones. The analysis of a probability problem usually begins with the sets of some rather small class $\mathscr{A}$, such as the class of subintervals of $(0, 1]$. As in Example 2.1, probabilistically natural constructions involving finite and countable operations can then lead to sets outside the initial class $\mathscr{A}$. This leads one to consider a class of sets that (i) contains $\mathscr{A}$ and (ii) is a σ-field; it is natural and convenient, as it turns out, to consider a class that has these two properties and that in addition (iii) is in a certain sense as small as possible. As will be shown, this class is the *intersection of all the σ-fields containing* $\mathscr{A}$; it is called the *σ-field generated by* $\mathscr{A}$ and is denoted by $\sigma(\mathscr{A})$.

There do exist σ-fields containing $\mathscr{A}$, the class of all subsets of Ω being one. Moreover, a completely arbitrary intersection of σ-fields (however many of them there may be) is itself a σ-field: Suppose that $\mathscr{F} = \bigcap_\theta \mathscr{F}_\theta$, where θ ranges over an arbitrary index set and each $\mathscr{F}_\theta$ is a σ-field. Then $\Omega \in \mathscr{F}_\theta$ for all θ, so that $\Omega \in \mathscr{F}$. And $A \in \mathscr{F}$ implies for each θ that $A \in \mathscr{F}_\theta$ and hence $A^c \in \mathscr{F}_\theta$, so that $A^c \in \mathscr{F}$. If $A_n \in \mathscr{F}$ for each n, then $A_n \in \mathscr{F}_\theta$ for each n and θ, so that $\bigcup_n A_n$ lies in each $\mathscr{F}_\theta$ and hence in $\mathscr{F}$.

Thus the intersection in the definition of $\sigma(\mathscr{A})$ is indeed a σ-field containing $\mathscr{A}$. It is as small as possible, in the sense that it is contained in every σ-field that contains $\mathscr{A}$: if $\mathscr{A} \subset \mathscr{G}$ and $\mathscr{G}$ is a σ-field, then $\mathscr{G}$ is one of

[†] If Ω is the unit interval, for example, take $A = (0, \frac{1}{2}]$, say. To show that the general uncountable Ω contains such an A requires the axiom of choice [A8]. As a matter of fact, to prove the existence of the sequence alluded to in Example 2.3 requires a form of the axiom of choice, as does even something so apparently down-to-earth as proving that a countable union of negligible sets is negligible. Most of us use the axiom of choice completely unaware of the fact. Even Borel and Lebesgue did; see WAGON, pp. 217 ff.

the σ-fields in the intersection defining $\sigma(\mathscr{A})$, so that $\sigma(\mathscr{A}) \subset \mathscr{G}$. Thus $\sigma(\mathscr{A})$ has these three properties:

(i) $\mathscr{A} \subset \sigma(\mathscr{A})$;

(ii) $\sigma(\mathscr{A})$ *is a σ-field;*

(iii) *if* $\mathscr{A} \subset \mathscr{G}$ *and* $\mathscr{G}$ *is a σ-field, then* $\sigma(\mathscr{A}) \subset \mathscr{G}$.

The importance of σ-fields will gradually become clear.

Example 2.5. If $\mathscr{F}$ is a σ-field, then obviously $\sigma(\mathscr{F}) = \mathscr{F}$. If $\mathscr{A}$ consists of the singletons, then $\sigma(\mathscr{A})$ is the σ-field in Example 2.4. If $\mathscr{A}$ is empty or $\mathscr{A} = \{\varnothing\}$ or $\mathscr{A} = \{\Omega\}$, then $\sigma(\mathscr{A}) = \{\varnothing, \Omega\}$. If $\mathscr{A} \subset \mathscr{A}'$, then $\sigma(\mathscr{A}) \subset \sigma(\mathscr{A}')$. If $\mathscr{A} \subset \mathscr{A}' \subset \sigma(\mathscr{A})$, then $\sigma(\mathscr{A}) = \sigma(\mathscr{A}')$. ∎

Example 2.6. Let $\mathscr{I}$ be the class of subintervals of $\Omega = (0, 1]$, and define $\mathscr{B} = \sigma(\mathscr{I})$. The elements of $\mathscr{B}$ are called the *Borel sets* of the unit interval. The field $\mathscr{B}_0$ of Example 2.2 satisfies $\mathscr{I} \subset \mathscr{B}_0 \subset \mathscr{B}$, and hence $\sigma(\mathscr{B}_0) = \mathscr{B}$.

Since $\mathscr{B}$ contains the intervals and is a σ-field, repeated finite and countable set-theoretic operations starting from intervals will never lead outside $\mathscr{B}$. Thus $\mathscr{B}$ contains the set (2.2) of normal numbers. It also contains for example the open sets in $(0, 1]$: If G is open and $x \in G$, then there exist rationals a_x and b_x such that $x \in (a_x, b_x] \subset G$. But then $G = \bigcup_{x \in G}(a_x, b_x]$; since there are only countably many intervals with rational endpoints, G is a *countable* union of elements of $\mathscr{I}$ and hence lies in $\mathscr{B}$.

In fact, $\mathscr{B}$ contains all the subsets of $(0, 1]$ actually encountered in ordinary analysis and probability. It is large enough for all "practical" purposes. It does not contain every subset of the unit interval, however; see the end of Section 3 (p. 45). The class $\mathscr{B}$ will play a fundamental role in all that follows. ∎

Probability Measures

A *set function* is a real-valued function defined on some class of subsets of Ω. A set function P on a field $\mathscr{F}$ is a *probability measure* if it satisfies these conditions:

(i) $0 \le P(A) \le 1$ for $A \in \mathscr{F}$;

(ii) $P(\varnothing) = 0$, $P(\Omega) = 1$;

(iii) if $A_1, A_2, \ldots$ is a disjoint sequence of $\mathscr{F}$-sets and if $\bigcup_{k=1}^{\infty} A_k \in \mathscr{F}$, then[†]

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k). \tag{2.3}$$

[†]As the left side of (2.3) is invariant under permutations of the A_n, the same must be true of the right side. But in fact, according to Dirichlet's theorem [A26], a nonnegative series has the same value whatever order the terms are summed in.

The condition imposed on the set function P by (iii) is called *countable additivity*. Note that, since $\mathscr{F}$ is a field but perhaps not a σ-field, it is necessary in (iii) to assume that $\bigcup_{k=1}^{\infty} A_k$ lies in $\mathscr{F}$. If $A_1, \ldots, A_n$ are disjoint $\mathscr{F}$-sets, then $\bigcup_{k=1}^{n} A_k$ is also in $\mathscr{F}$ and (2.3) with $A_{n+1} = A_{n+2} = \cdots = \varnothing$ gives

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k). \tag{2.4}$$

The condition that (2.4) holds for disjoint $\mathscr{F}$-sets is *finite additivity*; it is a consequence of countable additivity. It follows by induction on n that P is finitely additive if (2.4) holds for $n = 2$—if $P(A \cup B) = P(A) + P(B)$ for disjoint $\mathscr{F}$-sets A and B.

The conditions above are redundant, because (i) can be replaced by $P(A) \ge 0$ and (ii) by $P(\Omega) = 1$. Indeed, the weakened forms (together with (iii)) imply that $P(\Omega) = P(\Omega) + P(\varnothing) + P(\varnothing) + \cdots$, so that $P(\varnothing) = 0$, and $1 = P(\Omega) = P(A) + P(A^c)$, so that $P(A) \le 1$.

Example 2.7. Consider as in Example 2.2 the field $\mathscr{B}_0$ of finite disjoint unions of subintervals of $\Omega = (0, 1]$. The definition (1.3) assigns to each $\mathscr{B}_0$-set a number—the sum of the lengths of the constituent intervals—and hence specifies a set function P on $\mathscr{B}_0$. Extended inductively, (1.4) says that P is finitely additive. In Section 1 this property was deduced from the additivity of the Riemann integral (see (1.5)). In Theorem 2.2 below, the finite additivity of P will be proved from first principles, and it will be shown that P is, in fact, countably additive—is a probability measure on the field $\mathscr{B}_0$. The hard part of the argument is in the proof of Theorem 1.3, already done; the rest will be easy. ∎

If $\mathscr{F}$ is a σ-field in Ω and P is a probability measure on $\mathscr{F}$, the triple $(\Omega, \mathscr{F}, P)$ is called a *probability measure space*, or simply a *probability space*. A *support* of P is any $\mathscr{F}$-set A for which $P(A) = 1$.

Example 2.8. Let $\mathscr{F}$ be the σ-field of all subsets of a countable space Ω, and let $p(\omega)$ be a nonnegative function on Ω. Suppose that $\sum_{\omega \in \Omega} p(\omega) = 1$, and define $P(A) = \sum_{\omega \in A} p(\omega)$; since $p(\omega) \ge 0$, the order of summation is irrelevant by Dirichlet's theorem [A26]. Suppose that $A = \bigcup_{i=1}^{\infty} A_i$, where the A_i are disjoint, and let $\omega_{i1}, \omega_{i2}, \ldots$ be the points in A_i. By the theorem on nonnegative double series [A27], $P(A) = \sum_{ij} p(\omega_{ij}) = \sum_i \sum_j p(\omega_{ij}) = \sum_i P(A_i)$, and so P is countably additive. This $(\Omega, \mathscr{F}, P)$ is a *discrete probability space*. It is the formal basis for discrete probability theory. ∎

Example 2.9. Now consider a probability measure P on an arbitrary σ-field $\mathscr{F}$ in an arbitrary space Ω; P is a *discrete probability measure* if there exist finitely or countably many points ω_k and masses m_k such that $P(A) = \sum_{\omega_k \in A} m_k$ for A in $\mathscr{F}$. Here P is discrete, but the space itself may not be. In terms of indicator functions, the defining condition is $P(A) = \sum_k m_k I_A(\omega_k)$ for $A \in \mathscr{F}$. If the set $\{\omega_1, \omega_2, \ldots\}$ lies in $\mathscr{F}$, then it is a support of P.

If there is just one of these points, say ω_0, with mass $m_0 = 1$, then P is a *unit mass* at ω_0. In this case $P(A) = I_A(\omega_0)$ for $A \in \mathscr{F}$. ■

Suppose that P is a probability measure on a field $\mathscr{F}$, and that $A, B \in \mathscr{F}$ and $A \subset B$. Since $P(A) + P(B - A) = P(B)$, P is *monotone*:

$$P(A) \le P(B) \quad \text{if } A \subset B. \tag{2.5}$$

It follows further that $P(B - A) = P(B) - P(A)$, and as a special case,

$$P(A^c) = 1 - P(A). \tag{2.6}$$

Other formulas familiar from the discrete theory are easily proved. For example,

$$P(A) + P(B) = P(A \cup B) + P(A \cap B), \tag{2.7}$$

the common value of the two sides being $P(A \cup B^c) + 2P(A \cap B) + P(A^c \cap B)$. Subtraction gives

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{2.8}$$

This is the case $n = 2$ of the general *inclusion-exclusion formula*:

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). \tag{2.9}$$

To deduce this inductively from (2.8), note that (2.8) gives

$$P\left(\bigcup_{k=1}^{n+1} A_k\right) = P\left(\bigcup_{k=1}^n A_k\right) + P(A_{n+1}) - P\left(\bigcup_{k=1}^n (A_k \cap A_{n+1})\right).$$

Applying (2.9) to the first and third terms on the right gives (2.9) with $n+1$ in place of n.

If $B_1 = A_1$ and $B_k = A_k \cap A_1^c \cap \cdots \cap A_{k-1}^c$, then the B_k are disjoint and $\bigcup_{k=1}^n A_k = \bigcup_{k=1}^n B_k$, so that $P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(B_k)$. Since $P(B_k) \le P(A_k)$ by monotonicity, this establishes the *finite subadditivity* of P:

$$P\left(\bigcup_{k=1}^n A_k\right) \le \sum_{k=1}^n P(A_k). \tag{2.10}$$

Here, of course, the A_k need not be disjoint. Sometimes (2.10) is called *Boole's inequality*.

In these formulas all the sets are naturally assumed to lie in the field $\mathscr{F}$. The derivations above involve only the finite additivity of P. Countable additivity gives further properties:

Theorem 2.1. *Let P be a probability measure on a field $\mathscr{F}$.*

(i) *Continuity from below: If A_n and A lie in $\mathscr{F}$ and*[†] *$A_n \uparrow A$, then $P(A_n) \uparrow P(A)$.*

(ii) *Continuity from above: If A_n and A lie in $\mathscr{F}$ and $A_n \downarrow A$, then $P(A_n) \downarrow P(A)$.*

(iii) *Countable subadditivity: If $A_1, A_2, \ldots$ and $\bigcup_{k=1}^\infty A_k$ lie in $\mathscr{F}$ (the A_k need not be disjoint), then*

$$P\left(\bigcup_{k=1}^\infty A_k\right) \le \sum_{k=1}^\infty P(A_k). \tag{2.11}$$

PROOF. For (i), put $B_1 = A_1$ and $B_k = A_k - A_{k-1}$. Then the B_k are disjoint, $A = \bigcup_{k=1}^\infty B_k$, and $A_n = \bigcup_{k=1}^n B_k$, so that by countable and finite additivity, $P(A) = \sum_{k=1}^\infty P(B_k) = \lim_n \sum_{k=1}^n P(B_k) = \lim_n P(A_n)$. For (ii), observe that $A_n \downarrow A$ implies $A_n^c \uparrow A^c$, so that $1 - P(A_n) \uparrow 1 - P(A)$.

As for (iii), increase the right side of (2.10) to $\sum_{k=1}^\infty P(A_k)$ and then apply part (i) to the left side. ■

Example 2.10. In the presence of finite additivity, a special case of (ii) implies countable additivity. *If P is a finitely additive probability measure on the field $\mathscr{F}$, and if $A_n \downarrow \emptyset$ for sets A_n in $\mathscr{F}$ implies $P(A_n) \downarrow 0$, then P is countably additive.* Indeed, if $B = \bigcup_k B_k$ for disjoint sets B_k (B and the B_k in $\mathscr{F}$), then $C_n = \bigcup_{k>n} B_k = B - \bigcup_{k \le n} B_k$ lies in the field $\mathscr{F}$, and $C_n \downarrow \emptyset$. The hypothesis, together with finite additivity, gives $P(B) - \sum_{k=1}^n P(B_k) = P(C_n) \to 0$, and hence $P(B) = \sum_{k=1}^\infty P(B_k)$. ■

Lebesgue Measure on the Unit Interval

The definition (1.3) specifies a set function on the field $\mathscr{B}_0$ of finite disjoint unions of intervals in $(0, 1]$; the problem is to prove P countably additive. It will be convenient to change notation from P to λ, and to denote by $\mathscr{I}$ the class of subintervals $(a, b]$ of $(0, 1]$; then $\lambda(I) = |I| = b - a$ is ordinary length. Regard $\emptyset$ as an element of $\mathscr{I}$ of length 0. If $A = \bigcup_{i=1}^n I_i$, the I_i being

[†] For the notation, see [A4] and [A10].

disjoint $\mathscr{I}$-sets, the definition (1.3) in the new notation is

$$\lambda(A) = \sum_{i=1}^{n} \lambda(I_i) = \sum_{i=1}^{n} |I_i|. \tag{2.12}$$

As pointed out in Section 1, there is a question of uniqueness here, because A will have other representations as a finite disjoint union $\bigcup_{j=1}^{m} J_j$ of $\mathscr{I}$-sets. But $\mathscr{I}$ is closed under the formation of finite intersections, and so the finite form of Theorem 1.3(iii) gives

$$\sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n} \sum_{j=1}^{m} |I_i \cap J_j| = \sum_{j=1}^{m} |J_j|. \tag{2.13}$$

(Some of the $I_i \cap J_j$ may be empty, but the corresponding lengths are then 0.) The definition is indeed consistent.

Thus (2.12) defines a set function λ on $\mathscr{B}_0$, a set function called *Lebesgue measure*.

Theorem 2.2. *Lebesgue measure λ is a (countably additive) probability measure on the field $\mathscr{B}_0$.*

PROOF. Suppose that $A = \bigcup_{k=1}^{\infty} A_k$, where A and the A_k are $\mathscr{B}_0$-sets and the A_k are disjoint. Then $A = \bigcup_{i=1}^{n} I_i$ and $A_k = \bigcup_{j=1}^{m_k} J_{kj}$ are disjoint unions of $\mathscr{I}$-sets, and (2.12) and Theorem 1.3(iii) give

$$\begin{aligned} \lambda(A) &= \sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n} \sum_{k=1}^{\infty} \sum_{j=1}^{m_k} |I_i \cap J_{kj}| \\ &= \sum_{k=1}^{\infty} \sum_{j=1}^{m_k} |J_{kj}| = \sum_{k=1}^{\infty} \lambda(A_k). \end{aligned} \tag{2.14}$$ ■

In Section 3 it is shown how to extend λ from $\mathscr{B}_0$ to the larger class $\mathscr{B} = \sigma(\mathscr{B}_0)$ of Borel sets in $(0,1]$. This will complete the construction of λ as a probability measure (countably additive, that is) on $\mathscr{B}$, and the construction is fundamental to all that follows. For example, the set N of normal numbers lies in $\mathscr{B}$ (Example 2.6), and it will turn out that $\lambda(N) = 1$, as probabilistic intuition requires. (In Chapter 2, λ will be defined for sets outside the unit interval as well.)

It is well to pause here and consider just what is involved in the construction of Lebesgue measure on the Borel sets of the unit interval. That length defines a finitely additive set function on the class $\mathscr{I}$ of intervals in $(0,1]$ is a consequence of Theorem 1.3 for the case of only finitely many intervals and thus involves only the most elementary properties of the real number system. But proving countable additivity on $\mathscr{I}$ requires the deeper property of compactness (the Heine-Borel theorem). Once λ has been proved countably additive on $\mathscr{I}$, extending it to $\mathscr{B}_0$ by the definition (2.12) presents no real difficulty: the arguments involving (2.13) and (2.14) are easy. Difficulties again arise, however, in the further extension of λ from $\mathscr{B}_0$ to $\mathscr{B} = \sigma(\mathscr{B}_0)$, and here new ideas are again required. These ideas are the subject of Section 3, where it is shown that any probability measure on any field can be extended to the generated σ-field.

Sequence Space*

Let S be a finite set of points regarded as the possible outcomes of a simple observation or experiment. For tossing a coin, S can be $\{H, T\}$ or $\{0,1\}$; for rolling a die, $S = \{1, \dots, 6\}$; in information theory, S plays the role of a finite alphabet. Let $\Omega = S^\infty$ be the space of all infinite sequences

$$\omega = (z_1(\omega), z_2(\omega), \dots) \tag{2.15}$$

of elements of S: $z_k(\omega) \in S$ for all $\omega \in S^\infty$ and $k \geq 1$. The sequence (2.15) can be viewed as the result of repeating infinitely often the simple experiment represented by S. For $S = \{0,1\}$, the space S^∞ is closely related to the unit interval; compare (1.8) and (2.15).

The space S^∞ is an infinite-dimensional Cartesian product. Each $z_k(\cdot)$ is a mapping of S^∞ onto S; these are the *coordinate functions*, or the *natural projections*. Let $S^n = S \times \cdots \times S$ be the Cartesian product of n copies of S; it consists of the n-long sequences $(u_1, \dots, u_n)$ of elements of S. For such a sequence, the set

$$[\omega: (z_1(\omega), \dots, z_n(\omega)) = (u_1, \dots, u_n)] \tag{2.16}$$

represents the event that the first n repetitions of the experiment give the outcomes $u_1, \dots, u_n$ in sequence. A *cylinder of rank n* is a set of the form

$$A = [\omega: (z_1(\omega), \dots, z_n(\omega)) \in H], \tag{2.17}$$

where $H \subset S^n$. Note that A is nonempty if H is. If H is a singleton in S^n, (2.17) reduces to (2.16), which can be called a *thin* cylinder.

Let $\mathscr{C}_0$ be the class of cylinders of all ranks. Then $\mathscr{C}_0$ *is a field*: S^∞ and the empty set have the form (2.17) for $H = S^n$ and for $H = \emptyset$. If H is replaced by $S^n - H$, then (2.17) goes into its complement, and hence $\mathscr{C}_0$ is

*The ideas that follow are basic to probability theory and are used further on, in particular in Section 24 and (in more elaborate form) Section 36. On a first reading, however, one might prefer to skip to Section 3 and return to this topic as the need arises.

closed under complementation. As for unions, consider (2.17) together with

$$B = [\omega: (z_1(\omega), \dots, z_m(\omega)) \in I], \tag{2.18}$$

a cylinder of rank m. Suppose that $n \le m$ (symmetry); if H' consists of the sequences $(u_1, \dots, u_m)$ in S^m for which the truncated sequence $(u_1, \dots, u_n)$ lies in H, then (2.17) has the alternative form

$$A = [\omega: (z_1(\omega), \dots, z_m(\omega)) \in H']. \tag{2.19}$$

Since it is now clear that

$$A \cup B = [\omega: (z_1(\omega), \dots, z_m(\omega)) \in H' \cup I] \tag{2.20}$$

is also a cylinder, $\mathscr{C}_0$ is closed under the formation of finite unions and hence is indeed a field.

Let p_u, $u \in S$, be probabilities on S—nonnegative and summing to 1. Define a set function P on $\mathscr{C}_0$ (it will turn out to be a probability measure) in this way: For a cylinder A given by (2.17), take

$$P(A) = \sum_H p_{u_1} \cdots p_{u_n}, \tag{2.21}$$

the sum extending over all the sequences $(u_1, \dots, u_n)$ in H. As a special case,

$$P[\omega: (z_1(\omega), \dots, z_n(\omega)) = (u_1, \dots, u_n)] = p_{u_1} \cdots p_{u_n}. \tag{2.22}$$

Because of the products on the right in (2.21) and (2.22), P is called *product measure*; it provides a model for an infinite sequence of independent repetitions of the simple experiment represented by the probabilities p_u on S. In the case where $S = \{0, 1\}$ and $p_0 = p_1 = \frac{1}{2}$, it is a model for independent tosses of a fair coin, an alternative to the model used in Section 1.

The definition (2.21) presents a consistency problem, since the cylinder A will have other representations. Suppose that A is also given by (2.19). If $n = m$, then H and H' must coincide, and there is nothing to prove. Suppose then (symmetry) that $n < m$. Then H' must consist of those $(u_1, \dots, u_m)$ in S^m for which $(u_1, \dots, u_n)$ lies in H: $H' = H \times S^{m-n}$. But then

$$\sum_{H'} p_{u_1} \cdots p_{u_n} p_{u_{n+1}} \cdots p_{u_m} = \sum_H p_{u_1} \cdots p_{u_n} \sum_{S^{m-n}} p_{u_{n+1}} \cdots p_{u_m} = \sum_H p_{u_1} \cdots p_{u_n}. \tag{2.23}$$

The definition (2.21) is therefore consistent. And finite additivity is now easy: Suppose that A and B are disjoint cylinders given by (2.17) and (2.18).

Suppose that $n \le m$, and put A in the form (2.19). Since A and B are disjoint, H' and I must be disjoint as well, and by (2.20),

$$P(A \cup B) = \sum_{H' \cup I} p_{u_1} \cdots p_{u_m} = P(A) + P(B). \tag{2.24}$$

Taking $H = S^n$ in (2.21) shows that $P(S^\infty) = 1$. Therefore, (2.21) *defines a finitely additive probability measure on the field* $\mathscr{C}_0$.

Now, P is countably additive on $\mathscr{C}_0$, but this requires no further argument, because of the following completely general result.

Theorem 2.3. *Every finitely additive probability measure on the field $\mathscr{C}_0$ of cylinders in S^∞ is in fact countably additive.*

The proof depends on this fundamental fact:

Lemma. *If $A_n \downarrow A$, where the A_n are nonempty cylinders, then A is nonempty.*

PROOF OF THEOREM 2.3. Assume that the lemma is true, and apply Example 2.10 to the measure P in question: If $A_n \downarrow \emptyset$ for sets in $\mathscr{C}_0$ (cylinders) but $P(A_n)$ does *not* converge to 0, then $P(A_n) \ge \epsilon > 0$ for some ϵ. But then the A_n are nonempty, which by the lemma makes $A_n \downarrow \emptyset$ impossible. ∎

PROOF OF THE LEMMA.[†] Suppose that A_t is a cylinder of rank m_t, say

$$A_t = [\omega: (z_1(\omega), \dots, z_{m_t}(\omega)) \in H_t], \tag{2.25}$$

where $H_t \subset S^{m_t}$. Choose a point ω_n in A_n, which is nonempty by assumption. Write the components of the sequences in a square array:

$$\begin{array}{cccc} z_1(\omega_1) & z_1(\omega_2) & z_1(\omega_3) & \cdots \\ z_2(\omega_1) & z_2(\omega_2) & z_2(\omega_3) & \cdots \\ \vdots & \vdots & \vdots & \end{array} \tag{2.26}$$

The nth *column* of the array gives the components of ω_n.

Now argue by a modification of the diagonal method [A14]. Since S is finite, some element u_1 of S appears infinitely often in the first row of (2.26): for an increasing sequence $\{n_{1,k}\}$ of integers, $z_1(\omega_{n_{1,k}}) = u_1$ for all k. By the same reasoning, there exist an increasing subsequence $\{n_{2,k}\}$ of $\{n_{1,k}\}$ and an

[†]The lemma is a special case of Tychonov's theorem: If S is given the discrete topology, the topological product S^∞ is compact (and the cylinders are closed).

element u_2 of S such that $z_2(\omega_{n_{2,k}}) = u_2$ for all k. Continue. If $n_k = n_{k,k}$, then $z_r(\omega_{n_k}) = u_r$ for $k \ge r$, and hence $(z_1(\omega_{n_k}), \ldots, z_r(\omega_{n_k})) = (u_1, \ldots, u_r)$ for $k \ge r$.

Let ω° be the element of S^∞ with components u_r: $\omega^\circ = (u_1, u_2, \ldots) = (z_1(\omega^\circ), z_2(\omega^\circ), \ldots)$. Let t be arbitrary. If $k \ge t$, then (n_k is increasing) $n_k \ge t$ and hence $\omega_{n_k} \in A_{n_k} \subset A_t$. It follows by (2.25) that, for $k \ge t$, H_t contains the point $(z_1(\omega_{n_k}), \ldots, z_{m_t}(\omega_{n_k}))$ of S^{m_t}. But for $k \ge m_t$, this point is identical with $(z_1(\omega^\circ), \ldots, z_{m_t}(\omega^\circ))$, which therefore lies in H_t. Thus ω° is a point common to all the A_t. ∎

Let $\mathscr{C}$ be the σ-field in S^∞ generated by $\mathscr{C}_0$. By the general theory of the next section, the probability measure P defined on $\mathscr{C}_0$ by (2.21) extends to $\mathscr{C}$. The term *product measure*, properly speaking, applies to the extended P. Thus $(S^\infty, \mathscr{C}, P)$ is a probability space, one important in ergodic theory (Section 24).

Suppose that $S = \{0, 1\}$ and $p_0 = p_1 = \frac{1}{2}$. In this case, $(S^\infty, \mathscr{C}, P)$ is closely related to $((0,1], \mathscr{B}, \lambda)$, although there are essential differences. The sequence (2.15) can end in 0's, but (1.8) cannot. Thin cylinders are like dyadic intervals, but the sets in $\mathscr{C}_0$ (the cylinders) correspond to the finite disjoint unions of intervals with dyadic endpoints, a field somewhat smaller than $\mathscr{B}_0$. While nonempty sets in $\mathscr{B}_0$ (for example, $(\frac{1}{2}, \frac{1}{2} + 2^{-n}]$) can contract to the empty set, nonempty sets in $\mathscr{C}_0$ cannot. The lemma above plays here the role the Heine-Borel theorem plays in the proof of Theorem 1.3. The product probability measure constructed here on $\mathscr{C}_0$ (in the case $S = \{0, 1\}$, $p_0 = p_1 = \frac{1}{2}$, that is) is analogous to Lebesgue measure on $\mathscr{B}_0$. But a finitely additive probability measure on $\mathscr{B}_0$ can fail to be countably additive,† which cannot happen in $\mathscr{C}_0$.

Constructing σ-Fields*

The σ-field $\sigma(\mathscr{A})$ generated by $\mathscr{A}$ was defined from above or from the outside, so to speak, by intersecting all the σ-fields that contain $\mathscr{A}$ (including the σ-field consisting of all the subsets of Ω). Can $\sigma(\mathscr{A})$ somehow be constructed from the inside by repeated finite and countable set-theoretic operations starting with sets in $\mathscr{A}$?

For any class $\mathscr{H}$ of sets in Ω let $\mathscr{H}^*$ consist of the sets in $\mathscr{H}$, the complements of sets in $\mathscr{H}$, and the finite and countable unions of sets in $\mathscr{H}$. Given a class $\mathscr{A}$, put $\mathscr{A}_0 = \mathscr{A}$ and define $\mathscr{A}_1, \mathscr{A}_2, \ldots$ inductively by

$$(2.27) \qquad \mathscr{A}_n = \mathscr{A}^*_{n-1}.$$

That each $\mathscr{A}_n$ is contained in $\sigma(\mathscr{A})$ follows by induction. One might hope that $\mathscr{A}_n = \sigma(\mathscr{A})$ for some n, or at least that $\bigcup_{n=0}^\infty \mathscr{A}_n = \sigma(\mathscr{A})$. But this process applied to the class of intervals fails to account for all the Borel sets.

Let $\mathscr{I}_0$ consist of the empty set and the intervals in $\Omega = (0,1]$ with rational endpoints, and define $\mathscr{I}_n = \mathscr{I}^*_{n-1}$ for $n = 1, 2, \ldots$. *It will be shown that* $\bigcup_{n=0}^\infty \mathscr{I}_n$ *is strictly smaller than* $\mathscr{B} = \sigma(\mathscr{I}_0)$.

†See Problem 2.15.

*This topic may be omitted.

If a_n and b_n are rationals decreasing to a and b, then $(a, b] = \bigcup_m \bigcap_n (a_m, b_n] = \bigcup_m (\bigcup_n (a_m, b_n]^c)^c \in \mathscr{I}_4$. The result would therefore not be changed by including in $\mathscr{I}_0$ *all* the intervals in $(0, 1]$.

To prove $\bigcup_{n=0}^\infty \mathscr{I}_n$ smaller than $\mathscr{B}$, first put

$$(2.28) \qquad \Psi(A_1, A_2, \ldots) = A_1^c \cup A_2 \cup A_3 \cup A_4 \cup \cdots.$$

Since $\mathscr{I}_{n-1}$ contains $\Omega = (0, 1]$ and the empty set, every element of $\mathscr{I}_n$ has the form (2.28) for some sequence $A_1, A_2, \ldots$ of sets in $\mathscr{I}_{n-1}$. Let every positive integer appear exactly once in the square array

$$\begin{matrix} m_{11} & m_{12} & \cdots \\ m_{21} & m_{22} & \cdots \\ \vdots & \vdots & \end{matrix}$$

Inductively define

$$(2.29) \qquad \Phi_0(A_1, A_2, \ldots) = A_1,$$

$$\Phi_n(A_1, A_2, \ldots) = \Psi\big(\Phi_{n-1}(A_{m_{11}}, A_{m_{12}}, \ldots), \Phi_{n-1}(A_{m_{21}}, A_{m_{22}}, \ldots), \ldots\big), \qquad n = 1, 2, \ldots.$$

It follows by induction that every element of $\mathscr{I}_n$ has the form $\Phi_n(A_1, A_2, \ldots)$ for some sequence of sets in $\mathscr{I}_0$. Finally, put

$$(2.30) \qquad \Phi(A_1, A_2, \ldots) = \Phi_1(A_{m_{11}}, A_{m_{12}}, \ldots) \cup \Phi_2(A_{m_{21}}, A_{m_{22}}, \ldots) \cup \cdots.$$

Then every element of $\bigcup_{n=0}^\infty \mathscr{I}_n$ has the form (2.30) for some sequence $A_1, A_2, \ldots$ of sets in $\mathscr{I}_0$.

If $A_1, A_2, \ldots$ are in $\mathscr{B}$, then (2.28) is in $\mathscr{B}$; it follows by induction that each $\Phi_n(A_1, A_2, \ldots)$ is in $\mathscr{B}$ and therefore that (2.30) is in $\mathscr{B}$.

With each ω in $(0, 1]$ associate the sequence $(\omega_1, \omega_2, \ldots)$ of positive integers such that $\omega_1 + \cdots + \omega_k$ is the position of the kth 1 in the nonterminating dyadic expansion of ω (the smallest n for which $\sum_{j=1}^n d_j(\omega) = k$). Then $\omega \leftrightarrow (\omega_1, \omega_2, \ldots)$ is a one-to-one correspondence between $(0, 1]$ and the set of all sequences of positive integers. Let $I_1, I_2, \ldots$ be an enumeration of the sets in $\mathscr{I}_0$, put $\varphi(\omega) = \Phi(I_{\omega_1}, I_{\omega_2}, \ldots)$, and define $B = [\omega: \omega \notin \varphi(\omega)]$. It will be shown that B is a Borel set but is not contained in any of the $\mathscr{I}_n$.

Since ω lies in B if and only if ω lies outside $\varphi(\omega)$, $B \ne \varphi(\omega)$ for every ω. But every element of $\bigcup_{n=0}^\infty \mathscr{I}_n$ has the form (2.30) for some sequence in $\mathscr{I}_0$ and hence has the form $\varphi(\omega)$ for some ω. Therefore, B is not a member of $\bigcup_{n=0}^\infty \mathscr{I}_n$.

It remains to show that B is a Borel set. Let $D_k = [\omega: \omega \in I_{\omega_k}]$. Since $L_k(n) = [\omega: \omega_1 + \cdots + \omega_k = n] = [\omega: \sum_{j=1}^{n-1} d_j(\omega) < k = \sum_{j=1}^n d_j(\omega)]$ is a Borel set, so are $[\omega: \omega_k = n] = \bigcup_{m=1}^\infty L_{k-1}(m) \cap L_k(m+n)$ and

$$D_k = [\omega: \omega \in I_{\omega_k}] = \bigcup_n ([\omega: \omega_k = n] \cap I_n).$$

Suppose that it is shown that

$$[\omega: \omega \in \Phi_n(I_{\omega_{u_1}}, I_{\omega_{u_2}} \dots)] = \Phi_n(D_{u_1}, D_{u_2}, \dots) \tag{2.31}$$

for every n and every sequence $u_1, u_2, \dots$ of positive integers. It will then follow from the definition (2.30) that

$$B^c = [\omega: \omega \in \varphi(\omega)] = \bigcup_{n=1}^{\infty} [\omega: \omega \in \Phi_n(I_{\omega_{m_{n1}}}, I_{\omega_{m_{n2}}}, \dots)]$$

$$= \bigcup_{n=1}^{\infty} \Phi_n(D_{m_{n1}}, D_{m_{n2}}, \dots) = \Phi(D_1, D_2, \dots).$$

But as remarked above, (2.30), is a Borel set if the A_n are. Therefore, (2.31) will imply that B^c and B are Borel sets.

If $n = 0$, (2.31) holds because it reduces by (2.29) to $[\omega: \omega \in I_{\omega_{u_1}}] = D_{u_1}$. Suppose that (2.31) holds with $n-1$ in place of n. Consider the condition

$$\omega \in \Phi_{n-1}(I_{\omega_{u_{m_{k1}}}}, I_{\omega_{u_{m_{k2}}}}, \dots). \tag{2.32}$$

By (2.28) and (2.29), a necessary and sufficient condition for $\omega \in \Phi_n(I_{\omega_{u_1}}, I_{\omega_{u_2}}, \dots)$ is that either (2.32) is false for $k = 1$ or else (2.32) is true for some k exceeding 1. But by the induction hypothesis, (2.32) and its negation can be replaced by $\omega \in \Phi_{n-1}(D_{u_{m_{k1}}}, D_{u_{m_{k2}}}, \dots)$ and its negation. Therefore, $\omega \in \Phi_n(I_{\omega_{u_1}}, I_{\omega_{u_2}}, \dots)$ if and only if $\omega \in \Phi_n(D_{u_1}, D_{u_2}, \dots)$.

Thus $\bigcup_n \mathscr{I}_n \neq \mathscr{B}$, and there are Borel sets that cannot be arrived at from the intervals by any finite sequence of set-theoretic operations, each operation being finite or countable. It can even be shown that there are Borel sets that cannot be arrived at by any *countable* sequence of these operations. On the other hand, every Borel set can be arrived at by a countable *ordered set* of these operations if it is not required that they be performed in a simple *sequence*. The proof of this statement—and indeed even a precise explanation of its meaning—depends on the theory of infinite ordinal numbers.[†]

PROBLEMS

2.1. Define $x \vee y = \max\{x, y\}$, and for a collection $\{x_\alpha\}$ define $\bigvee_\alpha x_\alpha = \sup_\alpha x_\alpha$; define $x \wedge y = \min\{x, y\}$ and $\bigwedge_\alpha x_\alpha = \inf_\alpha x_\alpha$. Prove that $I_{A \cup B} = I_A \vee I_B$, $I_{A \cap B} = I_A \wedge I_B$, $I_{A^c} = 1 - I_A$, and $I_{A \Delta B} = |I_A - I_B|$, in the sense that there is equality at each point of Ω. Show that $A \subset B$ if and only if $I_A \leq I_B$ pointwise. Check the equation $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ and deduce the distributive law

[†]See Problem 2.22.

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. By similar arguments prove that

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

$$A \Delta C \subset (A \Delta B) \cup (B \Delta C),$$

$$\left(\bigcup_n A_n\right)^c = \bigcap_n A_n^c,$$

$$\left(\bigcap_n A_n\right)^c = \bigcup_n A_n^c.$$

2.2. Let $A_1, \dots, A_n$ be arbitrary events, and put $U_k = \dot{\bigcup}(A_{i_1} \cap \cdots \cap A_{i_k})$ and $I_k = \bigcap(A_{i_1} \cup \cdots \cup A_{i_k})$, where the union and intersection extend over all the k-tuples satisfying $1 \leq i_1 < \cdots < i_k \leq n$. Show that $U_k = I_{n-k+1}$.

2.3. (a) Suppose that $\Omega \in \mathscr{F}$ and that $A, B \in \mathscr{F}$ implies $A - B = A \cap B^c \in \mathscr{F}$. Show that $\mathscr{F}$ is a field.

(b) Suppose that $\Omega \in \mathscr{F}$ and that $\mathscr{F}$ is closed under the formation of complements and finite *disjoint* unions. Show that $\mathscr{F}$ need not be a field.

2.4. Let $\mathscr{F}_1, \mathscr{F}_2, \dots$ be classes of sets in a common space Ω.

(a) Suppose that $\mathscr{F}_n$ are fields satisfying $\mathscr{F}_n \subset \mathscr{F}_{n+1}$. Show that $\bigcup_{n=1}^{\infty} \mathscr{F}_n$ is a field.

(b) Suppose that $\mathscr{F}_n$ are σ-fields satisfying $\mathscr{F}_n \subset \mathscr{F}_{n+1}$. Show by example that $\bigcup_{n=1}^{\infty} \mathscr{F}_n$ need not be a σ-field.

2.5. The field $f(\mathscr{A})$ generated by a class $\mathscr{A}$ in Ω is defined as the intersection of all fields in Ω containing $\mathscr{A}$.

(a) Show that $f(\mathscr{A})$ is indeed a field, that $\mathscr{A} \subset f(\mathscr{A})$, and that $f(\mathscr{A})$ is minimal in the sense that if $\mathscr{G}$ is a field and $\mathscr{A} \subset \mathscr{G}$, then $f(\mathscr{A}) \subset \mathscr{G}$.

(b) Show that for nonempty $\mathscr{A}$, $f(\mathscr{A})$ is the class of sets of the form $\bigcup_{i=1}^{m} \bigcap_{j=1}^{n_i} A_{ij}$, where for each i and j either $A_{ij} \in \mathscr{A}$ or $A_{ij}^c \in \mathscr{A}$, and where the m sets $\bigcap_{j=1}^{n_i} A_{ij}$, $1 \leq i \leq m$, are disjoint. The sets in $f(\mathscr{A})$ can thus be explicitly presented, which is not in general true of the sets in $\sigma(\mathscr{A})$.

2.6. ↑ (a) Show that if $\mathscr{A}$ consists of the singletons, then $f(\mathscr{A})$ is the field in Example 2.3.

(b) Show that $f(\mathscr{A}) \subset \sigma(\mathscr{A})$, that $f(\mathscr{A}) = \sigma(\mathscr{A})$ if $\mathscr{A}$ is finite, and that $\sigma(f(\mathscr{A})) = \sigma(\mathscr{A})$.

(c) Show that if $\mathscr{A}$ is countable, then $f(\mathscr{A})$ is countable.

(d) Show for fields $\mathscr{F}_1$ and $\mathscr{F}_2$ that $f(\mathscr{F}_1 \cup \mathscr{F}_2)$ consists of the finite disjoint unions of sets $A_1 \cap A_2$ with $A_i \in \mathscr{F}_i$. Extend.

2.7. 2.5↑ Let H be a a set lying outside $\mathscr{F}$, where $\mathscr{F}$ is a field [or σ-field]. Show that the field [or σ-field] generated by $\mathscr{F} \cup \{H\}$ consists of sets of the form

$$(H \cap A) \cup (H^c \cap B), \qquad A, B \in \mathscr{F}. \tag{2.33}$$

2.8. Suppose for each A in $\mathscr{A}$ that A^c is a countable union of elements of $\mathscr{A}$. The class of intervals in $(0,1]$ has this property. Show that $\sigma(\mathscr{A})$ coincides with the smallest class over $\mathscr{A}$ that is closed under the formation of countable unions and intersections.

2.9. Show that, if $B \in \sigma(\mathscr{A})$, then there exists a countable subclass $\mathscr{A}_B$ of $\mathscr{A}$ such that $B \in \sigma(\mathscr{A}_B)$.

2.10. (a) Show that if $\sigma(\mathscr{A})$ contains every subset of Ω, then for each pair ω and ω' of distinct points in Ω there is in $\mathscr{A}$ an A such that $I_A(\omega) \neq I_A(\omega')$.

(b) Show that the reverse implication holds if Ω is countable.

(c) Show by example that the reverse implication need not hold for uncountable Ω.

2.11. A σ-field is *countably generated*, or *separable*, if it is generated by some countable class of sets.

(a) Show that the σ-field $\mathscr{B}$ of Borel sets is countably generated.

(b) Show that the σ-field of Example 2.4 is countably generated if and only if Ω is countable.

(c) Suppose that $\mathscr{F}_1$ and $\mathscr{F}_2$ are σ-fields, $\mathscr{F}_1 \subset \mathscr{F}_2$, and $\mathscr{F}_2$ is countably generated. Show by example that $\mathscr{F}_1$ may not be countably generated.

2.12. Show that a σ-field cannot be countably infinite—its cardinality must be finite or else at least that of the continuum. Show by example that a field can be countably infinite.

2.13. (a) Let $\mathscr{F}$ be the field consisting of the finite and the cofinite sets in an infinite Ω, and define P on $\mathscr{F}$ by taking $P(A)$ to be 0 or 1 as A is finite or cofinite. (Note that P is not well defined if Ω is finite.) Show that P is finitely additive.

(b) Show that this P is not countably additive if Ω is countably infinite.

(c) Show that this P is countably additive if Ω is uncountable.

(d) Now let $\mathscr{F}$ be the σ-field consisting of the countable and the cocountable sets in an uncountable Ω, and define P on $\mathscr{F}$ by taking $P(A)$ to be 0 or 1 as A is countable or cocountable. (Note that P is not well defined if Ω is countable.) Show that P is countably additive.

2.14. In $(0,1]$ let $\mathscr{F}$ be the class of sets that either (i) are of the first category [A15] or (ii) have complement of the first category. Show that $\mathscr{F}$ is a σ-field. For A in $\mathscr{F}$, take $P(A)$ to be 0 in case (i) and 1 in case (ii). Show that P is countably additive.

2.15. On the field $\mathscr{B}_0$ in $(0,1]$ define $P(A)$ to be 1 or 0 according as there does or does not exist some positive ϵ_A (depending on A) such that A contains the interval $(\frac{1}{2}, \frac{1}{2}+\epsilon_A]$. Show that P is finitely but not countably additive. No such example is possible for the field $\mathscr{C}_0$ in S^∞ (Theorem 2.3).

2.16. (a) Suppose that P is a probability measure on a field $\mathscr{F}$. Suppose that $A_t \in \mathscr{F}$ for $t > 0$, that $A_s \subset A_t$ for $s < t$, and that $A = \bigcup_{t>0} A_t \in \mathscr{F}$. Extend Theorem 2.1(i) by showing that $P(A_t) \uparrow P(A)$ as $t \to \infty$. Show that A necessarily lies in $\mathscr{F}$ if it is a σ-field.

(b) Extend Theorem 2.1(ii) in the same way.

2.17. Suppose that P is a probability measure on a field $\mathscr{F}$, that $A_1, A_2, \ldots$, and $A = \bigcup_n A_n$ lie in $\mathscr{F}$, and that the A_n are nearly disjoint in the sense that $P(A_m \cap A_n) = 0$ for $m \neq n$. Show that $P(A) = \sum_n P(A_n)$.

2.18. *Stochastic arithmetic.* Define a set function P_n on the class of all subsets of $\Omega = \{1, 2, \ldots\}$ by

$$P_n(A) = \frac{1}{n}\#[m: 1 \le m \le n,\ m \in A]; \tag{2.34}$$

among the first n integers, the proportion that lie in A is just $P_n(A)$. Then P_n is a discrete probability measure. The set A has *density*

$$D(A) = \lim_n P_n(A), \tag{2.35}$$

provided this limit exists. Let $\mathscr{D}$ be the class of sets having density.

(a) Show that D is finitely but not countably additive on $\mathscr{D}$.

(b) Show that $\mathscr{D}$ contains the empty set and Ω and is closed under the formation of complements, proper differences, and finite disjoint unions, but is not closed under the formation of countable disjoint unions or of finite unions that are not disjoint.

(c) Let $\mathscr{M}$ consist of the periodic sets $M_a = [ka: k = 1, 2, \ldots]$. Observe that

$$P_n(M_a) = \frac{1}{n}\left\lfloor \frac{n}{a} \right\rfloor \to \frac{1}{a} = D(M_a). \tag{2.36}$$

Show that the field $f(\mathscr{M})$ generated by $\mathscr{M}$ (see Problem 2.5) is contained in $\mathscr{D}$. Show that D is completely determined on $f(\mathscr{M})$ by the value it gives for each a to the event that m is divisible by a.

(d) Assume that $\sum p^{-1}$ diverges (sum over all primes; see Problem 5.20(e)) and prove that D, although finitely additive, is not countably additive on the field $f(M)$.

(e) Euler's function $\varphi(n)$ is the number of positive integers less than n and relatively prime to it. Let $p_1, \ldots, p_r$ be the distinct prime factors of n; from the inclusion-exclusion formula for the events $[m: p_i | m]$, (2.36), and the fact that the p_i divide n, deduce

$$\frac{\varphi(n)}{n} = \prod_{p|n}\left(1 - \frac{1}{p}\right). \tag{2.37}$$

(f) Show for $0 \le x \le 1$ that $D(A) = x$ for some A.

(g) Show that D is translation invariant: If $B = [m+1: m \in A]$, then B has a density if and only if A does, in which case $D(A) = D(B)$.

2.19. A probability measure space $(\Omega, \mathscr{F}, P)$ is *nonatomic* if $P(A) > 0$ implies that there exists a B such that $B \subset A$ and $0 < P(B) < P(A)$ (A and B in $\mathscr{F}$, of course).

(a) Assuming the existence of Lebesgue measure λ on $\mathscr{B}$, prove that it is nonatomic.

(b) Show in the nonatomic case that $P(A) > 0$ and $\epsilon > 0$ imply that there exists a B such that $B \subset A$ and $0 < P(B) < \epsilon$.

(c) Show in the nonatomic case that $0 \le x \le P(A)$ implies that there exists a B such that $B \subset A$ and $P(B) = x$. *Hint*: Inductively define classes $\mathscr{H}_n$, numbers h_n, and sets H_n by $\mathscr{H}_0 = \{\emptyset\} = \{H_0\}$, $\mathscr{H}_n = [H\colon\ H \subset A - \bigcup_{k<n} H_k,\ P(\bigcup_{k<n} H_k) + P(H) \le x]$, $h_n = \sup[P(H)\colon\ H \in \mathscr{H}_n]$, and $P(H_n) > h_n - n^{-1}$. Consider $\bigcup_k H_k$.

(d) Show in the nonatomic case that, if $p_1, p_2, \ldots$ are nonnegative and add to 1, then A can be decomposed into sets $B_1, B_2, \ldots$ such that $P(B_n) = p_n P(A)$.

2.20. Generalize the construction of product measure: For $n = 1, 2, \ldots$, let S_n be a finite space with given probabilities p_{nu}, $u \in S_n$. Let $S_1 \times S_2 \times \cdots$ be the space of sequences (2.15), where now $z_k(\omega) \in S_k$. Define P on the class of cylinders, appropriately defined, by using the product $p_{1u_1} \cdots p_{nu_n}$ on the right in (2.21). Prove P countably additive on $\mathscr{C}_0$, and extend Theorem 2.3 and its lemma to this more general setting. Show that the lemma fails if any of the S_n are infinite.

2.21. (a) Suppose that $\mathscr{A} = \{A_1, A_2, \ldots\}$ is a countable partition of Ω. Show (see (2.27)) that $\mathscr{A}_1 = \mathscr{A}_0^* = \mathscr{A}^*$ coincides with $\sigma(\mathscr{A})$. This is a case where $\sigma(\mathscr{A})$ can be constructed "from the inside."

(b) Show that the set of normal numbers lies in $\mathscr{I}_6$.

(c) Show that $\mathscr{H}^* = \mathscr{H}$ if and only if $\mathscr{H}$ is a σ-field. Show that $\mathscr{I}_{n-1}$ is strictly smaller than $\mathscr{I}_n$ for all n.

2.22. Extend (2.27) to infinite ordinals α by defining $\mathscr{A}_\alpha = (\bigcup_{\beta<\alpha} \mathscr{A}_\beta)^*$. Show that, if Ω is the first uncountable ordinal, then $\bigcup_{\alpha<\Omega} \mathscr{A}_\alpha = \sigma(\mathscr{A})$. Show that, if the cardinality of $\mathscr{A}$ does not exceed that of the continuum, then the same is true of $\sigma(\mathscr{A})$. Thus $\mathscr{B}$ has the power of the continuum.

2.23. ↑ Extend (2.29) to ordinals $\alpha < \Omega$ as follows. Replace the right side of (2.28) by $\bigcup_{n=1}^\infty (A_{2n-1} \cup A_{2n}^c)$. Suppose that Φ_β is defined for $\beta < \alpha$. Let $\beta_\alpha(1), \beta_\alpha(2), \ldots$ be a sequence of ordinals such that $\beta_\alpha(n) < \alpha$ and such that if $\beta < \alpha$, then $\beta = \beta_\alpha(n)$ for infinitely many even n and for infinitely many odd n; define

$$(2.38)\quad \Phi_\alpha(A_1, A_2, \ldots) = \Psi\big(\Phi_{\beta_\alpha(1)}(A_{m_{11}}, A_{m_{12}}, \ldots), \Phi_{\beta_\alpha(2)}(A_{m_{21}}, A_{m_{22}}, \ldots), \ldots\big).$$

Prove by transfinite induction that (2.38) is in $\mathscr{B}$ if the A_n are, that every element of $\mathscr{I}_\alpha$ has the form (2.38) for sets A_n in $\mathscr{I}_0$, and that (2.31) holds with α in place of n. Define $\varphi_\alpha(\omega) = \Phi_\alpha(I_{\omega_1}, I_{\omega_2}, \ldots)$, and show that $B_\alpha = [\omega\colon\ \omega \notin \varphi_\alpha(\omega)]$ lies in $\mathscr{B} - \mathscr{I}_\alpha$ for $\alpha < \Omega$. Show that $\mathscr{I}_\alpha$ is strictly smaller than $\mathscr{I}_\beta$ for $\alpha < \beta \le \Omega$.

SECTION 3. EXISTENCE AND EXTENSION

The main theorem to be proved here may be compactly stated this way:

Theorem 3.1. *A probability measure on a field has a unique extension to the generated σ-field.*

In more detail the assertion is this: Suppose that P is a probability measure on a field $\mathscr{F}_0$ of subsets of Ω, and put $\mathscr{F} = \sigma(\mathscr{F}_0)$. Then there

exists a probability measure Q on $\mathscr{F}$ such that $Q(A) = P(A)$ for $A \in \mathscr{F}_0$. Further, if Q' is another probability measure on $\mathscr{F}$ such that $Q'(A) = P(A)$ for $A \in \mathscr{F}_0$, then $Q'(A) = Q(A)$ for $A \in \mathscr{F}$.

Although the measure extended to $\mathscr{F}$ is usually denoted by the same letter as the original measure on $\mathscr{F}_0$, they are really different set functions, since they have different domains of definition. The class $\mathscr{F}_0$ is only assumed finitely additive in the theorem, but the set function P on it must be assumed countably additive (since this of course follows from the conclusion of the theorem).

As shown in Theorem 2.2, λ (initially defined for intervals as length: $\lambda(I) = |I|$) extends to a probability measure on the field $\mathscr{B}_0$ of finite disjoint unions of subintervals of $(0, 1]$. By Theorem 3.1, λ extends in a unique way from $\mathscr{B}_0$ to $\mathscr{B} = \sigma(\mathscr{B}_0)$, the class of Borel sets in $(0, 1]$. The extended λ is *Lebesgue measure* on the unit interval. Theorem 3.1 has many other applications as well.

The uniqueness in Theorem 3.1 will be proved later; see Theorem 3.3. The first project is to prove that an extension does exist.

Construction of the Extension

Let P be a probability measure on a field $\mathscr{F}_0$. The construction following extends P to a class that in general is much larger than $\sigma(\mathscr{F}_0)$ but nonetheless does not in general contain all the subsets of Ω.

For each subset A of Ω, define its *outer measure* by

$$(3.1)\qquad P^*(A) = \inf \sum_n P(A_n),$$

where the infimum extends over all finite and infinite sequences $A_1, A_2, \ldots$ of $\mathscr{F}_0$-sets satisfying $A \subset \bigcup_n A_n$. If the A_n form an efficient covering of A, in the sense that they do not overlap one another very much or extend much beyond A, then $\sum_n P(A_n)$ should be a good outer approximation to the measure of A if A is indeed to have a measure assigned it at all. Thus (3.1) represents a first attempt to assign a measure to A.

Because of the rule $P(A^c) = 1 - P(A)$ for complements (see (2.6)), it is natural in approximating A from the inside to approximate the complement A^c from the outside instead and then subtract from 1:

$$(3.2)\qquad P_*(A) = 1 - P^*(A^c).$$

This, the *inner measure* of A, is a second candidate for the measure of A.† A plausible procedure is to assign measure to those A for which (3.1) and (3.2)

†An idea which seems reasonable at first is to define $P_*(A)$ as the supremum of the sums $\sum_n P(A_n)$ for disjoint sequences of $\mathscr{F}_0$-sets in A. This will not do. For example, in the case where Ω is the unit interval, $\mathscr{F}_0$ is $\mathscr{B}_0$ (Example 2.2), and P is λ as defined by (2.12), the set N of normal numbers would have inner measure 0 because it contains no nonempty elements of $\mathscr{B}_0$; in a satisfactory theory, N will have both inner and outer measure 1.