trifling in the sense of Problem 1.3 if and only if $c^*(A)=0$. Define *inner content* by $c_*(A)=1-c^*(A^c)$. Show that $c_*(A)=\sup\Sigma_n|I_n|$, where the supremum extends over finite disjoint unions of intervals I_n contained in A (of course the analogue for λ_* fails). Show that $c_*(A)\le c^*(A)$; if the two are equal, their common value is taken as the *content* $c(A)$ of A, which is then *Jordan measurable*. Connect all this with Problem 3.6.

Show that $c^*(A)=c^*(A^-)$, where A^- is the closure of A (the analogue for λ^* fails).

A trifling set is Jordan measurable. Find (Problem 3.14) a Jordan measurable set that is not a Borel set.

Show that $c_*(A)\le\lambda_*(A)\le\lambda^*(A)\le c^*(A)$. What happens in this string of inequalities if A consists of the rationals in $(0,\frac{1}{2}]$ together with the irrationals in $(\frac{1}{2},1]$?

3.16. 1.5 ↑ Deduce directly by countable additivity that the Cantor set has Lebesgue measure 0.

3.17. From the fact that $\lambda(x\oplus A)=\lambda(A)$, deduce that sums and differences of normal numbers may be nonnormal.

3.18. Let H be the nonmeasurable set constructed at the end of the section.

(a) Show that, if A is a Borel set and $A\subset H$, then $\lambda(A)=0$—that is, $\lambda_*(H)=0$.

(b) Show that, if $\lambda^*(E)>0$, then E contains a nonmeasurable subset.

3.19. The aim of this problem is the construction of a Borel set A in $(0,1)$ such that $0<\lambda(A\cap G)<\lambda(G)$ for every nonempty open set G in $(0,1)$.

(a) It is shown in Example 3.1 how to construct a Borel set of positive Lebesgue measure that is nowhere dense. Show that every interval contains such a set.

(b) Let $\{I_n\}$ be an enumeration of the open intervals in $(0,1)$ with rational endpoints. Construct disjoint, nowhere dense Borel sets $A_1,B_1,A_2,B_2,\ldots$ of positive Lebesgue measure such that $A_n\cup B_n\subset I_n$.

(c) Let $A=\bigcup_k A_k$. A nonempty open G in $(0,1)$ contains some I_n. Show that $0<\lambda(A_n)\le\lambda(A\cap G)<\lambda(A\cap G)+\lambda(B_n)\le\lambda(G)$.

3.20. ↑ There is no Borel set A in $(0,1)$ such that $a\lambda(I)\le\lambda(A\cap I)\le b\lambda(I)$ for every open interval I in $(0,1)$, where $0<a\le b<1$. In fact prove:

(a) If $\lambda(A\cap I)\le b\lambda(I)$ for all I and if $b<1$, then $\lambda(A)=0$. *Hint:* Choose an open G such that $A\subset G\subset(0,1)$ and $\lambda(G)<b^{-1}\lambda(A)$; represent G as a disjoint union of intervals and obtain a contradiction.

(b) If $a\lambda(I)\le\lambda(A\cap I)$ for all I and if $a>0$, then $\lambda(A)=1$.

3.21. Show that not every subset of the unit interval is a Lebesgue set. *Hint:* Show that λ^* is translation-invariant on $2^{(0,1]}$; then use the first impossibility theorem (p. 45). Or use the second impossibility theorem.

SECTION 4. DENUMERABLE PROBABILITIES

Complex probability ideas can be made clear by the systematic use of measure theory, and probabilistic ideas of extramathematical origin, such as independence, can illuminate problems of purely mathematical interest. It is to this reciprocal exchange that measure-theoretic probability owes much of its interest.

The results of this section concern infinite sequences of events in a probability space.* They will be illustrated by examples in the *unit interval*. By this will always be meant the triple $(\Omega,\mathscr{F},P)$ for which Ω is $(0,1]$, $\mathscr{F}$ is the σ-field $\mathscr{B}$ of Borel sets there, and $P(A)$ is for A in $\mathscr{F}$ the Lebesgue measure $\lambda(A)$ of A. This is the space appropriate to the problems of Section 1, which will be pursued further. The definitions and theorems, as opposed to the examples, apply to *all* probability spaces. The unit interval will appear again and again in this chapter, and it is essential to keep in mind that there are many other important spaces to which the general theory will be applied later.

General Formulas

The formulas (2.5) through (2.11) will be used repeatedly. The sets involved in such formulas lie in the basic σ-field $\mathscr{F}$ by hypothesis. Any probability argument starts from given sets assumed (often tacitly) to lie in $\mathscr{F}$; further sets constructed in the course of the argument must be shown to lie in $\mathscr{F}$ as well, but it is usually quite clear how to do this.

If $P(A)>0$, the *conditional probability* of B given A is defined in the usual way as

$$P(B|A)=\frac{P(A\cap B)}{P(A)}. \tag{4.1}$$

There are the chain-rule formulas

$$\begin{aligned} P(A\cap B)&=P(A)P(B|A),\\ P(A\cap B\cap C)&=P(A)P(B|A)P(C|A\cap B),\\ &\ \,\vdots \end{aligned} \tag{4.2}$$

If $A_1,A_2,\ldots$ partition Ω, then

$$P(B)=\sum_n P(A_n)P(B|A_n). \tag{4.3}$$

*They come under what Borel in his first paper on the subject (see the footnote on p. 9) called *probabilités dénombrables*; hence the section heading.

Note that for fixed A the function $P(B|A)$ defines a probability measure as B varies over $\mathscr{F}$.

If $P(A_n) \equiv 0$, then by subadditivity $P(\bigcup_n A_n) = 0$. If $P(A_n) \equiv 1$, then $\bigcap_n A_n$ has complement $\bigcup_n A_n^c$ of probability 0. This gives two facts used over and over again:

If $A_1, A_2, \dots$ are sets of probability 0, *so is* $\bigcup_n A_n$. *If $A_1, A_2, \dots$ are sets of probability* 1, *so is* $\bigcap_n A_n$.

Limit Sets

For a sequence $A_1, A_2, \dots$ of sets, define a set

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \tag{4.4}$$

and a set

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \tag{4.5}$$

These sets† are the *limits superior* and *inferior* of the sequence $\{A_n\}$. They lie in $\mathscr{F}$ if all the A_n do. Now ω lies in (4.4) if and only if for each n there is some $k \geq n$ for which $\omega \in A_k$; in other words, ω lies in (4.4) if and only if it lies in *infinitely many* of the A_n. In the same way, ω lies in (4.5) if and only if there is some n such that $\omega \in A_k$ for all $k \geq n$; in other words, ω lies in (4.5) if and only if it lies in *all but finitely many* of the A_n.

Note that $\bigcap_{k=n}^{\infty} A_k \uparrow \liminf_n A_n$ and $\bigcup_{k=n}^{\infty} A_k \downarrow \limsup_n A_n$. For every m and n, $\bigcap_{k=m}^{\infty} A_k \subset \bigcup_{k=n}^{\infty} A_k$, because for $i \geq \max\{m, n\}$, A_i contains the first of these sets and is contained in the second. Taking the union over m and the intersection over n shows that (4.5) is a subset of (4.4). But this follows more easily from the observation that if ω lies in all but finitely many of the A_n then of course it lies in infinitely many of them. Facts about limits inferior and superior can usually be deduced from the logic they involve more easily than by formal set-theoretic manipulations.

If (4.4) and (4.5) are equal, write

$$\lim_n A_n = \liminf_n A_n = \limsup_n A_n. \tag{4.6}$$

To say that A_n has limit A, written $A_n \to A$, means that the limits inferior and superior do coincide and in fact coincide with A. Since $\liminf_n A_n \subset \limsup_n A_n$ always holds, to check whether $A_n \to A$ is to check whether $\limsup_n A_n \subset A \subset \liminf_n A_n$. From $A_n \in \mathscr{F}$ and $A_n \to A$ follows $A \in \mathscr{F}$.

†See Problems 4.1 and 4.2 for the analogy between set-theoretic and numerical limits superior and inferior.

Example 4.1. Consider the functions $d_n(\omega)$ defined on the unit interval by the dyadic expansion (1.7), and let $l_n(\omega)$ be the length of the run of 0's starting at $d_n(\omega)$: $l_n(\omega) = k$ if $d_n(\omega) = \cdots = d_{n+k-1}(\omega) = 0$ and $d_{n+k}(\omega) = 1$; here $l_n(\omega) = 0$ if $d_n(\omega) = 1$. Probabilities can be computed by (1.10). Since $[\omega: l_n(\omega) = k]$ is a union of 2^{n-1} disjoint intervals of length 2^{-n-k}, it lies in $\mathscr{F}$ and has probability 2^{-k-1}. Therefore, $[\omega: l_n(\omega) \geq r] = [\omega: d_i(\omega) = 0, n \leq i < n + r]$ lies also in $\mathscr{F}$ and has probability $\sum_{k \geq r} 2^{-k-1}$:

$$P[\omega: l_n(\omega) \geq r] = 2^{-r}. \tag{4.7}$$

If A_n is the event in (4.7), then (4.4) is the set of ω such that $l_n(\omega) \geq r$ for infinitely many n, or, n being regarded as a time index, such that $l_n(\omega) \geq r$ *infinitely often.* ■

Because of the theory of Sections 2 and 3, statements like (4.7) are valid in the sense of ordinary mathematics, and using the traditional language of probability—"heads," "runs," and so on—does not change this.

When n has the role of time, (4.4) is frequently written

$$\limsup_n A_n = [A_n \text{ i.o.}], \tag{4.8}$$

where "i.o." stands for "infinitely often."

Theorem 4.1. (i) *For each sequence* $\{A_n\}$,

$$P\left(\liminf_n A_n\right) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P\left(\limsup_n A_n\right). \tag{4.9}$$

(ii) *If $A_n \to A$, then $P(A_n) \to P(A)$.*

Proof. Clearly (ii) follows from (i). As for (i), if $B_n = \bigcap_{k=n}^{\infty} A_k$ and $C_n = \bigcup_{k=n}^{\infty} A_k$, then $B_n \uparrow \liminf_n A_n$ and $C_n \downarrow \limsup_n A_n$, so that, by parts (i) and (ii) of Theorem 2.1, $P(A_n) \geq P(B_n) \to P(\liminf_n A_n)$ and $P(A_n) \leq P(C_n) \to P(\limsup_n A_n)$. ■

Example 4.2. Define $l_n(\omega)$ as in Example 4.1, and let $A_n = [\omega: l_n(\omega) \geq r]$ for fixed r. By (4.7) and (4.9), $P[\omega: l_n(\omega) \geq r \text{ i.o.}] \geq 2^{-r}$. Much stronger results will be proved later. ■

Independent Events

Events A and B are *independent* if $P(A \cap B) = P(A)P(B)$. (Sometimes an unnecessary *mutually* is put in front of *independent*.) For events of positive probability, this is the same thing as requiring $P(B|A) = P(B)$ or $P(A|B) = P(A)$. More generally, a finite collection $A_1, \ldots, A_n$ of events is independent if

$$P(A_{k_1} \cap \cdots \cap A_{k_j}) = P(A_{k_1}) \cdots P(A_{k_j}) \tag{4.10}$$

for $2 \le j \le n$ and $1 \le k_1 < \cdots < k_j \le n$. Reordering the sets clearly has no effect on the condition for independence, and a subcollection of independent events is also independent. An infinite (perhaps uncountable) collection of events is defined to be independent if each of its finite subcollections is.

If $n = 3$, (4.10) imposes for $j = 2$ the three constraints

$$P(A_1 \cap A_2) = P(A_1)P(A_2), \quad P(A_1 \cap A_3) = P(A_1)P(A_3), \quad P(A_2 \cap A_3) = P(A_2)P(A_3), \tag{4.11}$$

and for $j = 3$ the single constraint

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3). \tag{4.12}$$

Example 4.3. Consider in the unit interval the events $B_{uv} = [\omega: d_u(\omega) = d_v(\omega)]$—the uth and vth tosses agree—and let $A_1 = B_{12}$, $A_2 = B_{13}$, $A_3 = B_{23}$. Then A_1, A_2, A_3 are *pairwise* independent in the sense that (4.11) holds (the two sides of each equation being $\frac{1}{4}$). But since $A_1 \cap A_2 \subset A_3$, (4.12) does *not* hold (the left side is $\frac{1}{4}$ and the right is $\frac{1}{8}$), and the events are not independent. ∎

Example 4.4. In the discrete space $\Omega = \{1, \ldots, 6\}$ suppose each point has probability $\frac{1}{6}$ (a fair die is rolled). If $A_1 = \{1,2,3,4\}$ and $A_2 = A_3 = \{4,5,6\}$, then (4.12) holds but none of the equations in (4.11) do. Again the events are not independent. ∎

Independence requires that (4.10) hold for each $j = 2, \ldots, n$ and each choice of $k_1, \ldots, k_j$, a total of $\sum_{j=2}^{n} \binom{n}{j} = 2^n - 1 - n$ constraints. This requirement can be stated in a different way: If $B_1, \ldots, B_n$ are sets such that for each $i = 1, \ldots, n$ either $B_i = A_i$ or $B_i = \Omega$, then

$$P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(B_1)P(B_2) \cdots P(B_n). \tag{4.13}$$

The point is that if $B_i = \Omega$, then B_i can be ignored in the intersection on the left and the factor $P(B_i) = 1$ can be ignored in the product on the right. For example, replacing A_2 by Ω reduces (4.12) to the middle equation in (4.11).

From the assumed independence of certain sets it is possible to deduce the independence of other sets.

Example 4.5. On the unit interval the events $H_n = [\omega: d_n(\omega) = 0]$, $n = 1, 2, \ldots$, are independent, the two sides of (4.10) having in this case value 2^{-j}. It seems intuitively clear that from this should follow the independence, for example, of $[\omega: d_2(\omega) = 0] = H_2$ and $[\omega: d_1(\omega) = 0,\ d_3(\omega) = 1] = H_1 \cap H_3^c$, since the two events involve disjoint sets of times. Further, any sets A and B depending, respectively, say, only on even and on odd times (like $[\omega: d_{2n}(\omega) = 0 \text{ i.o.}]$ and $[\omega: d_{2n+1}(\omega) = 1 \text{ i.o.}]$) ought also to be independent. This raises the general question of what it means for A to depend only on even times. Intuitively, it requires that knowing which ones among $H_2, H_4, \ldots$ occurred entails knowing whether or not A occurred—that is, it requires that the sets $H_2, H_4, \ldots$ "determine" A. The set-theoretic form of this requirement is that A is to lie in the σ-field generated by $H_2, H_4, \ldots$. From $A \in \sigma(H_2, H_4, \ldots)$ and $B \in \sigma(H_1, H_3, \ldots)$ it ought to be possible to deduce the independence of A and B. ∎

The next theorem and its corollaries make such deductions possible. Define *classes* $\mathscr{A}_1, \ldots, \mathscr{A}_n$ in the basic σ-field $\mathscr{F}$ to be independent if for each choice of A_i from $\mathscr{A}_i$, $i = 1, \ldots, n$, the events $A_1, \ldots, A_n$ are independent. This is the same as requiring that (4.13) hold whenever for each i, $1 \le i \le n$, either $B_i \in \mathscr{A}_i$ or $B_i = \Omega$.

Theorem 4.2. *If $\mathscr{A}_1, \ldots, \mathscr{A}_n$ are independent and each $\mathscr{A}_i$ is a π-system, then $\sigma(\mathscr{A}_1), \ldots, \sigma(\mathscr{A}_n)$ are independent.*

PROOF. Let $\mathscr{B}_i$ be the class $\mathscr{A}_i$ augmented by Ω (which may be an element of $\mathscr{A}_i$ to start with). Then each $\mathscr{B}_i$ is a π-system, and by the hypothesis of independence, (4.13) holds if $B_i \in \mathscr{B}_i$, $i = 1, \ldots, n$. For fixed sets $B_2, \ldots, B_n$ lying respectively in $\mathscr{B}_2, \ldots, \mathscr{B}_n$, let $\mathscr{L}$ be the class of $\mathscr{F}$-sets B_1 for which (4.13) holds. Then $\mathscr{L}$ is a λ-system containing the π-system $\mathscr{B}_1$ and hence (Theorem 3.2) containing $\sigma(\mathscr{B}_1) = \sigma(\mathscr{A}_1)$. Therefore, (4.13) holds if $B_1, B_2, \ldots, B_n$ lie respectively in $\sigma(\mathscr{A}_1), \mathscr{B}_2, \ldots, \mathscr{B}_n$, which means that $\sigma(\mathscr{A}_1), \mathscr{A}_2, \ldots, \mathscr{A}_n$ are independent. Clearly the argument goes through if 1 is replaced by any of the indices $2, \ldots, n$.

From the independence of $\sigma(\mathscr{A}_1), \mathscr{A}_2, \ldots, \mathscr{A}_n$ now follows that of $\sigma(\mathscr{A}_1), \sigma(\mathscr{A}_2), \mathscr{A}_3, \ldots, \mathscr{A}_n$, and so on. ∎

If $\mathscr{A} = \{A_1, \ldots, A_k\}$ is finite, then each A in $\sigma(\mathscr{A})$ can be expressed by a "formula" such as $A = A_2 \cap A_5^c$ or $A = (A_2 \cap A_7) \cup (A_3 \cap A_7^c \cap A_8)$. If $\mathscr{A}$ is infinite, the sets in $\sigma(\mathscr{A})$ may be very complicated; the way to make precise the idea that the elements of $\mathscr{A}$ "determine" A is not to require formulas, but simply to require that A lie in $\sigma(\mathscr{A})$.

Independence for an infinite collection of classes is defined just as in the finite case: $[\mathscr{A}_\theta: \theta \in \Theta]$ is independent if the collection $[A_\theta: \theta \in \Theta]$ of sets is independent for each choice of A_θ from $\mathscr{A}_\theta$. This is equivalent to the independence of each finite subcollection $\mathscr{A}_{\theta_1}, \ldots, \mathscr{A}_{\theta_n}$ of classes, because of the way independence for infinite classes of sets is defined in terms of independence for finite classes. Hence Theorem 4.2 has an immediate consequence:

Corollary 1. *If $\mathscr{A}_\theta$, $\theta \in \Theta$, are independent and each $\mathscr{A}_\theta$ is a π-system, then $\sigma(\mathscr{A}_\theta)$, $\theta \in \Theta$, are independent.*

Corollary 2. *Suppose that the array*

$$\begin{matrix} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \end{matrix} \tag{4.14}$$

of events is independent; here each row is a finite or infinite sequence, and there are finitely or infinitely many rows. If $\mathscr{F}_i$ is the σ-field generated by the ith row, then $\mathscr{F}_1, \mathscr{F}_2, \ldots$ are independent.

PROOF. If $\mathscr{A}_i$ is the class of all finite intersections of elements of the ith row of (4.14), then $\mathscr{A}_i$ is a π-system and $\sigma(\mathscr{A}_i) = \mathscr{F}_i$. Let I be a finite collection of indices (integers), and for each i in I let J_i be a finite collection of indices. Consider for $i \in I$ the element $C_i = \bigcap_{j \in J_i} A_{ij}$ of $\mathscr{A}_i$. Since every finite subcollection of the array (4.14) is independent (the intersections and products here extend over $i \in I$ and $j \in J_i$),

$$P\Big(\bigcap_i C_i\Big) = P\Big(\bigcap_i \bigcap_j A_{ij}\Big) = \prod_i \prod_j P(A_{ij}) = \prod_i P(\cap_j A_{ij})$$
$$= \prod_i P(C_i).$$

It follows that the classes $\mathscr{A}_1, \mathscr{A}_2, \ldots$ are independent, so that Corollary 1 applies. ∎

Corollary 2 implies the independence of the events discussed in Example 4.5. The array (4.14) in this case has two rows:

$$\begin{matrix} H_2 & H_4 & H_6 & \cdots \\ H_1 & H_3 & H_5 & \cdots \end{matrix}$$

Theorem 4.2 also implies, for example, that for independent $A_1, \ldots, A_n$,

$$P(A_1^c \cap \cdots \cap A_k^c \cap A_{k+1} \cap \cdots \cap A_n) = P(A_1^c) \cdots P(A_k^c) P(A_{k+1}) \cdots P(A_n). \tag{4.15}$$

To prove this, let $\mathscr{A}_i$ consist of A_i alone; of course, $A_i^c \in \sigma(\mathscr{A}_i)$. In (4.15) any subcollection of the A_i could be replaced by their complements.

Example 4.6. Consider as in Example 4.3 the events B_{uv} that, in a sequence of tosses of a fair coin, the uth and vth outcomes agree. Let $\mathscr{A}_1$ consist of the events B_{12} and B_{13}, and let $\mathscr{A}_2$ consist of the event B_{23}. Since these three events are pairwise independent, the classes $\mathscr{A}_1$ and $\mathscr{A}_2$ are independent. Since $B_{23} = (B_{12} \triangle B_{13})^c$ lies in $\sigma(\mathscr{A}_1)$, however, $\sigma(\mathscr{A}_1)$ and $\sigma(\mathscr{A}_2)$ are not independent. The trouble is that $\mathscr{A}_1$ is not a π-system, which shows that this condition in Theorem 4.2 is essential. ∎

Example 4.7. If $\mathscr{A} = \{A_1, A_2, \ldots\}$ is a finite or countable partition of Ω and $P(B|A_i) = p$ for each A_i of positive probability, then $P(B) = p$ and B is independent of $\sigma(\mathscr{A})$: If Σ' denotes summation over those i for which $P(A_i) > 0$, then $P(B) = \Sigma' P(A_i \cap B) = \Sigma' P(A_i)p = p$, and so B is independent of each set in the π-system $\mathscr{A} \cup \{\emptyset\}$. ∎

Subfields

Theorem 4.2 involves a number of σ-fields at once, which is characteristic of probability theory; measure theory not directed toward probability usually involves only one all-embracing σ-field $\mathscr{F}$. In proability, σ-fields in $\mathscr{F}$—that is, sub-σ-fields of $\mathscr{F}$—play an important role. To understand their function it helps to have an informal, intuitive way of looking at them.

A subclass $\mathscr{A}$ of $\mathscr{F}$ corresponds heuristically to *partial information*. Imagine that a point ω is drawn from Ω according to the probabilities given by P: ω lies in A with probability $P(A)$. Imagine also an observer who does not know which ω it is that has been drawn but who does know for each A in $\mathscr{A}$ whether $\omega \in A$ or $\omega \notin A$—that is, who does not know ω but does know the value of $I_A(\omega)$ for each A in $\mathscr{A}$. Identifying this partial information with the class $\mathscr{A}$ itself will illuminate the connection between various measure-theoretic concepts and the premathematical ideas lying behind them.

The set B is by definition independent of the class $\mathscr{A}$ if $P(B|A) = P(B)$ for all sets A in $\mathscr{A}$ for which $P(A) > 0$. Thus if B is independent of $\mathscr{A}$, then the observer's probability for B is $P(B)$ even after he has received the information in $\mathscr{A}$; in this case $\mathscr{A}$ contains no information about B. The point of Theorem 4.2 is that this remains true even if the observer is given the information in $\sigma(\mathscr{A})$, provided that $\mathscr{A}$ is a π-system. It is to be stressed that here *information*, like *observer* and *know*, is an informal, extramathematical term (in particular, it is not information in the technical sense of entropy).

The notion of partial information can be looked at in terms of partitions. Say that points ω and ω' are *$\mathscr{A}$-equivalent* if, for every A in $\mathscr{A}$, ω and ω' lie either both in A or both in A^c—that is, if

$$I_A(\omega) = I_A(\omega'), \qquad A \in \mathscr{A}. \tag{4.16}$$

This relation partitions Ω into sets of equivalent points; call this the $\mathscr{A}$-*partition*.

Example 4.8. If ω and ω' are $\sigma(\mathscr{A})$-equivalent, then certainly they are $\mathscr{A}$-equivalent. For fixed ω and ω', the class of A such that $I_A(\omega) = I_A(\omega')$ is a σ-field; if ω and ω' are $\mathscr{A}$-equivalent, then this σ-field contains $\mathscr{A}$ and hence $\sigma(\mathscr{A})$, so that ω and ω' are also $\sigma(\mathscr{A})$-equivalent. Thus $\mathscr{A}$-equivalence and $\sigma(\mathscr{A})$-equivalence are the same thing, and the $\mathscr{A}$-partition coincides with the $\sigma(\mathscr{A})$-partition. ■

An observer with the information in $\sigma(\mathscr{A})$ knows, not the point ω drawn, but only the equivalence class containing it. That is exactly the information he has. In Example 4.6, it is as though an observer with the items of information in $\mathscr{A}_1$ were unable to combine them to get information about B_{23}.

Example 4.9. If $H_n = [\omega: d_n(\omega) = 0]$ as in Example 4.5, and if $\mathscr{A} = \{H_1, H_3, H_5, \dots\}$, then ω and ω' satisfy (4.16) if and only if $d_n(\omega) = d_n(\omega')$ for all odd n. The information in $\sigma(\mathscr{A})$ is thus the set of values of $d_n(\omega)$ for n odd. ■

One who knows that ω lies in a set A has more information about ω the smaller A is. One who knows $I_A(\omega)$ for each A in a class $\mathscr{A}$, however, has more information about ω the larger $\mathscr{A}$ is. Furthermore, to have the information in $\mathscr{A}_1$ *and* the information in $\mathscr{A}_2$ is to have the information in $\mathscr{A}_1 \cup \mathscr{A}_2$, not that in $\mathscr{A}_1 \cap \mathscr{A}_2$.

The following example points up the informal nature of this interpretation of σ-fields as information.

Example 4.10. In the unit interval $(\Omega, \mathscr{F}, P)$ let $\mathscr{G}$ be the σ-field consisting of the countable and the cocountable sets. Since $P(G)$ is 0 or 1 for each G in $\mathscr{G}$, each set H in $\mathscr{F}$ is independent of $\mathscr{G}$. But in this case the $\mathscr{G}$-partition consists of the singletons, and so the information in $\mathscr{G}$ tells the observer exactly which ω in Ω has been drawn. (i) The σ-field $\mathscr{G}$ contains *no* information about H—in the sense that H and $\mathscr{G}$ are independent. (ii) The σ-field $\mathscr{G}$ contains *all* the information about H—in the sense that it tells the observer exactly which ω was drawn. ■

In this example, (i) and (ii) stand in apparent contradiction. But the mathematics is in (i)—H and $\mathscr{G}$ *are* independent—while (ii) only concerns heuristic interpretation. The source of the difficulty or apparent paradox here lies in the unnatural structure of the σ-field $\mathscr{G}$ rather than in any deficiency in the notion of independence.[†] The heuristic equating of σ-fields and information is helpful even though it sometimes breaks down, and of course proofs are indifferent to whatever illusions and vagaries brought them into existence.

[†]See Problem 4.10 for a more extreme example.

The Borel–Cantelli Lemmas

This is *the first Borel–Cantelli lemma*:

Theorem 4.3. *If $\sum_n P(A_n)$ converges, then $P(\limsup_n A_n) = 0$.*

PROOF. From $\limsup_n A_n \subset \bigcup_{k=m}^{\infty} A_k$ follows $P(\limsup_n A_n) \le P(\bigcup_{k=m}^{\infty} A_k) \le \sum_{k=m}^{\infty} P(A_k)$, and this sum tends to 0 as $m \to \infty$ if $\sum_n P(A_n)$ converges. ■

By Theorem 4.1, $P(A_n) \to 0$ implies that $P(\liminf_n A_n) = 0$; in Theorem 4.3 hypothesis and conclusion are both stronger.

Example 4.11. Consider the run length $l_n(\omega)$ of Example 4.1 and a sequence $\{r_n\}$ of positive reals. *If the series $\sum 1/2^{r_n}$ converges, then*

$$P[\omega: l_n(\omega) \ge r_n \text{ i.o.}] = 0. \tag{4.17}$$

To prove this, note that if s_n is r_n rounded up to the next integer, then by (4.7), $P[\omega: l_n(\omega) \ge r_n] = 2^{-s_n} \le 2^{-r_n}$. Therefore, (4.17) follows by the first Borel–Cantelli lemma.

If $r_n = (1+\epsilon)\log_2 n$ and ϵ is positive, there is convergence because $2^{-r_n} = 1/n^{1+\epsilon}$. Thus

$$P[\omega: l_n(\omega) \ge (1+\epsilon)\log_2 n \text{ i.o.}] = 0. \tag{4.18}$$

The limit superior of the ratio $l_n(\omega)/\log_2 n$ exceeds 1 if and only if ω belongs to the set in (4.18) for some positive rational ϵ. Since the union of this countable class of sets has probability 0,

$$P\left[\omega: \limsup_n \frac{l_n(\omega)}{\log_2 n} > 1\right] = 0. \tag{4.19}$$

To put it the other way around,

$$P\left[\omega: \limsup_n \frac{l_n(\omega)}{\log_2 n} \le 1\right] = 1. \tag{4.20}$$

Technically, the probability in (4.20) refers to Lebesgue measure. Intuitively, it refers to an infinite sequence of independent tosses of a fair coin. ■

In this example, whether $\limsup_n l_n(\omega)/\log_2 n \le 1$ holds or not is a property of ω, and the property in fact holds for ω in an $\mathscr{F}$-set of probability

1. In such a case the property is said to hold *with probability 1*, or *almost surely*. In nonprobabilistic contexts, a property that holds for ω outside a (measurable) set of measure 0 holds *almost everywhere*, or for *almost all* ω.

Example 4.12. The preceding example has an interesting arithmetic consequence. Truncating the dyadic expansion at n gives the standard $(n-1)$-place approximation $\sum_{k=1}^{n-1} d_k(\omega)2^{-k}$ to ω; the error is between 0 and 2^{-n+1}, and the error relative to the maximum is

$$e_n(\omega) = \frac{\omega - \sum_{k=1}^{n-1} d_k(\omega)2^{-k}}{2^{-n+1}} = \sum_{i=1}^{\infty} d_{n+i-1}(\omega)2^{-i}, \tag{4.21}$$

which lies between 0 and 1. The binary expansion of $e_n(\omega)$ begins with $l_n(\omega)$ 0's, and then comes a 1. Hence $.0\ldots01 \le e_n(\omega) \le .0\ldots0111\ldots$, where there are $l_n(\omega)$ 0's in the extreme terms. Therefore,

$$\frac{1}{2^{l_n(\omega)+1}} \le e_n(\omega) \le \frac{1}{2^{l_n(\omega)}}, \tag{4.22}$$

so that results on run length give information about the error of approximation.

By the left-hand inequality in (4.22), $e_n(\omega) \le x_n$ (assume that $0 < x_n \le 1$) implies that $l_n(\omega) \ge -\log_2 x_n - 1$; since $\sum 2^{-r_n} < \infty$ implies (4.17), $\sum x_n < \infty$ implies $P[\omega: e_n(\omega) \le x_n \text{ i.o.}] = 0$. (Clearly, $[\omega: e_n(\omega) \le x]$ is a Borel set.) In particular,

$$P[\omega: e_n(\omega) \le 1/n^{1+\epsilon} \text{ i.o.}] = 0. \tag{4.23}$$

Technically, this probability refers to Lebesgue measure; intuitively, it refers to a point drawn at random from the unit interval. ■

Example 4.13. The final step in the proof of the normal number theorem (Theorem 1.2) was a disguised application of the first Borel–Cantelli lemma. If $A_n = [\omega: |n^{-1}s_n(\omega)| \ge n^{-1/8}]$, then $\sum P(A_n) < \infty$, as follows by (1.29), and so $P[A_n \text{ i.o.}] = 0$. But for ω in the set complementary to $[A_n \text{ i.o.}]$, $n^{-1}s_n(\omega) \to 0$.

The proof of Theorem 1.6 is also, in effect, an application of the first Borel–Cantelli lemma. ■

This is *the second Borel–Cantelli lemma*:

Theorem 4.4. *If $\{A_n\}$ is an independent sequence of events and $\sum_n P(A_n)$ diverges, then* $P(\limsup_n A_n) = 1$.

PROOF. It is enough to prove that $P(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c) = 0$ and hence enough to prove that $P(\bigcap_{k=n}^{\infty} A_k^c) = 0$ for all n. Since $1 - x \le e^{-x}$,

$$P\left(\bigcap_{k=n}^{n+j} A_k^c\right) = \prod_{k=n}^{n+j}(1 - P(A_k)) \le \exp\left[-\sum_{k=n}^{n+j} P(A_k)\right].$$

Since $\sum_k P(A_k)$ diverges, the last expression tends to 0 as $j \to \infty$, and hence $P(\bigcap_{k=n}^{\infty} A_k^c) = \lim_j P(\bigcap_{k=n}^{n+j} A_k^c) = 0$. ■

By Theorem 4.1, $\limsup_n P(A_n) > 0$ implies $P(\limsup_n A_n) > 0$; in Theorem 4.4, the hypothesis $\sum_n P(A_n) = \infty$ is weaker but the conclusion is stronger because of the additional hypothesis of independence.

Example 4.14. Since the events $[\omega: l_n(\omega) = 0] = [\omega: d_n(\omega) = 1]$, $n = 1, 2, \ldots$, are independent and have probability $\frac{1}{2}$, $P[\omega: l_n(\omega) = 0 \text{ i.o.}] = 1$.

Since the events $A_n = [\omega: l_n(\omega) = 1] = [\omega: d_n(\omega) = 0,\ d_{n+1}(\omega) = 1]$, $n = 1, 2, \ldots$, are not independent, this argument is insufficient to prove that

$$P[\omega: l_n(\omega) = 1 \text{ i.o.}] = 1. \tag{4.24}$$

But the events $A_2, A_4, A_6, \ldots$ are independent (Theorem 4.2) and their probabilities form a divergent series, and so $P[\omega: l_{2n}(\omega) = 1 \text{ i.o.}] = 1$, which implies (4.24). ■

Significant applications of the second Borel–Cantelli lemma usually require, in order to get around problems of dependence, some device of the kind used in the preceding example.

Example 4.15. There is a complement to (4.17): *If r_n is nondecreasing and $\sum 2^{-r_n}/r_n$ diverges, then*

$$P[\omega: l_n(\omega) \ge r_n \ i.o.] = 1. \tag{4.25}$$

To prove this, note first that if r_n is rounded up to the next integer, then $\sum 2^{-r_n}/r_n$ still diverges and (4.25) is unchanged. Assume then that $r_n = r(n)$ is integral, and define $\{n_k\}$ inductively by $n_1 = 1$ and $n_{k+1} = n_k + r_{n_k}$, $k \ge 1$. Let $A_k = [\omega: l_{n_k}(\omega) \ge r_{n_k}] = [\omega: d_i(\omega) = 0,\ n_k \le i < n_{k+1}]$; since the A_k involve nonoverlapping sequences of time indices, it follows by Corollary 2 to Theorem 4.2 that $A_1, A_2, \ldots$ are independent. By the second Borel–Cantelli lemma, $P[A_k \text{ i.o.}] = 1$ if $\sum_k P(A_k) = \sum_k 2^{-r(n_k)}$ diverges. But since r_n is nondecreasing,

$$\sum_{k\ge1} 2^{-r(n_k)} = \sum_{k\ge1} 2^{-r(n_k)}r^{-1}(n_k)(n_{k+1} - n_k)$$
$$\ge \sum_{k\ge1}\sum_{n_k \le n < n_{k+1}} 2^{-r_n}r_n^{-1} = \sum_{n\ge1} 2^{-r_n}r_n^{-1}.$$

Thus the divergence of $\sum_n 2^{-r_n}r_n^{-1}$ implies that of $\sum_k 2^{-r(n_k)}$, and it follows that, with probability 1, $l_{n_k}(\omega) \ge r_{n_k}$ for infinitely many values of k. But this is stronger than (4.25).

The result in Example 4.2 follows if $r_n \equiv r$, but this is trivial. If $r_n = \log_2 n$, then $\sum 2^{-r_n}/r_n = \sum 1/(n \log_2 n)$ diverges, and therefore

$$P[\omega: l_n(\omega) \ge \log_2 n \text{ i.o.}] = 1. \tag{4.26}$$

By (4.26) and (4.20),

$$P\left[\omega: \limsup_n \frac{l_n(\omega)}{\log_2 n} = 1\right] = 1. \tag{4.27}$$

Thus for ω in a set of probability 1, $\log_2 n$ as a function of n is a kind of "upper envelope" for the function $l_n(\omega)$. ■

Example 4.16. By the right-hand inequality in (4.22), if $l_n(\omega) \geq \log_2 n$, then $e_n(\omega) \leq 1/n$. Hence (4.26) gives

$$P\left[\omega: e_n(\omega) \leq \frac{1}{n} \text{ i.o.}\right] = 1. \tag{4.28}$$

This and (4.23) show that, with probability 1, $e_n(\omega)$ has $1/n$ as a "lower envelope." The discrepancy between ω and its $(n-1)$-place approximation $\sum_{k=1}^{n-1} d_k(\omega)2^{-k}$ will fall infinitely often below $1/(n\,2^{n-1})$ but not infinitely often below $1/(n^{1+\epsilon}2^{n-1})$. ■

Example 4.17. Examples 4.12 and 4.16 have to do with the approximation of real numbers by rationals: Diophantine approximation. Change the $x_n = 1/n^{1+\epsilon}$ of (4.23) to $1/((n-1)\log 2)^{1+\epsilon}$. Then $\sum x_n$ still converges, and hence

$$P\left[\omega: e_n(\omega) \leq 1/\left(\log 2^{n-1}\right)^{1+\epsilon} \text{ i.o.}\right] = 0.$$

And by (4.28),

$$P\left[\omega: e_n(\omega) < 1/\log 2^{n-1} \text{ i.o.}\right] = 1.$$

The dyadic rational $\sum_{k=1}^{n-1} d_k(\omega)2^{-k} = p/q$ has denominator $q = 2^{n-1}$, and $e_n(\omega) = q(\omega - p/q)$. There is therefore probability 1 that, if q is restricted to the powers of 2, then $0 \leq \omega - p/q < 1/(q \log q)$ holds for infinitely many p/q but $0 \leq \omega - p/q \leq 1/(q \log^{1+\epsilon} q)$ holds only for finitely many.[†] But contrast this with Theorems 1.5 and 1.6: The sharp rational approximations to a real number come not from truncating its dyadic (or decimal) expansion, but from truncating its continued-fraction expansion; see Section 24. ■

The Zero–One Law

For a sequence $A_1, A_2, \ldots$ of events in a probability space $(\Omega, \mathscr{F}, P)$ consider the σ-fields $\sigma(A_n, A_{n+1}, \ldots)$ and their intersection

$$\mathscr{T} = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, \ldots). \tag{4.29}$$

[†]This ignores the possibility of even p (reducible p/q); but see Problem 1.11(b). And rounding ω up to $(p+1)/q$ instead of down to p/q changes nothing; see Problem 4.13.

This is the *tail σ-field* associated with the sequence $\{A_n\}$, and its elements are called *tail events*. The idea is that a tail event is determined solely by the A_n for arbitrarily large n.

Example 4.18. Since $\limsup_m A_m = \bigcap_{k \geq n} \bigcup_{i \geq k} A_i$ and $\liminf_m A_m = \bigcup_{k \geq n} \bigcap_{i \geq k} A_i$ are both in $\sigma(A_n, A_{n+1}, \ldots)$, the limits superior and inferior are tail events for the sequence $\{A_n\}$. ■

Example 4.19. Let $l_n(\omega)$ be the run length, as before, and let $H_n = [\omega: d_n(\omega) = 0]$. For each n_0,

$$\begin{aligned}[\omega: l_n(\omega) \geq r_n \text{ i.o.}] &= \bigcap_{n \geq n_0} \bigcup_{k \geq n} [\omega: l_k(\omega) \geq r_k] \\ &= \bigcap_{n \geq n_0} \bigcup_{k \geq n} H_k \cap H_{k+1} \cap \cdots \cap H_{k+r_k-1}.\end{aligned}$$

Thus $[\omega: l_n(\omega) \geq r_n \text{ i.o.}]$ is a tail event for the sequence $\{H_n\}$. ■

The probabilities of tail events are governed by *Kolmogorov's zero–one law:*[†]

Theorem 4.5. *If $A_1, A_2, \ldots$ is an independent sequence of events, then for each event A in the tail σ-field* (4.29), *$P(A)$ is either* 0 *or* 1.

PROOF. By Corollary 2 to Theorem 4.2, $\sigma(A_1), \ldots, \sigma(A_{n-1}), \sigma(A_n, A_{n+1}, \ldots)$ are independent. If $A \in \mathscr{T}$, then $A \in \sigma(A_n, A_{n+1}, \ldots)$ and therefore $A_1, \ldots, A_{n-1}, A$ are independent. Since independence of a collection of events is defined by independence of each finite subcollection, the sequence $A, A_1, A_2, \ldots$ is independent. By a second application of Corollary 2 to Theorem 4.2, $\sigma(A)$ and $\sigma(A_1, A_2, \ldots)$ are independent. But $A \in \mathscr{T} \subset \sigma(A_1, A_2, \ldots)$; from $A \in \sigma(A)$ and $A \in \sigma(A_1, A_2, \ldots)$ it follows that A is independent of itself: $P(A \cap A) = P(A)P(A)$. This is the same as $P(A) = (P(A))^2$ and can hold only if $P(A)$ is 0 or 1. ■

Example 4.20. By the zero–one law and Example 4.18, $P(\limsup_n A_n)$ is 0 or 1 if the A_n are independent. The Borel–Cantelli lemmas in this case go further and give a specific criterion in terms of the convergence or divergence of $\sum P(A_n)$. ■

Kolmogorov's result is surprisingly general, and it is in many cases quite easy to use it to show that the probability of some set must have one of the extreme values 0 and 1. It is perhaps curious that it should so often be very difficult to determine which of these extreme values is the right one.

[†]For a more general version, see Theorem 22.3

Example 4.21. By Kolmogorov's theorem and Example 4.19, $[\omega: l_n(\omega) \geq r_n \text{ i.o.}]$ has probability 0 or 1. Call the sequence $\{r_n\}$ an *outer boundary* or an *inner boundary* according as this probability is 0 or 1.

In Example 4.11 it was shown that $\{r_n\}$ is an outer boundary if $\Sigma 2^{-r_n} < \infty$. In Example 4.15 it was shown that $\{r_n\}$ is an inner boundary if r_n is nondecreasing and $\Sigma 2^{-r_n} r_n^{-1} = \infty$. By these criteria $r_n = \theta \log_2 n$ gives an outer boundary if $\theta > 1$ and an inner boundary if $\theta \leq 1$.

What about the sequence $r_n = \log_2 n + \theta \log_2 \log_2 n$? Here $\Sigma 2^{-r_n} = \Sigma 1/n(\log_2 n)^\theta$, and this converges for $\theta > 1$, which gives an outer boundary. Now $2^{-r_n} r_n^{-1}$ is of the order $1/n(\log_2 n)^{1+\theta}$, and this diverges if $\theta \leq 0$, which gives an inner boundary (this follows indeed from (4.26)). But this analysis leaves the range $0 < \theta \leq 1$ unresolved, although every sequence is either an inner or an outer boundary. This question is pursued further in Example 6.5. ■

PROBLEMS

4.1. 2.1 ↑ The limits superior and inferior of a numerical sequence $\{x_n\}$ can be defined as the supremum and infimum of the set of limit points—that is, the set of limits of convergent subsequences. This is the same thing as defining

$$\limsup_n x_n = \bigwedge_{n=1}^{\infty} \bigvee_{k=n}^{\infty} x_k \tag{4.30}$$

and

$$\liminf_n x_n = \bigvee_{n=1}^{\infty} \bigwedge_{k=n}^{\infty} x_k. \tag{4.31}$$

Compare these relations with (4.4) and (4.5) and prove that

$$I_{\limsup_n A_n} = \limsup_n I_{A_n}, \qquad I_{\liminf_n A_n} = \liminf_n I_{A_n}.$$

Prove that $\lim_n A_n$ exists in the sense of (4.6) if and only if $\lim_n I_{A_n}(\omega)$ exists for each ω.

4.2. ↑ (a) Prove that

$$\left(\limsup_n A_n\right) \cap \left(\limsup_n B_n\right) \supset \limsup_n (A_n \cap B_n),$$

$$\left(\limsup_n A_n\right) \cup \left(\limsup_n B_n\right) = \limsup_n (A_n \cup B_n),$$

$$\left(\liminf_n A_n\right) \cap \left(\liminf_n B_n\right) = \liminf_n (A_n \cap B_n),$$

$$\left(\liminf_n A_n\right) \cup \left(\liminf_n B_n\right) \subset \liminf_n (A_n \cup B_n).$$

Show by example that the two inclusions can be strict.

(b) The numerical analogue of the first of the relations in part (a) is

$$\left(\limsup_n x_n\right) \wedge \left(\limsup_n y_n\right) \geq \limsup_n (x_n \wedge y_n).$$

Write out and verify the numerical analogues of the others.

(c) Show that

$$\limsup_n A_n^c = \left(\liminf_n A_n\right)^c,$$

$$\liminf_n A_n^c = \left(\limsup_n A_n\right)^c,$$

$$\limsup_n A_n - \liminf_n A_n = \limsup_n (A_n \cap A_{n+1}^c)$$

$$= \limsup_n (A_n^c \cap A_{n+1}).$$

(d) Show that $A_n \to A$ and $B_n \to B$ together imply that $A_n \cup B_n \to A \cup B$ and $A_n \cap B_n \to A \cap B$.

4.3. Let A_n be the square $[(x, y): |x| \leq 1, |y| \leq 1]$ rotated through the angle $2\pi n\theta$. Give geometric descriptions of $\limsup_n A_n$ and $\liminf A_n$ in case

(a) $\theta = \frac{1}{8}$;

(b) θ is rational;

(c) θ is irrational. *Hint*: The $2\pi n\theta$ reduced modulo 2π are dense in $[0, 2\pi]$ if θ is irrational.

(d) When is there convergence is the sense of (4.6)?

4.4. Find a sequence for which all three inequalities in (4.9) are strict.

4.5. (a) Show that $\lim_n P(\liminf_k A_n \cap A_k^c) = 0$. *Hint*: Show that $\limsup_n \liminf_k A_n \cap A_k^c$ is empty.

Put $A^* = \limsup_n A_n$ and $A_* = \liminf_n A_n$.

(b) Show that $P(A_n - A^*) \to 0$ and $P(A_* - A_n) \to 0$.

(c) Show that $A_n \to A$ (in the sense that $A = A^* = A_*$) implies $P(A \Delta A_n) \to 0$.

(d) Suppose that A_n converges to A in the weaker sense that $P(A \Delta A^*) = P(A \Delta A_*) = 0$ (which implies that $P(A^* - A_*) = 0$). Show that $P(A \Delta A_n) \to 0$ (which implies that $P(A_n) \to P(A)$).

4.6. In a space of six equally likely points (a die is rolled) find three events that are not independent even though each is independent of the intersection of the other two.

4.7. For events $A_1, \ldots, A_n$, consider the 2^n equations $P(B_1 \cap \cdots \cap B_n) = P(B_1) \cdots P(B_n)$ with $B_i = A_i$ or $B_i = A_i^c$ for each i. Show that $A_1, \ldots, A_n$ are independent if all these equations hold.

4.8. For each of the following classes $\mathscr{A}$, describe the $\mathscr{A}$-partition defined by (4.16).

(a) The class of finite and cofinite sets.

(b) The class of countable and cocountable sets.

(c) A partition (of arbitrary cardinality) of Ω.

(d) The level sets of $\sin x$ ($\Omega = R^1$).

(e) The σ-field in Problem 3.5.

4.9. 2.9 2.10↑ In connection with Example 4.8 and Problem 2.10, prove these facts:

(a) Every set in $\sigma(\mathscr{A})$ is a union of $\mathscr{A}$-equivalence classes.

(b) If $\mathscr{A} = [A_\theta: \theta \in \Theta]$, then the $\mathscr{A}$-equivalence classes have the form $\bigcap_\theta B_\theta$, where for each θ, B_θ is A_θ or A_θ^c.

(c) Every finite σ-field is generated by a finite partition of Ω.

(d) If $\mathscr{F}_0$ is a field, then each singleton, even each finite set, in $\sigma(\mathscr{F}_0)$ is a countable intersection of $\mathscr{F}_0$-sets.

4.10. 3.2↑ There is in the unit interval a set H that is nonmeasurable in the extreme sense that its inner and outer Lebesgue measures are 0 and 1 (see (3.9) and (3.10)): $\lambda_*(H) = 0$ and $\lambda^*(H) = 1$. See Problem 12.4 for the construction.

Let $\Omega = (0,1]$, let $\mathscr{G}$ consist of the Borel sets in Ω, and let H be the set just described. Show that the class $\mathscr{F}$ of sets of the form $(H \cap G_1) \cup (H^c \cap G_2)$ for G_1 and G_2 in $\mathscr{G}$ is a σ-field and that $P[(H \cap G_1) \cup (H^c \cap G_2)] = \frac{1}{2}\lambda(G_1) + \frac{1}{2}\lambda(G_2)$ consistently defines a probability measure on $\mathscr{F}$. Show that $P(H) = \frac{1}{2}$ and that $P(G) = \lambda(G)$ for $G \in \mathscr{G}$. Show that $\mathscr{G}$ is generated by a countable subclass (see Problem 2.11). Show that $\mathscr{G}$ contains all the singletons and that H and $\mathscr{G}$ are independent.

The construction proves this: *There exist a probability space $(\Omega, \mathscr{F}, P)$, a σ-field $\mathscr{G}$ in $\mathscr{F}$, and a set H in $\mathscr{F}$, such that $P(H) = \frac{1}{2}$, H and $\mathscr{G}$ are independent, and $\mathscr{G}$ is generated by a countable subclass and contains all the singletons.*

Example 4.10 is somewhat similar, but there the σ-field $\mathscr{G}$ is not countably generated and each set in it has probability either 0 or 1. In the present example $\mathscr{G}$ is countably generated and $P(G)$ assumes every value between 0 and 1 as G ranges over $\mathscr{G}$. Example 4.10 is to some extent unnatural because the $\mathscr{G}$ there is not countably generated. The present example, on the other hand, involves the pathological set H. This example is used in Section 33 in connection with conditional probability; see Problem 33.11.

4.11. (a) If $A_1, A_2, \ldots$ are independent events, then $P(\bigcap_{n=1}^\infty A_n) = \prod_{n=1}^\infty P(A_n)$ and $P(\bigcup_{n=1}^\infty A_n) = 1 - \prod_{n=1}^\infty (1 - P(A_n))$. Prove these facts and from them derive the second Borel–Cantelli lemma by the well-known relation between infinite series and products.

(b) Show that $P(\limsup_n A_n) = 1$ if for each k the series $\sum_{n>k} P(A_n | A_k^c \cap \cdots \cap A_{n-1}^c)$ diverges. From this deduce the second Borel–Cantelli lemma once again.

(c) Show by example that $P(\limsup_n A_n) = 1$ does not follow from the divergence of $\sum_n P(A_n | A_1^c \cap \cdots \cap A_{n-1}^c)$ alone.

(d) Show that $P(\limsup_n A_n) = 1$ if and only if $\sum_n P(A \cap A_n)$ diverges for each A of positive probability.

(e) If sets A_n are independent and $P(A_n) < 1$ for all n, then $P[A_n \text{ i.o.}] = 1$ if and only if $P(\bigcup_n A_n) = 1$.

4.12. (a) Show (see Example 4.21) that $\log_2 n + \log_2 \log_2 n + \theta \log_2 \log_2 \log_2 n$ is an outer boundary if $\theta > 1$. Generalize.

(b) Show that $\log_2 n + \log_2 \log_2 \log_2 n$ is an inner boundary.

4.13. Let φ be a positive function of integers, and define B_φ as the set of x in $(0,1)$ such that $|x - p/2^i| < 1/2^i\varphi(2^i)$ holds for infinitely many pairs p, i. Adapting the proof of Theorem 1.6, show directly (without reference to Example 4.12) that $\sum_i 1/\varphi(2^i) < \infty$ implies $\lambda(B_\varphi) = 0$.

4.14. 2.19↑ Suppose that there are in $(\Omega, \mathscr{F}, P)$ independent events $A_1, A_2, \ldots$ such that, if $\alpha_n = \min\{P(A_n), 1 - P(A_n)\}$, then $\sum \alpha_n = \infty$. Show that P is nonatomic.

4.15. 2.18↑ Let F be the set of square-free integers—those integers not divisible by any perfect square. Let F_l be the set of m such that $p^2 | m$ for no $p \le l$, and show that $D(F_l) = \prod_{p \le l}(1 - p^{-2})$. Show that $P_n(F_l - F) \le \sum_{p > l} p^{-2}$, and conclude that the square-free integers have density $\prod_p (1 - p^{-2}) = 6/\pi^2$.

4.16. 2.18↑ Reconsider Problem 2.18(d). If D were countably additive on $f(\mathscr{M})$, it would extend to $\sigma(\mathscr{M})$. Use the second Borel–Cantelli lemma.

SECTION 5. SIMPLE RANDOM VARIABLES

Definition

Let $(\Omega, \mathscr{F}, P)$ be an arbitrary probability space, and let X be a real-valued function on Ω; X is a *simple random variable* if it has finite range (assumes only finitely many values) and if

$$[\omega: X(\omega) = x] \in \mathscr{F} \tag{5.1}$$

for each real x. (Of course, $[\omega: X(\omega) = x] = \emptyset \in \mathscr{F}$ for x outside the range of X.) Whether or not X satisfies this condition depends only on $\mathscr{F}$, not on P, but the point of the definition is to ensure that the probabilities $P[\omega: X(\omega) = x]$ are defined. Later sections will treat the theory of general random variables, of functions on Ω having arbitrary range; (5.1) will require modification in the general case.

The $d_n(\omega)$ of the preceding section (the digits of the dyadic expansion) are simple random variables on the unit interval: the sets $[\omega: d_n(\omega) = 0]$ and $[\omega: d_n(\omega) = 1]$ are finite unions of subintervals and hence lie in the σ-field $\mathscr{B}$ of Borel sets in $(0,1]$. The Rademacher functions are also simple random variables. Although the concept itself is thus not entirely new, to proceed further in probability requires a systematic theory of random variables and their expected values.

The run lengths $l_n(\omega)$ satisfy (5.1) but are not simple random variables, because they have infinite range (they come under the general theory). In a discrete space, $\mathscr{F}$ consists of all subsets of Ω, so that (5.1) always holds.

It is customary in probability theory to omit the argument ω. Thus X stands for a general value $X(\omega)$ of the function as well as for the function itself, and $[X = x]$ is short for $[\omega: X(\omega) = x]$