MATH 441

Analysis I: Measure Theory

C. J. Atkin, 2009

§0. Preliminaries (revision).

None of the material in this section is examinable.

Although all attempts to split mathematics up into parts are fuzzy round the edges — again and again it turns out that some combinatorics creeps into analysis, or algebra into topology or vice versa, or geometry into statistics, or whatever — nevertheless there is a crude characterization of analysis as that part of mathematics which deals with "limiting processes". In algebra, you may add or multiply two objects; in geometry, you may carry out a construction finitely many times; but only in analysis can you do something infinitely often and ask whether anything sensible results. And, in order to talk about limits in practice rather than in the abstract, you have first of all to decide what you mean by a real number.

I shall take the natural numbers $1, 2, 3, \ldots$ (forming the set $\mathbb{N}$) for granted. They are so familiar, and the properties we shall use so everyday, that deeper discussion is unnecessary.

The *rational numbers* (forming the set $\mathbb{Q}$) are conveniently constructed as equivalence classes of triples (m, n, p) of natural numbers under the relation $\sim$:

$$\text{“ } (m,n,p) \sim (m',n',p') \text{ ” means “ } mp' + n'p = m'p + np' \text{ ”.} \tag{1}$$

The idea behind this definition is that the triple (m, n, p) is to be a particular representative of the rational number which we should normally write as the fraction $\frac{m-n}{p}$; the other triples in the equivalence class represent the same number. By using triples, we obtain negative, zero, and positive rationals at one go. We define addition and multiplication in the way this suggests: for instance, writing square brackets for equivalence classes,

$$[(m,n,p)] + [(m',n',p')] \coloneqq [(mp' + m'p, np' + n'p, pp')]. \tag{2}$$

You must check that this addition is "well-defined". As it stands, the right-hand side of (2) denotes the equivalence class of a triple that depends on the choice of a specific representative for each class on the left-hand side. It is necessary to show that in fact the right-hand side (the equivalence class, that is) is the same no matter what choices of representatives on the left-hand side are made. One must show, therefore, that, if $(m, n, p) \sim (a, b, c)$ and $(m', n', n') \sim (a', b', c')$, then

$$(mp' + m'p, np' + n'p, pp') \sim (ac' + a'c, bc' + b'c, cc')\,.$$

This once established (by elementary algebra), (2) defines a genuine addition of equivalence classes. Similarly for multiplication. It is now straightforward but tedious to show that $\mathbb{Q}$ is a field, i.e. satisfies all the usual laws of arithmetic, including the existence of multiplicative

inverses for non-zero elements. The rational zero, written 0, is $[(1,1,1)]$, and the rational identity, written confusingly 1, is $[(1,0,1)]$. Notice that $0 \neq 1$.

However, $\mathbb{Q}$ is also an ordered field. We can define a partial order on the set of triples:

$$\text{“ } (m,n,p) \leq (m',n',p') \text{ ” means “ } mp' + n'p \leq m'p + np' \text{ ”.}$$

It is easily checked that this does define a partial order, and $(1,1,1) \leq (1,0,1)$. Since it is clear from (1) that the order carries over to equivalence classes, we get a partial order on $\mathbb{Q}$. It is even a *total* order (because of (1)), and, for any $\alpha, \beta, \gamma \in \mathbb{Q}$,

$$\begin{aligned} &\text{if } \alpha \leq \beta, \text{ then } \alpha + \gamma \leq \beta + \gamma, \quad \text{and} \\ &\text{if } \alpha \geq 0 \text{ and } \beta \geq 0, \text{ then } \alpha\beta \geq 0. \end{aligned}$$

I have in effect already noted that $0 < 1$, i.e. $0 \leq 1$ and $0 \neq 1$. These properties are what we mean when we say $\mathbb{Q}$ is an ordered field. Notice too that $\mathbb{Q}$ includes a copy of $\mathbb{N}$, consisting of the equivalence classes of triples $(n,0,1)$, and a copy of $\mathbb{Z}$, consisting of the equivalence classes of the triples $(n,m,1)$.

Whilst the treatment I have just sketched is fairly sleek and “modern” (equivalence relations, partial orders, and what-not), things that have all the properties of the rational numbers — or at least of the non-negative ones — have been around for a long time in the familiar guise of “fractions”. The Greeks, who tended to think of things geometrically, noticed that, because of Pythagoras’s theorem, not all “lengths” could be rational multiples of each other. Although the use of geometry begs several questions about the relation of diagrams to logic, this indicates that *rational numbers are not enough* for many purposes. On the other hand, we can certainly find rational numbers that approximate the ratio of any two lengths as closely as we wish, so that, in some vague sense, the possible lengths of line segments form “numbers” that fill out the rational numbers. These are the real numbers (forming the field $\mathbb{R}$).

As with the rational numbers, and even the natural numbers, there are several methods of defining things that have all the properties intuition requires of the real numbers. In the days when I taught the basic analysis course (MATH 205), I used to define them simply as infinite decimal expansions, with the usual rules for the ambiguous case, for addition and for multiplication. This is not really very satisfactory; apart from anything else, why *decimal* expansions? And I had to slide over division, which is messy, and *assert* that Dedekind’s axiom (below, 0.5) was a simple consequence (which it is, but the simplicity is not overwhelming).

There are two standard ways of defining real numbers; they were originally published by Dedekind[1] and by Cantor[2], both in 1872 (though Dedekind had invented his in 1858), and both can be seen as elaborations of ideas going back to the Greeks. Here is Dedekind’s, which in some rather Pickwickian sense is the simpler.

Definition 0.1. A subset A of $\mathbb{Q}$ is called a *Dedekind section* or *Dedekind cut* if

$$\left.\begin{aligned} &\textit{(i)} && \emptyset \neq A \neq \mathbb{Q}, \\ &\textit{(ii)} && \text{given } \alpha \in A, \text{ any } \beta \in \mathbb{Q} \text{ for which } \beta \geq \alpha \text{ also belongs to } A, \\ &\textit{(iii)} && \text{for any } \alpha \in A, \text{ there exists } \gamma \in A \text{ such that } \gamma < \alpha. \end{aligned}\right\} \tag{3}$$

We say the cut A is *non-negative* if it does not contain the rational number 0.

---

[1] *Stetigkeit und irrationale Zahlen*, Brunswick, 1872. It was republished in 1963 in New York, apart from appearing in his Collected Works, vol. 3.

[2] Math. Ann. V (1872), pp. 123–130.

Because of *(ii)*, a non-negative cut consists entirely of positive rationals. If A is a cut and $\beta \in \mathbb{Q} \setminus A$, then $A - \beta \coloneqq \{\alpha - \beta : \alpha \in A\}$ is a non-negative cut. Such a β exists, by *(i)*.

Given a cut A, let

$$-A \coloneqq \{-\beta : \beta \in \mathbb{Q} \setminus A \ \ \& \ \underbrace{(\exists \gamma \in \mathbb{Q} \setminus A)(\gamma > \beta)}_{(\dagger)}\} . \tag{4}$$

Then $-A$ is also a cut (notice that we have to include the condition (†) in (4) to ensure 0.1*(iii)* is satisfied. For instance, the set of all positive rationals is a cut, but its complement, the set of all non-negative rationals, has a largest member, namely 0; changing the sign of all its members gives a set with a least member, making *(iii)* false). The *zero cut* (for temporary convenience, I call it Z) is defined by

$$Z \coloneqq \{\alpha \in \mathbb{Q} : \alpha > 0\} ,$$

which is indeed a cut. More generally, if $\xi \in \mathbb{Q}$, let

$$S(\xi) \coloneqq \{\alpha \in \mathbb{Q} : \alpha > \xi\} ,$$

which is also a cut.

If A and B are both Dedekind cuts, we define

$$A + B \coloneqq \{\alpha + \beta : \alpha \in A \ \ \& \ \ \beta \in B\} .$$

It is easily seen that $A + B$ is also a cut, and that the "addition of cuts" thus defined is commutative and associative. Furthermore, $A + Z = A$ and $A + (-A) = Z$ for any cut A. (The last statement requires some argument.)

If A and B are non-negative cuts, we define

$$AB \coloneqq \{\alpha\beta : \alpha \in A \ \ \& \ \ \beta \in B\} , \tag{5}$$

which is also a cut. (This would definitely be false if A or B had a negative member). If A and B are general cuts, *choose* non-positive rationals $\lambda \in \mathbb{Q} \setminus A$, $\mu \in \mathbb{Q} \setminus B$, and then $A - \lambda$ and $B - \mu$ are non-negative cuts. Hence, $(A - \lambda)(B - \mu)$ is a cut, and so is

$$(A - \lambda)(B - \mu) + S(-\lambda)(-A) + S(-\mu)(-B) + S(\lambda\mu) . \tag{6}$$

It may be shown that this cut does not depend on the choice of λ and μ, and we can take it as the definition of AB for general cuts. (For non-negative cuts we could take $\lambda = \mu = 0$, and it is easily seen that (6) reduces to (5). This very awkward definition of multiplication of cuts is the principal defect of Dedekind's theory; in this respect Cantor's is superior, but pays for its superiority in other ways.) Finally, we say that, for cuts A and B,

$$\text{“}A \leq B\text{”} \quad \text{means} \quad \text{“}B \subseteq A\text{”}.$$

With all these definitions, it may be proved — not without effort — that the set of cuts forms an ordered field, which we agree to be $\mathbb{R}$. The cuts $S(\xi)$ for $\xi \in \mathbb{Q}$ form a subfield which is isomorphic to $\mathbb{Q}$; in effect, then, $\mathbb{R}$ includes a copy of $\mathbb{Q}$ (with the same addition, multiplication, and order). We denote these "rational reals" by the names of the corresponding rationals, and write $\mathbb{Q}$ for its copy in $\mathbb{R}$ (and $\mathbb{N}$ and $\mathbb{Z}$ for their copies in that copy of $\mathbb{Q}$).

The details of the theory can be varied in many ways, but its governing idea is that a real number should be defined, at least for mathematical purposes, as the set of rational numbers that "ought to be bigger" than that real number. For instance, $\sqrt{2}$ should be thought of as the

set of all positive rational numbers whose squares are greater than 2. The reason for the condition 0.1*(iii)*, which at first sight may seem superfluous, is that without it the rational real numbers would be represented twice; 3, for instance, would correspond not only to the genuine cut $S(3)$, but also to $\{\alpha \in \mathbb{Q} : \alpha \geq 3\}$, which does not satisfy *(iii)*.

I have commented that the definition of multiplication of cuts is messy. The reason for this is that the whole construction is founded on the order relation in $\mathbb{R}$, and the arithmetical operations have little to do with it. Indeed, if we had any reason to do so, we could introduce Dedekind cuts in any totally ordered set.

Apart from the basic idea, Dedekind's peculiar contribution was in indicating a property of $\mathbb{R}$ which can be used as the foundation of all analysis. This is the *Dedekind completeness axiom*, which fails for $\mathbb{Q}$. Again, it really only uses the order structure of $\mathbb{R}$.

Definition 0.2. Let $(T, \leq)$ be a partially ordered set, where the partial order $\leq$ is such that $(x \leq y \ \& \ y \leq x) \Longrightarrow x = y$. Let A be a subset of T. Suppose $a \in A$, $t \in T$.

(i) a is the *least element* of A if $(\forall x \in A)\ a \leq x$.
(ii) a is a *minimal element* of A if
$$(\forall x \in A)(x \leq a \Rightarrow x = a).$$
(iii) t is an *upper bound* for A if $(\forall x \in A)\ x \leq t$.
(iv) A is *bounded above* (in T) if it has an upper bound.

If $\leq$ is changed to $\geq$, one has the definitions of the *greatest* element, a *maximal* element, and a *lower* bound. A set is described as *bounded* if it is bounded both above and below.

If A has a greatest element a, then a is an upper bound for A, and, conversely, if an upper bound for A also belongs to A, then it must be the greatest element of A. But, for instance, the open interval $(0, 1)$ has no greatest element. (Whatever $a \in (0, 1)$ you take, $\frac{1}{2}(1 + a)$ is greater). It has many upper bounds in $\mathbb{R}$, such as 1 and 2.

There can be at most one least element, because if a, a' are both least elements, then $a \leq a'$ and $a' \leq a$, so they are equal.

It is obvious that a greatest element of A, if one exists, is also a maximal element. In a general partially ordered set, a maximal element need not be greatest; for instance, in the set $\{a, b, c\}$, with the ordering $a \leq b$, $a \leq c$ and nothing else, both b and c are maximal but neither is greatest.

On the other hand, when A is *totally* ordered by $\leq$, which is the case for $\mathbb{R}$, the distinction between a greatest element and a maximal element of A disappears. If a is maximal and $x \in A$, then $x \geq a$ is only possible if $x = a$ by the definition of maximality; thus, either $x = a$ or $x < a$, by the total ordering (by the way, we define $x \geq a$ to mean $a \leq x$, and $x < a$ to mean $x \leq a \neq x$); that is, $x \leq a$. This fact, that "maximal" elements and "greatest" elements are the same for subsets of $\mathbb{R}$, is the reason why we sometimes speak of the maximum value of a function rather than of the greatest value (supposing that one exists). Notice, though, that the function $f : (0, 1) \longrightarrow (0, 1) : x \mapsto x$ has no greatest value.

Definition 0.3. The partially ordered set T is *Dedekind-complete* (or *boundedly complete*) if every nonnull subset of T that is bounded below has a greatest lower bound; that is to say, if the set of lower bounds for A is non-null, then it has a *greatest* element. When such a "greatest lower bound" exists, it is commonly called the *infimum* of A and written $\inf A$.

In older books $\inf A$ is sometimes denoted g.l.b. A. The significance of the infimum is that, as remarked above, a non-null set that is bounded below, such as $(0,1)$, need not have a least element; the infimum in $\mathbb{R}$ is, as it were, the nearest approach to a least element that one can have in such a case. For $(0,1)$, the infimum is 0, since the lower bounds form the whole interval $(-\infty, 0]$.

Lemma 0.4. *If the partially ordered set T is boundedly complete, then any non-null subset A of T that is bounded above has a least upper bound.*

Proof. Let U be the set of upper bounds of A. By hypothesis, $U \neq \emptyset$, and U is bounded below (by any element of A). Thus U has a greatest lower bound u. However, any $a \in A$ is a lower bound for U; thus $u \geq a$. This shows that u is itself a upper bound for A, $u \in U$. As u is in U and is a lower bound for U, it is the least element of U. □

The asymmetry of Definition 0.3, in which I mentioned only sets bounded below and lower bounds, was, therefore, only apparent; the property would be the same if I used sets bounded above and upper bounds. (In $\mathbb{R}$, one can simply change the signs of all the numbers to see this, but it is true in a general partially ordered set).

The word "complete" is over-used in mathematics, and its principal meaning in this course is quite different; its use in Definition 0.3 is therefore qualified by "Dedekind" or "boundedly". It is necessary for our purposes to demand that A should be bounded below, since $\mathbb{R}$ itself has no infimum in $\mathbb{R}$. Less trivially, $\mathbb{Z}$ has no infimum in $\mathbb{R}$. However, in some partially ordered sets *all* non-null subsets have a least upper bound and a greatest lower bound; an example is $[0,1]$, with the usual partial order. A less banal example is this. Let Ω be any set, and take T to be the class of all subsets of Ω (that is, the "power class" $\mathcal{P}(\Omega)$ of Ω). There is a natural partial order in T: "$A \leq B$" means "$A \subseteq B$". Then any subset $\mathcal{Q}$ of T has both a supremum and an infimum. Indeed, $\sup \mathcal{Q} = \bigcup_{Q \in \mathcal{Q}} Q$ and $\inf \mathcal{Q} = \bigcap_{Q \in \mathcal{Q}} Q$.

We can now prove "Dedekind's axiom" for the real numbers.

Theorem 0.5. $\mathbb{R}$ *is Dedekind-complete.*

Proof. Let A be a non-empty subset of $\mathbb{R}$ that is bounded below, with a lower bound b. Now b is a Dedekind cut of $\mathbb{Q}$, as is any element $a \in A$, and $a \subseteq b$. Let $k := \bigcup_{a \in A} a$; thus $k \subseteq b \neq \mathbb{Q}$, and $k \neq \emptyset$ as each $a \neq \emptyset$ by 0.1*(i)*. Hence, k satisfies 0.1*(i)*. If $\alpha \in k$, there is some $a \in A$ with $\alpha \in a$, and, if $\beta \geq \alpha$ in $\mathbb{Q}$, then $\beta \in a$ too; so $\beta \in k$. This means that k satisfies 0.1*(ii)*. Similarly, there exists some $\gamma \in a$ with $\gamma < \alpha$, and, as $\gamma \in k$, k satisfies 0.1*(iii)*. k is a Dedekind section of $\mathbb{Q}$; and, by definition, k is the smallest set which includes all $a \in A$. It is, therefore, the greatest lower bound of A. □

For any *finite* subset $\{a_1, a_2, \ldots, a_k\}$ of $\mathbb{R}$, or of $\mathbb{Q}$, or of any *totally ordered* set, one may find the greatest element by comparing elements in pairs, and this greatest element is denoted $\max\{a_1, a_2, \ldots, a_k\}$. It is of course the supremum of the subset. For totally ordered sets, it is only infinite subsets that may not have suprema.

The common, and not altogether false, impression that people have of analysis is that it is full of ϵs and δs. They are related to the preceding remarks by

Lemma 0.6. *Let A be a non-null subset of $\mathbb{R}$, and let $x \in \mathbb{R}$. Then x is the supremum of A if and only if*

(i) *for every $a \in A$, $a \leq x$, and*
(ii) *for every $\epsilon > 0$, there exists some $a \in A$ such that $x - \epsilon < a$.*

Proof. *(i)* clearly says that x is an upper bound for A. Suppose that x is the least upper bound. Then, for any $\epsilon > 0$, $x - \epsilon$ (being *less* than x) cannot be an upper bound for A — which means that there is some $a \in A$ with $a \not\leq x - \epsilon$, or $x - \epsilon < a$. On the other hand, if *(ii)* is satisfied but x is not the least upper bound, there is some upper bound y for A with $y < x$. Take $\epsilon := x - y > 0$; then *(ii)* says there is some $a \in A$ with $y < a$, and this contradicts the assumption that y is an upper bound. Consequently x must be the least upper bound. □

Lemma 0.7. *Let A be a non-null subset of $\mathbb{R}$, and let $x \in \mathbb{R}$. Then x is the infimum of A if and only if*

(i) *for every $a \in A$, $a \geq x$, and*
(ii) *for every $\epsilon > 0$, there exists some $a \in A$ such that $a < x + \epsilon$.* □

The proof of this Lemma may be by the obvious modification of the previous proof, or by means of the following

Lemma 0.8. *Let A be a non-null subset of $\mathbb{R}$, and let $-A := \{-a : a \in A\}$. Then $x \in \mathbb{R}$ is the supremum of A if and only if $-x$ is the infimum of $-A$.* □

Lemmas 0.6 and 0.7 constitute a link between the order structure of $\mathbb{R}$ and its "metric structure". I shall have more to say about this link later, because it has an influence on integration theory.

Definition 0.9. Let Ω be a set. A *metric* (or *distance function*) on Ω is a function $d : \Omega \times \Omega \longrightarrow \mathbb{R}$ such that, for any $x, y, z \in \Omega$,

(a) $d(x, y) = 0$ if and only if $x = y$, and
(b) $d(x, z) \leq d(x, y) + d(z, y)$.

The pair (Ω, d) is called a *metric space*. Very often, when the metric d has been unambiguously fixed, one speaks of "the metric space Ω".

The definition is often stated in slightly different, less concise, and perhaps more natural forms. Taking $x = y$ in *(b)* and applying *(a)*, $d(y, z) \leq d(z, y)$ for any $y, z \in \Omega$; since y and z may be swapped, we deduce $d(y, z) = d(z, y)$ always. Taking $x = z$, we find similarly that $0 \leq 2d(x, y)$, so that d only takes non-negative values.

(Note for those who have had some contact with these matters: theoretical physicists, and some differential geometers, use the word "metric" to denote not the actual distance function on a manifold but its "infinitesimal" version, which is a structure in the tangent bundle.)

In $\mathbb{R}$ and in $\mathbb{Q}$, there is a standard metric given in each case by

$$d(x, y) := |x - y| := \max(x - y, y - x).$$

(It is not really essential for the definition to demand that metrics take values in $\mathbb{R}$. In the case of $\mathbb{Q}$, the metric just defined takes only rational values.)

Definition 0.10. Let (x_n) be a sequence in the metric space (Ω, d), and let $x \in \Omega$. We say that x_n tends to x as n tends to infinity, or that (x_n) converges to x, or (briefly) that $x_n \to x$, if, for any positive real number ϵ, there exists some natural number N such that $d(x_n, x) < \epsilon$ whenever $n \geq N$. If there is some $x \in \Omega$ such that $x_n \to x$, we say that (x_n) is convergent or that the limit of x_n is x, $\lim_{n\to\infty} x_n = x$.

It is clear that one could allow for ϵ only positive rational values, or even numbers of the form $1/k$ for $k \in \mathbb{N}$, without affecting the meaning of the definition. I am assuming here that a "sequence" is indexed by some subset of $\mathbb{N} \cup \{0\}$. The conventional warning is worth repeating: the sense of the word "convergent" depends on the context; convergence of series is a different concept from convergence of sequences or from convergence of integrals.

Unfortunately, Definition 0.10 is not helpful in many situations of real practical importance, because very often you do not know in advance what the limit is. For instance, very often you try to find a solution of some equation by a procedure of successive approximation, without knowing in advance that there is a solution. How can you say the sequence of approximations converges to a solution?

Definition 0.11. A sequence (x_n) in the metric space (Ω, d) is *Cauchy* (that is, it is a *Cauchy sequence*) if, for any positive real number ϵ, there exists some natural number N such that $d(x_m, x_n) < \epsilon$ whenever $m \geq N$ and $n \geq N$.

Again, it would be enough to consider values $1/k$, for $k \in \mathbb{N}$, for ϵ. I leave it to you to make the necessary changes in the arguments.

Lemma 0.12. *A convergent sequence in any metric space is Cauchy.* □

Theorem 0.13. *A Cauchy sequence in $\mathbb{R}$ is convergent in $\mathbb{R}$.*

Several proofs are possible. (It is a consequence of Dedekind-completeness; indeed, the two properties are in a sense equivalent.) By Lemma 0.12 and Theorem 0.11, Cauchy sequences and convergent sequences of real numbers in $\mathbb{R}$ are the same; this is called the *General Principle of Convergence* in some old textbooks. However, a Cauchy sequence of *rational* numbers need not have a limit in $\mathbb{Q}$. A familiar example is the sequence defined inductively by

$$a_1 = 1, \quad a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right),$$

which converges in $\mathbb{R}$ to $\sqrt{2}$, so is Cauchy, but is a sequence in $\mathbb{Q}$ whose limit is not in $\mathbb{Q}$.

Definition 0.14. The metric space (Ω, d) is *complete* if every Cauchy sequence in Ω converges in Ω.

This is the commonest meaning of the word "complete" as far as we are concerned; in situations where it is ambiguous, one might say "metrically complete".

§1. Introductory remarks.

Again, this material is not examinable.

A. The general problem.

In any elementary treatment of integration (for instance in MATH 113), one meets two notions of the integral of a function $f : [a, b] \longrightarrow \mathbb{R}$ (where $a < b$ in $\mathbb{R}$). The first, customarily called an *indefinite integral*, is a *function* $g : [a, b] \longrightarrow \mathbb{R}$ such that $g'(t) = f(t)$ at every point $t \in [a, b]$, it being understood that $g'(a)$ means the right derivative at a and $g'(b)$ means the left derivative at b. This concept goes back to Newton; I shall sometimes call it the Newton integral. It obviously arises naturally from the notion of the derivative, and has the further advantage that, for many functions that are useful in applications, we can find explicit indefinite integrals; indeed, much of 113 was about methods of doing so.

The second notion, usually called the *definite integral*, and essentially due to Leibniz, is merely a *number*; speaking intuitively, it is the area under the graph of f over the interval $[a, b]$. (More exactly, that is the definite integral when f takes only non-negative values — if f were to change in sign, its definite integral over $[a, b]$ would be the difference of the area above the interval and below the graph and the area below the interval and above the graph). I shall sometimes call it the Leibniz integral. Leibniz and Newton, of course, thought of a function as something defined by a rather simple formula.

Unfortunately, as I used to stress when I taught these things, there is a difficulty with the Leibniz integral: it takes for granted that we know what is meant by "area". There are two aspects of this question, the philosophical/psychological (what is the basic concept of area, where does it come from, and why do we believe there should be such a thing?) and the computational (how can we calculate the area of a specific figure, supposing such a number to exist?). As very often in mathematics, we can, for mathematical purposes, fudge the philosophy and psychology by defining the area of a figure to be the result, *when it exists*, of a suitable procedure of calculation. For simple figures like rectangles or polygons, there are more or less mechanical rules for calculating their areas directly (split them up into triangles and add the areas of the triangles). It is less clear what to do for figures like circles, or more generally anything with a curved boundary, although our intuition — whatever its origin — certainly does not balk at the idea that such figures should have areas.

This was already a problem for the Greek mathematicians, and one of their great achievements was the calculation of areas and volumes of some curved regions. Their results are now easy exercises in integration, but, of course, they did not have a satisfactory algebraic notation to help. It is a historical commonplace that modern mathematics really started with the development of good notation. Anyway, the Greek idea was that the area of a curvilinear region in two dimensions should be calculated by, and therefore *is*, the result of approximating the region by polygons; and that is essentially the idea which everybody has followed since.

There are, however, figures in $\mathbb{R}^2$, even 'regions' under the graphs of functions (I shall temporarily say 'region' to mean 'set in $\mathbb{R}^2$'), for which the very notion of area seems strange. The hackneyed example is the 'region' under the graph of the function $f : [0, 1] \longrightarrow \mathbb{R}$, where $f(t) = 1$ when t is rational and $f(t) = 0$ when t is irrational. (Some people call this the "Dirichlet function"). The graph of this function would appear to the naked eye as a line segment at height 1 above the t-axis, actually consisting of the points at that height over rational values of x, and a second line segment on the axis, really containing only the irrational points; thus neither is a genuine segment at all. Any polygon

which includes the whole 'region' has area at least 1, and any polygon included in the 'region' has zero area. Approximation of the 'region' by polygons is, therefore, not really possible. There are many other similar constructions.

Such examples suggest serious difficulties with the concept of area. The practical man will say, of course, that our example is merely artificial, and that, in the real world, such functions and such 'regions' will never arise. There is quite a large grain of truth in this, for it is difficult to see how a description of the real world might have to call on them directly; but what if the mathematical techniques you then apply lead to these "impractical" functions? Our function f, for example, is a the limit (pointwise) of a sequence of continuous functions. It is, therefore, very difficult to see a good reason why a function like f should be considered illegitimate; its definition seems logically unassailable.

We are therefore faced with a definite, if imprecise, question: for what sorts of set in $\mathbb{R}^2$ can we reasonably speak of their "area"? To make the question precise, we should first have to specify what we require of "area". We could, for instance, arbitrarily assign "area" 1 to all non-null sets in $\mathbb{R}^2$; such a definition would have no useful properties at all. Presumably we want our "areas" either to agree with ordinary area on polygons, or at least to have properties as good as ordinary area on polygons.

Once we agree a method of defining areas for some class of subsets of $\mathbb{R}^2$, we shall automatically acquire a definition of the Leibniz integral for a corresponding class of functions, namely those for which the 'regions' between the graph and the axis belong to the class of subsets in question. That is obvious; and, suitably interpreted, it is also true the other way round: a reasonable definition of the Leibniz integral will lead to a definition of area for a corresponding class of regions. In short: theories of (Leibniz) integration and theories of area are more or less equivalent.

The Newton integral ought to reappear here, as a possible method of defining definite integrals and therefore areas. But it, too, is subject to a rather surprising objection: given a real-valued function on an interval in $\mathbb{R}$, there is at present no way of determining whether it is the derivative of a differentiable function; the class of functions which have Newton integrals has never been intrinsically characterized. The derivative of a differentiable function is known to have many special properties, but no necessary and sufficient condition for a function to be a derivative has yet been discovered. One may presume that any such condition would have to be rather messy. In practice, you know a function has a Newton integral if you know what that Newton integral is.

The important fact is, of course, that any *continuous* function is a derivative. This is one of the assertions of Cauchy's "fundamental theorem of calculus". To prove it, one makes use of an integral of the Leibniz type, namely the Riemann integral. However, a discontinuous function as simple as $f(x) = 0$ for $x \leq 0$, $f(x) = 1$ for $x > 0$, cannot be Newton-integrable, though it is Leibniz-integrable on any interval $[a, b]$.

It is possible to extend the Newton integral in a simple way so as to encompass all the functions of a single real variable that one needs for elementary purposes. The idea is to require that the indefinite integral g should be continuous, and that it should satisfy the equality $g'(x) = f(x)$ *except at countably many points*. (Non-trivial argument is needed to show that this definition has some of the properties one expects of an integral). This allows f to have plenty of discontinuities and still be "Newton-integrable" in this extended sense, but the class of such "integrable" functions remains obscure. In practice, we can make more progress by studying the Leibniz integral first.

For the sake of its appeal to intuition, I have talked about the problem of assigning an "area" to a figure in $\mathbb{R}^2$. By extension, there is a like problem in $\mathbb{R}^n$ for $n \geq 3$. Once the difficulty is recognized, however, it is clear that there is an analogous problem even in $\mathbb{R}$. An interval has a well-defined length, and so has any set which can be expressed as the union of a finite class of intervals. When one considers more complicated sets, is there any reasonable sense in which one can talk about their "length"? Intuition suggests, rather unreliably, that maybe there ought to be, because some sets (even amongst subsets of $\mathbb{R}$ that have the same cardinality) seem "thinner" or "sparser" than others.

The first aim of the course is to construct a general theory which will enable us to speak of the "length" of a set in $\mathbb{R}$, or the "area" of a set in $\mathbb{R}^2$, or the "volume" of a set in $\mathbb{R}^3$, with the confidence that the numbers we are assigning to these sets behave in a way appropriate to our intuition. This is a surprisingly difficult task. For historical reasons, to emphasize the theoretical framework, and to avoid terms like "length" or "area" that are felt to be specific to a given dimension, the general word actually employed is "measure"; the subject is called "measure theory", and embraces not only the ordinary lengths, areas, and volumes, which correspond to "Lebesgue measure" in $\mathbb{R}^n$, but also generalizations both to other spaces and to other ways of describing the size of sets in $\mathbb{R}^n$. (As an example: the "area" of a straight line segment in $\mathbb{R}^2$ is zero, but it is possible to define a measure in $\mathbb{R}^2$ which assigns to each straight line segment its length and takes the value ∞ on 'two-dimensional regions'). The theory as we have it, basically Lebesgue's, was introduced in his thesis in 1902, but our treatment follows some improvements, though not the nomenclature, due to Carathéodory in his 1918 book on the subject. After him the theory became more abstract and general, and much of the terminology now commonly used, which I shall reproduce, was invented by Halmos (his book came out in 1950). It remains true that the vocabulary of the subject is not entirely fixed, but I have tried to follow the commonest usages.

With the abstract idea of a *measure* in any set whatever, we shall be able to associate an *integral* of the Leibniz type: the so-called "Lebesgue integral". This procedure does not require the measure to be Lebesgue measure in $\mathbb{R}^n$, or the domain of the integrand to be a subset of $\mathbb{R}^n$. Furthermore, the definition, as we shall present it, is quite intuitive. Indeed, the only really serious reason why the Lebesgue integral is not mentioned in earlier courses is the difficulty of explaining what a measure is and of constructing interesting measures. (The Lebesgue integral does involve other technical problems, as we shall see, but they are no worse than those for the Riemann integral.)

The Lebesgue integral is the single most essential tool of modern analysis, and I am about to devote some space to praising it. Nevertheless, it is only after seeing what results from it — it is more satisfactory in almost every way than the Riemann integral — that you will begin to grasp why it is so important.

Lebesgue measure in $\mathbb{R}^n$ assigns a "measure" to *all sets in* $\mathbb{R}^n$ *that one can meet in practice*, and has all the properties expected of a "volume" (and some more). It follows — almost — that the Lebesgue integral, unlike Riemann's, enables us to integrate all functions that we can reasonably hope to. (There are functions that we should not expect to be able to integrate: for instance, the function $f : \mathbb{R} \longrightarrow \mathbb{R}$, where

$$f(x) := \begin{cases} 1 & \text{when } x \leq 0\,, \\ -1 & \text{when } x > 0\,.) \end{cases} \tag{7}$$

As this example suggests, the integrals we can define are "absolutely convergent". Other

integrals have been invented since to permit integration of certain functions whose integrals are only conditionally convergent, such as $\sin(x^2)$ over the whole of $\mathbb{R}$, but they have to do so by abandoning some of the desirable properties of the Lebesgue integral and usually by restricting attention to $\mathbb{R}$. One relatively recent theory (48 years old or more by now) begins by a clever modification of the definition of the Riemann integral, and there are other definitions of an integral that have similarly been invented with the aim of simplifying the construction, of handling special situations, and, in particular, of cutting short the discussion of measure, but their final results have always been more or less expressible in terms of the Lebesgue theory and they have often been less generally applicable.

In short, the Lebesgue integral, understood after Carathéodory as the integral with respect to a measure that may not necessarily be Lebesgue measure in $\mathbb{R}^n$, seems in a sense to be *the* integral. You can trim it round the edges or put fringes on it, but any essential change will lose you some useful property. This is not mere prejudice; *once the existence of a measure is granted*, the construction of the Lebesgue integral is strikingly "natural". Indeed, the most sternly "practical" applied mathematicians, who still regard anything beyond the Riemann integral as an idle theoretical subtlety, tend nevertheless tacitly to assume the Lebesgue integral exists and has the properties they want.

I have presented the case for the Lebesgue theory of integration on the basis of the "geometrical" problem "what is area?", but there is another reason. Kolmogorov pointed out in 1932 that a rigorous mathematical theory of probability, which had been conspicuously lacking until then, could be founded on the idea that probability is in fact a measure defined on the "events" in "sample space" (which need not be $\mathbb{R}^n$), that "expected values" are in fact integrals, and so on. This was a conceptual breakthrough, because, despite a great deal of probabilistic knowledge, no-one had really had a satisfactory notion what the logical basis of probability should be. (I do not mean here the quite different problem of finding the probability of a specific event).

My treatment of the theory is thoroughly old-fashioned. I shall begin by constructing a large class of measures, of which Lebesgue measure is overwhelmingly the most important example, and then discuss integration with respect to a measure, which need not be Lebesgue measure. There is a good reason for proceeding in this way: almost all the concepts we shall meet have applications outside their immediate context. The abstract concept of a measure crops up not only in probability theory but in logic, number theory, theoretical physics, and, indeed, in almost all of modern mathematics. Furthermore, as I have already hinted, by sticking to the main road of measure theory I shall be using the methods and nomenclature that are most commonly met. Indeed, I make a deliberate effort to conform to the conventions of the more recent writers on related subjects.

There is a wealth of books on measure and integration, and quite a number of them are recommendable. I hope my notes will be more or less sufficient, but the three books I customarily suggest as supplementary sources are

S. Saks, *Theory of the Integral*.

I give the title of the second edition, which is available in the Library in the Dover reprint; the first "published" edition, in French, is also there (the first edition of all consisted of lecture notes in Polish). This is a great classic, which remains the principal reference for many of its unfashionable later topics. We are only concerned with the first three chapters, which Saks rushes through with great clarity and efficiency. You are warned, though, that his notation is old-fashioned; he writes unions as sums, intersections as products, and so on. His terminology has also been in some respects superseded.

M. E. Munroe, *Introduction to the theory of measure and integration*.

Not a classic, but a very superior American textbook, not infrequently cited as a reference. Its great merits are that it is readable, assumes remarkably little of the reader, and covers in a very clear and accessible fashion several topics often slighted nowadays; furthermore, it takes the space to discuss some important side-issues and applications, which more ambitious and advanced books, such as Saks or Halmos, often understandably ignore. Its defects are, again, that it uses rather singular and outdated terminology, that it takes a very long time to come to the integral — for example, the very extended first chapter consists of general mathematical culture, very interesting in itself, but only occasionally needed later — and that Munroe is perhaps not a very good mathematician. There are parts of the book where he loses his way and makes very heavy weather of easy facts; in other places he seems to miss an essential point; and the first printing of the first edition even contained a whopping error in the main text, which cannot possibly be described as a slip. The second corrected printing still has minor inaccuracies. But, taking it as a whole, it remains one of the most informative and most interesting elementary introductions to the subject.

P. R. Halmos, *Measure Theory*.

Probably still the standard reference, and not only in English. As I have said above, it was responsible for fixing quite a lot of the terminology we shall use; and it was the first really coherent presentation of the subject in "modern" guise. It is also both clear and very readable, as usual with Halmos. As a textbook, however, it is defective; it tends to relegate concrete examples, no matter how important, to exercises, and — for quite defensible reasons — develops the foundations at infuriating length and in a generality which is not really desirable for a first course. The definition of the integral arrives irritatingly late and is expressed in a rather odd way; some important topics are omitted altogether; and the last three chapters, irrelevant for us, expound a rather eccentric version of the theory they discuss.

Several other books are also acceptable, but I cannot say much about them. There is one by Berberian, which seems very close to the spirit of the course; one by Williamson, restricted to the integral in $\mathbb{R}^n$, but pleasingly concise; one by Zaanen, which includes some valuable material neglected elsewhere; one by Burkill. The book by Kolmogorov and Fomin, which has been published in English in several versions, is a very readable discussion of a alrge amount of material in which two chapters are relevant to our course.

At a rather higher level, there is a superb book by Dudley, which includes most of the things I should have put in a book if I had written one, but is too compressed to be a good introduction. An enormous number of books on functional analysis, harmonic analysis, probability theory, etc., begin with abbreviated treatments of integration more or less exactly on our lines (for example, Dunford and Schwartz's *Linear Operators*, Zaanen's earlier book *Linear Analysis*, Loève's *Probability Theory*, Federer's *Geometric Measure Theory*), and there are other books such as Rudin's *Real and Complex Analysis* which include good discussions of measure and integration. Generally speaking, any book with 'real analysis' in its title will do so.

Avoid Bourbaki, who is heavy going and takes a different and less practical route from ours. The first part of Riesz and Nagy's *Functional Analysis*, though very readable and informative (Riesz wrote it), also takes a rather unusual approach to the subject that scarcely touches ours at all, and is less general. The "generalized Riemann integral" or "Henstock-Kurzweil integral" is discussed in books by Henstock, McLeod, and Bartle. The central

position of our subject means that many approaches are possible, and if you have some special application in mind you may prefer some unusual derivation of the main results.

§B. Conventions.

To avoid later explanations, here are a few remarks on phrases and what-not that will appear again and again.

Definition 1.1. Let E be a set. We say that E is *countable* if there is a one-to-one correspondence between E and a *subset* of the set $\mathbb{N}$ of natural numbers.

Notice that an infinite subset F of $\mathbb{N}$ is necessarily in one-to-one correspondence with the whole of $\mathbb{N}$. The correspondence may carry the least element of F to 1, the next least element of F to 2, and so on. Thus any set E which is both countable and infinite is in one-to-one correspondence with $\mathbb{N}$ itself.

My use of "countable" means "either finite or denumerably infinite". Rather often, I shall casually neglect the possibility that a countable class may be finite, and it is to be understood that in such a situation any argument, if I were to give it, would be unchanged in the finite case except for minor alterations in notation. For instance, I might write $\bigcup_{i=1}^{\infty} F_i$, as if there were infinitely many indices i, even though in fact there may be only finitely many.

Lemma 1.2. *If E is a countable set and $E \supseteq F$, then F is countable; if E and H are both countable sets, then so is $E \times H$; if $\mathcal{A}$ is a countable class whose members are countable sets, then the union of the members of $\mathcal{A}$ is also a countable set.* □

Make sure you understand what is being said here. As an example, if E_i is a countable set for each of the indices $i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} E_i$ is also a countable set.

There are a number of phrases we shall naturally use whose literal meaning is debatable, or even definitely different from their customary meaning.

If I speak of a 'finite union of sets from a class $\mathcal{A}$', what is meant is a set E which can be expressed in the form $E = \bigcup_{K \in \mathcal{B}} K$, where $\mathcal{B}$ is a finite subclass of $\mathcal{A}$. In other words, E is called a 'finite union' because it can be written as the union of finitely many sets, not because it is itself a finite set. This is a quite standard turn of phrase, and one talks likewise of 'finite intersections', 'finite Cartesian products', 'countable unions', and so on. With this convention, the last assertion of 1.2 is often briefly stated in the apparently tautologous form: a countable union of countable sets is countable.

Similarly, if, again, $\mathcal{A}$ is a class of sets, we describe it as a 'disjoint class' when any two of its members are either the same or disjoint (or both). That is: $\mathcal{A}$ is a disjoint class if, for any $A, B \in \mathcal{A}$, either $A = B$ or $A \cap B = \emptyset$. In the same way, a sequence of sets $(E_n)_{n=1}^{\infty}$ will be called a 'disjoint sequence' if $E_m \cap E_n = \emptyset$ whenever $n \neq m$. (In such a sequence, any repeats must be null.) Some people describe such sequences as 'pairwise disjoint', for obvious reasons.

The *set difference* $E \setminus F$ of two sets E, F is $\{x : x \in E \ \&\ x \notin F\}$. I shall customarily write it with a 'backslash' or 'slant' to distinguish it from arithmetical subtraction, since we shall often be using both. Many authors use an ordinary minus sign.

We shall also occasionally use the *symmetric difference* of the two sets E and F, which is customarily written $E\Delta F$ and is defined to be $(E \setminus F) \cup (F \setminus E)$. It is the set of elements belonging to exactly one of E and F.

As these remarks have probably already suggested, a great deal of the course will involve manipulations (unions, intersections, differences) with sets. They will be made more confusing by the fact that often whole sequences of sets, rather than one or two, will be in play. So, firstly, do be on your guard — it is very easy even for experienced mathematicians to be deceived by complicated formulæ of set unions and differences into supposing various equalities to hold that are in fact wrong. As an example: it is quite tempting to suppose on the basis of the notation that, for any sequences $(E_n), (F_n)$ of sets,

$$\bigcup_{n=1}^{\infty}(E_n \setminus F_n) = \left(\bigcup_{n=1}^{\infty} E_n\right) \setminus \left(\bigcup_{n=1}^{\infty} F_n\right),$$

but, if you think about it more carefully, you will see that equality does not hold in general. (Take, for instance, $E_n = [n, n+1)$ and $F_n = [n+1, n+2)$.) The equality is true if the E_n are disjoint and $F_n \subseteq E_n$ for each n, by the way.

This may seem to mean that every formula, other than the simplest, needs scrupulous analysis; but fortunately there is a simple way of deciding, at least in the easier cases, what ought to be true: namely, Venn diagrams. The second piece of advice is therefore not to despise Venn diagrams, which are often the quickest way of grasping set-theoretic formulæ. They cannot, of course, constitute a proof in themselves, and they have to be drawn carefully, but very often they suggest how a proof might go.

The algebra of sets uses various operations such as $\cup$, $\cap$, $\setminus$, Δ. As far as I know, there is no 'official' order of precedence amongst these operations, such as we are taught in school for the arithmetical operations $+$, $-$, $\times$, $\div$ of ordinary algebra. Thus $A \cap B \cup C$ is literally meaningless. In Saks's day $\cup$ was written as $+$ and $\cap$ as multiplication, ., and the precedence of multiplication was observed; $AB + C$ meant $(A \cap B) \cup C$. (This notation was perhaps dropped later because of such odd-looking statements as the distributive law $(A + C)(B + C) = AB + C$). Today, however, some care is necessary in setting out set-theoretical formulæ. I shall try to be careful to insert brackets where they are logically necessary for the sense to be unambiguous, but many other authors (including Munroe and Halmos) rely only on the typeface and on common sense; thus, in an expression like

$$A \cup \bigcap_{n=1}^{\infty} E_n,$$

the size of the 'intersection' sign, and its affixes, are considered to make the interpretation of the formula certain, despite the absence of the parentheses that I should consider desirable.

Whatever the formal prerequisites, this course is, I think, surprisingly near to being self-contained in terms of the concepts required, although you do need a certain familiarity with abstract argument. We shall use only the basic facts about the algebra of sets, convergence of series, and so on. Nevertheless, I shall occasionally refer to some elementary concepts of topology such as 'open sets', 'closed sets', 'compactness', and so on, and I shall not discuss them at any length when this happens.

§C. Jordan content.

This subsection may be ignored if you are in a hurry, and is in any case not central. It outlines the theory of area which corresponds to the Riemann integral and was superseded by Lebesgue's ideas. This theory (in essence due to Camille Jordan, some time after Riemann's death) makes an immediate appeal to our intuition; it is, in fact, more or less the idea the ancient Greeks worked with; and its construction requires only a remarkably slight modification to yield Lebesgue's *definition* of measure in $\mathbb{R}^n$. (The deeper properties of Lebesgue measure cannot, however, be established so easily). There is no point in developing Jordan's theory in detail, since Lebesgue's results are better in every respect, but my sketch will perhaps explain where Lebesgue's work began.

Let a, b, c be points in $\mathbb{R}^2$. The *open triangle* with vertices a, b, c is the set

$$\Delta(a,b,c) := \{\lambda a + \mu b + \nu c \,:\, \lambda, \mu, \nu \in \mathbb{R} \;\&\; \lambda > 0 \;\&\; \mu > 0 \;\&\; \nu > 0 \;\&\; \lambda + \mu + \nu = 1\}\,.$$

If a, b, c are collinear, the triangle is called *degenerate*. The *closed triangle* with the same vertices is

$$\bar{\Delta}(a,b,c) := \{\lambda a + \mu b + \nu c \,:\, \lambda, \mu, \nu \in \mathbb{R} \;\&\; \lambda \geq 0 \;\&\; \mu \geq 0 \;\&\; \nu \geq 0 \;\&\; \lambda + \mu + \nu = 1\}\,.$$

The words 'open' and 'closed' in these definitions are merely conventional, and you should not think of them as directly related to the topological terms 'open' and 'closed'.

The geometrical meaning of the definitions is important. To begin with a simpler instance, the set $\{\lambda a + \mu b \,:\, \lambda > 0 \;\&\; \mu > 0 \;\&\; \lambda + \mu = 1\}$ is precisely the line segment between a and b, excluding the points a and b themselves — unless $a = b$, when it is just the singleton $\{a\}$. In the same way, $\Delta(a,b,c)$ is the "interior" (again, in a geometrical rather than topological sense) of the triangle with vertices a, b, c, as long as a, b, c are the vertices of a genuine triangle. If a, b, c are all on a single straight line ℓ but not all the same, $\Delta(a,b,c)$ is the open segment of the line ℓ between the two furthest separated points of a, b, c. If a, b, c all coincide, $\Delta(a,b,c)$ is the singleton $\{a, b, c\}$. For $\bar{\Delta}(a,b,c)$, change the word "open" to "closed". It is geometrically obvious (I shan't give a formal proof) that, when a, b, c are not collinear, they are the only possible vertices for $\Delta(a,b,c)$. (To avoid bias towards dimension 2, $\Delta(a,b,c)$ is often called the open 2-simplex spanned by $\{a, b, c\}$; there are corresponding constructions in higher dimensions).

It is never necessary in what follows to talk explicitly about closed triangles. The reason is that $\bar{\Delta}(a,b,c)$ is a disjoint union of open triangles, namely of some selection of $\Delta(a,b,c)$, $\Delta(b,c,b)$, $\Delta(c,a,c)$, $\Delta(a,b,a)$, $\Delta(a,a,a)$, $\Delta(b,b,b)$, $\Delta(c,c,c)$. If a, b, c are not collinear, all these open triangles are disjoint and have union $\bar{\Delta}(a,b,c)$; but, of course, only the first is non-degenerate.

Let E be an open triangle in $\mathbb{R}^2$. If E is a subset of a line, set $a(E) := 0$; if not, let its vertices be $a = (a_1, a_2)$, $b = (b_1, b_2)$, $c = (c_1, c_2)$, and then define

$$a(E) := \tfrac{1}{2}|b_1c_2 - b_2c_1 + c_1a_2 - c_2a_1 + a_1b_2 - a_2b_1|\,.$$

(The expression inside the absolute value signs is the third coordinate of the vector product $((c_1, c_2, 0) - (a_1, a_2, 0)) \times ((b_1, b_2, 0) - (a_1, a_2, 0))$, which has 0 as both its first and its second coordinate. Thus its magnitude is just the absolute value of the third coordinate. But the magnitude of the vector product is $\|c - a\|\|b - a\| \sin\theta$, where θ is the angle between the vectors $c - a$ and $b - a$. This is twice the 'ordinary' area of the triangle with adjacent

sides described by the vectors $c-a, b-a$. So our formula expresses the area of the triangle with vertices a, b, c).

Let us say that a set P in $\mathbb{R}^2$ is a *polygon* if it is a finite union of open triangles. It follows immediately that the union of two polygons is still a polygon.

There are two less obvious facts. Firstly, a polygon may be written as a finite *disjoint* union of open triangles (many of them degenerate, of course). Secondly, the difference of two polygons is still a polygon — notice that, in particular, the null set is a polygon, $\Delta(a,a,a)$. You may easily convince yourself of the truth of these statements by drawing simple pictures; the formal proofs must involve induction, beginning from an argument to show that the difference of two open triangles is a finite disjoint union of open triangles. This, incidentally, is the reason why it is convenient to use *open* triangles.

Let P be a polygon in $\mathbb{R}^2$. Express it as a disjoint union of open triangles, and take the sum $a(P)$ of the areas of these triangles. This sum does not depend on the way P is expressed as a disjoint union of open triangles, because, if you so express it in two different ways, you can split it up in a third way more finely so that every triangle of either of the first two decompositions is a finite disjoint union of triangles of the third, and its area is the corresponding sum of the areas of the smaller triangles. (This, whilst obvious, is *not* easy to prove, although the proof is not subtle or clever; it is just long and not very interesting). Thus $a(P)$ may naturally be described as the area of P. This concept has the obvious property, which we should expect of any notion of area, that

$$a(P_1 \cup P_2) = a(P_1) + a(P_2)$$

whenever the polygons P_1 and P_2 are disjoint.

Thus we have a class $\mathcal{P}$ of sets in $\mathbb{R}^2$, namely the class of polygons, which has the property that, when $P_1, P_2 \in \mathcal{P}$, both $P_1 \cup P_2$ and $P_1 \setminus P_2$ also belong to $\mathcal{P}$ (and, as a consequence, $P_1 \cap P_2 = P_1 \setminus (P_1 \setminus P_2)$ also belongs to $\mathcal{P}$). Furthermore, the area function $a : \mathcal{P} \longrightarrow \mathbb{R}$ is such that

$$\left.\begin{array}{ll} \textit{(i)} & a(\emptyset) = 0\,, \\ \textit{(ii)} & a(P_1) \geq a(P_2) \quad \text{for any } P_1, P_2 \in \mathcal{P} \text{ such that } P_2 \subseteq P_1\,, \text{ and} \\ \textit{(iii)} & a(P_1 \cup P_2) = a(P_1) + a(P_2) \quad \text{if } P_1, P_2 \in \mathcal{P},\ P_1 \cap P_2 = \emptyset,\ P_1 \cup P_2 \in \mathcal{P}\,. \end{array}\right\} \qquad (8)$$

Condition *(iii)* on its own is described by saying that a is *additive* on $\mathcal{P}$. (The hypothesis in *(iii)* that $P_1 \cup P_2 \in \mathcal{P}$ is automatically satisfied for the class of polygons.)

The additivity of a derives most of its force from the set-theoretical properties of $\mathcal{P}$. As an extreme instance, one might have a function a on some class $\mathcal{P}$ which satisfied *(i)*, *(ii)*, and *(iii)*, but for which *(iii)* held only because there were no disjoint pairs of non-null sets in $\mathcal{P}$ whose union is also in $\mathcal{P}$.

For the class of polygons, on the other hand, *(iii)* has its full value, because for any $P_1, P_2 \in \mathcal{P}$ one has a disjoint pair in $\mathcal{P}$ consisting of P_1 and $P_2 \setminus P_1$, and its union is $P_1 \cup P_2 \in \mathcal{P}$. Incidentally, $P_2 \setminus P_1$ and $P_1 \cap P_2$ form another disjoint pair, so we get

$$a(P_1 \cup P_2) = a(P_1) + a(P_2 \setminus P_1) \quad \text{and} \quad a(P_2) = a(P_2 \setminus P_1) + a(P_1 \cap P_2)\,.$$

The properties *(i)*, *(ii)*, *(iii)* are perhaps the natural requirements for an "area". So far, then, we can define the area of any *polygon* in a way that agrees with our intuition and satisfies (as one would expect) the requirements one would like to impose on any idea of area.

The elementary notion of area that one learns in school is based on rectangles, not triangles. This leads to a smaller class of sets for which areas are defined, as above, by taking finite unions. Even triangles cannot be expressed as *finite* unions of disjoint rectangles.

Let $E \subseteq \mathbb{R}^2$. Define the *outer Jordan content* of E to be

$$c^*(E) \coloneqq \inf\{a(P) : E \subseteq P \in \mathcal{P}\}\,.$$

If E is a bounded set in $\mathbb{R}^2$, it is included in a sufficiently large triangle, so that $c^*(E)$ is finite. Conversely, if E is unbounded, it cannot be included in any polygon (all polygons are bounded), so that, symbolically (see §2), $c^*(E) = \infty$. It is evident from (8)*(ii)* that c^* agrees with a on polygons, and, unlike a, it is defined for any set E. Moreover, it satisfies the properties corresponding to (8)*(i)* and (8)*(ii)*. Unfortunately, it does not automatically satisfy (8)*(iii)*. For instance, if

$$E_1 \coloneqq \{(x,y) \,:\, 0 \le x \le 1\,,\, 0 \le y \le 1\,, x \in \mathbb{Q}\}\,,$$
$$E_2 \coloneqq \{(x,y) \,:\, 0 \le x \le 1\,,\, 0 \le y \le 1\,, x \notin \mathbb{Q}\}\,,$$

then $E_1 \cup E_2$ is the unit square, so that $c^*(E_1 \cup E_2) = 1$, and $E_1 \cap E_2 = \emptyset$, but both $c^*(E_1)$ and $c^*(E_2)$ are 1. (This corresponds to the 'Dirichlet function' of (7)).

The question therefore arises whether the analogue of (8)*(iii)* for c^* is true for any interesting class of sets. The problem with the example just given is — intuitively speaking — that the sets E_1, E_2 are geometrically very peculiar; each has the whole unit square as its topological boundary. There is a natural way of eliminating such sets from consideration. We define the *inner Jordan content* of a set E to be

$$c_*(E) \coloneqq \sup\{a(P) : E \supseteq P \in \mathcal{P}\}\,,$$

and say that a set E is *Jordan-measurable* if it is bounded and $c_*(E) = c^*(E)$. (The idea is that a Jordan-measurable set can be squeezed between two polygons whose areas differ by an arbitrarily small number. Thus it is desirable to restrict attention to bounded sets; otherwise the outer content would be ∞, and there are examples of nasty unbounded sets whose inner Jordan content is also ∞. For instance, take the union of the left half-plane with E_1.)

It may be shown rather easily that a bounded set is Jordan-measurable if and only if its topological frontier has outer Jordan content zero.

When E is Jordan-measurable, the common value of the inner and outer Jordan contents of E is called simply the Jordan content of E, and I denote it by $c(E)$.

Theorem 1.3. *Let $\mathcal{J}$ denote the class of Jordan-measurable sets in $\mathbb{R}^2$. Then $\mathcal{P} \subseteq \mathcal{J}$, so that $\emptyset \in \mathcal{J}$, and $c(\emptyset) = 0$; more generally, for any $P \in \mathcal{P}$, $c(P) = a(P)$; for any $J_1, J_2 \in \mathcal{J}$, both $J_1 \setminus J_2 \in \mathcal{J}$ and $J_1 \cup J_2 \in \mathcal{J}$; if $J_1, J_2 \in \mathcal{J}$ and $J_1 \cap J_2 = \emptyset$, then $c(J_1 \cup J_2) = c(J_1) + c(J_2)$; and $c(J) \ge 0$ for any $J \in \mathcal{J}$.* □

In effect, all the properties which we noted for the area function a and the class of polygons extend to the Jordan content function on the class of Jordan-measurable sets, and this class is much more inclusive. It contains all "elementary geometrical figures", even with curved boundaries. (This statement cannot be *proved* without a more precise notion of "elementary geometrical figure"). Jordan content is precisely the notion of "area" which corresponds to the Riemann integral.

The definition of outer Jordan content may be reformulated as follows. Given a set E, $c^*(E)$ is the infimum of the sums of the areas of finite collections of triangles that cover E. (Indeed, if the triangles are disjoint, their union is a polygon including E whose area is just the sum of the areas of the triangles; if they are not disjoint, the sum of their areas is not less than the area of their union, which is itself a disjoint union). The devastating insight of Lebesgue was to substitute *countable* collections of triangles in this definition; the corresponding construct is the *Lebesgue 2-dimensional outer measure* of E. Since limiting processes are essential at every stage of his construction, for instance in summing countably many areas, nothing is gained here by using triangles — the same outer measure results from taking coverings by countable collections of coordinate rectangles, which was in fact Lebesgue's definition. The change in the definition from Jordan's version may well seem nugatory (the limiting procedure has just been shifted to an earlier stage of the construction), but it turns out to have remarkable consequences.

The original treatment of Lebesgue followed Jordan's arguments rather closely; in particular, inner (Lebesgue) measure was also defined for bounded sets, and such a set was defined to be *measurable* if its inner and outer measures agreed. The extension to unbounded sets was achieved by splitting them up as unions of countably many bounded sets. A rather similar procedure was followed in defining the integral: first integrate bounded functions on bounded sets, then extend to unbounded functions, then to unbounded sets. In all these respects, Lebesgue's original theory, like many other theories in their first versions, was rather messy. Some 15 years later, Carathéodory proposed simplifications which removed the necessity of introducing inner measure; he could define measurable sets in an ingenious fashion requiring no boundedness assumption. The theory becomes a little less intuitive, because Carathéodory's definition of measurability is rather unexpected, and the analogy with Jordan content is less transparent, but there is a great gain in generality and elegance. Crudely speaking, one is no longer tied to $\mathbb{R}^n$. This is the version of the theory that we shall study.

Even in $\mathbb{R}^n$, Lebesgue's theory is a great improvement on Jordan's, because *all reasonable sets in* $\mathbb{R}^n$ *are measurable*. In everyday language, any set you can actually get hold of has an area (or volume, or whatever you call it in higher dimensions) which agrees with the ordinary idea of area of a triangle (or cube or whatever) and has the properties (8) you hope for, and even some better ones. Since every useful set has an 'area', the corresponding notion of integration will handle every useful function we might hope to integrate. (As I pointed out in §A, there are, nonetheless, simple functions which must be non-integrable).

There is, however, a thorny point here. In talking of 'reasonable' or 'useful' sets, I am obviously being very vague. Any set which is obtained from polygons (or polyhedra) by the standard procedures of analysis, all of which are "countable" in character, will be Lebesgue-measurable, and these are the sets I have in mind. But it is not clear whether all sets can be constructed in this way, and, in fact, if one assumes the Axiom of Choice, which most people very sensibly prefer to do, one can show that there must be some sets in $\mathbb{R}^n$ which are not Lebesgue-measurable. Being constructed by means of the Axiom of Choice, they are 'all in the mind' and cannot be specified in any explicit way. However, if we are to allow that they exist, then in all our later work we must be careful either to impose the hypothesis that the sets we deal with are measurable, or, where appropriate, to prove that they are. In more general spaces and for measures other than Lebesgue measure, there may be certain sets which are clearly non-measurable. This is the reason why the later part of the theory is framed by the rather complicated apparatus of σ-algebras and so on; we must take account of the possibility that some sets do not fit the theory.

§2. The extended real numbers.

In first-year tutorials, I have been in the habit of telling students "infinity is not a number", to discourage those who want to write things like $(5 \times \infty)/\infty = 5$. More exactly, what I have in mind is that, since the symbol ∞ cannot be related to any idea of counting or mensuration, it should not be subjected to the usual laws of arithmetic, which ultimately derive from such ideas. In most undergraduate courses, the only serious use we make of ∞ is as a symbol in expressions like "$a_n \to \infty$ as $n \to \infty$", whose meaning, as usually defined, does not have anything to do with a real object called ∞.

In measure theory, on the other hand, there is good reason to allow ∞ as a 'formal' value of some measures. (The word 'formal' in this context is mathematical jargon for 'having no natural interpretation'.) An attractive instance is that the area of the whole plane obviously "must" be ∞ if we want to speak of it at all. But, if we wish to allow infinite values, we also need to decide what arithmetical rules, if any, these infinite values will obey. It is not difficult to settle on a suitable list; I shall explain the motives behind it once I have given all the rules in question, but it is important to grasp that, ∞ being just a symbol, we are free to make our own rules for using it, and we choose rules that are appropriate to our purposes. They are not absolute rules established for ever. (For instance, the symbol ∞ is also used in complex analysis, where the appropriate rules are substantially different).

The use of ∞ and $-\infty$ is by no means a universal convention; in particular, Saks is reluctant to employ it. It has to be admitted that it often forces us to divide proofs into two cases, a trivial one where infinities occur and a serious one where they don't. Nevertheless, I think most people implicitly or explicitly follow the line I shall take, and, despite the minor complications it introduces, it does tend to simplify the statements of theorems.

So we agree to proceed as follows.

Definition 2.1. Let $\infty, \overline{\infty}$ be two elements different from each other and from all elements of $\mathbb{R}$. The set $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{\infty, \overline{\infty}\}$ is called the set of *extended real numbers*. We order $\overline{\mathbb{R}}$ by the relation $\ll$, defined by

$$\overline{\infty} \ll a \text{ and } a \ll \infty \quad \text{for any } a \in \mathbb{R}, \text{ and } \overline{\infty} \ll \infty,$$
$$\text{and, when } a, b \in \mathbb{R},\ a \ll b \text{ if and only if } a < b.$$

Then $\ll$ is a total order on $\overline{\mathbb{R}}$, such that $\overline{\infty}$ is the least element and ∞ the greatest.

Next, we define algebraic operations in $\overline{\mathbb{R}}$. For any $x \in \mathbb{R}$,

(i) $\infty + \infty \coloneqq \infty$, $\overline{\infty} + \overline{\infty} \coloneqq \overline{\infty}$, $\infty.\infty \coloneqq \infty$,
$\infty.\overline{\infty} \coloneqq \overline{\infty}$, $\overline{\infty}.\infty \coloneqq \overline{\infty}$, $\overline{\infty}.\overline{\infty} \coloneqq \infty$.

(ii) $x + \infty = \infty + x \coloneqq \infty$, $x + \overline{\infty} = \overline{\infty} + x \coloneqq \overline{\infty}$,
$x - \overline{\infty} = \infty - x \coloneqq \infty$, $x - \infty = \overline{\infty} - x \coloneqq \overline{\infty}$.

(iii) If $x > 0$, $x.\infty = \infty.x \coloneqq \infty$ and $x.\overline{\infty} = \overline{\infty}.x \coloneqq \overline{\infty}$;
if $x < 0$, $x.\infty = \infty.x \coloneqq \overline{\infty}$ and $x.\overline{\infty} = \overline{\infty}.x \coloneqq \infty$.

(iv) $0.\infty = \infty.0 \coloneqq 0$, $0.\overline{\infty} = \overline{\infty}.0 \coloneqq 0$.

Here I have listed the definitions of operations which involve infinities. For elements of $\mathbb{R}$ itself, which are called the "finite" elements of $\overline{\mathbb{R}}$, the algebraic operations are to be the usual

ones. There are various slightly different formulations of the rules; for instance, I have not listed the 'unary' operation of reversal of sign, which we might define as subtraction from 0, so that, by *(ii)*, $-\infty = \overline{\infty}$ and $-\overline{\infty} = \infty$. Others might take $x \mapsto -x$ as the primary operation, and *define* subtraction from it, as reversal of sign followed by addition.

Notice that the domain of the operation $+$ is not the whole of $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$, since we have not specified values for $\infty + \overline{\infty}$ and $\overline{\infty} + \infty$. Hence $\overline{\mathbb{R}}$ is not an abelian group with respect to $+$, or any other kind of familiar algebraic object.

The set $\overline{\mathbb{R}}$, with the order $\ll$ and the operations of addition, subtraction, and multiplication defined as above, is called the *extended real number system*. We usually write $<$ instead of $\ll$, and $-\infty$ instead of $\overline{\infty}$, and say that ∞ is positive or has positive sign, whilst $-\infty$ is negative or has negative sign. (In this section I shall often retain $\overline{\infty}$ for the sake of clarity). We say that ∞ and $\overline{\infty}$ are *opposite infinities*.

The rules *(i)–(iii)* are precisely what you expect on the basis of well-known theorems about limits of sequences. For instance, if $x_n \to x < 0$ (where x and each x_n are in $\mathbb{R}$) and $y_n \to \infty$ (where each y_n is in $\mathbb{R}$), then $x_n y_n \to -\infty$, and this indicates that we want $x.\infty = \overline{\infty}$. The notable absences from the list correspond to cases where no such theorem is true. As an example, if $x_n \to \infty$ and $y_n \to \infty$, nothing general can be said about the behaviour of $x_n - y_n$, so the list does not mention $\infty - \infty$; we say that $\infty - \infty$ is "undefined" or "does not make sense", in common with other binary expressions that are not explicitly given a value in the definition.

On the basis just suggested, the rules *(iv)* may seem rather surprising — there are no theorems about sequences to which they correspond; if $x_n \to 0$ and $y_n \to \infty$, no general conclusion can be drawn about the behaviour of $x_n y_n$. One would therefore expect that $0.\infty$ and $0.\overline{\infty}$ should be undefined. In the context of measure theory, however, the stated rules are natural, for a rectangle of length ∞ and height 0, such as the x-axis in $\mathbb{R}^2$, is expected to have zero two-dimensional area. We shall see that *(iv)* leads to results that are consistent with intuition.

It may also seem odd that there is no mention of division. Of course, if $x, y \in \mathbb{R}$ and $y \neq 0$, we know how to interpret $x \div y$, and $\infty \div y$ may be interpreted as $\infty.y^{-1}$; thus those divisions are already implicit. Furthermore, ∞/∞ and $x \div 0$ have no reasonable values. However, I have not listed "for any $x \in \mathbb{R}$, $x \div \infty = 0$ and $x \div \overline{\infty} = 0$" as rules. There is no deep reason for the omission; the proposed rules would lead to no logical difficulties; but they are also quite useless for our purposes.

Note 2.2. There is an important convention to which we shall adhere throughout. A statement of equality between two expressions involving extended real numbers is understood to include the assertion that either both expressions make sense or both are undefined. Thus, if I write

$$a = b + c\,,$$

this states both that $b + c$ makes sense and that its value is a. In detail, b and c cannot be opposite infinities; if they are both infinite, they are equal to each other and to a. If only one of them is infinite, it is equal to a.

Remark 2.3. Since $+$ and $-$ (and .) are binary operations, we cannot, in principle, write $a + b + c$ or $a.b.c$; brackets should be written in certain places, according to well-known rules of mathematical grammar, to indicate the order of carrying out the operations,

$(a+b)+c$ or $a+(b+c)$. However, it is clear after a moment's thought that, for any finite sequence $a_1, a_2, \ldots, a_r$ of extended real numbers,

(i) any grammatical method of inserting brackets in the product $a_1.a_2.\ldots.a_r$ will give an expression that makes sense, and its value will not depend on the bracketing. If all the a's are finite, this is just the usual statement of associativity of multiplication in $\mathbb{R}$. If any of the a's is 0, the product is 0. If none is 0 but some are infinite, the product is ∞ when the number of negative factors — including $\overline{\infty}$'s — is even, or $\overline{\infty}$ when the number of negative factors is odd. Also,

(ii) the sum $a_1 + a_2 + \ldots + a_r$ either makes sense in any grammatical bracketing (which occurs when all the terms a_i that are infinite, if there are any, have the same sign), and then the value of the bracketed sum does not depend on the bracketing; or it does not make sense in any grammatical bracketing, which occurs when and only when two of the terms are opposite infinities.

The general principle, then, is that we may omit brackets exactly as for sums in $\mathbb{R}$, the only added complication being that the sum may not make sense (no matter how it is bracketed). Multiplication is distributive over addition in the strongest possible sense:

$$a.(b+c) = a.b + a.c\,,$$

where, in accordance with the convention of 2.2, each side makes sense if and only if the other does. (Check the possible cases).

Lemma 2.4. *Every subset of $\overline{\mathbb{R}}$ has a least upper bound and a greatest lower bound.*

Proof. Let A be a non-null subset of $\overline{\mathbb{R}}$. Then A is bounded both above and below in $\overline{\mathbb{R}}$, for ∞ is an upper bound and $\overline{\infty}$ is a lower bound. If $\infty \notin A \neq \{\overline{\infty}\}$, and $A \cap \mathbb{R}$ is bounded above in $\mathbb{R}$, then Dedekind's axiom says $A \cap \mathbb{R} \neq \emptyset$ has a supremum in $\mathbb{R}$, which is clearly also the supremum of A in $\overline{\mathbb{R}}$ (the set of upper bounds of A in $\overline{\mathbb{R}}$ consists of the upper bounds of $A \cap \mathbb{R}$ in $\mathbb{R}$ plus ∞). If $A \cap \mathbb{R} \neq \emptyset$ but A is not bounded above in $\mathbb{R}$, its only upper bound in $\overline{\mathbb{R}}$ must be ∞, which is therefore its supremum. If $\infty \in A$, ∞ is trivially the supremum of A. The only remaining possibilities are that $A = \{\overline{\infty}\}$ or that $A = \emptyset$, when any element of $\overline{\mathbb{R}}$ is an upper bound and $\overline{\infty}$ is the supremum. There are corresponding arguments for infima. □

In particular, the supremum of $\emptyset$ is $\overline{\infty}$, whilst its infimum is ∞; hence, for the null set (and for no other), the infimum is larger than the supremum.

Lemma 2.5. *Let A be a non-empty subset of a totally ordered set T. An element $t \in T$ is the supremum of A in T if and only if it is an upper bound for A and, for any $s < t$, there exists some $a \in A$ such that $s < a$.*

Proof. If t is the supremum and $s < t$, then s is not an upper bound, and there is some element $a \in A$ not less than or equal to s. The order being total, $a > s$. Conversely, if t is an upper bound but not the supremum, then there is some $s < t$ which is also an upper bound, which contradicts the proposed condition. □

Of course this Lemma really repeats 0.6, except for the use of subtraction.

If (x_n) is a sequence in a partially ordered set X, we describe it as *increasing* if $x_n \leq x_{n+1}$ for all n; as *strictly increasing* if $x_n \leq x_{n+1} \neq x_n$ for all n; as *decreasing* if

$x_n \geq x_{n+1}$ for all n; as *strictly decreasing* if $x_n \geq x_{n+1} \neq x_n$ for all n; as *monotonic* [or *strictly monotonic*] if it is either increasing [or strictly increasing] or decreasing [or strictly decreasing].

Many authors write "monotonic increasing" etc., instead of just "increasing".

Definition 2.6. Let (a_n) be a sequence in $\overline{\mathbb{R}}$. We define the convergence of such a sequence to a limit in $\overline{\mathbb{R}}$ as follows.

(i) $a_n \to \infty$ if, for any $K \in \mathbb{R}$, there exists $N \in \mathbb{N}$ such that $a_n \geq K$ whenever $n \geq N$. [This is exactly the usual definition; but, of course, a_n is now allowed to take the values ∞ and $\overline{\infty}$. If $a_n \to \infty$, there can be only finitely many indices n for which $a_n = \overline{\infty}$, and, if $\overline{\infty} < a_n < \infty$ for infinitely many indices n, the subsequence given by those indices tends to ∞.]

(ii) $a_n \to \overline{\infty}$ as $n \to \infty$ if, for any $K \in \mathbb{R}$, there exists $N \in \mathbb{N}$ such that $a_n \leq K$ whenever $n \geq N$. [It follows that $a_n \to \overline{\infty}$ if and only if $-a_n \to \infty$.]

(iii) If $a \in \mathbb{R}$, then $a_n \to a$ as $n \to \infty$ if, for any positive real number ϵ, there exists $N \in \mathbb{Z}$ such that $|a_n - a| < \epsilon$ whenever $n \geq N$. [In particular, this means that a_n must be finite except for a finite number of values of n.]

The *ad hoc* definitions of the various kinds of convergence given above are exactly the standard ones, with the one extension that a_n is allowed to take infinite values. In fact, the definition 2.6 can be derived as convergence with respect to a suitable metric on $\overline{\mathbb{R}}$. One such metric is ρ, where, for any $x, y \in \mathbb{R}$,

$$\begin{aligned}
\rho(x,y) = \rho(y,x) &= \left| \frac{x}{1+|x|} - \frac{y}{1+|y|} \right|, \\
\rho(x,\infty) = \rho(\infty,x) &= \left| \frac{x}{1+|x|} - 1 \right|, \\
\rho(x,\overline{\infty}) = \rho(\overline{\infty},x) &= \left| \frac{x}{1+|x|} + 1 \right|, \\
\rho(\infty,\overline{\infty}) = \rho(\overline{\infty},\infty) &= 2\,.
\end{aligned}$$

(Another possible choice of metric is $\rho'(x,y) := |\tan^{-1}x - \tan^{-1}y|$, with

$$\rho'(\infty,x) = \rho'(x,\infty) = \left|\tfrac{1}{2}\pi - \tan^{-1}x\right|, \quad \rho'(\overline{\infty},x) = \rho'(x,\overline{\infty}) = \left|\tfrac{1}{2}\pi + \tan^{-1}x\right|.$$

However, this definition uses the transcendental function $\tan^{-1}$ instead of the elementary functions used above.) In effect, this metric results from mapping $\overline{\mathbb{R}}$ on to the closed interval $[-1,1]$ by the mapping ϕ which takes ∞ into 1, $x \in \mathbb{R}$ into $x/(1+|x|)$, and $\overline{\infty}$ into -1. Then, for any $\alpha, \beta \in \overline{\mathbb{R}}$, $\rho(\alpha,\beta) = |\phi(\alpha) - \phi(\beta)|$.

Throughout the course I shall tend, as above, to omit the phrase "as $n \to \infty$", at least when it is obvious that the convergence in question must be as $n \to \infty$. I shall have some deeper remarks about Definition 2.6 later.

Lemma 2.7. *Any convergent sequence in $\mathbb{R}$ is bounded both above and below. An increasing sequence in $\mathbb{R}$ converges in $\mathbb{R}$ if and only if its terms are bounded above, and its limit in that case is the supremum. Any increasing sequence in $\overline{\mathbb{R}}$ converges in $\overline{\mathbb{R}}$ to its supremum. In either case, if the original increasing sequence converges, any infinite subsequence also converges to the same limit; if the original increasing sequence (in $\mathbb{R}$) does not converge in $\mathbb{R}$, then neither does any subsequence.*

Proof. Suppose (y_n) is a sequence in $\mathbb{R}$ convergent to y. Then, taking "ϵ" to be 1 in Definition 0.10, there exists $N \in \mathbb{N}$ such that $n \geq N \Longrightarrow y - 1 < y_n < y + 1$. Therefore,

$$(\forall n \in \mathbb{N}) \quad \min(y-1, y_1, y_2, \ldots, y_{N-1}) \leq y_n \leq \max(y+1, y_1, y_2, \ldots, y_{N-1}),$$

which means that the sequence has both a lower and an upper bound in $\mathbb{R}$.

Now, let (x_n) be increasing in $\mathbb{R}$ and bounded above in $\mathbb{R}$, with supremum x. Given $\epsilon > 0$, Lemma 0.6 shows there exists some N for which $x_N > x - \epsilon$. Whenever $n \geq N$,

$$x - \epsilon < x_N \leq x_n \leq x$$

(since $x_n \geq x_N$ because the sequence is increasing and $x_n \leq x$ because x is an upper bound). Hence $n \geq N \Longrightarrow |x_n - x| < x$, as required.

Let (z_n) be increasing in $\overline{\mathbb{R}}$. It has a supremum q in $\overline{\mathbb{R}}$. There are three cases. If $q = \overline{\infty}$, then $z_n = \overline{\infty}$ for all n, so $z_n \to \overline{\infty}$. If $\overline{\infty} < q < \infty$, then, by 2.5, there is some $M \in \mathbb{N}$ such that $z_n > \overline{\infty}$ for $n \geq M$; so $(z_n)_{n \geq M}$ is an increasing sequence bounded above in $\mathbb{R}$, with supremum q; thus it converges in $\mathbb{R}$ to q. Recall 2.6*(iii)*. If $q = \infty$, then, for any $K \in \mathbb{R}$, there exists L such that $x_L > K$, by 2.5; hence $z_n > K$ for $n \geq L$, so that 2.6*(i)* is satisfied.

Finally, notice that the set of upper bounds for the whole sequence is the same as the set of upper bounds for any infinite subsequence, so they have the same supremum. □

If you know about nets, you can formulate and prove a version of the above argument for nets instead of sequences. This makes the discussion of unordered summation a little shorter.

Definition 2.8. Let $(x_n)_{n=1}^{\infty}$ be a sequence in $\overline{\mathbb{R}}$. Then the series $\sum_{n=1}^{\infty} x_n$ *converges to the sum* $X \in \overline{\mathbb{R}}$, which is expressed by the expression $X = \sum_{n=1}^{\infty} x_n$, if each partial sum $\sigma_p := \sum_{n=1}^{p} x_n$ is defined in $\overline{\mathbb{R}}$ (for $p = 1, 2, \ldots$), and the sequence (σ_p) converges in $\overline{\mathbb{R}}$ to X. We say that the series $\sum_{n=1}^{\infty} x_n$ *converges* in $\overline{\mathbb{R}}$ if there is some element $X \in \overline{\mathbb{R}}$ such that the series converges to the sum X.

If $(x_n)_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$, the series $\sum_{n=1}^{\infty} x_n$ *converges to the sum* $X \in \mathbb{R}$, and one writes $X = \sum_{n=1}^{\infty} x_n$, if the sequence (σ_p) of partial sums of the series converges in $\mathbb{R}$ to X. We say that the series $\sum_{n=1}^{\infty} x_n$ *converges* in $\mathbb{R}$ if there is some element $X \in \mathbb{R}$ such that the series converges to the sum X.

The most important difference between series in $\mathbb{R}$ and in $\overline{\mathbb{R}}$ is that in $\overline{\mathbb{R}}$ there is the possibility that partial sums may not be defined. This only occurs if both ∞ and $\overline{\infty}$ occur as terms of the series. In fact, if only one infinite value appears (possibly repeatedly) as a term of the series, then the series sums to that infinite value. If, on the other hand, all terms of the series are finite, then it converges in $\overline{\mathbb{R}}$ if the sequence of partial sums converges in $\overline{\mathbb{R}}$.

As I warned you, the word "convergence" has a different meaning for series from the previous one in Definition 0.10. Perhaps for this reason, students tend to have difficulty with it, although there is another reason: the whole concept of a "series" is odd, and, I think, is rarely defined in precise language. One might say, as one of several possible definitions not usually given, that a series *is* just a sequence whose terms are specified indirectly as partial sums of a related sequence. It would be tedious to keep repeating such an explanation, and so we write something like "the series $\sum_{n=1}^{\infty} x_n$" as a shorthand reminder that the series is the sequence of partial sums of the sequence (x_n). The series converges if $\sum_{n=1}^{k} x_n$ (which is its kth term) has a limit as $k \to \infty$. But we also want a notation for this limit when it exists, and it is unfortunate that the natural choice is $\sum_{n=1}^{\infty} x_n$, which thus comes to denote both the

series itself and the limit; and the limit is equally naturally called the sum. It would be possible to avoid these ambiguities by writing $\sum x_n$ for the series and $\sum_{n=1}^{\infty} x_n$ for its "sum", but no-one consistently does so that I know of.

Since we need partial sums, it is only reasonable to talk of series whose terms belong to some object in which they may be added. Although, for the moment, we are only interested in $\mathbb{R}$ or $\overline{\mathbb{R}}$, one can also have series in $\mathbb{C}$, in $\mathbb{R}^n$ or $\mathbb{C}^n$, or in a more general normed space.

Definition 2.9. The series $\sum_{n=1}^{\infty} y_n$ is a *rearrangement* of the series $\sum_{n=1}^{\infty} x_n$ if there is a one-to-one and onto mapping $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ such that $y_n = x_{\sigma(n)}$ for each n.

A one-to-one and onto mapping is often called a *bijection*.

There are fairly obvious modifications of Definitions 0.10, 2.6, 2.8, 2.9 that apply when the sequence or series is indexed by $\mathbb{N} \cup \{0\}$ or an infinite subset thereof, and I shall use them without comment.

Definition 2.10. A series $\sum x_n$ in $\mathbb{R}$ [or in $\overline{\mathbb{R}}$, or more generally] is *unconditionally convergent in* $\mathbb{R}$ [or in $\overline{\mathbb{R}}$, etc] if all rearrangements of $\sum x_n$ converge in $\mathbb{R}$ [or in $\overline{\mathbb{R}}$, etc] to the same sum. A series $\sum x_n$ in $\mathbb{R}$ [or in a normed space] is *absolutely convergent* if $\sum_{n=1}^{\infty} |x_n|$ [or $\sum_{n=1}^{\infty} \|x_n\|$] converges in $\mathbb{R}$.

Definition 2.11. Let S be a set and $f : S \longrightarrow \mathbb{R}$ a function. f has the *unordered sum* $x \in \mathbb{R}$ if, for any $\epsilon > 0$, there exists a finite subset F of S with the property that, whenever G is a finite subset of S including F, $|x - \sum_{s \in G} f(s)| < \epsilon$.

Lemma 2.12. *Let $f, g : S \longrightarrow \mathbb{R}$ be functions which have the unordered sums $x, y \in \mathbb{R}$. Then the function $f + g$ has the unordered sum $x + y$.*

Proof. Given $\epsilon > 0$, there are finite subsets F_1, F_2 of S such that, for any finite subset G of S, $G \supseteq F_1 \Longrightarrow |x - \sum_{s \in G} f(s)| < \frac{1}{2}\epsilon$ and $G \supseteq F_2 \Longrightarrow |y - \sum_{s \in G} g(s)| < \frac{1}{2}\epsilon$. Take $F := F_1 \cup F_2$; then $G \supseteq F \Longrightarrow |(x + y) - \sum_{s \in G}(f(s) + g(s))| < \epsilon$. □

Recall that a real sequence (x_n) is just a function $x : \mathbb{N} \longrightarrow \mathbb{R}$. I shall say that a series $\sum x_n$ in $\mathbb{R}$ is *unorderedly convergent in* $\mathbb{R}$ if the sequence (x_n) has an unordered sum in $\mathbb{R}$.

Lemma 2.13. *If the series $\sum x_n$ in $\mathbb{R}$ is unorderedly convergent with sum x in $\mathbb{R}$, it is unconditionally convergent with sum x.*

Proof. Given $\epsilon > 0$, there is a finite subset F of $\mathbb{N}$ such that, whenever G is a finite subset including F, $|x - \sum_{n \in G} x_n| < \epsilon$. Take $N := \max F \in \mathbb{N}$. If $n \geq N$, the finite set $\{1, 2, \ldots, n\}$ includes F, and so $|x - \sum_{k=1}^{n} x_n| < \epsilon$. Thus the series is convergent in the given ordering. However, unordered convergence and the unordered sum are clearly unaffected by rearrangement. □

The converse is also true, but is left as an *exercise*.

Theorem 2.14. *A series in $\mathbb{R}$ whose terms are all non-negative is convergent in $\mathbb{R}$ if and only if its partial sums are bounded above; in that case it is unorderedly convergent, and so unconditionally convergent, in $\mathbb{R}$. The sum is strictly decreased if the values of any non-null*

set of non-zero terms are strictly diminished. Any series in $\overline{\mathbb{R}}$ whose terms are all non-negative is unconditionally convergent in $\overline{\mathbb{R}}$.

Proof. For a series (x_n) of non-negative terms in $\overline{\mathbb{R}}$, all partial sums are defined. The sequence of partial sums increases; apply Lemma 2.7 to get the assertions about convergence.

Suppose B is an upper bound in $\mathbb{R}$ for the partial sums. For any finite $F \subseteq \mathbb{N}$,

$$\sum\nolimits_{k\in F} x_k \leq \sum\nolimits_{k=1}^{\max F} x_k \leq B$$

since all x_k are non-negative. Thus B is also an upper bound for the finite sums $\sum_{k\in F} x_k$. Conversely, any upper bound for the finite sums must also be an upper bound for the partial sums, which are a special kind of finite sum. Let Λ be the common least upper bound. For any $\epsilon > 0$, there must be some finite $F \subseteq \mathbb{N}$ such that $\sum_{k\in F} x_k > \Lambda - \epsilon$, since otherwise $\Lambda - \epsilon$ would be an upper bound; but then, if $G \subseteq \mathbb{N}$ is also finite and $G \supseteq F$,

$$\Lambda \geq \sum\nolimits_{k\in G} x_k \geq \sum\nolimits_{k\in F} x_k > \Lambda - \epsilon\,,$$

so that $|\Lambda - \sum_{k\in G} x_k| < \epsilon$. This proves unordered convergence to Λ. Apply 2.13.

The effect of changing any non-zero term x_p to $x'_p \in (0, x_p)$ is to reduce all the later partial sums by at least $x_p - x'_p$. Thus $\Lambda - (x_p - x'_p)$ will be an upper bound for all the partial sums, and their supremum cannot exceed it.

The case remaining is where the partial sums of the original series have supremum ∞ in $\overline{\mathbb{R}}$. For any $K \in \mathbb{R}$, there must be a partial sum $\sum_{k=1}^{m} x_k > K$. If $\sum x_{\sigma(k)}$ is a rearrangement of the series, choose $N := \max\{\sigma^{-1}(1), \sigma^{-1}(2), \ldots, \sigma^{-1}(m)\}$; then the indices $1, 2, \ldots, m$ all crop up among $\sigma(1), \sigma(2), \ldots, \sigma(N)$, and, whenever $n \geq N$,

$$\sum\nolimits_{k=1}^{n} x_{\sigma(k)} \geq \sum\nolimits_{k=1}^{N} x_{\sigma(k)} \geq \sum\nolimits_{k=1}^{m} x_k > K\,.$$

This shows that the rearranged series also converges to ∞. □

The last paragraph of the proof could be substituted by an argument from 2.13, if I had done 2.13 for series in $\overline{\mathbb{R}}$. In fact, unordered convergence is really a form of convergence of nets, as I remarked after 2.7.

The result I really need is a rather more general version of the above. For series of non-negative terms in $\overline{\mathbb{R}}$, the sum is unaffected not only by rearrangement, as just shown, but also by "grouping of terms". It is also unaffected by the insertion or removal of any number of terms whose value is 0; I leave it to you to work out precisely what that means, and to prove it. To make the statement true without exception, one needs to give their customary meaning to finite sums and to regard the "empty sum" (the sum of no terms) as 0.

Theorem 2.15. *Suppose that, for each $i \in \mathbb{N}$, $(x_{ni})_{n=1}^{\infty}$ is a sequence of non-negative terms in $\overline{\mathbb{R}}$. Let $q : \mathbb{N} \longrightarrow \mathbb{N} \times \mathbb{N}$ be a one-to-one correspondence, and set $y_j = x_{q(j)}$. Then*

$$\sum\nolimits_{j=1}^{\infty} y_j = \sum\nolimits_{i=1}^{\infty} \Big(\sum\nolimits_{k=1}^{\infty} x_{ki} \Big).$$

Since all the series have only non-negative terms, they converge unconditionally in $\overline{\mathbb{R}}$.

Proof. Let us suppose first that $\sum_{k=1}^{\infty} x_{ki} < \infty$ for each i. Consider a "finite grouping" $\sum_{i=1}^{\infty}\left(\sum_{k=1}^{n(i)} x_{ki}\right)$. The partial sums of this series form a subsequence of the partial sums of the series $x_{11} + x_{21} + \cdots + x_{n(1),1} + x_{12} + x_{22} + \cdots + x_{n(2),2} + x_{13} + \cdots$, which is a rearrangement of $\sum y_j$ with certain terms reduced to 0. By 2.14, then,

$$\sum_{j=1}^{\infty} y_j \geq \sum_{i=1}^{\infty}\left(\sum_{k=1}^{n(i)} x_{ki}\right).$$

Take any $K < \sum_{j=1}^{\infty} y_j$. Then there exists some N such that $\sum_{j=1}^{N} y_j > K$. Each y_j is an x_{ki} for some suitable choice of k and i; thus

$$K < \sum_{j=1}^{N} y_j \leq \sum_{i=1}^{M}\left(\sum_{k=1}^{n(i)} x_{ki}\right) \leq \sum_{i=1}^{\infty}\left(\sum_{k=1}^{n(i)} x_{ki}\right).$$

for some suitable $M \in \mathbb{N}$ and some suitable $N(i)$ for each i between 1 and M. By 2.14 again, $\sum_{i=1}^{\infty}\left(\sum_{k=1}^{n(i)} x_{ki}\right) \leq \sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} x_{ki}\right)$. Hence $K < \sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} x_{ki}\right)$, which implies that $\sum_{j=1}^{\infty} y_j \leq \sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} x_{ki}\right)$; otherwise take $K := \sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} x_{ki}\right)$ to get a contradiction.

Now consider $\sum_{i=1}^{L}(\sum_{k=1}^{\infty} x_{ki})$. If all the internal series converge in $\mathbb{R}$, it is a familiar fact that $\sum_{i=1}^{L}(\sum_{k=1}^{\infty} x_{ki}) = \sum_{k=1}^{\infty}\left(\sum_{i=1}^{L} x_{ki}\right)$, and this remains true even if one of the internal series sums to ∞, by 2.14. The partial sums of this last series form a subsequence of the partial sums of a rearrangement of $\sum y_j$ with some terms reduced to 0; hence, from 2.14,

$$\sum_{i=1}^{L}\left(\sum_{k=1}^{\infty} x_{ki}\right) \leq \sum y_j,$$

which is, therefore, an upper bound for the partial sums of $\sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} x_{ki}\right)$. So

$$\sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} x_{ki}\right) \leq \sum y_j.$$

This completes the proof of the required equality. □

The proof could probably be presented more economically, but most authors seem to treat it as obvious and assume it without comment.

We can now begin the serious development.

§3. Outer measures.

In §0C, I introduced the ideas of 'outer Jordan content' and of 'inner Jordan content' in $\mathbb{R}^2$. They are based on the area of polygons, and a moment's thought will convince you that the outer Jordan content of a bounded set E could be defined as the infimum of the numbers obtained as the sums of the areas of finite systems of triangles covering E. The idea of Lebesgue was to use *countable* systems of triangles instead — or rather countable systems of rectangles, which after a good deal of work can be seen to give the same answer. Subsequently, Carathéodory observed that only the *outer* construction is really necessary.

Definition 3.1. Let Ω be any set. The *power set* $\mathcal{P}(\Omega)$ of Ω is the set whose members are all the subsets of Ω (including $\emptyset$ and Ω itself): $\mathcal{P}(\Omega) := \{E : E \subseteq \Omega\}$.

Logicians sometimes call the power set 2^Ω. Topologists tend to use 2^Ω to denote the class of *closed* subsets of a topological space, so I prefer $\mathcal{P}(\Omega)$. Incidentally, that $\mathcal{P}(\Omega)$ is also a set is an axiom of formal set theory.

Definition 3.2. Given the set Ω, a *weighting function* in Ω is a mapping $\tau : \mathcal{C} \longrightarrow \overline{\mathbb{R}}$, where $\mathcal{C}$ is a subset of $\mathcal{P}(\Omega)$ containing $\emptyset$ and

$$\textit{(a)} \quad \tau(\emptyset) = 0\,, \qquad \textit{(b)} \quad (\forall C \in \mathcal{C})\ \ \tau(C) \geq 0\,.$$

For instance, $\mathcal{C}$ might be the class of triangles and τ the area. I may call $\tau(C)$ the *weight* of C, but neither this nor my name "weighting function" for τ are in common use.

Definition 3.3. If $\tau : \mathcal{C} \longrightarrow \overline{\mathbb{R}}$ is a weighting function in Ω, and $E \in \mathcal{P}(\Omega)$, let

$$\tau^\dagger(E) := \inf\Big\{ \textstyle\sum_{k=1}^\infty \tau(C_k) \ :\ (C_k)_{k=1}^\infty \text{ is a sequence in } \mathcal{C} \text{ and } E \subseteq \bigcup_{k=1}^\infty C_k \Big\} \in \overline{\mathbb{R}}\,.$$

In words, $\tau^\dagger(E)$ is the infimum (in $\overline{\mathbb{R}}$) of the sums of the weights of sequences in $\mathcal{C}$ that cover E. The set whose infimum is taken will commonly be an uncountable class (there being often uncountably many ways of covering E by sequences of members of $\mathcal{C}$) of non-negative extended real numbers. It may well be the case, however, that E *cannot* be covered by countably many sets from $\mathcal{C}$ — $\mathcal{C}$ might, for instance, consist entirely of countable sets, although E is uncountable. In that case $\tau^\dagger(E) = \infty$, as $\inf \emptyset = \infty$. That understood, $\tau^\dagger$ becomes a function $\mathcal{P}(\Omega) \longrightarrow \overline{\mathbb{R}}$.

In general, there is no reason to suppose that the function $\tau^\dagger$ resembles τ in any significant way. It is easy, for instance, to give examples where $\tau^\dagger$ is always zero when τ is not. The really important thing is that $\tau^\dagger(E)$ is defined for *any* $E \subseteq \Omega$.

Lemma 3.4. *Given a weighting function τ in Ω, $\tau^\dagger$ has the properties*

(a) $\tau^\dagger(\emptyset) = 0$,
(b) *if* $M, N \in \mathcal{P}(\Omega)$ *and* $M \subseteq N$, *then* $\tau^\dagger(M) \le \tau^\dagger(N)$,
(c) *for any sequence* $(E_n)_{n=1}^\infty$ *in* $\mathcal{P}(\Omega)$,

$$\tau^\dagger\Big(\bigcup_{n=1}^\infty E_n\Big) \le \sum_{n=1}^\infty \tau^\dagger(E_n)\,.$$

Proof. *(a)* and *(b)* are trivial. For *(c)*, there are two cases. If $\sum_{n=1}^\infty \tau^\dagger(E_n) = \infty$, there is nothing to prove. Suppose that $\sum_{n=1}^\infty \tau^\dagger(E_n) < \infty$, and, consequently, $\tau^\dagger(E_n) < \infty$ for each n. Let $\epsilon > 0$. For each n, there exists by 0.7 a sequence $(C_{ni})_{i=1}^\infty$ in $\mathcal{C}$ such that $E_n \subseteq \bigcup_{i=1}^\infty C_{ni}$ and

$$\tau^\dagger(E_n) \le \sum_{i=1}^\infty \tau(C_{ni}) \le \tau^\dagger(E_n) + 2^{-n}\epsilon\,, \quad \text{and so}$$
$$\sum_{n=1}^\infty \Big(\sum_{i=1}^\infty \tau(C_{ni})\Big) \le \sum_{n=1}^\infty (\tau^\dagger(E_n) + 2^{-n}\epsilon) = \Big(\sum_{n=1}^\infty \tau^\dagger(E_n)\Big) + \epsilon\,. \tag{9}$$

Rearrange the double sequence (C_{ni}) as a single sequence (D_j). Then

$$\bigcup_{n=1}^\infty E_n \subseteq \bigcup_{n=1}^\infty \Big(\bigcup_{i=1}^\infty C_{ni}\Big) = \bigcup_{n,i=1}^\infty C_{ni} = \bigcup_{j=1}^\infty D_j \quad \text{and}$$
$$\tau^\dagger\Big(\bigcup_{n=1}^\infty E_n\Big) \le \sum_{j=1}^\infty \tau(D_j) \quad \text{by definition.}$$

By 2.15, $\sum_{j=1}^\infty \tau(D_j) = \sum_{n=1}^\infty (\sum_{i=1}^\infty \tau(C_{ni}))$. Hence, by (9),

$$\tau^\dagger\Big(\bigcup_{n=1}^\infty E_n\Big) \le \Big(\sum_{n=1}^\infty \tau^\dagger(E_n)\Big) + \epsilon\,. \tag{10}$$

But $\sum_{n=1}^\infty \tau^\dagger(E_n) < \infty$, and ϵ is arbitrary; hence, in fact,

$$\tau^\dagger\Big(\bigcup_{n=1}^\infty E_n\Big) \le \sum_{n=1}^\infty \tau^\dagger(E_n)\,.$$

(If this were untrue, we could take $\epsilon := \frac{1}{2}\big(\tau^\dagger(\bigcup_{n=1}^\infty E_n) - \big(\sum_{n=1}^\infty \tau^\dagger(E_n)\big)\big) > 0$ and get a contradiction to (10)). □

It appears at first sight that from τ, whose properties were feeble in the extreme, we have constructed a function $\tau^\dagger$ which is not only defined for all subsets of Ω but also has far stronger properties. However, we are not really getting something for nothing; the problem is that $\tau^\dagger$ may be quite mysterious or quite uninteresting. But let us continue.

Definition 3.5. An *outer measure* in a set Ω is a function $\mu^* : \mathcal{P}(\Omega) \longrightarrow \overline{\mathbb{R}}$ such that

(a) $\mu^*(\emptyset) = 0$,
(b) if $M, N \in \mathcal{P}(\Omega)$ and $M \subseteq N$, then $\mu^*(M) \le \mu^*(N)$,
(c) if (M_n) is any sequence in $\mathcal{P}(\Omega)$, then

$$\mu^*\Big(\bigcup_{n=1}^\infty M_n\Big) \le \sum_{n=1}^\infty \mu^*(M_n)\,.$$

Notice that *(a)* and *(b)* ensure $\mu^*(M) \geq 0$ for all $M \subseteq \Omega$, so that the sum in *(c)* must make sense. *(b)* is sometimes expressed by the statement that μ^* is *nondecreasing* (as a function on $\mathcal{P}(\Omega)$). *(c)* says that μ^* is *countably subadditive*.

3.4 proves that $\tau^\dagger$ is an outer measure.

3.3 is the standard elementary construction of an outer measure. The question arises whether every outer measure can arise in this way, or whether outer measures constructed from weighting functions are somehow special. The answer is uninteresting:

Lemma 3.6. *If μ^* is an outer measure in Ω, $\mathcal{C} \coloneqq \mathcal{P}(\Omega)$, and $\tau \coloneqq \mu^*$, then $\tau^\dagger = \mu^*$.*

Proof. If $E \subseteq \Omega$, and (E_n) is a sequence in $\mathcal{P}(\Omega)$ such that $E \subseteq \bigcup_{n=1}^\infty E_n$, then

$$\mu^*(E) \leq \sum\nolimits_{n=1}^\infty \mu^*(E_n) = \sum\nolimits_{n=1}^\infty \tau(E_n)$$

by 3.5*(c)*. So $\mu^*(E)$ is a lower bound for the sums that define $\tau^\dagger(E)$ (see 3.3), and $\mu^*(E) \leq \tau^\dagger(E)$. On the other hand, the sequence $E, \emptyset, \emptyset, \ldots$ covers E, and thus

$$\tau^\dagger(E) \leq \mu^*(E) + \mu^*(\emptyset) + \mu^*(\emptyset) + \cdots = \mu^*(E) + 0 + 0 + \cdots = \mu^*(E)\,.$$

Putting the two inequalities together, $\tau^\dagger(E) = \mu^*(E)$, as required. □

Lemma 3.7. (a) *Suppose that $(\mu_\alpha^*)_{\alpha \in A}$ is any family of outer measures in Ω. Define*

$$(\forall E \subseteq \Omega) \quad \mu^*(E) \coloneqq \sup\{\mu_\alpha^*(E) : \alpha \in A\}\,.$$

Then μ^* is an outer measure in Ω.

(b) *If μ_1^*, μ_2^* are outer measures in Ω, so is $\mu_1^* + \mu_2^*$, defined by*

$$(\forall E \subseteq \Omega) \quad (\mu_1^* + \mu_2^*)(E) \coloneqq \mu_1^*(E) + \mu_2^*(E)\,.$$

(c) *If $(\mu_k^*)_{k=1}^\infty$ is a sequence of outer measures in Ω, define*

$$(\forall E \subseteq \Omega) \quad \mu^*(E) \coloneqq \sum\nolimits_{k=1}^\infty \mu_k^*(E)\,.$$

Then, again, μ^* is an outer measure in Ω. □

These statements are not really remarkable, for the conditions 3.5 are not very demanding, and there are many uninteresting outer measures. For instance, the zero function is an outer measure, as is the function which is 0 on the null set and ∞ on every other set.

§4. Sets measurable with respect to an outer measure.

The substance of this section was invented by Carathéodory; as I remarked in §1A, he pointed out that a satisfactory theory does not need "inner measure". The essential step is Definition 4.1, which is both unexpected and states an extremely demanding condition.

Definition 4.1. Let μ^* be an outer measure in the set Ω. A set $M \in \mathcal{P}(\Omega)$ is said to be *measurable with respect to μ^**, or *μ^*-measurable*, if, for every $A \in \mathcal{P}(\Omega)$,

$$\mu^*(A) = \mu^*(A \setminus M) + \mu^*(A \cap M)\,. \tag{11}$$

μ^*-measurability of M is thus not an "internal" property of M, but rather describes how M behaves in society: it splits *every* set $A \subseteq \Omega$ "additively". I may occasionally refer to A in (11) as a "test set". In the interesting cases, $\mathcal{P}(\Omega)$ has very many members, so 4.1 is, in principle, very unlikely to be true. The surprise is that there are interesting outer measures for which there are many measurable sets.

On the whole I shall write $\Omega \setminus E$ for the complement of E in Ω. Other notations that are in use are E^c and $\mathcal{C}E$. Both of them assume the set Ω is fixed.

There is also a point of vocabulary. For the rest of this section, only one outer measure μ^* will be considered. In such a situation, one might well write "measurable" instead of μ^*-measurable; and in many books, especially older ones by authors who were brought up on Lebesgue's original theory in $\mathbb{R}^n$, the word "measurable" always means "measurable with respect to Lebesgue outer measure" (which we shall construct later) and "outer measure" always means Lebesgue outer measure. Unfortunately, there is a more modern usage we shall soon meet, and to avoid confusion I shall, therefore, retain the μ^*.

Lemma 4.2. *Given the outer measure μ^* in Ω,*

(a) *$\emptyset$ is μ^*-measurable,*
(b) *if $E \in \mathcal{P}(\Omega)$ is μ^*-measurable, so is $\Omega \setminus E$.*

Proof. *(a)* holds as $\mu^*(\emptyset) = 0$, and *(b)* since (11) is symmetrical between E and $\Omega \setminus E$. □

Remark 4.3. In fact, if $E \in \mathcal{P}(\Omega)$ and $\mu^*(E) = 0$, then E is μ^*-measurable. Given a test set A, $\mu^*(A \cap E) \leq \mu^*(E) = 0$. But $\mu^*(A) \leq \mu^*(A \setminus E) + \mu^*(A \cap E)$ by 3.5*(a)*, *(c)*, whilst $\mu^*(A \setminus E) \leq \mu^*(A)$ by 3.5*(b)*. So 4.1 is satisfied.

Lemma 4.4. *Given the outer measure μ^* in Ω, let $E_1, E_2, \ldots, E_n$ be μ^*-measurable subsets of Ω. Then $\bigcup_{k=1}^n E_k$ is also μ^*-measurable.*

Proof. It will suffice to show that, if E and F are μ^*-measurable, so is $E \cup F$; the result will follow by induction. Let $A \subseteq \Omega$ be a test set. As E is μ^*-measurable,

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E) \qquad \text{and} \tag{12}$$

$$\begin{aligned} \mu^*(A \cap (E \cup F)) &= \mu^*(A \cap (E \cup F) \cap E) + \mu^*((A \cap (E \cup F)) \setminus E) \\ &= \mu^*(A \cap E) + \mu^*(A \cap (F \setminus E)), \end{aligned} \tag{13}$$

the latter equalities arising from the test set $A \cap (E \cup F)$. However, F is also μ^*-measurable, and so, taking $A \setminus E$ as a test set,

$$\begin{aligned} \mu^*(A \setminus E) &= \mu^*((A \setminus E) \cap F) + \mu^*((A \setminus E) \setminus F) \\ &= \mu^*(A \cap (F \setminus E)) + \mu^*(A \setminus (E \cup F)). \end{aligned} \tag{14}$$

Adding $\mu^*(A \cap (F \setminus E))$ to both sides of (12), and applying (13) and (14) in succession,

$$\begin{aligned} \mu^*(A) + \mu^*(A \cap (F \setminus E)) &= \mu^*(A \cap E) + \mu^*(A \setminus E) + \mu^*(A \cap (F \setminus E)) \\ &= \mu^*(A \cap (E \cup F)) + \mu^*(A \setminus E) \\ &= \mu^*(A \cap (E \cup F)) + \mu^*(A \cap (F \setminus E)) + \mu^*(A \setminus (E \cup F)). \end{aligned} \tag{15}$$

We wish to "cancel $\mu^*(A \cap (F \setminus E))$". However, there may be infinite values, so a little care

is needed. If $\mu^*(A \cap (F \setminus E)) = \infty$, then 3.5*(b)* implies that $\mu^*(A \cap (E \cup F)) = \infty$ and $\mu^*(A) = \infty$. On the other hand, if $\mu^*(A \cap (F \setminus E)) < \infty$, either both sides of (15) are infinite, in which case omitting $\mu^*(A \cap (F \setminus E))$ on either side will still leave ∞, or both sides are finite, when we may apply subtraction in $\mathbb{R}$. In all cases, then,

$$\mu^*(A) = \mu^*(A \cap (E \cup F)) + \mu^*(A \setminus (E \cup F)) .$$ □

Lemma 4.5. *If E and F are μ^*-measurable subsets of Ω, so are $E \cap F$ and $E \setminus F$.*

Proof. By 4.2*(b)*, $\Omega \setminus E$ and $\Omega \setminus F$ are μ^*-measurable; by 4.4, so is their union. By 4.2*(b)* again, $E \cap F = \Omega \setminus ((\Omega \setminus E) \cup (\Omega \setminus F))$ is μ^*-measurable. And $E \setminus F = E \cap (\Omega \setminus F)$ is μ^*-measurable in the same way. □

Notice that 3.5*(a)* was used in 4.2*(a)*, and 3.5*(b)* in 4.4. The next Lemma, however, depends on 3.5*(c)*. It is a curious result, stating a property of μ^*-measurable sets that we shall never need in its full strength.

Lemma 4.6. *Let (E_n) be a disjoint sequence of μ^*-measurable sets in Ω, and let $A \subseteq \Omega$. Set $E := \bigcup_{n=1}^{\infty} E_n$. Then*

$$\mu^*(A \cap E) = \sum\nolimits_{n=1}^{\infty} \mu^*(A \cap E_n) .$$

Proof. Let $F_k := \bigcup_{n=1}^{k} E_n$, for $k = 1, 2, 3, \ldots$. Suppose that, for a given k,

$$\mu^*(A \cap F_k) = \sum\nolimits_{n=1}^{k} \mu^*(A \cap E_n) . \tag{16}$$

(This is certainly true when $k = 1$). By 4.4, F_k is measurable, and so

$$\begin{aligned} \mu^*(A \cap F_{k+1}) &= \mu^*(A \cap F_{k+1} \cap F_k) + \mu^*((A \cap F_{k+1}) \setminus F_k) \\ &= \mu^*(A \cap F_k) + \mu^*(A \cap E_{k+1}) \\ &= \sum\nolimits_{n=1}^{k+1} \mu^*(A \cap E_n) \qquad \text{because of (16).} \end{aligned}$$

Thus, (16) holds for all $k \in \mathbb{N}$. As $F_k \subseteq E$ for each k, 3.5*(b)* and (16) give

$$\mu^*(A \cap E) \geq \mu^*(A \cap F_k) = \left(\sum\nolimits_{n=1}^{k} \mu^*(A \cap E_n) \right) ;$$

this holds for all k, so

$$\mu^*(A \cap E) \geq \sup_k \sum\nolimits_{n=1}^{k} \mu^*(A \cap E_n) := \sum\nolimits_{n=1}^{\infty} \mu^*(A \cap E_n) .$$

However, the opposite inequality is assured by 3.5*(c)*. □

Remark 4.7. If (E_n) is a sequence of subsets of Ω, there is a standard procedure for obtaining a related disjoint sequence. For lack of a better name I have sometimes called it *disjunctification*. Define $E'_1 := E_1$ and $E'_{k+1} := E_{k+1} \setminus \left(\bigcup_{n=1}^{k} E_n \right)$. Then $E'_k \subseteq E_k$ for each k and the sequence (E'_n) is disjoint (by construction, E'_n is disjoint from E_m for all $m < n$). But also

$$\bigcup_{n=1}^{k} E'_n = \bigcup_{n=1}^{k} E_n \tag{17}$$

for any $k \in \mathbb{N} \cup \{\infty\}$. The inclusion $\subseteq$ is obvious; but any element x of the right-hand side must belong to an E_n of smallest possible index, and then $x \in E'_n$ too.

As a consequence of this procedure, the requirement 3.5*(c)* of the definition of outer measure could be (and sometimes is) stated in the apparently weaker form

if (M_n) is any *disjoint* sequence in $\mathcal{P}(\Omega)$, then $\mu^*\Big(\bigcup_{n=1}^{\infty} M_n\Big) \leq \sum_{n=1}^{\infty} \mu^*(M_n)$.

Lemma 4.8. *Let (E_n) be a sequence of μ^*-measurable subsets of Ω. Then $E := \bigcup_{n=1}^{\infty} E_n$ is also μ^*-measurable.*

Proof. Since each finite union $F_k := \bigcup_{n=1}^{k} E_n$ is μ^*-measurable by 4.4 and 4.5 ensures the difference is μ^*-measurable, $E'_{k+1} := E_{k+1} \setminus \Big(\bigcup_{n=1}^{k} E_n\Big)$ is μ^*-measurable for each $k \geq 1$, and $E'_1 := E_1$ is too. In view of (17), it will suffice to consider the disjoint sequence (E'_n).

Take any test set $A \in \mathcal{P}(\Omega)$. Then, for any $k \in \mathbb{N}$,

$$\begin{aligned}
\mu^*(A) &= \mu^*(A \cap F_k) + \mu^*(A \setminus F_k) \\
&= \Big(\sum_{n=1}^{k} \mu^*(A \cap E'_n)\Big) + \mu^*(A \setminus F_k) \quad \text{by 4.6 (or (16))} \\
&\geq \Big(\sum_{n=1}^{k} \mu^*(A \cap E'_n)\Big) + \mu^*(A \setminus E) \quad \text{by 3.5\textit{(b)}.}
\end{aligned}$$

This inequality holds for all k, and therefore

$$\begin{aligned}
\mu^*(A) &\geq \Big(\sum_{n=1}^{\infty} \mu^*(A \cap E'_n)\Big) + \mu^*(A \setminus E) \\
&\geq \mu^*(A \cap E) + \mu^*(A \setminus E) \qquad \text{by 3.5\textit{(c)}.}
\end{aligned}$$

However, the opposite inequality also follows from 3.5*(a)* and 3.5*(c)*. (Apply 3.5*(c)* to the sequence $A \cap E, A \setminus E, \emptyset, \emptyset, \emptyset, \ldots$). □

It is a curiosity of this lemma that it employs only the "finite version" (16) of 4.6. And the culminating Theorem below only uses the case $A = \Omega$.

Theorem 4.9. (Carathéodory). *Let μ^* be an outer measure in the set Ω, and let Σ be the class of μ^*-measurable sets in Ω. Then Σ has the following properties.*

(a) $\emptyset \in \Sigma$ *and* $\Omega \in \Sigma$.

(b) *If* $E, F \in \Sigma$, *then* $E \setminus F \in \Sigma$.

(c) *If* $(E_n)_{n=1}^{\infty}$ *is any sequence of elements of* Σ, *then* $\bigcup_{n=1}^{\infty} E_n \in \Sigma$ *too.*

Furthermore, if $(F_n)_{n=1}^{\infty}$ *is a disjoint sequence of members of* Σ, *then*

$$\mu^*\Big(\bigcup_{n=1}^{\infty} F_n\Big) = \sum_{n=1}^{\infty} \mu^*(F_n).$$

Proof. *(a)* is 4.2; *(b)* is part of 4.5; *(c)* is 4.8. The last assertion is the case $A = \Omega$ of 4.6. □

The class Σ of μ^*-measurable sets in Ω has, therefore, quite remarkable properties. Unfortunately, it is quite possible — indeed it is "usually" the case — that Σ has only the two elements $\emptyset$ and Ω, so that the properties in question reduce to trivialities. We need further information about μ^* to ensure that Σ is large enough for the Theorem to be interesting, but we shall first digress to look at the properties in more detail.

The last assertion of 4.9 is often expressed by saying that μ^* is *countably additive* on Σ.

The class of μ^*-measurable sets satisfies even stronger conditions than those given in 4.9. Suslin, in 1917, invented a method of combining sets which is more general than taking countable unions or intersections. (He was hoping, wrongly, that he could construct all Lebesgue-measurable sets in this way). If you apply the "Suslin operation" to a system of μ^*-measurable sets, the result is also μ^*-measurable. This is proved on pp. 47–50 of Saks; much more information on Suslin's construction can be found in Kuratowski's or Hausdorff's books. However, the Suslin operation appears only rather rarely in mainstream analysis.

§5. Rings, fields, measures.

Although outer measures are, both historically and intuitively, perhaps the most natural way of constructing measures (there are other ways), the further theory scarcely notices them. The properties of Σ listed in 4.9 turn out to be more important than the way in which they arose.

Definition 5.1. Let Ω be any set. A *ring in* Ω (or, more precisely, a *ring of subsets of* Ω) is a subset $\mathcal{R}$ of $\mathcal{P}(\Omega)$ such that

(i) $\emptyset \in \mathcal{R}$;
(ii) if $E, F \in \mathcal{R}$, then $E \setminus F \in \mathcal{R}$; and
(iii) if $E, F \in \mathcal{R}$, then $E \cup F \in \mathcal{R}$.

Granted *(ii)*, *(i)* says only that $\mathcal{R} \neq \emptyset$ (for, if $E \in \mathcal{R}$, then *(ii)* implies $\emptyset = E \setminus E \in \mathcal{R}$). And *(ii)* also entails that $E \cap F = E \setminus (E \setminus F) \in \mathcal{R}$.

The word "ring" is used because of a very imprecise algebraic analogy (union $\approx$ sum, intersection $\approx$ product). As far as I know, Halmos is responsible for the term. A ring $\mathcal{R}$ is precisely what you need if the disjunctification trick of 4.7, applied to a sequence of members of $\mathcal{R}$, is to construct a new sequence of members of $\mathcal{R}$.

A typical example of a ring is the class $\mathcal{P}$ of polygons in $\mathbb{R}^2$, introduced in §1C. There was a problem in calculating the area of a polygon: we had to express it as a finite *disjoint* union of open triangles. This, and similar examples, suggest the idea of a "semiring", which, as it were, generates a ring.

Definition 5.2. A *semiring in the set* Ω (more precisely, a *semiring of subsets of* Ω) is a subset $\mathcal{S}$ of $\mathcal{P}(\Omega)$ such that

(i) $\emptyset \in \mathcal{S}$; and
(ii) if $E, F \in \mathcal{S}$, then $E \setminus F$ is a finite *disjoint* union of members of $\mathcal{S}$.

Lemma 5.3. (a) *A ring of subsets of* Ω *is always a semiring.*

(b) *Let* $\mathcal{S}$ *be a semiring in* Ω. *Then the class of finite disjoint unions of members of* $\mathcal{S}$ *constitutes a ring in* Ω. *In particular, it is the same as the class of arbitrary finite unions.*

Proof. *(a)* is obvious, and I think it is appropriate to leave *(b)* as an *exercise*. It is proved by several tedious inductions. □

Remark 5.4. In linear algebra, any subset A of a vector space V spans or "generates" a linear subspace that we may call $\operatorname{span}(A)$ (also denoted $\operatorname{lin}(A)$ by some). There are two ways of defining $\operatorname{span}(A)$. On the one hand, it is the set of all vectors that may be obtained as linear combinations of elements of A; this is the "internal" or "constructive" definition, which builds $\operatorname{span}(A)$ from A by telling us exactly what the elements of $\operatorname{span}(A)$ look like. On the other hand, $\operatorname{span}(A)$ is the smallest linear subspace of V that includes A. This is the "external" or "implicit" definition, which tells us nothing at all about the individual elements of $\operatorname{span}(A)$.

Both definitions require supplementary facts. For the first, one needs the trivial statement that the linear combinations of elements of A form a linear subspace, whilst for the second, one must prove (which is easy) that the class of linear subspaces including A, ordered by inclusion, has a least element.

In group theory, similarly, any subset A of a group G generates a subgroup, which may be described "internally" or "constructively" as the set of elements of G obtained by evaluating words on the elements of A, and "externally" as the least subgroup of G that includes A.

In algebraic situations like these, we are dealing with finitary operations. As I remarked at the start, algebra is in some sense the study of finitary operations — mostly of binary ones.

Let $\mathcal{A}$ be any subclass of the set $\mathcal{P}(\Omega)$. Take $\mathcal{A}_0 := \mathcal{A} \cup \{\emptyset\}$, and, if $\mathcal{A}_n$ is constructed, let $\mathcal{A}_{n+1}$ consist of all subsets of Ω that may be obtained as either the union or the difference of two members of $\mathcal{A}_n$:

$$\mathcal{A}_{n+1} := \{E \cup F,\ E \setminus F : E, F \in \mathcal{A}_n\}.$$

Lemma 5.5. *With the notation just established,*

(a) $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ *for each* $n \in \mathbb{N} \cup \{0\}$,
(b) $\bigcup_{n=0}^{\infty} \mathcal{A}_n$ *is a ring in* Ω *that includes* $\mathcal{A}$,
(c) *if* $\mathcal{R}$ *is any ring in* Ω *that includes* $\mathcal{A}$, *then* $\mathcal{R} \supseteq \bigcup_{n=0}^{\infty} \mathcal{A}_n$.

Proof. Easy exercise. □

It follows immediately that $\bigcup_{n=0}^{\infty} \mathcal{A}_n$ is the smallest ring of subsets of Ω that includes $\mathcal{A}$, which we may denote $\mathcal{R}(\mathcal{A})$. Any member of $\mathcal{R}(\mathcal{A})$ belongs to $\mathcal{A}_n$ for some n, and thus is the result of finitely many operations of taking the union or the difference of two sets, applied in the beginning to members of $\mathcal{A}$. When $\mathcal{S}$ is a semiring, $\mathcal{R}(\mathcal{S})$ is just the class of finite unions of members of $\mathcal{S}$, by 5.3.

Definition 5.6. Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. A *finitely additive signed measure* (fasm) on $\mathcal{A}$ is a function $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$ such that

(a) $\mu(\emptyset) = 0$, if $\emptyset \in \mathcal{A}$; and

(b) if $A \in \mathcal{A}$ and $A_1, A_2, \ldots, A_n$ is a pairwise disjoint finite sequence of members of $\mathcal{A}$ such that $A = \bigcup_{k=1}^{n} A_k$, then

$$\mu(A) = \sum\nolimits_{k=1}^{n} \mu(A_k)\,. \tag{18}$$

A *finitely additive measure* (fam) on $\mathcal{A}$ is a finitely additive signed measure which takes only non-negative values.

Recall 2.2: the equality (18) means that, in the given circumstances, the right-hand side always makes sense in $\overline{\mathbb{R}}$ and has value equal to $\mu(A)$. However, there is no reason why any member of $\mathcal{A}$ should be a non-trivial finite disjoint union of other members. The condition *(b)* may sometimes be vacuously true.

If μ takes only finite values (quite a useful possibility — consider the areas of triangles), or only non-negative values, the sum is always defined. More generally, one may consider "finitely additive measures" with values in any abelian group. There are many minor modifications that may be made to the theory without serious change to the proofs.

Lemma 5.7. *Let $\mathcal{S}$ be a semiring in Ω and let $\sigma : \mathcal{S} \longrightarrow \overline{\mathbb{R}}$ be a fasm on $\mathcal{S}$ that takes at most one infinite value. Then there is a unique fasm $\widetilde{\sigma} : \mathcal{R}(\mathcal{S}) \longrightarrow \overline{\mathbb{R}}$ such that $\widetilde{\sigma}|\mathcal{S} = \sigma$. If σ is a fam, so is $\widetilde{\sigma}$.*

Proof. By 5.3*(b)*, any member $R \in \mathcal{R}(\mathcal{S})$ is a finite disjoint union $\bigcup_{q=1}^{r} S_q$, for some $n \in \mathbb{N}$ and some (disjoint) members S_q of $\mathcal{S}$. So $\widetilde{\sigma}(R)$ must be $\sum_{q=1}^{r} \sigma(S_q)$. The only problem with *defining* $\widetilde{\sigma}(R)$ by this formula is that R may be expressible as a finite disjoint union of members of $\mathcal{S}$ in more than one way (again, think of $\mathcal{S}$ as the set of triangles in $\mathbb{R}^2$).

Let $(C_i)_{i=1}^{m}$ and $(D_j)_{j=1}^{n}$ be finite disjoint sequences in $\mathcal{S}$ such that

$$\bigcup\nolimits_{i=1}^{m} C_i = \bigcup\nolimits_{j=1}^{n} D_j\,.$$

For each choice of i, j, $C_i \cap D_j$ is in $\mathcal{R}(\mathcal{S})$, and, by 5.3*(b)*, is a finite disjoint union of members of $\mathcal{S}$: for some $p(i,j) \in \mathbb{N}$ and some members B_{ijk} of $\mathcal{S}$, $1 \le k \le p(i,j)$,

$$C_i \cap D_j = \bigcup\nolimits_{k=1}^{p(i,j)} B_{ijk}\,,$$

where the B_{ijk} for different k are (pairwise) disjoint. It follows that $B_{ijk} \cap B_{i'j'k'} = \emptyset$ whenever the triples (i,j,k) and (i',j',k') differ; for instance, if $j \neq j'$, then $B_{ijk} \subseteq D_j$ and $B_{i'j'k'} \subseteq D_{j'}$, but $D_j \cap D_{j'} = \emptyset$. However,

$$C_i = C_i \cap \Big(\bigcup\nolimits_{j=1}^{n} D_j\Big) = \bigcup\nolimits_{j=1}^{n} (C_i \cap D_j) = \bigcup\nolimits_{j=1}^{n} \Big(\bigcup\nolimits_{k=1}^{p(i,j)} B_{ijk}\Big)\,,$$

and this is a finite disjoint union of members of $\mathcal{S}$. Similarly, $D_j = \bigcup_{i=1}^{m} \Big(\bigcup_{k=1}^{p(i,j)} B_{ijk}\Big)$, also a finite disjoint union.

Now, σ is a fasm. So

$$\sigma(C_i) = \sum_{j=1}^{n} \sum_{k=1}^{p(i,j)} \sigma(B_{ijk}), \quad \sigma(D_j) = \sum_{i=1}^{m} \sum_{k=1}^{p(i,j)} \sigma(B_{ijk}), \quad \text{and}$$
$$\sum_{i=1}^{m} \sigma(C_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p(i,j)} \sigma(B_{ijk})$$
$$= \sum_{j=1}^{n} \sum_{i=1}^{m} \sum_{k=1}^{p(i,j)} \sigma(B_{ijk}) = \sum_{j=1}^{n} \sigma(D_j).$$

Thus, the manner in which $R \in \mathcal{R}$ is expressed as a finite disjoint union of members of $\mathcal{S}$ does not affect the sum of the corresponding values of σ, and we may unambiguously define

$$\widetilde{\sigma}(R) = \sum_{q=1}^{r} \sigma(S_q) \quad \text{when} \quad R = \bigcup_{q=1}^{r} S_q,$$

for any choice of a finite disjoint sequence $S_1, S_2, \ldots, S_r$ of members of $\mathcal{S}$ whose union is R.

It remains to check that $\widetilde{\sigma}$, so defined, is a fasm on $\mathcal{R}$; which is trivial. □

The condition that σ takes at most one infinite value is implicit in the proof, because I have casually assumed that all the sums make sense. It is easy to give examples where fasms on a semiring do not extend to the generated ring because the condition is not satisfied. Indeed, it is a necessary as well as sufficient condition:

Lemma 5.8. *Let $\mathcal{R}$ be a ring in Ω, and let $\sigma : \mathcal{R} \longrightarrow \overline{\mathbb{R}}$ be a fasm. Then σ can take at most one infinite value; that is, if $E_1, E_2 \in \mathcal{R}$ and $\sigma(E_1) = \infty$, then $\sigma(E_2) > -\infty$. Furthermore, if $E_1 \subseteq E_2$ and $\sigma(E_1) = \pm\infty$, then $\sigma(E_2) = \sigma(E_1)$.*

Proof. Suppose $E_1, E_2 \in \mathcal{R}$ and $\sigma(E_1) = \infty$. By hypothesis $E_1 \setminus E_2, E_1 \cap E_2 \in \mathcal{R}$, and

$$\sigma(E_1) = \sigma(E_1 \setminus E_2) + \sigma(E_1 \cap E_2),$$

which means that either $\sigma(E_1 \setminus E_2) = \infty$ or $\sigma(E_1 \cap E_2) = \infty$. In the first case, $E_1 \cup E_2 \in \mathcal{R}$ and $\sigma(E_1 \cup E_2) = \sigma(E_1 \setminus E_2) + \sigma(E_2)$, and this excludes the possibility that $\sigma(E_2) = -\infty$, which would make the right-hand side undefined. In the second case, $E_2 \supseteq E_1 \cap E_2$, and $\sigma(E_2) = \sigma(E_2 \setminus E_1) + \sigma(E_1 \cap E_2)$, which forces $\sigma(E_2) = \infty$. (This argument also proves the last assertion of the Lemma.) The proof with $-\infty$ instead of ∞ is identical. □

This is a typical example of arguments about infinities such as I warned of in the introduction to §2. The result has no substantial "mathematical content"; it says merely that, if we want to allow infinite values for fasms on a ring, consistency demands that we only have one. As I have commented before, it does simplify proofs if one forbids infinite values.

Some authors speak of "set functions", meaning functions defined on classes of sets. This has the advantage that a "fam" is a *nonnegative finitely additive set function*, so that the more specialized notion is defined, as one would normally expect, by adding restrictive adjectives to the more general one (whereas the phrase "finitely additive measure" is linguistically perverse — it need not, in principle, be a *measure* in the usual sense 5.10 at all). However, I think my terminology is the common one.

Lemma 5.9. *If M and N are members of the ring $\mathcal{R}$ in Ω, and μ is a fam on $\mathcal{R}$, and $M \subseteq N$, then $\mu(M) \leq \mu(N)$.*

Proof. This is much the same argument as 5.8, with $\mu(N \setminus M)$ now non-negative:

$$N \setminus M \in \mathcal{R} \quad \text{and} \quad \mu(N) = \mu(M) + \mu(N \setminus M) \geq \mu(M)\,. \qquad \square$$

This property of μ is sometimes expressed by saying that μ is an "increasing set function".

Definition 5.10. Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ as before, and $\mu : \mathcal{A} \longrightarrow \overline{\mathbb{R}}$. μ is described as *countably additive* on $\mathcal{A}$ if, whenever $(A_i)_{i=1}^{\infty}$ is a pairwise disjoint sequence of elements of $\mathcal{A}$ whose union is also in $\mathcal{A}$, $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, then

$$\mu(A) = \sum\nolimits_{i=1}^{\infty} \mu(A_i)\,. \tag{19}$$

A *signed measure* on $\mathcal{A}$ is a countably additive fasm on $\mathcal{A}$. A *measure* on $\mathcal{A}$ is a signed measure that takes only non-negative values; that is, a countably additive fam on $\mathcal{A}$.

Thus, as I have already remarked, a "signed measure" need not be a measure, and a "fasm" need not be a signed measure. I have stated explicitly above that a signed measure must be finitely as well as countably additive; in practice, one usually has $\emptyset \in \mathcal{A}$ and $\mu(\emptyset) = 0$, and then countable additivity *implies* finite additivity. (19) is to be understood, as always, as including the assertion that the sum on the right-hand side makes sense in $\overline{\mathbb{R}}$.

The phrase "countably additive" is often abbreviated to "σ-additive". Some people (e.g. Munroe) call it "completely additive", or "totally additive". Since the fundamental fact of the course is that measures can be defined on rather large classes $\mathcal{A}$, it is worth pausing to point out the reason why we stop at *countable* additivity. The equality (19) is only admissible because we have a satisfactory notion of "addition" for countable classes of non-negative numbers (see 2.15, for instance; but notice that the sum in (19) must be unconditionally convergent, since A is also the union of any rearrangement of the sequence (A_i)). Tempting as it is to try to construct "uncountably additive" measures, the idea is unsatisfactory in practice because *any* subset of $\mathbb{R}$ is a possibly *uncountable* disjoint union of singletons. If (19) were often to hold, and uncountably many singletons were to have positive measure, this would mean that unacceptably many sets would have infinite measure; if only countably many singletons had positive measure, unacceptably many sets would have zero measure. The usefulness of Lebesgue measure, on the other hand, is related to the profusion of sets with positive finite Lebesgue measure.

As with 5.6, the definitions above lack content if elements of $\mathcal{A}$ are never countable disjoint unions of sequences in $\mathcal{A}$. The "natural" domain for a fasm is a ring, in which the set operations that enable us to extract the full value of the additivity condition (18) do not take us outside the domain of μ; for signed measures, the corresponding natural domain is a σ-ring.

Definition 5.11. A σ-ring in the set Ω is a subset Σ of $\mathcal{P}(\Omega)$ such that

(a) $\emptyset \in \Sigma$; *(b)* if $E, F \in \Sigma$, then $E \setminus F \in \Sigma$;
(c) if $(E_i)_{i=1}^{\infty}$ is any sequence in Σ, then $\bigcup_{i=1}^{\infty} E_i \in \Sigma$.

Lemma 5.12. *A σ-ring is a ring. If $E, F \in \Sigma$, then $E \cap F \in \Sigma$; if $(E_i)_{i=1}^{\infty}$ is any sequence in Σ, then $\bigcap_{i=1}^{\infty} E_i \in \Sigma$.*

Proof. The first statement: take $E_1 := E$, $E_2 := F$, $\cdots = E_4 = E_3 := \emptyset$, in *(c)*. Then, as in the remarks after 5.1, $E \cap F = E \setminus (E \setminus F) \in \Sigma$ when $E, F \in \Sigma$.

For the last statement: take $Q := \bigcup_{i=1}^{\infty} E_i \in \Sigma$, by *(c)*. Then, by *(b)*, $Q \setminus E_i \in \Sigma$ for each i. Applying *(c)*, $\bigcup_{i=1}^{\infty}(Q \setminus E_i) \in \Sigma$. By *(b)*, $\bigcap_{i=1}^{\infty} E_i = Q \setminus (\bigcup_{i=1}^{\infty}(Q \setminus E_i)) \in \Sigma$. □

In short, a σ-ring is a ring in which countable as well as finite unions and intersections are permitted. A simple example is the class of countable sets in $\mathbb{R}$, which is a σ-ring by 1.2. The question arises what σ-rings look like in general.

Lemma 5.13. *Suppose $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. There exists a smallest σ-ring in Ω that includes $\mathcal{A}$: that is, there is a σ-ring which itself includes $\mathcal{A}$ and is included in any σ-ring that includes $\mathcal{A}$.*

Proof. Let $\mathcal{R}$ denote the class of all σ-rings in Ω that include $\mathcal{A}$. $\mathcal{R}$ is not empty, for $\mathcal{P}(\Omega)$ itself is a member of $\mathcal{R}$. Define

$$\Sigma'(\mathcal{A}) := \bigcap_{\Sigma \in \mathcal{R}} \Sigma. \tag{20}$$

It is easy to check that, as 5.11*(a)*, *(b)*, and *(c)* hold for each $\Sigma \in \mathcal{R}$, they hold also for $\Sigma'(\mathcal{A})$, which is therefore a σ-ring. It obviously includes $\mathcal{A}$, and is included in any σ-ring that includes $\mathcal{A}$. □

An exactly similar argument can be given to prove the existence of a "least ring including $\mathcal{A}$", or of a "least vector subspace including a given subset of a vector space", and so on. Notice that $\mathcal{R} \in \mathcal{P}(\mathcal{P}(\mathcal{P}(\Omega)))$, for each σ-ring is a member of $\mathcal{P}(\mathcal{P}(\Omega))$.

One might try to *construct* $\Sigma'(\mathcal{A})$ as follows, imitating 5.5. Let $\mathcal{A}_0 := \mathcal{A} \cup \{\emptyset\}$, and, when $\mathcal{A}_n$ is known, let $\mathcal{A}_{n+1}$ consist of all the subsets of Ω that may be obtained either as the countable union of sets in $\mathcal{A}_n$ or as the difference of two sets in $\mathcal{A}_n$. Having thus defined $\mathcal{A}_n$ inductively for $n \in \mathbb{N}$, take $\mathcal{A}_\infty := \bigcup_{n=0}^{\infty} \mathcal{A}_n$. One expects that $\Sigma'(\mathcal{A}) = \mathcal{A}_\infty$.

Unfortunately, it is impossible to prove that $\mathcal{A}_\infty$ satisfies 5.11*(c)*, and there are cases where it does not ($\mathcal{A}$ may be the class of open sets in $\mathbb{R}$. The proofs are non-trivial; you can find them in Hausdorff or Kuratowski). The problem is that an infinite sequence of members of $\mathcal{A}_\infty$ may take its nth term from $\mathcal{A}_n \setminus \mathcal{A}_{n-1}$. In fact, the procedure does work if you continue the inductive construction to transfinite ordinals and stop, not, as we did, at the first infinite ordinal, but at the first uncountable ordinal. However, in doing so, you lose any explicit description of the members of $\Sigma'(\mathcal{A})$, so there is not much point.

The moral is that, for most practical purposes, results about σ-rings have to be proved by "implicit" methods. Here is a simple example.

Lemma 5.14. *Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. Then, for any $B \in \Sigma'(\mathcal{A})$, there is a sequence $(A_i)_{i=1}^{\infty}$ of members of $\mathcal{A}$ such that $B \subseteq \bigcup_{i=1}^{\infty} A_i$.*

Proof. Let $\mathcal{B}$ be the class of *all* subsets of Ω which are included in *some* countable union of members of $\mathcal{A}$. Clearly $\mathcal{B} \supseteq \mathcal{A}$, and $\mathcal{B}$ is trivially a σ-ring. Hence $\mathcal{B} \supseteq \Sigma'(\mathcal{A})$. But this is just the assertion of the Lemma. □

It should be emphasized that the sequence (A_i) "depends on B" — it is definitely not always true that there is a single sequence that works for all members of $\Sigma'(\mathcal{A})$. A very simple example is when $\mathcal{A}$ is the class of singletons in $\mathbb{R}$; then $\Sigma'(\mathcal{A})$ is exactly the class of

countable subsets of $\mathbb{R}$, each of which can be covered by some sequence of singletons, but (as $\mathbb{R}$ is uncountable) they cannot all be covered by the same sequence of singletons.

Proposition 5.15. *Let μ^* be an outer measure in Ω. Then the class Σ of μ^*-measurable sets in Ω is a σ-ring in Ω, and $\mu^*|\Sigma : \Sigma \longrightarrow \overline{\mathbb{R}}$ is a measure.*

Proof. This is a restatement of (part of) Carathéodory's Theorem 4.9. □

§6. Countable additivity.

In this section we can at last discern a practical construction of measures.

Proposition 6.1. *Let $\mathcal{R}$ be a ring in Ω and $\mu : \mathcal{R} \longrightarrow \overline{\mathbb{R}}$ a fam. Regard μ as a weighting function in Ω (that is, take $\mathcal{C} \coloneqq \mathcal{R}$ and $\tau \coloneqq \mu$ in* 3.2*). Then $\Sigma'(\mathcal{R})$ (the σ-ring generated by $\mathcal{R}$, cf.* (20) *of* 5.13*) consists of $\mu^\dagger$-measurable sets.*

Proof. Take a test set $A \in \mathcal{P}(\Omega)$ and $M \in \mathcal{R}$. If $(M_i)_{i=1}^\infty$ is a sequence in $\mathcal{R}$ covering A, then $(M_i \setminus M)$ is a sequence in $\mathcal{R}$ covering $A \setminus M$, and $(M_i \cap M)$ is a sequence in $\mathcal{R}$ covering $A \cap M$. Hence, from the definition of $\mu^\dagger$ (see 3.3),

$$\begin{aligned}\mu^\dagger(A \setminus M) + \mu^\dagger(A \cap M) &\leq \sum_{i=1}^\infty \mu(M_i \setminus M) + \sum_{i=1}^\infty \mu(M_i \cap M)\\ &= \sum_{i=1}^\infty (\mu(M_i \setminus M) + \mu(M_i \cap M)) \quad \text{by 2.15}\\ &= \sum_{i=1}^\infty \mu(M_i) \quad \text{as } \mu \text{ is a fam on } \mathcal{R}.\end{aligned}$$

Taking the infimum over all such sequences (M_i), we find that

$$\mu^\dagger(A \setminus M) + \mu^\dagger(A \cap M) \leq \mu^\dagger(A)\,.$$

The opposite inequality comes from 3.4*(a), (c)*. It follows that M is $\mu^\dagger$-measurable.

By Carathéodory's theorem 4.9, the class Σ of $\mu^\dagger$-measurable sets is a σ-ring. I have just shown that $\Sigma \supseteq \mathcal{R}$. Hence $\Sigma \supseteq \Sigma'(\mathcal{R})$, as asserted. □

This result gives a situation of obvious practical significance (one could take $\mathcal{R}$ to be the class of polygons in $\mathbb{R}^2$ and μ to be the area — see §2C) in which an outer measure has a large class of measurable sets. However, it is *not* asserted that $\Sigma'(\mathcal{R})$ is exactly the class of $\mu^\dagger$-measurable sets; in fact, there are usually very many more $\mu^\dagger$-measurable sets than belong to $\Sigma'(\mathcal{R})$. The other defect of the Proposition is that, so far, we know nothing about the values of the measure $\mu^\dagger$ on $\Sigma'(\mathcal{R})$. It might even be identically zero; or its values might, in practice, be extremely difficult to calculate, which would make it awkward to integrate with respect to. We can, however, find a simple criterion that fixes some values of $\mu^\dagger$.

Lemma 6.2. *In* 6.1, *$\mu^\dagger|\mathcal{R} = \mu$ if and only if μ is countably additive on $\mathcal{R}$.*

Proof. By Carathéodory's theorem 4.9, $\mu^\dagger$ is countably additive on the class of $\mu^\dagger$-measurable sets; by 6.1, this class includes $\mathcal{R}$. So, if $\mu^\dagger|\mathcal{R} = \mu$, τ must be σ-additive on $\mathcal{R}$.

Suppose that μ is σ-additive on $\mathcal{R}$ and $A \in \mathcal{R}$. As A is covered by the sequence $A, \emptyset, \emptyset, \dots$ in $\mathcal{R}$, clearly $\mu^\dagger(A) \le \mu(A)$. (21)

Let $(M_i)_{i=1}^\infty$ be any sequence in $\mathcal{R}$ that covers A; then set $M_i' := M_i \cap A \in \mathcal{R}$ for each i, and disjunctify as at 4.7:

$$N_1 := M_1', \quad N_{i+1} := M_{i+1}' \setminus \left(\bigcup_{k=1}^i M_k'\right) \quad \text{for } i \in \mathbb{N}.$$

Then (N_i) is a disjoint sequence in $\mathcal{R}$ (as I remarked after 5.1), $N_i \subseteq M_i$ for each i, and

$$A = \bigcup_{i=1}^\infty M_i' = \bigcup_{i=1}^\infty N_i.$$

By the hypothesis that μ is countably additive on $\mathcal{R}$, $\mu(A) = \sum_{i=1}^\infty \mu(N_i)$. However, for each i, $\mu(N_i) \le \mu(M_i)$ by 5.9; hence $\mu(A) \le \sum_{i=1}^\infty \mu(M_i)$. This holds for any sequence (M_i) in $\mathcal{R}$ that covers A, and so $\mu(A) \le \mu^\dagger(A)$. With (21), this completes the proof. □

Theorem 6.3. *Let $\mathcal{R}$ be a ring in Ω and $\mu : \mathcal{R} \longrightarrow \overline{\mathbb{R}}$ a measure. Then there is a measure $\widehat{\mu}$ on $\Sigma(\mathcal{R})$ such that $\widehat{\mu}|\mathcal{R} = \mu$, namely $\widehat{\mu} := \mu^\dagger|\Sigma(\mathcal{R})$.*

Proof. 4.9, 6.1, and 6.2. □

Lemma 6.4. *Let $\mathcal{S}$ be a semiring in Ω, and let $\mu : \mathcal{S} \longrightarrow \overline{\mathbb{R}}$ be a fasm on $\mathcal{S}$ which takes only one infinite value. The induced fasm $\tilde{\mu} : \mathcal{R}(\mathcal{S}) \longrightarrow \overline{\mathbb{R}}$ (see* 5.7*) is σ-additive on $\mathcal{R}(\mathcal{S})$ if and only if μ is σ-additive on $\mathcal{S}$.*

Proof. Exercise. □

We now have a method for constructing measures. In effect, all we have to do is to define a countably additive measure on a semiring $\mathcal{S}$, which (provided it takes at most one infinite value) will then extend to a measure on the whole of the generated σ-ring. However, there are a number of loose ends. Perhaps the most conspicuous is that the class $\mathcal{M}$ of $\mu^\dagger$-measurable sets always contains Ω itself, by 4.9*(a)*, whereas $\Sigma'(\mathcal{S})$ may not — $\mathcal{S}$ may, for instance, be the class of singletons in $\mathbb{R}$, and μ might take the value 1 on each singleton.

It is also clear from 4.9 that, if $M \in \mathcal{M}$ and $\mu^\dagger(M) = 0$, then *any* subset of M also belongs to $\mathcal{M}$. This property (it is, unfortunately, usually called "completeness" of the measure) need not be true for $\Sigma'(\mathcal{R})$. A subtler objection is that, as I observed after 4.9, the Suslin operation applied to members of $\mathcal{M}$ yields members of $\mathcal{M}$, and it is more general than countable unions; one might therefore suppose that the results should be formulated for classes of subsets of Ω closed under the Suslin operation.

The course I have taken is, however, justifiable. To deal with the last objection first — I have already remarked that the Suslin operation is rarely used in mainstream analysis, and it will turn out that for all of this course, and for most purposes outside it, we need only deal with differences and countable unions. We should not, therefore, demand that our measures have properties stronger than we need for our later theory, since we want our theorems to be as general as possible. The same applies to completeness; indeed, probabilists absolutely must work with incomplete measures much of the time.

The objection that Ω must be $\mu^\dagger$-measurable, but may not be in $\Sigma'(\mathcal{R})$, is in a way more serious. Broadly speaking, the reason I have used $\Sigma'(\mathcal{R})$ above is that there is a uniqueness

theorem: if $\mu : \mathcal{R} \longrightarrow \overline{\mathbb{R}}$ is a measure on $\mathcal{R}$, and satisfies a "σ-finiteness condition", then it has only one possible extension to a measure $\widehat{\mu} : \Sigma'(\mathcal{R}) \longrightarrow \overline{\mathbb{R}}$. This is not true if one allows the extended measure to have a larger domain. For this reason, σ-rings have an important part in the theory, and Halmos goes so far as to develop everything in terms of them. But I shall not follow him. He is almost the only writer who has so consistently avoided the use of σ-algebras, and the conclusion of his efforts — after considerable striving — is that, in practice, there is little reason to do so. The uniqueness theorem is the only point where σ-rings really make a difference; and, to save time, I shall not discuss it here anyway.

Definition 6.5. A class Σ of subsets of Ω is a *field* or *algebra* in Ω (a *field* or *algebra of subsets of* Ω) if it is a ring in Ω and in addition $\Omega \in \Sigma$. It is a *σ-field* or *σ-algebra* in Ω if it is a σ-ring in Ω and in addition $\Omega \in \Sigma$.

Probabilists tend mostly to speak of σ-fields and analysts more often of σ-algebras, but there is no rule. Notice that Carathéodory's theorem 4.9 says that the class of sets measurable with respect to an outer measure is a σ-algebra, not just a σ-ring.

Exactly as in §5, one may construct, for any $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, a least σ-field in Ω that includes $\mathcal{A}$; it is simply the intersection of all the σ-fields that include $\mathcal{A}$. I shall denote it by $\Sigma(\mathcal{A})$.

Lemma 6.6. *For any* $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, $\Sigma(\mathcal{A}) = \{Y \subseteq \Omega : Y \in \Sigma'(\mathcal{A}) \text{ or } \Omega \setminus Y \in \Sigma'(\mathcal{A})\}$.

Proof. Firstly, any σ-field Σ which includes $\mathcal{A}$ is a σ-ring, so $\Sigma \supseteq \Sigma'(\mathcal{A})$; and, as $\Omega \in \Sigma$, $\Omega \setminus Y \in \Sigma$ for any $Y \in \Sigma$, and in particular for any $Y \in \Sigma'(\mathcal{A})$. Hence

$$\Sigma \supseteq \mathcal{B} := \{Y \subseteq \Omega : Y \in \Sigma'(\mathcal{A}) \text{ or } \Omega \setminus Y \in \Sigma'(\mathcal{A})\}. \tag{22}$$

On the other hand, whenever $B \in \mathcal{B}$, then $\Omega \setminus B \in \mathcal{B}$ too; as $\mathcal{B} \supseteq \Sigma'(\mathcal{A})$, necessarily $\emptyset \in \mathcal{B}$, and therefore $\Omega \in \mathcal{B}$. Now suppose $E, F \in \mathcal{B}$. If both belong to $\Sigma'(\mathcal{A})$, then $E \setminus F \in \Sigma'(\mathcal{A})$ because $\Sigma'(\mathcal{A})$ is a σ-ring. If $\Omega \setminus E, \Omega \setminus F \in \Sigma'(\mathcal{A})$, then

$$E \setminus F = (\Omega \setminus F) \setminus (\Omega \setminus E) \in \Sigma'(\mathcal{A}).$$

If $\Omega \setminus E, F \in \Sigma'(\mathcal{A})$, then $(\Omega \setminus E) \cup F \in \Sigma'(\mathcal{A})$ by the definition, 5.11, and so

$$E \setminus F = \Omega \setminus ((\Omega \setminus E) \cup F) \in \mathcal{B},$$

whilst $F \setminus E = (\Omega \setminus E) \cap F \in \Sigma'(\mathcal{A})$ by 5.12.

Finally, suppose (M_i) is a sequence in $\mathcal{B}$. We may split it into a sequence (E_m) in $\Sigma'(\mathcal{A})$, and a sequence (F_n) such that $\Omega \setminus F_n \in \Sigma'(\mathcal{A})$ for each n. Either or both of these sequences may be a finite sequence, or indeed have no terms at all. Now

$$\bigcup\nolimits_m E_m \in \Sigma'(\mathcal{A}) \quad \text{by 5.11,} \qquad \bigcap\nolimits_n (\Omega \setminus F_n) \in \Sigma'(\mathcal{A}) \quad \text{by 5.12,}$$

and so, if the sequence (E_m) is non-empty, $\bigcup_i M_i = (\bigcup_m E_m) \setminus (\bigcap_n (\Omega \setminus F_n)) \in \Sigma'(\mathcal{A})$, whilst otherwise $\Omega \setminus (\bigcup_i M_i) = \bigcap_n (\Omega \setminus F_n) \in \Sigma'(\mathcal{A})$.

Hence, $\mathcal{B}$ is a σ-field. From (22), it is the least σ-field including $\mathcal{A}$. □

In the cases that will mostly interest us, $\Sigma(\mathcal{A}) = \Sigma'(\mathcal{A})$ anyway.

Lemma 6.7. *Suppose that* $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ *is such that, for some sequence* $(A_i)_{i=1}^{\infty}$ *in* $\mathcal{A}$, $\Omega = \bigcup_{i=1}^{\infty} A_i$. *Then* $\Sigma(\mathcal{A}) = \Sigma'(\mathcal{A})$.

Proof. Indeed, $\Omega \in \Sigma'(\mathcal{A})$, as it is a countable union of members of $\mathcal{A} \subseteq \Sigma'(\mathcal{A})$.

Definition 6.8. Let Ω be a topological space with topology $\mathcal{T}$. The *Borel σ-algebra* of Ω, often denoted $\mathcal{B}(\Omega)$, is $\Sigma(\mathcal{T})$. The members of $\Sigma(\mathcal{T})$ are the *Borel sets* of Ω (with respect to the topology $\mathcal{T}$).

Since I shall later be taking Ω to be an interval in $\mathbb{R}$, it is worth recalling that the *subspace topology* or *relative topology* in a subset X of a topological space $(\Omega, \mathcal{T})$ is the class

$$\{X \cap U : U \in \mathcal{T}\}$$

of subsets of X. That is, we *define* a set to be open in X if it is the intersection with X of a set open in Ω.

§7. Convergence ideas.

Definition 7.1. Let $(M_n)_{n=1}^{\infty}$ be a sequence in $\mathcal{P}(\Omega)$. Define

$$\limsup_{n\to\infty} M_n = \limsup_{n\to\infty} M_n = \limsup M_n = \overline{\lim}\, M_n := \bigcap_{k=1}^{\infty} \left(\bigcup_{n=k}^{\infty} M_n \right),$$

the *upper limit* or *limes superior* of the sequence, and

$$\liminf_{n\to\infty} M_n = \liminf_{n\to\infty} M_n = \liminf M_n = \underline{\lim}\, M_n := \bigcup_{k=1}^{\infty} \left(\bigcap_{n=k}^{\infty} M_n \right),$$

the *lower limit* or *limes inferior* of the sequence.

Lemma 7.2. $\limsup M_n$ is the set of all elements of Ω that are in M_n for infinitely many indices n, whilst $\liminf M_n$ is the set of elements of Ω that belong to M_n for all n with finitely many exceptions: that is[3], if $\nu_1(x) := \#(\{n \in \mathbb{N} : x \in M_n\}) \in \mathbb{N} \cup \{\infty\}$ and $\nu_2(x) := \#(\{n \in \mathbb{N} : x \notin M_n\}$,

$$\limsup M_n = \{x : \nu_1(x) = \infty\}, \quad \liminf M_n = \{x : \nu_2(x) < \infty\}. \qquad \square$$

Corollary 7.3. *For any sequence (M_n) in $\mathcal{P}(\Omega)$, $\underline{\lim}\, M_n \subseteq \overline{\lim}\, M_n$ and*

$$\Omega \setminus \overline{\lim}\, M_n = \underline{\lim}\, (\Omega \setminus M_n). \qquad \square$$

[3] I use the "hash" sign # to mean "the number of elements in (the set in question)", understood as being either a non-negative integer or ∞. There are several reasons. The notation $|A|$ you may be more accustomed to is a little confusing in our context, where absolute values of real numbers and moduli of complex numbers can also appear; this is the same as my reason for preferring $\setminus$ to $-$ for set difference. But also I suspect the notation $|A|$ usually denotes "the cardinal of A"; for infinite A, it may be many different infinities.

Definition 7.4. Let $(\xi_n)_{n=1}^{\infty}$ be a sequence in $\overline{\mathbb{R}}$. Define

$$\left.\begin{aligned} \limsup_{n\to\infty} \xi_n = \varlimsup_{n\to\infty} \xi_n = \limsup \xi_n = \overline{\lim}\, \xi_n &:= \inf_{k\in\mathbb{N}} \sup_{n\geq k} \xi_n\,, \\ \liminf_{n\to\infty} \xi_n = \varliminf_{n\to\infty} \xi_n = \liminf \xi_n = \underline{\lim}\, \xi_n &:= \sup_{k\in\mathbb{N}} \inf_{n\geq k} \xi_n\,. \end{aligned}\right\} \tag{23}$$

Definition 7.5. Given the sequence (ξ_n) in $\overline{\mathbb{R}}$, say that $u \in \overline{\mathbb{R}}$ is an *upper number* [or *lower number*] for the sequence if the inequality $u < \xi_n$ [or $u > \xi_n$] holds for only finitely many indices n.

Clearly, if u is an upper number for the sequence and $v > u$, then v is also an upper number. However, there may or may not be a *least* upper number, so that they need not form a Dedekind cut (cf. 0.1*(iii)*).

Lemma 7.6. (a) *$\overline{\lim}\, \xi_n$ is the greatest extended real number which is the limit in $\overline{\mathbb{R}}$ of an infinite subsequence of (ξ_n). It is also the infimum in $\overline{\mathbb{R}}$ of the set of upper numbers for (ξ_n); thus any number greater than $\overline{\lim}\, \xi_n$ is an upper number.*

(b) *$\underline{\lim}\, \xi_n$ is the least extended real number which is the limit in $\overline{\mathbb{R}}$ of an infinite subsequence of (ξ_n). It is also the supremum in $\overline{\mathbb{R}}$ of the set of lower numbers for (ξ_n); thus any number less than $\underline{\lim}\, \xi_n$ is a lower number.*

(c) $\underline{\lim}\, \xi_n \leq \overline{\lim}\, \xi_n$.

Proof. Firstly, *(c)*. For any $k, l \in \mathbb{N}$, $\inf_{n\geq k} \xi_n \leq \xi_{\max(k,l)} \leq \sup_{n\geq l} \xi_n$. So $\sup_{n\geq l} \xi_n$ (for a given l) is an upper bound for $\{\inf_{n\geq k} \xi_n : k \in \mathbb{N}\}$, and

$$\sup_{n\geq l} \xi_n \geq \sup_{k\in\mathbb{N}} \inf_{n\geq k} \xi_n = \underline{\lim}\, \xi_n\,.$$

As this is true for any l, we find similarly that $\overline{\lim}\, \xi_n = \inf_{l\in\mathbb{N}} \sup_{n\geq l} \xi_n \geq \underline{\lim}\, \xi_n$.

Now for *(a)*. Let c be an upper number for the sequence. There exists k such that $\xi_n \leq c$ for $n \geq k$; so, for $l \geq k$, $\sup_{n\geq l} \xi_n \leq \sup_{n\geq k} \xi_n \leq c$, and $\overline{\lim}\, \xi_n \leq c$.

Conversely, if $c > \overline{\lim}\, \xi_n := \inf_{k\in\mathbb{N}} \sup_{n\geq k} \xi_n$, by 0.6 there exists some k such that $\sup_{n\geq k} \xi_n < c$, and consequently $\xi_n < c$ whenever $n \geq k$. So c is an upper number. This proves that $\overline{\lim}\, \xi_n$ is the infimum of the upper numbers. Notice too that if $c' > \overline{\lim}\, \xi_n$, we can take $c \in (\overline{\lim}\, \xi_n, c')$, and then c is an upper number; this implies that no subsequence of (ξ_n) can converge to c'.

Suppose $n(1), n(2), \ldots, n(r)$ have been chosen. Take $c := \overline{\lim}\, \xi_n + 2^{-r-1}$, and then we have just seen that $\xi_n < c$ for all $n \geq k$, for some suitable k; take $d := \overline{\lim}\, \xi_n - 2^{-r-1}$, and then $d < \inf_k \sup_{n\geq k} \xi_n$, $\sup_{n\geq \max(k,n(r)+1)} \xi_n > d$, and there must be some $n(r+1) \geq \max(k, n(r)+1)$ for which $d < \xi_{n(r+1)} < c$. In this way we can construct, inductively, a subsequence $(\xi_{n(r)})_{r=1}^{\infty}$ such that, for each r,

$$\overline{\lim}\, \xi_n - 2^{-r} < \xi_{n(r)} < \overline{\lim}\, \xi_n + 2^{-r}\,.$$

It therefore converges to $\overline{\lim}\, \xi_n$. As shown above, no *larger* number can be the limit of a subsequence. This proves *(a)*; *(b)* is proved similarly (or by considering $(-\xi_n)$). □

It is clear that there is a similarity between the set-theoretic meaning of $\overline{\lim}$ and its meaning for real-number sequences. It is possible to give a common formulation.

Suppose that $(T, \leq)$ is a partially ordered set. We may say it is *countably order-complete* if any countable subset has an infimum and a supremum. Then any sequence (t_n) in T has upper and lower limits defined by (23). It is an exercise to show that $\underline{\lim}\, t_n \leq \overline{\lim}\, t_n$.

The two cases studied above are $T := \mathcal{P}(\Omega)$, with "$A \leq B$" meaning "$A \subseteq B$", and $T := \overline{\mathbb{R}}$, with the usual order. Both of them are *order-complete* in the sense that *all* subsets have a supremum and an infimum, not just countable sets. For a subset $\mathcal{A}$ of $\mathcal{P}(\Omega)$,

$$\sup \mathcal{A} = \bigcup_{A \in \mathcal{A}} A, \qquad \inf \mathcal{A} = \bigcap_{A \in \mathcal{A}} A.$$

(I leave it as an exercise to check these assertions).

Lemma 7.7. *Let the partially ordered set T be countably order-complete. Suppose that (t_n) and (s_n) are sequences in T.*

(a) *If $s_n \leq t_n$ except for finitely many indices n, then $\overline{\lim}\, s_n \leq \overline{\lim}\, t_n$ and $\underline{\lim}\, s_n \leq \underline{\lim}\, t_n$.*

(b) *If $(t_{n(k)})_{k=1}^{\infty}$ is an infinite subsequence of (t_n), then*

$$\liminf_{n\to\infty} t_n \leq \liminf_{k\to\infty} t_{n(k)} \leq \limsup_{k\to\infty} t_{n(k)} \leq \limsup_{n\to\infty} t_n .$$

(c) Both $\overline{\lim}\, t_n$ and $\underline{\lim}\, t_n$ are unchanged if finitely many terms are omitted from the sequence. □

For $\mathbb{R}$, there is the added complication that it is only boundedly order-complete (see 0.3); this means that, to make the definitions 7.4, it would be necessary to hypothesize from the start that (ξ_n) is bounded above and below, and then the upper and lower limits in $\overline{\mathbb{R}}$ are the same as those in $\mathbb{R}$.

Definition 7.8. The sequence (t_n) in the countably order-complete partially ordered set $(T, \leq)$ is *order-convergent* if $\underline{\lim}\, t_n = \overline{\lim}\, t_n$. If that is so, the common value of $\underline{\lim}\, t_n$ and $\overline{\lim}\, t_n$ is called the *order-limit* of (t_n), which is said to *order-converge* to it. In particular,

Definition 7.9. A sequence (M_n) of subsets of a set Ω is *convergent* if $\underline{\lim}\, M_n = \overline{\lim}\, M_n$. In that case, the common value of the limits is called the *limit* of the sequence.

This notion of convergence of a sequence of sets was, I believe, originally proposed by Fréchet; as far as we are concerned, it is the only one of interest, which is why it is not usually called "order-convergence". By 7.2, one may characterize it as follows: (M_n) is convergent if and only if every element of Ω that belongs to M_n for infinitely many indices actually belongs to M_n for all indices with finitely many exceptions. This is clearly a very strong condition, but it is satisfied in some useful situations.

Definition 7.10. Let T be a partially ordered set. A sequence $(t_n)_{n\in\mathbb{N}}$ in T is *increasing* (we write $t_n \uparrow$) if, for all $n \in \mathbb{N}$, $t_n \leq t_{n+1}$; it is *decreasing* (or $t_n \downarrow$) if $t_n \geq t_{n+1}$ for all $n \in \mathbb{N}$; it is *monotonic* if it is either increasing or decreasing.

Lemma 7.11. *A monotonic sequence in a countably order-complete partially ordered set is order-convergent. An increasing sequence order-converges to its supremum and a decreasing sequence order-converges to its infimum.*

Proof. If $t_n \uparrow$, then, for all k, $\sup_{n\geq k} t_n = \sup_{n\in\mathbb{N}} t_n$, whilst $\inf_{n\geq k} t_n = t_k$; hence,

$$\overline{\lim}\, t_n = \inf_{k\in\mathbb{N}} \sup_{n\in\mathbb{N}} t_n = \sup_{n\in\mathbb{N}} t_n = \sup_{k\in\mathbb{N}} t_k = \sup_{k\in\mathbb{N}} \inf_{n\geq k} t_n\,.$$

Similarly if $t_n \downarrow$. □

We have now had two distinct notions of convergence for sequences of real numbers: metric convergence, 0.10, and order-convergence, 7.8 (strictly speaking, either for *bounded* sequences only, or for sequences in $\overline{\mathbb{R}}$). They are equivalent. This fact is one of the central properties of real numbers, like Dedekind-completeness, metric completeness, or the compactness of closed bounded intervals; each implies the others, in the sense that, once you have defined $\mathbb{R}$ in a fashion that makes one of them true, the others may be deduced without further use of the definition. For this reason, there are many different ways of proving all three facts.

Proposition 7.12. *A sequence (ξ_n) in $\overline{\mathbb{R}}$ converges to $\xi \in \overline{\mathbb{R}}$ if and only if it order-converges to ξ. A sequence (ξ_n) in $\mathbb{R}$ converges to $\xi \in \mathbb{R}$ if and only if it is bounded and order-converges to ξ.*

Proof. The definitions 0.10 and 2.6 may be consolidated: (ξ_n) converges to ξ if and only if, for any $a, b \in \overline{\mathbb{R}}$ such that $a < \xi < b$, there is $N \in \mathbb{N}$ for which $n \geq N \Longrightarrow a < \xi_n < b$. (If ξ is infinite, for instance ∞, no b exists and the second inequality holds vacuously.) It follows that b is an upper number and a is a lower number for the sequence, and so, by 7.6,

$$a \leq \underline{\lim}\, \xi_n \leq \overline{\lim}\, \xi_n \leq b\,. \tag{24}$$

This holds for any $a < \xi < b$, and it follows that, necessarily, $\underline{\lim}\, \xi_n = \xi = \overline{\lim}\, \xi_n$. (Suppose, for example, that $\xi < \overline{\lim}\, \xi_n$; we could take b such that $\xi < b < \overline{\lim}\, \xi_n$, and this would contradict (24)).

Now suppose the sequence order-converges to ξ. Let $a < \xi < b$. By 7.6, b is an upper number and a a lower number, so there exist N_1, N_2 such that

$$n \geq N_1 \Longrightarrow \xi_n \leq b\,, \quad n \geq N_2 \Longrightarrow \xi_n \geq b\,.$$

Take $N := \max(N_1, N_2)$, and then $n \geq N \Longrightarrow a \leq \xi_n \leq b$, so that ξ_n converges to ξ. □

I can now give a simple proof of Cauchy's "General Principle of Convergence", 0.11.

Theorem 7.13. *A sequence in $\mathbb{R}$ is convergent (in $\mathbb{R}$) if and only if it is Cauchy.*

Proof. Suppose $\xi_n \to \xi \in \mathbb{R}$. Then, for any $\epsilon > 0$, there exists N such that $|\xi_n - \xi| < \frac{1}{2}\epsilon$ whenever $n \geq N$. Hence, if $m, n \geq N$, $|\xi_m - \xi_n| \leq |\xi_n - \xi| + |\xi_m - \xi| < \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon$. That is, the sequence is Cauchy. (The statement 0.12 in a general metric space has the same proof, using the triangle inequality for the metric).

Now suppose that the sequence (ξ_n) is Cauchy. Firstly, take "$\epsilon := 1$" in the definition; there is N such that $|\xi_n - \xi_m| < 1$ for all $m, n \geq N$. This implies (fix m to be N) that

$$\xi_N - 1 \leq \xi_n \leq \xi_N + 1 \quad \text{for all } n \geq N\,.$$

The sequence is therefore bounded (an upper bound is $\max\{\xi_1, \xi_2, \ldots, \xi_{N-1}, \xi_N + 1\}$, and a lower bound is $\min\{\xi_1, \xi_2, \ldots, \xi_{N-1}, \xi_N - 1\}$). So $\underline{\lim}\, \xi_n, \overline{\lim}\, \xi_n$ are defined in $\mathbb{R}$.

Now suppose ϵ is an arbitrary positive number. Then, there exists N' such that

$$n \geq N' \Longrightarrow \xi_{N'} - \epsilon < \xi_n < \xi_{N'} + \epsilon\,.$$

Hence, $\xi_{N'} - \epsilon$ is a lower number, $\xi_{N'} + \epsilon$ is an upper number, and so, by 7.6,

$$\xi_{N'} - \epsilon \leq \underline{\lim}\,\xi_n \leq \overline{\lim}\,\xi_n \leq \xi_{N'} + \epsilon\,,$$

from which $0 \leq \overline{\lim}\,\xi_n - \underline{\lim}\,\xi_n \leq 2\epsilon$. This conclusion must hold for any $\epsilon > 0$; so the sequence is order-convergent. Apply 7.12. □

I have developed the proofs above in a rather leisurely fashion in the hope of emphasizing the concepts involved, for we shall have other uses for upper and lower limits. But there are many other possible proofs of 0.11. For instance, it is easy to prove that, if a Cauchy sequence (in a metric space) has a *subsequence* convergent to a point x, then the whole sequence converges to x. As above, a real Cauchy sequence is bounded; so one need only prove that

$$\text{a bounded sequence has a convergent subsequence.} \tag{25}$$

This is also included in 7.6. (If one started from a different point, for instance from Cantor's definition of the reals, (25) might be a consequence of the Bolzano-Weierstraß theorem).

Definition 7.14. Let Ω be any set, and $f : \Omega \longrightarrow [0, \infty]$ a nonnegative extended-real-valued function. The *ordinate sets* of f are

$$\underline{\Gamma}(f) \coloneqq \{(x, \xi) \in \Omega \times [0, \infty] : 0 \leq \xi < f(x)\}\,,$$
$$\overline{\Gamma}(f) \coloneqq \{(x, \xi) \in \Omega \times [0, \infty] : 0 \leq \xi \leq f(x)\}\,.$$

In other words, $\underline{\Gamma}(f)$ is the set of points *strictly underneath* the graph of f, and $\overline{\Gamma}(f)$ the set of points *under or on* the graph of f. These are not standard notations.

Remark 7.15. The class $\mathcal{F}$ of functions $f : \Omega \longrightarrow \overline{\mathbb{R}}$ can itself be given a partial order, if we define "$f \leq g$" to mean that $(\forall x \in \Omega)\ f(x) \leq g(x)$. (This may be called the 'pointwise order'; notice that it is *not* a total order, unless Ω is a singleton or $\emptyset$.) Then $\mathcal{F}$ is order-complete. It is easy to see that, if $\mathcal{X}$ is a subset of $\mathcal{F}$, its supremum is the function g, where

$$(\forall x \in \Omega) \quad g(x) \coloneqq \sup\{f(x) : f \in \mathcal{X}\}\,.$$

(g is the 'pointwise supremum' of $\mathcal{X}$.) The infimum of $\mathcal{X}$ is similarly constructed pointwise.

We may therefore define the upper and lower limits of sequences of functions $\Omega \longrightarrow \overline{\mathbb{R}}$ according to the order in the class of functions, and they agree with the "pointwise limits". Indeed, it is only necessary to know for this purpose that *countable* classes of functions have suprema and infima, a remark of some later significance.

Lemma 7.16. *Let (f_n) be a sequence of nonnegative extended-real-valued functions. Then*

$$\underline{\Gamma}(\liminf f_n) \subseteq \liminf \underline{\Gamma}(f_n)\,, \quad \limsup \overline{\Gamma}(f_n) \subseteq \overline{\Gamma}(\limsup f_n)\,.$$

Proof.
$$\begin{aligned}(x, \xi) \in \underline{\Gamma}(\underline{\lim}\, f_n) &\Longleftrightarrow 0 \leq \xi < (\underline{\lim}\, f_n)(x) \quad (\underline{\lim} \text{ in } \mathcal{F}) \\ &\Longleftrightarrow 0 \leq \xi < \underline{\lim}\,(f_n(x)) \quad (\underline{\lim} \text{ in } \overline{\mathbb{R}}).\end{aligned}$$

By 7.6, $\xi < \underline{\lim}\,(f_n(x))$ means that $\frac{1}{2}(\xi + \underline{\lim}\,(f_n(x)))$ is a lower number for $(f_n(x))$, which in turn means (by the definition 7.5) that there exists N for which $f_n(x) \geq \frac{1}{2}(\xi + \underline{\lim}\,(f_n(x)))$ whenever $n \geq N$. Hence, for $n \geq N$,

$$f_n(x) \geq \tfrac{1}{2}(\xi + \underline{\lim}\,(f_n(x))) > \xi\,, \quad (x,\xi) \in \underline{\Gamma}(f_n)\,.$$

That is, $(x,\xi) \in \underline{\Gamma}(f_n)$ except for finitely many indices, $(x,\xi) \in \liminf \underline{\Gamma}(f_n)$. This shows that $\underline{\Gamma}(\underline{\lim} f_n) \subseteq \underline{\lim}\,\underline{\Gamma}(f_n)$. The other result is proved similarly (or by "duality"). □

§8. Measurable and measure spaces.

Definition 8.1. A *measurable space* is a pair (Ω, Σ) consisting of a set Ω and a σ-algebra Σ of subsets of Ω. Σ may be described as the *measurable structure* on Ω. A subset E of Ω is described as *measurable* with respect to the given measurable structure Σ, or as Σ-*measurable*, if it is a member of Σ; and Σ may also be described as the σ-algebra (or algebra) of measurable sets in Ω.

This is the modern convention: we specify a σ-field of "measurable" sets, as it were by decree. Of course one important way of doing so is via an outer measure μ^* in Ω, but I have been careful to speak in that case of μ^*-measurable sets. Very often, the σ-field Σ is fixed by the context and one talks of measurable sets without further qualification; for instance, when dealing with sets in $\mathbb{R}^n$, one often describes them as measurable without specifying that they are measurable with respect to Lebesgue outer measure.

The definition is analogous to the definition of a topological space, and the ambiguities of terminology are also similar. The *idea* of a topological space arises from the notion of "open set" in $\mathbb{R}^n$, but in the modern version "open sets" are just members of a "topology" — in effect, they are open by decree. There are thus two conventions in operation, a historical one and a modern one. In practice this rarely causes difficulty, but you should be aware of it.

Definition 8.2. A *measure space* is a triple (Ω, Σ, μ), where Σ is a σ-field of subsets of Ω and $\mu : \Sigma \longrightarrow \overline{\mathbb{R}}$ is a measure. A *signed measure space* is a triple (Ω, Σ, σ), where Σ is a σ-field of subsets of Ω and $\sigma : \Sigma \longrightarrow \overline{\mathbb{R}}$ is a signed measure.

Much of the following theory requires only that a fixed measure space should be given, although in probability theory the measurable structure on Ω can sometimes be variable. Of course, we often say "let Ω be a measure space", taking the notations Σ and μ as read.

Lemma 8.3. If Ω is a measure space and $(E_i)_{i=1}^{\infty}$ is a sequence in Σ, then

$$\mu\Big(\bigcup_{i=1}^{\infty} E_i\Big) \leq \sum_{i=1}^{\infty} \mu(E_i)\,.$$

Proof. "Disjunctify" (M_i) to (N_i), as at 4.7, so that $N_i \subseteq M_i$ for each i and $\bigcup_{i=1}^{\infty} N_i = \bigcup_{i=1}^{\infty} M_i$. By 5.9 and 2.14, the N_i being pairwise disjoint,

$$\mu\Big(\bigcup_{i=1}^{\infty} M_i\Big) = \mu\Big(\bigcup_{i=1}^{\infty} N_i\Big) = \sum_{i=1}^{\infty} \mu(N_i) \leq \sum_{i=1}^{\infty} \mu(M_i)\,. \qquad \square$$

This is of course the same reasoning as I presented in 4.7, but with slightly different hypotheses. (We could have extended μ to $\mu^\dagger$ and applied the remark in 4.7 instead.)

Remark 8.4. Suppose that (Ω, Σ, σ) is a signed measure space, and let $A, B \in \Sigma$, where $B \subseteq A$. As σ is finitely additive, $\sigma(A) = \sigma(B) + \sigma(A \setminus B)$. If $\sigma(B)$ is finite, one may deduce from this that $\sigma(A \setminus B) = \sigma(A) - \sigma(B)$, this including the possibility that $\sigma(A)$ may be infinite (in which case $\sigma(A \setminus B) = \sigma(A)$). If $\sigma(B)$ is infinite, then $\sigma(A) = \sigma(B)$ and the only information on $\sigma(A \setminus B)$ is that it cannot be the opposite infinity (see 5.8).

Now recall 7.11.

Proposition 8.5. *Let (Ω, Σ, σ) be a signed measure space. Suppose that $(M_n)_{n=1}^\infty$ is a sequence in Σ.*

(a) *If (M_n) is increasing, then, as $n \to \infty$, $\sigma(M_n) \to \sigma(\lim M_n)$.*

(b) *If (M_n) is decreasing and, for some index k, $\sigma(M_k)$ is finite, then $\sigma(M_n) \to \sigma(\lim M_n)$ as $n \to \infty$.*

The conclusion in both cases is that "σ commutes with limits": $\sigma(\lim M_n) = \lim \sigma(M_n)$. The finiteness restriction in *(b)* is unavoidable; consider $M_n := [n, \infty)$, which has Lebesgue measure (length) ∞ for all n, but whose limit is $\emptyset$, of measure 0.

Proof. *(a)* Disjunctify, setting $N_1 := M_1$, $N_{n+1} := M_{n+1} \setminus M_n$. Then, as (N_n) is a disjoint sequence in Σ and σ is countably additive,

$$\begin{aligned}\sigma(\lim M_n) &= \sigma\Big(\bigcup_{n=1}^\infty M_n\Big) = \sigma\Big(\bigcup_{n=1}^\infty N_n\Big) = \sum_{n=1}^\infty \sigma(N_n)\\ &= \lim_{r\to\infty} \sum_{n=1}^r \sigma(N_n) \quad \text{by the definition of the "sum"}\\ &= \lim_{r\to\infty} \sigma(M_r) \qquad\qquad \text{again as } \sigma \text{ is } \sigma\text{-additive.}\end{aligned}$$

(b) The sequence $(M_k \setminus M_r)_{r=k}^\infty$ is increasing, with limit (in Σ)

$$\bigcup_{r=k}^\infty (M_k \setminus M_r) = M_k \setminus \Big(\bigcap_{r=k}^\infty M_r\Big),$$

and so, by *(a)*,

$$\sigma\Big(M_k \setminus \Big(\bigcap_{r=k}^\infty M_r\Big)\Big) = \lim_{r\to\infty} \sigma(M_k \setminus M_r). \tag{26}$$

As $\sigma(M_k)$ is finite, 5.8 shows that, for any $r \geq k$, $\sigma(M_r)$, $\sigma(M_k \setminus M_r)$, $\sigma(\bigcap_{r=k}^\infty M_r)$ are all finite. From 8.4,

$$\begin{aligned}\sigma(M_k \setminus M_r) &= \sigma(M_k) - \sigma(M_r) \quad \text{for each } r \geq k, \text{ and}\\ \sigma\Big(M_k \setminus \Big(\bigcap_{r=k}^\infty M_r\Big)\Big) &= \sigma(M_k) - \sigma\Big(\bigcap_{r=k}^\infty M_r\Big). \quad \text{Apply (21):}\\ \sigma(M_k) - \sigma\Big(\bigcap_{r=k}^\infty M_r\Big) &= \lim_{r\to\infty} (\sigma(M_k) - \sigma(M_r)).\end{aligned}$$

Since all the terms are finite, it follows that $\sigma(M_r)$ converges to $\sigma(\bigcap_{r=k}^\infty M_r)$. However, $\bigcap_{r=k}^\infty M_r = \bigcap_{r=1}^\infty M_r = \lim M_n$. □

The asymmetry between the increasing and the decreasing cases has substantial consequences in the later development.

Lemma 8.6. *Suppose* (Ω, Σ, μ) *is a measure space, and* (M_n) *is a sequence in* Σ*. Then*

$$\mu(\liminf M_n) \leq \liminf \mu(M_n)\,.$$

Proof. For each $k \in \mathbb{N}$ and each $r \geq k$, $\bigcap_{n=k}^{\infty} M_n \subseteq M_r$, so $\mu(\bigcap_{n=k}^{\infty} M_n) \leq \mu(M_r)$ by 5.9. But, therefore, $\mu(\bigcap_{n=k}^{\infty} M_n) \leq \inf_{r\geq k} \mu(M_r)$. On the other hand, $(\bigcap_{n=k}^{\infty} M_n)_{k=1}^{\infty}$ is an increasing sequence in Σ, so, by 8.5,

$$\begin{aligned}\mu(\liminf M_n) &= \mu\Big(\bigcup_{k=1}^{\infty}\Big(\bigcap_{n=k}^{\infty} M_n\Big)\Big) = \lim \mu\Big(\bigcap_{n=k}^{\infty} M_n\Big)\\ &= \sup_{k\in\mathbb{N}} \mu\Big(\bigcap_{n=k}^{\infty} M_n\Big) \leq \sup_{k\in\mathbb{N}} \inf_{r\geq k} \mu(M_r)\\ &= \liminf \mu(M_n)\,, \quad \text{as asserted.}\end{aligned}$$

□

Lemma 8.7. *Suppose in* 8.6 *that there exists an index* N *and some set* $A \in \Sigma$ *for which* $\mu(A) < \infty$ *and* $M_n \subseteq A$ *whenever* $n \geq N$*. Then*

$$\limsup \mu(M_n) \leq \mu(\limsup M_n)\,.$$

Proof. By 7.7, I may omit the terms $M_1, M_2, \ldots, M_{N-1}$ without affecting the limits, and so I may assume without loss of generality that $M_n \subseteq A$ for all n. Then, using 7.3,

$$\begin{aligned}\limsup \mu(M_n) &= \mu(A) - \liminf (\mu(A) - \mu(M_n))\\ &= \mu(A) - \liminf \mu(A \setminus M_n)\\ &\leq \mu(A) - \mu(\liminf (A \setminus M_n)) \quad \text{by 8.6}\\ &= \mu(A \setminus \liminf (A \setminus M_n)) = \mu(\limsup M_n)\,.\end{aligned}$$

□

Corollary 8.8. *Let* (Ω, Σ, μ) *be a measure space, and* (M_n) *a sequence in* Σ *such that* $\mu(\bigcup_{n=1}^{\infty} M_n) < \infty$*. If the sequence* (M_n) *is convergent, then* $\mu(\lim M_n) = \lim \mu(M_n)$*.*

Proof.
$$\begin{aligned}\mu(\underline{\lim}\, M_n) &\leq \underline{\lim}\, \mu(M_n) \quad \text{by 8.6}\\ &\leq \overline{\lim}\, \mu(M_n) \quad \text{by 7.6}\\ &\leq \mu(\overline{\lim}\, M_n) \quad \text{by 8.7,}\end{aligned}$$

and the hypothesis says the ends of this chain of inequalities are equal. □

The finiteness hypothesis is needed (why?). However, one obvious way of satisfying it leads to an unexpectedly strong conclusion. This is the *first Borel-Cantelli lemma*. (The second Borel-Cantelli lemma involves the notion of independence, so is more explicitly probabilistic.)

Lemma 8.9. *Let* (Ω, Σ, μ) *be a measure space, and suppose that* (M_n) *is a sequence in* Σ *such that* $\sum_{n=1}^{\infty} \mu(M_n) < \infty$*. Then* $\mu(\overline{\lim}\, M_n) = 0$*.*

Proof. Given $\epsilon > 0$, there exists N such that $\sum_{n=N}^{\infty} \mu(M_n) < \epsilon$. But now

$$\mu(\overline{\lim}\, M_n) = \mu\Big(\bigcap_{k=1}^{\infty}\Big(\bigcup_{n=k}^{\infty} M_n\Big)\Big) \le \mu\Big(\bigcup_{n=N}^{\infty} M_n\Big) \le \sum_{n=N}^{\infty} \mu(M_n) < \epsilon\,,$$

so that $0 \le \mu(\overline{\lim}\, M_n) < \epsilon$, for any $\epsilon > 0$. Hence $\mu(\overline{\lim}\, M_n) = 0$. □

This result is not surprising if one recalls that $\overline{\lim}\, M_n$ consists of the points that appear in infinitely many M_n, and so are "counted infinitely often" in $\sum_{n=1}^{\infty} \mu(M_n)$.

§9. Lebesgue-Stieltjes measures in one dimension.

The last two sections, although they have introduced a number of ideas that will be significant later, have not contributed anything directly to the main question whether we can find interesting examples of measures — that is, measures defined on large σ-fields and having plenty of finite positive values. Here, at last, we shall construct a substantial class of such measures in $\mathbb{R}$ (or in a subinterval of $\mathbb{R}$, though that is a rather trifling generalization).

J will denote a non-null interval in $\mathbb{R}$ that is open on the left; that is, of the form (a, b) or $(a, b]$, where $b > a$ and a may be $-\infty$. (For $(a, b]$, we also assume $b < \infty$).

Definition 9.1. Let $f : J \longrightarrow \mathbb{R}$ be a function, and $a \in J$. f is *continuous on the right at* a (or *right-continuous at* a) when

$$(\forall \epsilon > 0)(\exists \delta > 0) \quad x \in J \cap (a, a + \delta) \Longrightarrow |f(x) - f(a)| < \epsilon\,.$$

Equivalently, f is right-continuous at a if and only if *either* a is the right-hand end-point of J (there may be no right-hand end-point), *or* $\lim_{x \downarrow a} f(x)$ is defined and equal to $f(a)$.

This is the usual definition of continuity at a point a, except that attention is restricted to values of x to the right of a.

Definition 9.2. A function $f : J \longrightarrow \mathbb{R}$ is a *Lebesgue-Stieltjes distribution function on* J if

(a) it is increasing; that is, whenever $x, y \in J$ and $x < y$, then $f(x) \le f(y)$; and

(b) it is right-continuous at each point of J (or "right-continuous on J").

I shall abbreviate "Lebesgue-Stieltjes distribution function" to d.f.

Example 9.3. *(a)* The *Lebesgue distribution function* is given by $(\forall x \in \mathbb{R})\ f(x) = x$. This is, of course, overwhelmingly the most important example. Other uncomplicated examples are

$$f(x) = x^3\,, \quad f(x) = \exp x\,, \quad f(x) = \tan^{-1} x\,.$$

These are all *strictly* increasing, continuous, and even differentiable on the whole of $\mathbb{R}$.

(b) Set $f(x) := -\dfrac{1}{x}$ for $x > 0$. This is a d.f. on $(0, \infty)$, which cannot be extended further to the left. It is strictly increasing and differentiable.

(c) Set $f(x) = 0$ for $x \in (-1, 0)$ and $f(x) = 1$ for $x \in [0, 1]$. This is a d.f. on $(-1, 1]$. It is only non-decreasing, and is discontinuous at 0.

(d) Set $f(x) = [x]$, the integer part of x. This is a d.f. on $\mathbb{R}$. It is important to notice that an integer is its own integer part, so that $[x]$ is a right-continuous function.

Definition 9.4. Let $\mathcal{S}(J)$ be the class of bounded subintervals of J of the form $(a, b]$, where a is greater than the left-hand end-point of J; that is, they are to be closed on the right and open on the left, and their closures in $\mathbb{R}$ are to be included in J.

$\mathcal{S}(J)$ is a semiring in J. Indeed, the difference of two members of $\mathcal{S}(J)$ is the disjoint union of at most two members of $\mathcal{S}(J)$.

Definition 9.5. Let f be a Lebesgue-Stieltjes distribution function on J. The corresponding *Lebesgue-Stieltjes weighting function* $\tau_f : \mathcal{S} \longrightarrow \mathbb{R}$ is defined by

$$(\forall (a, b] \in \mathcal{S}) \quad \tau_f((a, b]) = f(b) - f(a).$$

As f is increasing, τ_f is nonnegative-valued. Then $\tau_f^\dagger$ (see 3.3) is the *Lebesgue-Stieltjes outer measure in J induced by f*. When f is the Lebesgue d.f., $\tau_f^\dagger$ is called the *Lebesgue outer measure in J*. The measure space (J, Σ_f, μ_f) in which Σ_f is the σ-field of $\tau_f^\dagger$-measurable subsets of J and $\mu_f := \tau_f^\dagger | \Sigma_f$ is the *Lebesgue-Stieltjes measure space induced by the d.f. f*.

The σ-algebra Σ_f does depend on f. For instance, it is easily proved that, for the d.f.s of 9.3*(c)* and *(d)*, $\Sigma_f = \mathcal{P}(J)$. We shall see that this cannot be true for the Lebesgue d.f. On the other hand, τ_f may easily be seen to be a fam on $\mathcal{S}$ (see 5.6), and therefore by 6.1 $\Sigma_f \supseteq \Sigma'(\mathcal{R}(\mathcal{S})) = \Sigma'(\mathcal{S})$. Thus *all* Lebesgue-Stieltjes measures are defined on $\Sigma'(\mathcal{S})$, which is in fact the Borel σ-field in J (see below, 9.10). This raises the possibility of comparing them, as measures on $\mathcal{B}(J)$.

The thing still lacking is information on the *values* of μ_f. This will be provided by 6.3 and 6.4, if we can prove that τ_f is countably additive on $\mathcal{S}$ and not just finitely additive. Here is where the right-continuity of f is required, and the theory at last comes down to earth.

Theorem 9.6. *Let $a, b \in \mathbb{R}$ and $a \leq b$. Suppose given any class $\{(c_\alpha, d_\alpha) : \alpha \in A\}$ of open intervals such that $[a, b] \subseteq \bigcup_{a \in A} (c_\alpha, d_\alpha)$. Then there exists a finite subset B of the index set A such that $[a, b] \subseteq \bigcup_{a \in B} (c_\alpha, d_\alpha)$.*

A more modern formulation is that *any covering of a bounded closed interval by open intervals admits a finite subcovering*. In intuitive terms, most — all but a finite number — of the open intervals (c_α, d_α) are redundant for the purpose of covering $[a, b]$. This is the original observation of Heine, when he proved that a continous function on a closed bounded interval is uniformly continuous; Borel was the first to state it explicitly, and, in a generalized statement, it is the *Heine-Borel property* of certain subsets of a metric space. Finally it became the definition of a *compact set* in a topological space: a subset X of the topological space Ω is *compact* if any covering of X by open sets of Ω admits a finite subcovering (i.e. only finitely many of the open sets in the covering are really needed). It is an extremely important property with many equivalent forms, as you will know if you have done 312. Its importance lies in its being a "topological version of finiteness", and the curious thing, of course, is that interesting compact sets (that is, ones that are not finite) should exist at all.

Undergraduate folklore when I was a student said the proof I am about to give was invented by a candidate in an examination, faced by a question that expected the "standard"

proof presented in lectures. It is, indeed, rather simple by comparison with the somewhat messy arguments that are given in old textbooks.

Proof. Say that a point $x \in [a,b]$ is *reachable* if there is a finite subset X of A for which $[a,x] \subseteq \bigcup_{\alpha\in X}(c_\alpha, d_\alpha)$. ("You only need finitely many of the given open intervals to reach x from a".) Evidently a is reachable (a belongs to *one* of the open intervals). So the set of reachable points is non-empty and bounded above by b. By Dedekind's axiom, it has a supremum $q \in [a,b]$. Now there is some index $\alpha \in A$ for which $q \in (c_\alpha, d_\alpha)$. Since $c_\alpha < q$, 0.6*(ii)* applies, and there is a reachable point r for which $c_\alpha < r \leq q$.

Let $q' := \frac{1}{2}(q + \min(b, d_\alpha)) \in (c_\alpha, d_\alpha) \cap [a,b]$. Then q' is also reachable (r may be reached by finitely many of the given intervals, and the one additional interval (c_α, d_α) suffices to reach q'), and $q' \geq q$. As q was the supremum of the reachable points, necessarily $q = q'$, which, since $d_\alpha > q$ and $b \geq q$, can only occur if $q = q' = b$. We conclude that b is reachable, which is just what is desired. □

Proposition 9.7. *Let $f : J \longrightarrow \mathbb{R}$ be a Lebesgue-Stieltjes distribution function. Then the associated weighting function $\tau_f : \mathcal{S} \longrightarrow \mathbb{R}$ is countably additive on $\mathcal{S}$.*

Proof. As already remarked (after 9.5), τ_f is a fam on $\mathcal{S}$.

Suppose that $(a,b] \in \mathcal{S}$ is expressible as the disjoint union of a sequence of sets in $\mathcal{S}$:

$$(a,b] = \bigcup_{k=1}^{\infty}(a_k, b_k]\,.$$

Here $a < b$ and $a_k < b_k$ for each k. For any $n \in \mathbb{N}$, $(a,b] \supseteq \bigcup_{k=1}^{n}(a_k, b_k]$, and if necessary we may re-index these intervals so that $a \leq a_1 < a_2 < a_3 < \cdots < a_n$. (It is not possible for two left-hand end-points to coincide, if the corresponding intervals are disjoint). But then, to ensure disjointness, $a_1 < b_1 \leq a_2 < b_2 \leq a_3 < \cdots < b_{n-1} \leq a_n < b_n \leq b$, and, as f is increasing,

$$\begin{aligned}\tau_f((a,b]) &= f(b) - f(a) \geq f(b_n) - f(a_1)\\ &\geq f(b_n) - f(a_n) + f(b_{n-1}) - \cdots - f(a_2) + f(b_1) - f(a_1)\end{aligned}$$

(because $f(a_n) \geq f(b_{n-1}), \ldots, f(a_2) \geq f(b_1)$). Of course the re-indexing does not affect the sum, so, for any n,

$$\tau_f((a,b]) \geq \sum\nolimits_{k=1}^{n} \tau_f((a_k, b_k])\,.$$

This being so for all n, we also have (the sum is the supremum over n)

$$\tau_f((a,b]) \geq \sum\nolimits_{k=1}^{\infty} \tau_f((a_k, b_k])\,. \tag{27}$$

The difficulty in establishing the contrary inequality is that one cannot usually reorder *all* the end-points as a monotonic infinite sequence. For instance, one might have the intervals $(2^{-n} + 2^{-n-m}, 2^{-n} + 2^{-n-m+1}]$ for $m, n \in \mathbb{N}$; they constitute a countable disjoint covering of $(0,1]$, but their end-points cannot be set out in a monotonic sequence. (Whether it is to be a decreasing or an increasing sequence, it will never go below the largest number 2^{-n} that is less than the first term listed.)

Suppose that $\epsilon > 0$. As f is right-continuous at a, there exists $a' \in (a,b)$ such that $f(a') - f(a) < \frac{1}{2}\epsilon$, and, for each k such that b_k is not the right-hand point of J, there exists $b_k' > b_k$, with $b_k' \in J$, such that $f(b_k') < f(b_k) + 2^{-k-1}\epsilon$. If b_k should be the right-hand

end-point of J, one may extend the definition of f so that, for any $x > b_k$, $f(x) = f(b_k)$, and let b'_k be any number greater than b_k. Now

$$[a', b] \subseteq (a, b] = \bigcup_{k=1}^{\infty} (a_k, b_k] \subseteq \bigcup_{k=1}^{\infty} (a_k, b'_k) .$$

By the Heine-Borel property 9.6, only finitely many of the open intervals (a_k, b'_k) are needed to cover $[a', b]$. Let N be the largest index that appears in this finite list:

$$[a', b] \subseteq \bigcup_{k=1}^{N} (a_k, b'_k) .$$

Reject successively from the sequence $(a_k, b'_k)_{k=1}^{N}$ of intervals any term whose intersection with $[a', b]$ is included in the union of those remaining. After finitely many steps, what is left is an "irredundant" covering, from which no further member can be removed without leaving some point of $[a', b]$ uncovered. (Convince yourself this procedure can be performed with the stated effect). Suppose this has been done; I may still denote the number of remaining intervals by N. Reorder them so that $a_1 \leq a_2 \leq \cdots \leq a_N$. Then

$$\left.\begin{array}{c} a_1 < a' \leq a_2 < a_3 < \cdots < a_N , \\ b'_1 < b'_2 < \cdots < b'_{N-1} \leq b < b'_N , \\ a_2 < b'_1 ,\ a_3 < b'_2 , \ldots ,\ a_N < b'_{N-1} . \end{array}\right\} \tag{28}$$

Each of these conditions must be satisfied if the intervals form an irredundant covering.

As f is increasing, the conditions (28) imply that

$$\begin{aligned} \sum_{k=1}^{N} (f(b'_k) - f(a_k)) &= f(b'_N) - f(a_1) + \sum_{k=1}^{N-1} (f(b'_k) - f(a_{k+1})) \\ &\geq f(b'_N) - f(a_1) \geq f(b) - f(a') . \end{aligned}$$

Hence,

$$\begin{aligned} f(b) - f(a) &< f(b) - f(a') + \tfrac{1}{2}\epsilon \\ &\leq \sum_{k=1}^{N} (f(b'_k) - f(a_k)) + \tfrac{1}{2}\epsilon \\ &< \sum_{k=1}^{N} (f(b_k) + 2^{-k-1}\epsilon - f(a_k)) + \tfrac{1}{2}\epsilon \\ &< \sum_{k=1}^{N} (f(b_k) - f(a_k)) + \epsilon \\ &\leq \sum_{k=1}^{\infty} (f(b_k) - f(a_k)) + \epsilon . \end{aligned}$$

But ϵ was an *arbitrary* positive number, so

$$\tau_f((a, b]) = f(b) - f(a) \leq \sum_{k=1}^{\infty} (f(b_k) - f(a_k)) = \sum_{k=1}^{N} \tau_f((a_k, b_k]) . \tag{29}$$

The inequalities (27) and (29) prove the result. □

In a vague philosophical sense, the result just given is the heart of the matter. It is the compactness of bounded closed intervals that allows us to prove countable additivity, and thus to construct measures whose values on intervals are known.

Theorem 9.8. *Given a Lebesgue-Stieltjes distribution function $f : J \longrightarrow \mathbb{R}$, there is a measure space (J, Σ_f, μ_f) such that $\mathcal{S} \subseteq \Sigma_f$ and $\mu_f((a, b]) = f(b) - f(a)$ for any $(a, b] \in \mathcal{S}$.*

Proof. 9.7, 6.4, and 6.3. □

Certainly the σ-field Σ_f in J includes the σ-field $\Sigma(\mathcal{S})$ in J generated by $\mathcal{S}$. Recall that $\Sigma'(\mathcal{S})$ denotes the *σ-ring* generated by $\mathcal{S}$, and that $\mathcal{B}(J)$ is the Borel σ-field in J (the σ-field in J generated by the relatively open sets in J).

Lemma 9.9. *Any (relatively) open set in J is a countable disjoint union of (relatively) open intervals in J.*

Proof. Let U be a (relatively) open set in J. By definition, each rational point $\xi \in U$ belongs to a (relatively) open interval included in U. The union of all such intervals must be the greatest possible relatively open interval $I(\xi)$ containing ξ and included in U. Distinct intervals $I(\xi)$ are disjoint, for if two of them met, their union would be a relatively open interval bigger than either. Any $x \in U$ is contained in some relatively open interval $I_1(x)$ included in U; $I_1(x)$ must contain a rational ξ, and then $I_1(x) \subseteq I(\xi)$. Therefore, $U = \bigcup_{\xi \in U \cap \mathbb{Q}} I(\xi)$, which means U is a countable disjoint union of open intervals. □

Lemma 9.10. $\mathcal{B}(J) = \Sigma'(\mathcal{S}) = \Sigma(\mathcal{S})$.

Proof. Firstly, $\mathcal{S} \subseteq \mathcal{B}(J)$. Take a typical member $(a, b]$ of $\mathcal{S}$. Then

$$(a, b] = \bigcap_{k=1}^{\infty} (J \cap (a, b + 2^{-k})),$$

so that $(a, b]$ is a countable intersection of open sets in J. As $\mathcal{B}(J)$ is a σ-algebra, it follows that $\Sigma'(\mathcal{S}) \subseteq \Sigma(\mathcal{S}) \subseteq \mathcal{B}(J)$.

For the contrary inclusion, it will suffice to show that any open set in the subspace topology on J, including J itself, belongs to $\Sigma'(\mathcal{S})$. Then the σ-ring generated by the open sets will automatically be a σ-algebra, and must be $\mathcal{B}(J)$; and, as $\Sigma'(\mathcal{S})$ is a σ-ring, $\Sigma'(\mathcal{S}) \supseteq \mathcal{B}(J)$. For simplicity I shall deal only with one case, where $J := (c, \infty)$ and $\infty > c > -\infty$; the other possibilities may be dealt with very similarly (with minor complications), or by proving some general theorems (some of these are in the exercise sets).

Because of 9.9, I need only prove that any open *interval* $(a, b) \subseteq J$ belongs to $\Sigma'(\mathcal{S})$.

Firstly, suppose $b < \infty$. Then $(a, b) = \bigcup_{k=1}^{\infty} (a + 2^{-k}, b - 2^{-k}]$, where, of course, some of the intervals may be null, but they all belong to $\mathcal{S}$.

If $b = \infty$, $(a, \infty) = \bigcup_{k=1}^{\infty} (a + 2^{-k}, 2^k]$, where again all the intervals belong to $\mathcal{S}$.

Thus, in fact, any open set in J may be expressed as a countable union of elements of $\mathcal{S}$; which is sufficient. □

The point of this Lemma is that all Lebesgue-Stieltjes measures on J are defined on the Borel σ-algebra, which is very large — as we shall see, it contains "all the sets one normally needs". In particular, it contains all singletons and all intervals (open, closed, or half-open).

Definition 9.11. A measure which is defined on the Borel σ-algebra in a topological space Ω is called a *Borel measure* in Ω.

Thus $\mu_f \big| \mathcal{B}(J)$ is a Borel measure in J, for any d.f. f in J. "Borel measure" in older books means the restriction of Lebesgue measure to the Borel sets. This is for historical reasons — Borel noticed that the idea of "length" could be extended step by step to ever more complicated sets, before Lebesgue produced a more general procedure.

Remark 9.12. In metric space topology as it was developed until about 1940, and as it is expounded in Kuratowski's book or Hausdorff's, the notion of metric ball took precedence over the notion of a topology, and the class of open sets (i.e. the topology) in a metric space Ω was often denoted $\mathcal{G}$. The class of closed sets was written $\mathcal{F}$. (Munroe suggests, not implausibly, that $\mathcal{G}$ was suggested by "Gebiet" and $\mathcal{F}$ by "fermé".) Then the subscripts σ and δ were used to suggest countable unions and intersections.

Given $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, $\mathcal{A}_\delta$ denotes the class of subsets of Ω that may be expressed as the intersection of a sequence of sets all belonging to $\mathcal{A}$, and $\mathcal{A}_\sigma$ the class of subsets that may be expressed as the union of a sequence of sets all belonging to $\mathcal{A}$. Evidently $\mathcal{G} = \mathcal{G}_\sigma$ and $\mathcal{F} = \mathcal{F}_\delta$, and it is not difficult to show that, in a metric space, $\mathcal{G} \subseteq \mathcal{F}_\sigma$ and $\mathcal{F} \subseteq \mathcal{G}_\delta$. Much less obviously, the sequence $\mathcal{G}, \mathcal{G}_\delta, (\mathcal{G}_\delta)_\sigma, ((\mathcal{G}_\delta)_\sigma)_\delta, \dots$ (it is customary to abbreviate to $\mathcal{G}_{\delta\sigma}, \mathcal{G}_{\delta\sigma\delta}, \dots$) is *strictly* increasing in important cases. This is the background to my remarks after 5.13; the whole Borel σ-algebra $\mathcal{B}(\Omega)$ cannot be constructed by this inductive procedure (unless you pass to transfinite induction).

Borel's idea was to define the measure of sets in the classes $\mathcal{G}, \mathcal{G}_\delta, \mathcal{G}_{\delta\sigma}, \mathcal{G}_{\delta\sigma\delta}, \dots$ inductively by limiting procedures like 8.5.

Lemma 9.13. *Let f be a Lebesgue-Stieltjes distribution function in J. Then, if $a \in J$,*

$$\mu_f(\{a\}) = f(a) - \lim_{x \uparrow a} f(x)\,.$$

Proof. By 8.5*(b)*, as $\{a\} = \bigcap_{k=1}^{\infty} (J \cap (a - 2^{-k}, a])$ and, once k is large enough to ensure that $a - 2^{-k} \in J$, $\mu_f((a - 2^{-k}, a]) = f(a) - f(a - 2^{-k}) < \infty$, therefore

$$\begin{aligned}\mu_f(\{a\}) &= \lim \mu_f((a - 2^{-k}, a]) = \lim(f(a) - f(a - 2^{-k})) \\ &= f(a) - \lim f(a - 2^{-k}) = f(a) - \lim_{x \uparrow a} f(x)\,.\end{aligned}$$

(as f is increasing, it is easily seen that $\lim_{x \uparrow a} f(x)$ exists and is the same as $\lim_{n\to\infty} f(\xi_n)$ for any sequence $\xi_n \uparrow a$). □

Together with 9.8, this enables us to determine the value of μ_f on any interval. Notice that it tells us that the μ_f-measure of the singleton $\{a\}$ will be positive if and only if a is a point of discontinuity of f.

Remark 9.14. The Lebesgue-Stieltjes measure μ_f is defined at least on $\mathcal{B}(J)$. Quite often the fact that the construction of μ_f defines it on a significantly larger σ-field Σ_f is tacitly ignored; this slovenly custom has the justification that that larger σ-field is merely the "completion" of $\mathcal{B}(J)$ with respect to μ_f (although I have not proved it). It has also the crucial property that, for any closed bounded interval $[a, b] \subseteq J$,

$$\mu_f([a, b]) = f(b) - \lim_{x \uparrow a} f(x) < \infty\,.$$

(In topological terms, μ_f is finite on compact sets in J.)

It is not difficult to show that, since f is increasing, it has only countably many points of discontinuity. It is possible to carry out the construction of $\tau_f^\dagger$ without assuming right-continuity of f; then 9.7 fails. Suppose, in fact, that μ is any Borel measure in J which is finite on all closed bounded intervals in J, and, given $a \in J$, set

$$f(x) \coloneqq \begin{cases} \mu((a,x]) & \text{for } a \le x \in J\,, \\ -\mu((x,a]) & \text{for } a > x \in J\,. \end{cases}$$

This is a distribution function because of 8.5, and by the uniqueness theorem (which also I have not proved!) and 9.10, μ_f and μ agree on $\mathcal{B}(J)$. This is the reason for imposing the right-continuity condition in the definition of a d.f.

Overwhelmingly the most important example is Lebesgue measure. Notice that our arguments have proved that *there is a measure* λ (and by the uniqueness theorem only one) *defined for all Borel sets in* $\mathbb{R}$ *which agrees with ordinary length on intervals*. It is really rather startling that this is possible for a *countably additive* λ, without inconsistencies. For instance, the set $\mathbb{Q}$, which Jordan's theory could not handle, is a countable union of singletons, each of measure 0, so $\lambda(\mathbb{Q}) = \sum_{a\in\mathbb{Q}} \lambda(\{a\}) = \sum_{a\in\mathbb{Q}} 0 = 0$.

It is possible to extend the idea of Lebesgue-Stieltjes measures to signed measures, but the corresponding d.f.s and the whole construction require more subtlety.

§10. Non-measurable sets.

It is by no means clear which sets are $\tau_f^\dagger$-measurable for non-trivial distribution functions f. All Borel sets are, and it seems at least conceivable that all sets are, without restriction. In 1905, however, Vitali gave a simple construction of a set that is not Lebesgue-measurable. In 1908, Bernstein gave a much more demanding construction of a subset that is not measurable for *any* non-trivial Lebesgue-Stieltjes outer measure vanishing on singletons. It is on p. 422 of Kuratowski's vol. I (1958 French edition) or in Oxtoby's little book, p. 23. Here is Vitali's construction, which is sufficient for most purposes. Σ denotes the class of Lebesgue-measurable sets in $\mathbb{R}$, λ is Lebesgue measure.

Let me define, for any $x \in \mathbb{R}$, a mapping $R_x : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$(\forall t \in \mathbb{R}) \quad R_x(t) \coloneqq x + t\,.$$

R_x is usually called *translation by* x. It is clearly one-one and onto, with inverse R_{-x}.

In the construction of Lebesgue measure, every step is "translation-invariant". That is, for any $x, a, b \in \mathbb{R}$ and any $E \subseteq \mathbb{R}$,

$$\tau_f(R_x(a,b]) = \tau_f((a,b]) \quad \text{and so} \quad \tau_f^\dagger(R_x(E)) = \tau_f^\dagger(E)\,,$$

and, therefore, E is Lebesgue-measurable if and only if $R_x(E)$ is Lebesgue-measurable, and then they have the same Lebesgue measure. This is a property specific to Lebesgue measure.

Let $J \coloneqq (0,1]$, and, for any $x \in \mathbb{R}$, define $T_x : J \longrightarrow J$ by

$$(\forall t \in J) \quad T_x(t) \coloneqq x + t - [x+t] = (R_x(t))\,,$$

where $[x+t]$ denotes the integer part of $x+t$ and $(R_x(t))$ denotes the "fractional part" of $R_x(t)$. Thus T_x is "reduction of R_x modulo $\mathbb{Z}$". It is easily checked that

$$T_{x+y} = T_x \circ T_y\,, \quad T_0 = I \text{ (the identity map).}$$

[In effect, we are looking at the quotient group $\mathbb{R}/\mathbb{Z}$ and translations in it.]

Lemma 10.1. *Let* $J \supseteq A \in \Sigma$, $x \in J$. *Then* $T_x(A) \in \Sigma$ *and* $\lambda(T_x(A)) = \lambda(A)$.

Proof. A is the disjoint union $(A \cap (0, 1-x]) \cup (A \cap (1-x, 1])$, and since intervals are in Σ, both $A_1 \coloneqq A \cap (0, 1-x] \in \Sigma$ and $A_2 \coloneqq A \cap (1-x, 1] \in \Sigma$. However,

$$\begin{aligned} &\text{for } t \in A_1, \quad T_x(t) = x + t, \quad \text{whilst} \\ &\text{for } t \in A_2, \quad T_x(t) = x + t - 1. \end{aligned}$$

$T_x(A_1)$ and $T_x(A_2)$ are both Lebesgue-measurable, and they are disjoint, for

$$T_x(A_1) \subseteq T_x((0, 1-x]) = (x, 1], \quad T_x(A_2) \subseteq T_x((1-x, 1]) = (0, x].$$

Hence

$$\begin{aligned} T_x(A) = T_x(A_1 \cup A_2) = T_x(A_1) \cup T_x(A_2) \in \Sigma \quad &\text{and} \\ \lambda(T_x(A)) = \lambda(T_x(A_1)) + \lambda(T_x(A_2)) = \lambda(A_1) + \lambda(A_2) = \lambda(A). \quad &\square \end{aligned}$$

We can introduce an equivalence relation in J, writing $x \sim y$ when there is a *rational* number ξ such that $T_\xi x = y$. Notice that $T_\xi x = T_\eta x$ only when ξ, η differ by an integer. Each equivalence class is countable, as $\mathbb{Q}$ is. [In effect, the equivalence classes are the cosets of the subgroup $\mathbb{Q}/\mathbb{Z}$ in the group $\mathbb{R}/\mathbb{Z}$.] As J is uncountable, the number of equivalence classes must be uncountable (because of 1.2).

Choose one element from each equivalence class. Let the set thus constituted be W. (An algebraist would call it a transversal for the action of $\mathbb{Q}/\mathbb{Z}$ on $\mathbb{R}/\mathbb{Z}$, with one representative for each orbit — the orbits being the cosets.)

Lemma 10.2. $W \notin \Sigma$.

Proof. I assert that the sets $T_\xi(W)$ for $\xi \in \mathbb{Q} \cap J$ are all disjoint.

If $a \in T_\xi(W) \cap T_\eta(W)$, then $a = T_\xi(w) = T_\eta(w')$ for some $w, w' \in W$, and therefore $w' = T_{\xi-\eta}(w)$. That implies $w \sim w'$, which, by the definition of W, is only possible if $w = w'$. In turn, $T_\xi w = T_\eta w$ only if $\xi - \eta \in \mathbb{Z}$, which is impossible for $\xi, \eta \in \mathbb{Q} \cap J$ unless $\xi = \eta$. That is, if $T_\xi(W) \cap T_\eta(W) \neq \emptyset$, then necessarily $T_\xi(W) = T_\eta(W)$.

If $W \in \Sigma$, then each translate $T_\xi(W)$ is also in Σ. Hence, λ being σ-additive on Σ,

$$\lambda(J) = \lambda\Big(\bigcup\nolimits_{\xi \in \mathbb{Q} \cap J} T_\xi(W)\Big) = \sum\nolimits_{\xi \in \mathbb{Q} \cap J} \lambda(T_\xi(W)) = \sum\nolimits_{\xi \in \mathbb{Q} \cap J} \lambda(W). \tag{30}$$

There are two possibilities. If $\lambda(W) = 0$, (30) shows that $\lambda(J) = 0$; whilst, if $\lambda(W) > 0$, (30) shows that $\lambda(J) = \infty$. However, $\lambda(J) = 1$. The contradiction proves that W cannot be Lebesgue-measurable. $\square$

Remark 10.3. The argument shows that W, as constructed, cannot belong to any σ-field of sets in $\mathbb{R}$ that is invariant under rational translations, includes $\mathcal{B}(\mathbb{R})$, and admits a rational-translation-invariant measure that is positive and finite on $(0, 1]$. Since the Lebesgue-measurable sets form such a σ-algebra, W cannot be Lebesgue-measurable.

The first, and obvious, comment is that we have made essential use of the translation-invariance of the Lebesgue construction. This is less limiting than it may seem — once a non-measurable set has been found for Lebesgue measure, it may be manipulated in various ways to yield examples for other suitable measures. But Bernstein's construction is in this respect

much more general, being based on other (namely topological) properties of Lebesgue-Stieltjes measures; unfortunately, it is also far less straightforward.

Secondly, as $W \notin \Sigma$, one may extend Σ to a larger σ-field $\Sigma_1 := \Sigma(\Sigma \cup \{W\})$. It is possible to define a measure μ on Σ_1 such that $\mu|\Sigma = \lambda$. (This assertion is not entirely trivial, though not profound either.) The extension will have to lose the desirable properties that made the Vitali and Bernstein examples possible.

Remark 10.4. The ideal would be to have, for "useful" spaces Ω, "natural" measures defined on the whole of the power class $\mathcal{P}(\Omega)$. This ideal fails if the measure is to be translation-invariant, as a consequence of Vitali's example. However, one might still hope that (at least for Lebesgue measure in $\mathbb{R}$) the tedious apparatus of σ-fields might be avoided to some extent: maybe an "unnatural", non-translation-invariant, measure on $\mathcal{P}(\mathbb{R})$ could be constructed by extending λ step by step, as suggested in 10.3; all its practical applications would only involve Lebesgue measure. Unfortunately, there is a famous theorem of Ulam (1930) which says, in its basic version, that *if Ω is of cardinality $\aleph_1$, a measure defined on all of $\mathcal{P}(\Omega)$ and vanishing on singletons must be identically* 0. So, if we assume the *continuum hypothesis* that the cardinality of $\mathbb{R}$ is $\aleph_1$, *any* extension of λ to all of $\mathcal{P}(\mathbb{R})$ would have to be identically 0.

Vitali's construction was criticized very early. The offensive step was the definition of W by choosing one element from each equivalence class. Since there are uncountably many equivalence classes and no visible method to pick a special element from any of them, this must involve some version of the *Axiom of Choice*, which, in a fairly strong formulation, asserts that, if $\mathcal{C}$ is any set whose members are nonnnull sets, there exists a set E consisting of exactly one element from each $C \in \mathcal{C}$.

These days the Axiom of Choice is relatively uncontroversial, principally because Gödel proved in 1940 that it is consistent with the other usual axioms of the set theory he was using; that is to say, if the set theory itself is consistent (leads to no contradictions), adding the axiom of choice to the theory will not allow you to derive any contradictions either. (Mendelson in 1958 showed that the denial of the Axiom is also consistent with set theory.) But previously the Axiom was regarded with serious, and not entirely unfounded, suspicion. The reason was, I suppose, that it seemed to have consequences that are in a sense too good to be true, for instance the theorem that any vector space, over any field, has a basis. Now $\mathbb{R}$ is a vector space over $\mathbb{Q}$, but it seems quite impossible to specify, or even to imagine, a basis. Oversimplifying the matter, one might say that the problem is "naming"; any such basis must be uncountable, we lack *names* for so many objects, and other kinds of specification one can think of seem unlikely to work.

The effects of this suspicion can be seen in many older books. Littlewood always took care, in proofs where infinitely many choices were required, to prescribe (if possible) how to make them; see his lectures on the foundations of analysis. In Zaanen's "Linear Analysis" (first published in 1953), the author comments on p. 148 that he avoids using the Axiom of Choice because it is controversial. (He changed his mind later.)

I mentioned the continuum hypothesis above. It was proved by Cohen in 1963–4 that the generalized continuum hypothesis and its negation are also consistent with the axioms of set theory, even if they are enriched by the Axiom of Choice. It was known earlier (Sierpiński) that the denial of the GCH *implies* the Axiom of Choice. The GCH, however, though not devoid of consequences (see Ulam's theorem above), is a rather different matter from the AC; it has no intuitive appeal.

It is not really necessary for our purpose to go deeply into the various forms of the Axiom, its applications, or its relations with other suggested axioms. I used to hand out notes on it,

but these days, now that we have an undergraduate logic course, they are probably unnecessary. Nevertheless, one or two further comments may be helpful.

My own view, for what it is worth, is hesitantly Platonist. I think that mathematical concepts do exist, in some unclear absolute sense — in fact, I can see no reason for studying mathematics if you seriously believe it is just a game with rather bizarre rules, quite unrelated to reality. Indeed, I agree with the usual jocular statement that "mathematics is independent of its foundations". The foundations postdate the familiar theorems, and were constructed for the sake of clarifying the logical structure; if they were unsatisfactory, either by leading to a contradiction or by failing to imply a standard theorem, we should not throw away mathematics in the mass, but rather modify the foundations to preserve the mathematics, albeit attempting to preserve our intuition about the concepts we are dealing with. This process has really happened once, when Russell pointed out the inadequacy of Frege's set theory. The axioms of set theory were adjusted so that Russell's paradox ceased to hold. It is conceivable that a contradiction might arise that could not be so easily resolved — the Intuitionists came close to saying that about Russell's paradox —, but in that case we should also have to reconsider the validity of our whole system of thought (as, indeed, the Intuitionists claimed). I am not suggesting that the distinction between "constructive" and "nonconstructive" proofs that they introduced is a silly one, only that a nonconstructive proof ought still to be a proof.

In short, the foundations, however dubious in their details, are there to support an edifice most of which is already built and should not be demolished except under extreme necessity. By and large, mathematics *works* very well, both in itself and as a means of analyzing the real world; it would be silly to demolish it for an airy-fairy thing like Russell's paradox, which clearly does not deal with "practical" constructs.

The Axiom of Choice seems to me manifestly "true" at the level of intuition. From a strictly logical point of view, one could divide it into various cases: the *finite* axiom of choice (when the set $\mathcal{C}$ is finite) is actually a theorem of the standard set theories, the *countable* axiom of choice (for $\mathcal{C}$ countable) is already unprovable but seems relatively unexceptionable, and so on. If very large sets are to be allowed, it seems clear that we should allow "choice functions" in all these cases, just as another axiom specifically allows "power sets". The only serious objection to doing so is the fear that an axiom that, as it turns out, has such sweeping consequences, and is apparently unprovable from previous axioms, might be inconsistent with the rest of set theory; and once Gödel had shown it was not, denying it loses all point. Why should we not happily accept that every vector space does have a basis, if the statement both agrees with our intuition and entails no contradiction? But — and this is where things get messy — we must also accept the existence of Lebesgue non-measurable sets as part of the package, and, therefore, our exposition of measures in $\mathbb{R}$ or in $\mathbb{R}^n$ will have to assume they are defined only on σ-fields. (By the way, Bernstein's construction also uses the Axiom of Choice, albeit in a far less elementary fashion involving transfinite arithmetic.) In any case, there are other situations, not just Lebesgue outer measure in $\mathbb{R}$ and its relatives in $\mathbb{R}^n$ but completely different spaces and measures, where measures, to be interesting, must be defined on σ-fields smaller than a whole power set. Ulam's theorem mentioned above is a weak example of this.

As I said above, the reason why the Axiom of Choice appears at all is that without it we cannot handle very large sets, although our set theory demands they should exist. It is a paradox of sorts (not a *logical* paradox but a "semantic paradox", a linguistic curiosity) that we can set up theories which discuss, and require the existence of, uncountable sets, despite the countability of the set of symbols at our disposal. The oddity is no greater than that involved in the statements people occasionally make — on what precise grounds I have no idea — that the total number of fundamental particles in the universe is less than (say) 10^{256}

or whatever number they use, so that you can describe, by finitely many symbols and quite unambiguously, a number (say $10^{10^{257}}$) that you believe to have no physical correspondent at all. At any rate, most applications of the Axiom of Choice have to do with the existence of something too large to be described explicitly, like a basis of $\mathbb{R}$ over $\mathbb{Q}$. In this sense, the constructs provided by the Axiom are "all in the mind", and their practical significance, at any rate once one goes beyond the countable axiom of choice, is nil. To put this idea more precisely: the effects of the Axiom tend to be in the unrestricted statement of many results which would otherwise be true with some provisos that are satisfied "in all cases of practical importance" anyway. That the vector space $\mathbb{R}$ has an algebraic basis over the field $\mathbb{Q}$, or that a function space of infinite dimension has an algebraic basis, are assertions that have no "practical" consequences.

This does not exclude some oddities. There are a couple of notorious theorems (the existence of Haar measure and of the Shilov boundary) where the Axiom was used to prove the existence of something which was subsequently shown to be unique, so that no "choice" is really present. In one of these cases (Haar measure), an alternative proof of equal generality was subsequently found. For the Shilov boundary, it remains a puzzle why, or whether, the Axiom is needed. The whole idea is very abstract, and maybe the Axiom can be avoided "in all practical cases"; but it is not clear what "practical" would mean, and as far as I know no adequately general proof has been found that does not use the Axiom.

The question arises whether the Axiom is absolutely necessary for the existence of a Lebesgue non-measurable set in $\mathbb{R}$. This was (almost) resolved by Solovay in 1970. Provided that the existence of an inaccessible cardinal is consistent with set theory (which has not been proved, although it seems to be generally believed for reasons I don't understand), there is a model for set theory (even adding in a form of the *countable* axiom of choice, the so-called "principle of dependent choices") in which every set of real numbers is Lebesgue-measurable. So there seems to be little point in hunting non-measurable sets without the Axiom of Choice. Denying the (uncountable) Axiom and assuming that all sets in $\mathbb{R}$ are measurable might seem a useful possibility, but to do so you would also have to abandon other consequences of the Axiom, many of which are extremely convenient; furthermore, it would not follow that *any* measure you wanted to use in an *arbitrary* space Ω could be defined on $\mathcal{P}(\Omega)$. So we are in a slightly uncomfortable position: we are convinced that any set we can actually define in any "practical" way must be Lebesgue-measurable, but, to be consistent, we must still prove it in each case. Fortunately, the proof is usually fairly trivial, granted the standard properties of Lebesgue-measurable sets.

§11. Lebesgue-Stieltjes measures in higher dimensions.

Definition 11.1. Let $a \coloneqq (a_1, a_2, \dots, a_n)$, $b \coloneqq (b_1, b_2, \dots, b_n)$ be points of $\mathbb{R}^n$. Define

$$\begin{aligned} a < b \quad &\text{to mean} \quad (\forall i,\ 1 \le i \le n)\ a_i < b_i\,, \\ a \le b \quad &\text{to mean} \quad (\forall i,\ 1 \le i \le n)\ a_i \le b_i\,, \\ (a, b] &\coloneqq \{x \in \mathbb{R}^n : a < x \le b\}\,, \\ [a, b] &\coloneqq \{x \in \mathbb{R}^n : a \le x \le b\}\,, \\ (a, b) &\coloneqq \{x \in \mathbb{R}^n : a < x < b\}\,. \end{aligned}$$

Notice that $\le$ is *not* a total order in $\mathbb{R}^n$, and that $<$ does *not* mean "$\le$ and $\ne$". The "multi-intervals" $(a, b], [a, b], (a, b)$ may be described as *half-open on the left, closed,* or *open*; $(a, b]$ is empty if, for any index i, $a_i \ge b_i$. The formal values $a_i = -\infty$ and $b_i = \infty$ may also be allowed when the intervals are open at the corresponding ends.

Definition 11.2. Suppose $J \coloneqq (a, b]$ as above. For any subset P of the integer interval

$$\langle 1, n\rangle \coloneqq \{1, 2, \dots, n\}\,,$$

where $P \coloneqq \{j_1, j_2, \dots, j_p\}$ and $j_1 < j_2 < \cdots < j_p$, let

$$\begin{aligned} a^P &\coloneqq (a_1, a_2, \dots, \widehat{a}_{j_1}, a_{j_1+1}, \dots, \widehat{a}_{j_2}, \dots, \dots, \widehat{a}_{j_p}, a_{j_p+1}, \dots, a_n)\,, \\ b^P &\coloneqq (b_1, b_2, \dots, \widehat{b}_{j_1}, b_{j_1+1}, \dots, \widehat{b}_{j_2}, \dots, \dots, \widehat{b}_{j_p}, b_{j_p+1}, \dots, b_n)\,, \end{aligned}$$

where the 'hats' indicate that the terms they distinguish are omitted. (This is a common convention.) Then let $J^P \coloneqq (a^P, b^P] \subseteq \mathbb{R}^{n-p}$. In words: J^P is the set of points in $\mathbb{R}^{n-p}$ whose coordinates are obtained from those of a point in J by omitting those indexed by members of P.

Definition 11.3. Suppose $J \coloneqq (a, b]$ as above; let $i \in \langle 1, n\rangle$ and $a_i < \alpha \le \beta \le b_i$. If $f : J \longrightarrow \mathbb{R}$, define $\delta^{i,\langle 1,n\rangle}_{\alpha\beta} f : J^{\{i\}} \longrightarrow \mathbb{R}$ by

$$\begin{aligned} (x_1, x_2, \dots, x_{n-1}) \mapsto\ & f(x_1, x_2, \dots, x_{i-1}, \beta, x_i, x_{i+1}, \dots, x_{n-1}) \\ & - f(x_1, x_2, \dots, x_{i-1}, \alpha, x_i, x_{i+1}, \dots, x_{n-1})\,. \end{aligned}$$

If $a < c \coloneqq (c_1, c_2, \dots, c_n) \le d \coloneqq (d_1, d_2, \dots, d_n) \le b$, define

$$\Delta_{cd} f \coloneqq \delta^{1,\{1\}}_{c_1 d_1} \delta^{2,\{1,2\}}_{c_2,d_2} \cdots \delta^{n-1,\langle 1,n-1\rangle}_{c_{n-1},d_{n-1}} \delta^{n,\langle 1,n\rangle}_{c_n,d_n} f\,. \tag{31}$$

This is in effect a number, the last step being the subtraction of two values of a function $(a_1, b_1] \longrightarrow \mathbb{R}$, whether or not you interpret $J^{\langle 1,n\rangle}$ as a singleton.

The formula (31) is only one of several possible ways of expressing Δ_{cd}. There are commutation relations among the δs. If $1 < i < j \le n$ and $\alpha \le \beta$, $\gamma \le \delta$ in J,

$$\delta_{\alpha\beta}^{j,\langle 1,n-1\rangle}\delta_{\gamma\delta}^{i,\langle 1,n\rangle} = \delta_{\gamma\delta}^{i-1,\langle 1,n-1\rangle}\delta_{\alpha\beta}^{j,\langle 1,n\rangle}\,.$$

(In fact, the δs are coface operators corresponding to the boundary operators for cubical homology.) There is, therefore, no need to begin in (31) with the nth coordinate.

Remark 11.4. The idea behind (31) is this. We think of the space $\mathbb{R}^n$ as occupied by matter of varying density whose total mass over all space is finite. Then $f(c)$ is the mass of the matter occupying the set $\{x \in \mathbb{R}^n : x \le c\}$ of all points as it were "below and to the left" of c; each δ thus selects the mass of a "slice" of the previous set. In dimension 2, $\delta_{c_2,d_2}^{2,\{1,2\}} f(x_1)$ is the mass of the strip

$$\{(\xi_1,\xi_2) : c_2 < \xi_2 \le d_2\,,\ \xi_1 \le x_1\}\,,$$

and $\Delta_{cd} f$ selects the mass of the rectangle $\{(\xi_1,\xi_2) : c_2 < \xi_2 \le d_2\,,\ c_1 < \xi_1 \le d_1\}$. Alternatively, one can consider f as a cumulative probability distribution in the same way.

Definition 11.5. Let $f : J \longrightarrow \mathbb{R}$ as above. f is *separately right-continuous* on J when, for any $c := (c_1, c_2, \ldots, c_n) \in J$ and any $i \in \langle 1, n\rangle$, and for any $\epsilon > 0$, there exists $\delta > 0$ such that, whenever $c_i < x < c_i + \delta$ and $x \le b_i$,

$$|f(c_1, c_2, \ldots, c_{i-1}, x, c_{i+1}, \ldots, c_n) - f(c_1, c_2, \ldots, c_{i-1}, c_i, c_{i+1}, \ldots, c_n)| < \epsilon\,.$$

That is, all the functions of one variable (defined on the intervals $(a_i, b_i]$ for the various i) that are obtained from f by fixing all coordinates but i are right-continuous on $(a_i, b_i]$. Less formally, f is right-continuous in any one coordinate when the others are fixed.

f would be *jointly* right-continuous if, for any $c \in (a, b]$ and $\epsilon > 0$, there were $\delta > 0$ such that, for any $x \in (a, b]$ for which, for each i, $c_i < x_i < c_i + \delta$, then $|f(x) - f(c)| < \epsilon$. (An equivalent formulation is that there is some $d > c$ such that, whenever $x \in J \cap (c, d)$, then $|f(x) - f(c)| < \epsilon$.)

These are the definitions of separate and of joint right-continuity on the whole of J, and it is obvious how to define separate and joint right-continuity at an individual point of J. There are also definitions of left and of two-sided continuity. Joint two-sided continuity is, in effect, just "continuity" on J, in the usual sense for functions of several variables.

It should be emphasized that joint continuity (of any kind) is genuinely a much stronger condition than separate continuity of the corresponding kind. This is less obvious than it might be because one usually considers rather simple functions.

Definition 11.6. Let $J := (a, b] \subseteq \mathbb{R}^n$ as above. The function $f : J \longrightarrow \mathbb{R}$ is a *Lebesgue-Stieltjes distribution function in* J if

(a) whenever $a < c \le d \le b$, $\Delta_{cd} f \ge 0$, and
(b) f is separately right-continuous on J.

Condition *(a)* reduces in one dimension to 9.2*(a)*. The reason for it is Remark 11.4; it seems difficult to express the idea in any more direct way. The curious aspect of *(b)* is that only separate right-continuity is required, but it will result from the theory (and can easily be proved directly) that, in the presence of *(a)*, *(b)* implies joint right continuity.

If μ is a measure defined on the Borel sets in $\mathbb{R}^n$, and finite on bounded intervals, define $f(a) := \mu((0, a])$ whenever $a > 0$. It follows that $\Delta_{yx} f = \mu((y, x])$ when $0 < y < x$,

which explains *(a)*. On the other hand, *(b)* follows from 8.5. Indeed, even joint right-continuity would follow from 8.5.

It is not easy to recognize examples of d.f.s in $\mathbb{R}^n$, because of *(a)*. Nevertheless, they abound. Suppose f_i is a 1-dimensional d.f. in $(a_i, b_i]$, for $1 \leq i \leq n$. Then define

$$f : J \longrightarrow \mathbb{R} : x = (x_1, x_2, \ldots, x_n) \mapsto f_1(x_1) f_2(x_2) \cdots f_n(x_n) .$$

f will be a d.f. in J, the *product distribution function* or the *product* of the d.f.s $f_1, \ldots, f_n$. The most important example is the n-dimensional *Lebesgue distribution function*

$$f(x_1, x_2, \ldots, x_n) = x_1 x_2 \cdots x_n ,$$

for which $\Delta_{cd} f$ is just the n-dimensional volume of the multi-interval $(c, d]$.

Definition 11.7. Let $J \coloneqq (a, b]$ be a multi-interval in $\mathbb{R}^n$, and let $f : J \longrightarrow \mathbb{R}^n$ be a Lebesgue-Stieltjes distribution function. Define

$$\mathcal{S} \coloneqq \{(c, d] \subseteq \mathbb{R}^n : a < c \leq d \leq b\} ,$$

and, for each $(c, d] \in \mathcal{S}$, let $\tau_f((c, d]) \coloneqq \Delta_{cd} f$.

Lemma 11.8. *$\mathcal{S}$ is a semiring in J; τ_f is countably additive on $\mathcal{S}$.* □

Neither the finite nor the countable additivity is entirely trivial. One needs the fact that a bounded closed multi-interval $[c, d]$ is compact (a finite covering of it by open multi-intervals admits a finite sub-covering). The proof is otherwise much as before, 9.7, except for non-trivial technical changes.

Definition 11.9. Given the d.f. f as above, $\tau_f^\dagger$ is the *Lebesgue-Stieltjes outer measure in J induced by f*; if the d.f. is the Lebesgue d.f., the outer measure is the Lebesgue outer measure in J. The resulting measure space (J, Σ_f, μ_f) is the Lebesgue-Stieltjes (or Lebesgue) measure space on J.

The crucial fact is of course that $\tau_f((c, d]) = \Delta_{cd} f$ for any $(c, d] \in \mathcal{S}$.

We now have a very substantial stock of interesting measure spaces. It is possible to carry the study of measures a great deal further, but for the moment we shall change tack and discuss integration. Since we cannot assume our measures are defined on all subsets of the domain Ω because of §10, we must first study the class of functions that are in some sense adapted to the σ-algebra on which the measure will be defined.

§12. Measurable functions.

Definition 12.1. Let $\mathcal{F}$ be a field in Ω. A function $f : \Omega \longrightarrow \overline{\mathbb{R}}$ is *measurable with respect to* $\mathcal{F}$, or *$\mathcal{F}$-measurable*, or (if there is no ambiguity) *measurable*, if, for every $\alpha \in \overline{\mathbb{R}}$, the set

$$f^{-1}[\alpha, \infty] = \{x \in \Omega : f(x) \geq \alpha\}$$

is a member of $\mathcal{F}$.

This is not the only definition of a measurable function that you will find in the literature, but it is convenient for our purposes and consistent with other definitions.

Lemma 12.2. *Let Σ be a σ-field in Ω.*

(a) *Let $f : \Omega \longrightarrow \overline{\mathbb{R}}$ be Σ-measurable. Then, for any $\alpha, \beta \in \overline{\mathbb{R}}$, each of the sets $f^{-1}[\alpha, \beta]$, $f^{-1}(\alpha, \beta]$, $f^{-1}[\alpha, \beta)$, $f^{-1}(\alpha, \beta)$ belongs to Σ.*

(b) *$f : \Omega \longrightarrow \overline{\mathbb{R}}$ is Σ-measurable if and only if the sets $f^{-1}\{-\infty\}$, $f^{-1}\{\infty\}$ and $f^{-1}(\alpha, \infty)$ belong to Σ for every $\alpha \in \mathbb{R}$.*

Proof. *(a)* If $\beta > -\infty$, $f^{-1}(\beta, \infty] = \bigcup_{k=1}^{\infty} f^{-1}[\beta + \frac{1}{k}, \infty] \in \Sigma$. If $\beta = -\infty$, then $f^{-1}(\beta, \infty] = f^{-1}(-\infty, \infty] = \bigcup_{k=1}^{\infty} f^{-1}[-k, \infty] \in \Sigma$. Hence, $f^{-1}(\beta, \infty] \in \Sigma$ whenever $\beta \in \overline{\mathbb{R}}$. Then $f^{-1}[\alpha, \beta] = f^{-1}[\alpha, \infty] \setminus f^{-1}(\beta, \infty] \in \Sigma$. And so on.

(b) $f^{-1}[\alpha, \infty] = \bigcap_{k=1}^{\infty} f^{-1}(\alpha - \frac{1}{k}, \infty) \cup f^{-1}\{\infty\} \in \Sigma$, and so on. □

Proposition 12.3. *Let Σ be a σ-field in Ω and let $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ be Σ-measurable functions. Then the sets*

$$\{x \in \Omega : f(x) > g(x)\}, \ \{x \in \Omega : f(x) \geq g(x)\}, \ \{x \in \Omega : f(x) = g(x)\}$$

all belong to Σ.

Proof. Any non-empty open interval in $\overline{\mathbb{R}}$ contains a (finite) rational, and $\mathbb{Q}$ is countable. So

$$\begin{aligned}\{x \in \Omega : f(x) > g(x)\} &= \bigcup\nolimits_{\alpha \in \mathbb{Q}} \{x \in \Omega : f(x) \geq \alpha > g(x)\} \\ &= \bigcup\nolimits_{\alpha \in \mathbb{Q}} (\{x : f(x) \geq \alpha\} \setminus \{x : g(x) \geq \alpha\}) \in \Sigma.\end{aligned}$$

But then

$$\begin{aligned}&\{x \in \Omega : f(x) \leq g(x)\} = \Omega \setminus \{x : f(x) > g(x)\} \in \Sigma \quad \text{and} \\ &\{x \in \Omega : f(x) = g(x)\} = \{x : f(x) \leq g(x)\} \cap \{x : g(x) \leq f(x)\} \in \Sigma.\end{aligned}$$ □

Lemma 12.4. *Any constant function $f : \Omega \longrightarrow \overline{\mathbb{R}}$ is $\mathcal{F}$-measurable, for any field $\mathcal{F}$ in Ω.*

Proof. Indeed, if $f(x) = c$ for all $x \in \Omega$, then $f^{-1}([\alpha, \infty]) = \emptyset$ when $c < \alpha$, and also $f^{-1}([\alpha, \infty]) = \Omega$ when $c \geq \alpha$. □

Remark 12.5. In the next proposition, I have to mention the sum $f + g$ (and difference $f - g$) of two Σ-measurable functions f and g. In principle this means the pointwise sum:

$(f+g)(x) \coloneqq f(x)+g(x)$ for all $x \in \Omega$. However, I am allowing f and g to take infinite values, so that $f+g$, so understood, may be undefined at some points, namely where $f(x)$ and $g(x)$ are opposite infinities. For the sake of a simple statement of the proposition, let us agree that, at such points, $(f+g)(x)$ is understood to be 0, and, likewise, that $(f-g)(x)$ is understood to be 0 at points where $f(x)$ and $g(x)$ are the same infinity. Similarly, in part *(d)*, let us say that $|f(x)|^a$ means $+\infty$ when $a>0$ and $f(x)=\pm\infty$, means 1 when $a=0$ (even if $f(x)=0$; it is a curious point that 0^0 is not usually defined, but x^0 in formulæ is commonly understood as meaning 1 even when $x=0$) and means 0 when $a<0$ and $f(x)=\pm\infty$. These are merely *ad hoc* conventions, not intended to supersede the general rules of 2.1. In the later development they will scarcely be needed.

Proposition 12.6. *Let Σ be a σ-field in Ω, and let $f,g:\Omega\longrightarrow\overline{\mathbb{R}}$ be Σ-measurable. Then*

(a) *the pointwise sum $f+g$, defined as at* 12.5*, is Σ-measurable;*
(b) *the pointwise maximum $\max(f,g)$ is Σ-measurable,*
(c) *for any constant $a\in\mathbb{R}$, the function af is Σ-measurable,*
(d) *for any $a\in\mathbb{R}\setminus\{0\}$, $|f|^a$ (defined pointwise as in* 12.5*) is Σ-measurable,*
(e) *the pointwise product fg is Σ-measurable.*

Proof. *(a)* Take $\gamma\in\overline{\mathbb{R}}$, and consider $E_\gamma \coloneqq \{x\in\Omega : f(x)+g(x)\geq\gamma\}$. There are various cases. Of course $E_{-\infty}=\Omega\in\Sigma$. Next,

$$\begin{aligned} E_\infty &= \{x : f(x)=\infty,\ g(x)>-\infty\}\cup\{x : f(x)>-\infty,\ g(x)=\infty\} \\ &= \big(f^{-1}(\{\infty\})\cap g^{-1}((-\infty,\infty])\big)\cup\big(g^{-1}(\{\infty\})\cap f^{-1}((-\infty,\infty])\big)\in\Sigma \end{aligned}$$

by 12.2. Now consider the case when γ is finite and *positive*. Then $\gamma\leq f(x)+g(x)<\infty$ is only possible when both $f(x)$ and $g(x)$ are both finite, and, for any $\alpha\in\overline{\mathbb{R}}$,

$$\{x:\gamma-g(x)\geq\alpha\}=\{x:g(x)\leq\gamma-\alpha\}=g^{-1}([-\infty,\gamma-\alpha])\in\Sigma,$$

again by 12.2. Hence, the function $h:\Omega\longrightarrow\overline{\mathbb{R}} : x\mapsto\gamma-g(x)$ is defined at every point of Ω and is also measurable, and

$$\begin{aligned} E_\gamma &= E_\infty\cup\{x:0<\gamma\leq f(x)+g(x)<\infty\}, \\ &= E_\infty\cup\{x:f(x)\geq h(x)\}\in\Sigma, \end{aligned}$$

the second set of the union being in Σ by 12.3.

If $-\infty<\gamma\leq 0$, then the conventions of 12.5 must be taken into consideration, and

$$\begin{aligned} E_\gamma = E_\infty\cup\{x:f(x)\geq h(x)\}\cup(\{x:f(x)=\infty\}\cap\{x:g(x)=-\infty\}) \\ \cup(\{x:f(x)=-\infty\}\cap\{x:g(x)=\infty\}). \end{aligned}$$

This shows that, once again, $E_\gamma\in\Sigma$. All cases have now been discussed.

(b) For any $\alpha\in\overline{\mathbb{R}}$,

$$\{x\in\Omega:\max(f(x),g(x))\geq\alpha\}=f^{-1}([\alpha,\infty])\cup g^{-1}([\alpha,\infty])\in\Sigma.$$

(c) If $a>0$, then $\{x:af(x)\geq\alpha\}=\{x:f(x)\geq a^{-1}\alpha\}\in\Sigma$. Likewise, if $a<0$, $\{x:af(x)\geq a\}=\{x:f(x)\leq a^{-1}\alpha\}\in\Sigma$, by 12.2. If $a=0$, then af is the constant function "zero", so is measurable by 12.4.

(d) Examine the cases that arise, much as in *(c)*.

(e) This can be done in more than one way, but after *(d)* it is perhaps easiest to proceed as follows. Let $h(x) := \frac{1}{4}\{|f(x)+g(x)|^2 - |f(x)-g(x)|^2\}$, with the conventions of 12.5. Define $\Omega_1 := \{x \in \Omega : f(x), g(x) \in \mathbb{R}\} \in \Sigma$. If $x \in \Omega_1$, $f(x)g(x) = h(x)$; this is definitely untrue at some points off Ω_1. For instance, if $f(x) = \infty$ and $g(x) = 1$, $h(x) = 0$ by 12.5, but $f(x)g(x) = \infty$.

h is measurable by *(a), (c),* and *(d)*. Thus, given $\alpha \in \overline{\mathbb{R}}$,

$$\Omega_1 \cap \{x \in \Omega : f(x)g(x) \geq \alpha\} = \Omega_1 \cap \{x \in \Omega : h(x) \geq \alpha\} \in \Sigma\,.$$

If $\alpha > 0$, then

$$\begin{aligned}\{x : f(x)g(x) \geq \alpha\} \setminus \Omega_1 = (&\{x : f(x) = \infty\} \cap \{x : g(x) > 0\}) \\ &\cup (\{x : f(x) > 0\} \cap \{x : g(x) = \infty\}) \\ &\cup (\{x : f(x) < 0\} \cap \{x : g(x) = -\infty\}) \\ &\cup (\{x : f(x) = -\infty\} \cap \{x : g(x) < 0\})\,,\end{aligned}$$

which is certainly in Σ. If $-\infty < \alpha \leq 0$, change each $<$ to $\leq$ and each $>$ to $\geq$. So, for any $\alpha > -\infty$,

$$\{x : f(x)g(x) \geq \alpha\} = (\{x : f(x)g(x) \geq \alpha\} \cap \Omega_1) \cup (\{x : f(x)g(x) \geq \alpha\} \setminus \Omega_1) \in \Sigma\,.$$

Finally, if $\alpha = -\infty$, $\{x : f(x)g(x) \geq -\infty\} = \Omega$. □

The difficulties of this proof arise almost wholly from the presence of infinite values, but the next Lemma should demonstrate why it is convenient to allow them.

Lemma 12.7. *Let Σ be a σ-field in Ω. Suppose that (f_n) is a sequence of Σ-measurable functions $\Omega \longrightarrow \overline{\mathbb{R}}$. The functions $\inf_n f_n$, $\sup_n f_n$, $\liminf f_n$, $\limsup f_n$ are Σ-measurable.*

Proof. Given $a \in \overline{\mathbb{R}}$,

$$\{x \in \Omega : (\inf f_n)(x) \geq \alpha\} = \{x \in \Omega : \inf(f_n(x)) \geq \alpha\} = \bigcap\nolimits_{n=1}^{\infty} \{x : f_n(x) \geq \alpha\} \in \Sigma\,,$$

so $\inf f_n$ is Σ-measurable; then $\sup f_n = -\inf(-f_n)$ is Σ-measurable. The rest follows. □

Recall from 7.15 that the infima or suprema here can be described either as defined pointwise or in terms of the partial order on the functions $\Omega \longrightarrow \overline{\mathbb{R}}$. The Lemma then says that the induced partial order on the subset of measurable functions is countably order-complete. (It is not usually order-complete, for a non-Lebesgue-measurable function on $\mathbb{R}$, for instance, is the supremum of *uncountably* many functions that are non-zero except on a singleton, and each such "singleton function" is Lebesgue-measurable.)

In summary: all the usual operations of analysis, when applied to Σ-measurable functions, yield Σ-measurable functions.

Lemma 12.8. *Suppose Σ is a σ-field in Ω and $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous and $f : \Omega \longrightarrow \mathbb{R}$ is Σ-measurable. Then $\phi \circ f$ is Σ-measurable.*

Proof. Take $\alpha \in \mathbb{R}$. As ϕ is continuous, $\phi^{-1}((\alpha, \infty))$ is open in $\mathbb{R}$. Hence it is a countable unsion of open intervals (this was proved in 9.9), say $\phi^{-1}((\alpha, \infty)) = \bigcup_{k=1}^{\infty} (c_k, d_k)$. Then

$$(\phi \circ f)^{-1}((\alpha, \infty)) = \bigcup_{k=1}^{\infty} f^{-1}((c_k, d_k)) \in \Sigma\,,$$

as each $f^{-1}((c_k, d_k)) \in \Sigma$ by 12.2*(a)*. Since $\phi \circ f$ does not take infinite values, $(\phi \circ f)^{-1}(-\infty) = (\phi \circ f)^{-1}(\infty) = \emptyset$, and 12.2*(b)* establishes Σ-measurability. □

This result has been stated for finite-valued functions because I have not discussed continuity for extended-real-valued functions, but it is true more generally. However, it is *not* usually true that $\phi \circ f$ is measurable for Lebesgue-measurable $\phi : \mathbb{R} \longrightarrow \mathbb{R}$. This is inconvenient for some purposes, and probabilists in particular often restrict attention to Borel-measurable functions; that is to say, they specify that, for functions $\mathbb{R} \longrightarrow \mathbb{R}$ (or indeed more generally), measurability is understood in terms of the Borel σ-field $\mathcal{B}(\mathbb{R})$. If f is measurable $\Omega \longrightarrow \mathbb{R}$ and ϕ is Borel-measurable $\mathbb{R} \longrightarrow \mathbb{R}$, then $\phi \circ f$ is measurable $\Omega \longrightarrow \mathbb{R}$. (Exercise.)

Definition 12.9. Let $A \in \mathcal{P}(\Omega)$. The *indicator function* of the set A is the function $\mathbf{1}_A : \Omega \longrightarrow \mathbb{R}$ defined by

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

This concept seems quite obvious. It is clear that it is just the adaptation to values in $\mathbb{R}$ of Cantor's idea that we use to show that $\mathcal{P}(\Omega)$ has cardinality $2^{\#(\Omega)}$. However, it is said that it was first explicitly defined as late as 1915 by de la Vallée Poussin.

Both the name and the notation are disputed. Analysts often called it the *characteristic function* of A, and denoted it by χ_A or c_A. Unfortunately, probabilists, before their subject was properly grounded in analysis, grew accustomed to use the phrase "characteristic function" of something else (the Fourier transform, in fact), and as a consequence preferred the name "indicator function". It is sometimes denoted by $\boldsymbol{I}_A$ or even by A. (!)

Lemma 12.10. *Let $\mathcal{F}$ be a field in Ω. $\mathbf{1}_A$ is $\mathcal{F}$-measurable if and only if $A \in \mathcal{F}$.* □

Definition 12.11. A function $f : \Omega \longrightarrow X$ (where X may be $\mathbb{R}$ or $\overline{\mathbb{R}}$) is called *simple with respect to the field $\mathcal{F}$*, or *$\mathcal{F}$-simple*, or (when there is no ambiguity) *simple*, if it is $\mathcal{F}$-measurable and assumes only finitely many values from X.

f is called *elementary* with respect to the σ-field Σ, or Σ-elementary, if it is Σ-measurable and assumes only countably many values.

Recall that a "linear combination" of a set of functions with values in $\mathbb{R}$ is understood pointwise, and is by definition a *finite* linear combination — i.e. it is a sum of finitely many terms, each a constant multiple of one of the functions of the set.

Lemma 12.12. *Given a field $\mathcal{F}$ in Ω, any real linear combination of real-valued $\mathcal{F}$-simple functions is $\mathcal{F}$-simple; the pointwise product and the maximum of two real-valued (or extended-real-valued) $\mathcal{F}$-simple functions are $\mathcal{F}$-simple.*

Proof. In each case, it is easy to see that the new function has only finitely many possible values, and 12.6 shows that it is measurable. □

Lemma 12.13. *Any $\mathcal{F}$-simple function $f : \Omega \longrightarrow \mathbb{R}$ may be expressed as a real linear combination of the indicator functions of pairwise disjoint sets of $\mathcal{F}$ whose union is Ω.*

Proof. Let the distinct values of f be $b_1, b_2, \dots, b_r \in \mathbb{R}$. Take, for $k = 1, 2, \dots, r$,

$$E(r) := f^{-1}(\{b_r\}) \in \mathcal{F}.$$

Then $E(i) \cap E(j) = \emptyset$ if $i \neq j$; if $x \in E(i) \cap E(j)$, $f(x) = b_i$ and $f(x) = b_j$ simultaneously, which is impossible as $b_i \neq b_j$. Also, if $y \in \Omega$, $f(y)$ must be one of the bs, so that $y \in E(k)$ for some k between 1 and r; thus $\Omega = \bigcup_{k=1}^{r} E(k)$. And, finally,

$$f = \sum\nolimits_{k=1}^{r} b_k \mathbf{1}_{E(k)}, \tag{32}$$

because, for any $z \in \Omega$, there is exactly one l for which $z \in E(l)$, which means that $f(z) = b_l$, and that is precisely $\sum b_k \mathbf{1}_{E(k)}(z)$. □

I shall, for convenience, describe a real linear combination of the indicator functions of pairwise disjoint measurable sets whose union is Ω as a *standard form* of the simple function that is its sum. (Notice that (32) is, in addition, a special kind of standard form, because f takes a different value on each of the $E(k)$).

Simple functions are useful because of the next proposition.

Proposition 12.14. *Let $\mathcal{F}$ be a field in Ω, and let $f : \Omega \longrightarrow \overline{\mathbb{R}}$ be an $\mathcal{F}$-measurable function taking non-negative values* [for brevity, we often speak of a "non-negative measurable function"]. *Then there is a sequence $(g_n)_{n=1}^{\infty}$ of non-negative real-valued $\mathcal{F}$-simple functions which is (pointwise) increasing and converges pointwise to f.*

Proof. Define

$$g_n(x) := \begin{cases} n & \text{for each } x \in \Omega \text{ such that } f(x) > n, \\ 0 & \text{when } f(x) = 0, \text{ and} \\ \dfrac{i-1}{2^n} & \text{when } \dfrac{i-1}{2^n} < f(x) \leq \dfrac{i}{2^n}, \text{ for integers } i \text{ such that } 0 < i \leq 2^n n. \end{cases}$$

The verification that the sequence has the desired properties is routine. (To understand what is going on, try to think of it in terms of the graph of f.) □

§13. Integration of simple functions.

As I remarked in §1A, there are many approaches to the integral. The one I shall present has the advantage of needing little further preparation and being rather "natural". It has one sticky point, where we need to appeal to the countable additivity of the measure.

If f is a real-valued $\mathcal{F}$-simple function (see 12.11), it has a standard form $\sum_{k=1}^{r} a_k \mathbf{1}_{E(k)}$, as in 12.13. In principle it may have many standard forms; only one of them will have the property that all the coefficients a_k are different. However, it is not really desirable to impose this as a further condition, because it may be destroyed if two simple functions are added.

Definition 13.1. Let $\mathcal{F}$ be a field in Ω, and let μ be a fam on $\mathcal{F}$ [or let σ be a fasm on $\mathcal{F}$]. Let f be an $\mathcal{F}$-simple function with value in $\mathbb{R}$ or $\overline{\mathbb{R}}$. It will be convenient to set

$$Q(f) := \{x \in \Omega : f(x) \neq 0\}\,.$$

f is *integrable* with respect to μ [or σ] if $\mu(Q(f))) < \infty$ [or if $-\infty < \sigma(Q(f)) < \infty$].

Lemma 13.2. *Given the field $\mathcal{F}$ in Ω, with a fam μ and a fasm σ, let f be an $\mathcal{F}$-simple function $\Omega \longrightarrow \overline{\mathbb{R}}$, and let $\sum_{i=1}^m a_i \mathbf{1}_{E(i)}$ and $\sum_{j=1}^n b_j \mathbf{1}_{F(j)}$ be standard forms for f. Then, if f is non-negative,*

$$\sum\nolimits_{i=1}^m a_i\,\mu(E(i)) = \sum\nolimits_{j=1}^n b_j\,\mu(F(j))\,,$$

where the sums make sense in $\overline{\mathbb{R}}$, whilst, if $f : \Omega \longrightarrow \mathbb{R}$ is integrable with respect to σ,

$$\sum\nolimits_{i=1}^m a_i\,\sigma(E(i)) = \sum\nolimits_{j=1}^n b_j\,\sigma(F(j))\,,$$

where the sums make sense in $\mathbb{R}$.

Proof. Set $H(i,j) := E(i) \cap F(j)$; the $H(i,j)$ are pairwise disjoint and belong to $\mathcal{F}$. But

$$E(i) = E(i) \cap \Omega = E(i) \cap \Big(\bigcup\nolimits_{j=1}^n F(j)\Big) = \bigcup\nolimits_{j=1}^n (E(i) \cap F(j)) = \bigcup\nolimits_{j=1}^n H(i,j)$$

for each choice of i, and similarly $F(j) = \bigcup_{i=1}^m H(i,j)$ for each j. These unions are disjoint, so that $\mu(E(i)) = \sum_{j=1}^n \mu(H(i,j))$ and $\mu(F(j)) = \sum_{i=1}^m \mu(H(i,j))$. Hence

$$\sum\nolimits_{i=1}^m a_i\,\mu(E(i)) = \sum\nolimits_{i=1}^m \sum\nolimits_{j=1}^n a_i\,\mu(H(i,j)) = \sum\nolimits_{i,j} a_i\,\mu(H(i,j))\,,$$
$$\sum\nolimits_{j=1}^n b_j\,\mu(F(j)) = \sum\nolimits_{i=n}^n \sum\nolimits_{i=1}^m b_j\,\mu(H(i,j)) = \sum\nolimits_{i,j} b_j\,\mu(H(i,j))\,.$$

If $\mu(H(i,j)) = 0$, then $a_i\,\mu(H(i,j)) = b_j\,\mu(H(i,j)) = 0$. If $\mu(H(i,j)) \neq 0$, certainly $H(i,j) \neq \emptyset$; if $x \in H(i,j)$, $a_i = f(x) = b_j$, and so again $a_i\,\mu(H(i,j)) = b_j\,\mu(H(i,j))$. Hence, $\sum_{i=1}^m a_i\,\mu(E(i)) = \sum_{i,j} a_i\,\mu(H(i,j)) = \sum_{i,j} b_j\,\mu(H(i,j)) = \sum_{j=1}^n b_j\,\mu(F(j))$.

The argument for the signed measure σ is identical in form. In the previous paragraph, the sums considered are defined because all their terms are non-negative (the coefficients a_i, b_j are values of f, and so non-negative), whereas, in the case of σ, the sums are defined *and finite* because all the terms are finite. Specifically, if $a_i \neq 0$, then $E(i) \subseteq Q(f)$, and, as f is σ-integrable, 5.8 ensures that $\sigma(E(i)) \in \mathbb{R}$; the term $a_i\,\sigma(E(i))$ is finite. If $a_i = 0$, the convention 2.1*(iv)* ensures that $a_i\,\sigma(E(i)) = 0$. Similar arguments apply to the other sums that appear. (Compare 5.7.) □

Definition 13.3. When $\mathcal{F}$ is a field in Ω, $f : \Omega \longrightarrow \overline{\mathbb{R}}$ is a non-negative $\mathcal{F}$-simple function, and μ is a fam on $\mathcal{F}$, define the *pre-integral of f with respect to μ* to be the sum

$$\mathcal{S}(f, \mathcal{F}, \mu) := \sum\nolimits_{i=1}^m a_i\,\mu(E(i))\,, \tag{33}$$

where $\sum_{i=1}^m a_i\,\mathbf{1}_{E(i)}$ is a standard form for f. By 13.2, the sum is defined in $\overline{\mathbb{R}}$ and does not depend on the choice of standard form; nor is it necessary to assume f is integrable with respect to μ.

Similarly, if σ is a fasm on $\mathcal{F}$ and $f : \Omega \longrightarrow \mathbb{R}$ is an $\mathcal{F}$-simple function integrable with respect to σ, the *pre-integral of f with respect to σ* is the sum

$$\mathcal{S}(f, \mathcal{F}, \sigma) := \sum\nolimits_{i=1}^{m} a_i \, \sigma(E(i)) \,,$$

which is defined in $\mathbb{R}$ (and independent of the choice of standard form for f) because f is integrable with respect to σ.

The term "pre-integral" is an *ad hoc* invention. Most authors would call it the "integral" of f, with respect to μ or σ as the case may be; it is the only reasonable value for the "integral" of an $\mathcal{F}$-simple function f. I shall abbreviate the notation to $\mathcal{S}(f)$ when the other data have been fixed. The problem lies in extending the idea of the integral to more general functions, and that is where the σ-additivity of the measure will come in.

Lemma 13.4. *Let $\mathcal{F}$ be a field in Ω, μ a fam on $\mathcal{F}$, $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ non-negative finite-valued $\mathcal{F}$-simple functions, and α, β non-negative extended real numbers. Then*

(a) $\mathcal{S}(f) \geq 0$ *;*
(b) *if f is finite-valued, $\mathcal{S}(f) < \infty$ if and only if f is integrable;*
(c) $\mathcal{S}(\alpha f + \beta g) = \alpha \mathcal{S}(f) + \beta \mathcal{S}(g)$ *; in particular, $\mathcal{S}(0) = 0$;*
(d) *if $f \leq g$ (pointwise), then $\mathcal{S}(f) \leq \mathcal{S}(g)$.*

Proof. Each term in (33) is non-negative, so *(a)* follows. If a_i is finite, the term $a_i \, \mu(E(i))$ will be finite if and only if *either* a_i is 0 *or* the measure $\mu(E(i)) < \infty$. This proves *(b)*.

Suppose in *(c)* that $\sum_{i=1}^{m} a_i \, \mathbf{1}_{E(i)}$ is a standard form for f and $\sum_{i=1}^{n} b_j \, \mathbf{1}_{F(j)}$ a standard form for g. Define $H(i,j) := E(i) \cap F(j)$, and then $\sum_{1 \leq i \leq m,\, 1 \leq j \leq n} (\alpha a_i + \beta b_j) \, \mathbf{1}_{H(i,j)}$ is a standard form for $\alpha f + \beta g$. It follows (compare 13.2) that

$$\begin{aligned} \mathcal{S}(\alpha f + \beta g) &:= \sum\nolimits_{i,j} (\alpha a_i + \beta b_j) \, \mu(H(i,j)) \\ &= \alpha \sum\nolimits_i a_i \Big(\sum\nolimits_j \mu(H(i,j)) \Big) + \beta \sum\nolimits_j b_j \Big(\sum\nolimits_i \mu(H(i,j)) \Big) \\ &= \alpha \sum\nolimits_i a_i \, \mu(E(i)) + \beta \sum\nolimits_j b_j \, \mu(F(j)) = \alpha \mathcal{S}(f) + \beta \mathcal{S}(g) \,. \end{aligned}$$

If $f \leq g$, then, for each pair (i,j) such that $H(i,j) \neq 0$, $a_i \leq b_j$. So

$$\mathcal{S}(f) := \sum\nolimits_{i,j} a_i \, \mu(H(i,j)) \leq \sum\nolimits_{i,j} b_j \, \mu(H(i,j)) = \mathcal{S}(g) \,, \qquad \square$$

Part *(b)* is still true if one considers the pre-integrals with respect to a fasm σ on $\mathcal{F}$ of $\mathcal{F}$-simple functions integrable with respect to σ.

Corollary 13.5. *Suppose $\mathcal{F}$ is a field in Ω, μ a fam on $\mathcal{F}$, (f_n) be an increasing sequence of non-negative $\mathcal{F}$-simple functions $\Omega \longrightarrow \overline{\mathbb{R}}$. Then $(\mathcal{S}(f_n, \mathcal{F}, \mu))$ is an increasing sequence in $\overline{\mathbb{R}}$, and so converges in $\overline{\mathbb{R}}$ by* 2.7. $\square$

In view of 12.14, this immediately suggests a way of defining the "integral" of a non-negative Σ-measurable function, where Σ is a Σ-field in Ω, that is reasonably consonant with our idea of the integral as the "area under the graph". The only problem is that a non-negative Σ-measurable function may be the limit of many different increasing sequences of non-negative Σ-simple functions. Here, for the first time, we make essential use of σ-additivity.

Proposition 13.6. *Let (Ω, Σ, μ) be a measure space, and let (f_n) be an increasing sequence of non-negative Σ-simple functions $\Omega \longrightarrow \overline{\mathbb{R}}$. Suppose that g is also a non-negative Σ-simple function, and that $\lim f_n \geq g$. Then $\lim \mathcal{S}(f_n) \geq \mathcal{S}(g)$.*

Proof. There are two cases: when g is or is not integrable. Suppose firstly that it is not integrable. Since g has only finitely many values anyway, there must be some $a > 0$ such that $\mu(g^{-1}(\{a\})) = \infty$. Define

$$E(n) \coloneqq \{x \in \Omega : f_n(x) \geq \tfrac{1}{2}a\}. \tag{34}$$

Since $f_n \uparrow$, clearly $E(n) \uparrow$ too. As $\lim f_n \geq g$, $\bigcup_{n=1}^{\infty} E(n) \supseteq g^{-1}(\{a\})$. By 8.5*(a)*, $\mu(E(n)) \uparrow \mu(\bigcup_{n=1}^{\infty} E(n)) \geq \mu(g^{-1}(\{a\}) = \infty$. However, (34) shows that $f_n \geq \frac{1}{2}a\mathbf{1}_{E(n)}$, and so $\mathcal{S}(f_n) \geq \mathcal{S}(\frac{1}{2}a\mathbf{1}_{E(n)}) = \frac{1}{2}a\mu(E(n))$ by 13.4*(d)*; and so $\mathcal{S}(f_n) \uparrow \infty = \mathcal{S}(g)$.

Now, suppose that g is integrable. (There is a special case when $\mu(Q(g)) = 0$; then $\mathcal{S}(g) = 0$, so the result is immediate; but the following argument still works.) Since g is simple, it has only finitely many values. Let M be its largest value.

Take any $\lambda \in (0, 1)$. For each $x \in Q(g)$, $g(x) > \lambda g(x)$. Define — for this fixed λ —

$$F(n) \coloneqq \{x \in Q(g) : f_n(x) \geq \lambda g(x)\} \subseteq Q(g). \tag{35}$$

As at (34), $F(n) \uparrow$, because $f_n \uparrow$; in this case, however, $\lim F(n) = Q(g)$. However,

$$g = g.\mathbf{1}_{F(n)} + g.\mathbf{1}_{Q(g)\setminus F(n)},$$

and $g.\mathbf{1}_{F(n)}$, $g.\mathbf{1}_{Q(g)\setminus F(n)}$ are both non-negative integrable Σ-simple functions. By 13.4*(c)*,

$$\mathcal{S}(g) = \mathcal{S}(g.\mathbf{1}_{F(n)}) + \mathcal{S}(g.\mathbf{1}_{Q(g)\setminus F(n)}). \tag{36}$$

Each of the terms on the right of (36) is finite and non-negative, and from (35)

$$g.\mathbf{1}_{Q(g)\setminus F(n)} \leq M\mathbf{1}_{Q(g)\setminus F(n)}, \quad \mathcal{S}(g.\mathbf{1}_{Q(g)\setminus F(n)}) \leq M\mu(Q(g) \setminus F(n)).$$

Putting these facts together with 13.4*(c)*, *(d)*,

$$\begin{aligned} \mathcal{S}(f_n) &\geq \mathcal{S}(\lambda g.\mathbf{1}_{F(n)}) \qquad \text{by (35)} \\ &= \lambda\mathcal{S}(g.\mathbf{1}_{F(n)}) \\ &= \lambda\mathcal{S}(g) - \lambda\mathcal{S}(g.\mathbf{1}_{Q(g)\setminus F(n)}) \\ &= \lambda\mathcal{S}(g) - \lambda M\mu(Q(g) \setminus F(n)). \end{aligned}$$

Now, however, $F(n) \uparrow Q(g)$ and $\mu(Q(g)) < \infty$, so, by 8.5*(b)*, $\mu(Q(g) \setminus F(n)) \downarrow 0$. Hence, $\lim \mathcal{S}(f_n) \geq \lambda\mathcal{S}(g)$. However, λ could have been any number in $(0, 1)$; we conclude that $\lim \mathcal{S}(f_n) \geq \mathcal{S}(g)$. □

Theorem 13.7. *Let (Ω, Σ, μ) be a measure space, and suppose that (f_n) and (g_n) are increasing sequences of non-negative simple functions $\Omega \longrightarrow \overline{\mathbb{R}}$ that have the same limit. Then the sequences $(\mathcal{S}(f_n, \Sigma, \mu))$ and $(\mathcal{S}(g_n, \Sigma, \mu))$ have the same limit in $\overline{\mathbb{R}}$.*

That is: if, for each $x \in \Omega$, $f_n(x) \uparrow h(x)$ and $g_n(x) \uparrow h(x)$, then the corresponding pre-integrals have the same limit: $\lim \mathcal{S}(f_n) = \lim \mathcal{S}(g_n)$.

Proof. Fix k. Then $g_k \leq \lim f_n$, and so, by 13.6, $\mathcal{S}(g_k) \leq \lim \mathcal{S}(f_n)$. This holds for each k; ergo, $\lim \mathcal{S}(g_n) \leq \lim \mathcal{S}(f_n)$. The converse inequality must also hold, by symmetry. □

§14. The integral in general.

Throughout this section (and, indeed, until further notice) (Ω, Σ, μ) is a fixed measure space.

Definition 14.1. Let f be a non-negative Σ-measurable function $\Omega \longrightarrow \overline{\mathbb{R}}$. By 12.14, there is an increasing sequence (f_n) of non-negative Σ-simple functions that converges pointwise to f. Define the *integral of f with respect to the measure μ*, $\int f\, d\mu$, by

$$\int f\, d\mu := \lim_n \mathcal{S}(f_n)\,.$$

The limit exists by 13.5, and, by 13.7, it does not depend on the choice of the sequence (f_n). (It would be possible to *define* the integral using the specific sequence given in 12.14, but 13.7 is essential to show that it has useful properties). It will be convenient to write "isnsf" for "increasing sequence of non-negative simple functions".

The notation is very variable, according to the data that are fixed. For the time being, we may write just $\int f$. But if the space, the σ-algebra, the measure, or the "variable of integration" are in doubt (for f may involve several parameters), one may see

$$\int_\Omega f\,,\ \int_\Omega f\, d\mu\,,\ \int_{\Omega,\Sigma} f\, d\mu\,,\ \int f(x)\, d\mu(x)\,,\ \int_{\Omega,\Sigma} f(\omega)\, d\mu(\omega)\,.$$

The "d" in these expressions has no independent meaning — it is a historical survival from Leibniz's notation for the integral, which was itself illogical; but it has the merit of corresponding to the phrase "with respect to". A slightly less irrational notation that is also in use is $\int f(x)\, \mu(dx)$, where "$\mu(dx)$" does at least vaguely indicate the idea of assigning "mass" to "small bits dx of the domain Ω". Another notation that is occasionally met with is $\mu(f)$. The probabilistic notation is quite different, and I shall explain it later.

Notice that, for a non-negative measurable function f, $\int f$ is defined and is in $\overline{\mathbb{R}}$. If f is, in fact, a non-negative simple function, then $\int f = \mathcal{S}(f)$, since the chosen sequence may consist of f alone. This is why $\mathcal{S}(f)$ is usually called the "integral" of the simple function f.

Lemma 14.2. *Suppose f, g are non-negative measurable functions and $f \le g$ (that is, for each $x \in \Omega$, $f(x) \le g(x)$; see* 7.15*). Then $\int f \le \int g$.*

Proof. Take isnsfs $f_n \uparrow f$, $g_n \uparrow g$. Then, for any fixed k, $f_k \le g = \lim g_n$. Apply 13.6; $\mathcal{S}(f_k) \le \lim \mathcal{S}(g_n) = \int g$, and so, letting $k \to \infty$, $\int f = \lim \mathcal{S}(f_k) \le \int g$. □

Definition 14.3. The non-negative measurable function f is *integrable* if $\int f < \infty$.

This is consistent with the terminology for finite-valued simple functions, by 13.4*(b)*. Notice, however, that for a non-negative measurable function the integral is always *defined*; integrability means something else. For this reason, some authors use a different convention. What we call "integrable", they call "summable", both here and later.

I recall the abbreviation (introduced for simple functions, but useful in general)

$$Q(f) \coloneqq \{x \in \Omega : f(x) \neq 0\}\,.$$

There are, vaguely speaking, two ways in which a non-negative measurable function may be non-integrable. Its *values* may be "too large"; or $Q(f)$ may be too large. To make this a little more precise:

Definition 14.4. A set $A \subseteq \Omega$ is *σ-finite* with respect to (Σ, μ) if there is a sequence (A_i) in Σ such that $A \subseteq \bigcup_{i=1}^{\infty} E_i$ and, for each i, $\mu(E_i) < \infty$.

(The phrase "σ-finite" occurs in several contexts; for instance, μ might be substituted by a signed measure on Σ).

Lemma 14.5. *If the non-negative measurable function $f : \Omega \longrightarrow \overline{\mathbb{R}}$ is integrable, then*

(a) *$Q(f)$ is σ-finite,* (b) $\mu(\{x : f(x) = \infty\}) = 0$.

Proof. *(a)* Let (f_n) be an isnsf tending to f. Then, by definition, $\mathcal{S}(f_n) \uparrow \int f$, and so $\mathcal{S}(f_n) < \infty$ for each n. By 13.4*(a)*, $\mu(Q(f_n)) < \infty$. But, clearly,

$$Q(f) = \bigcup_{n=1}^{\infty} Q(f_n)\,,$$

and, consequently, $Q(f)$ is σ-finite.

(b) Let $A \coloneqq \{x \in \Omega : f(x) = \infty\} \in \Sigma$. Then, for any $n \in \mathbb{N}$, $f \geq n\mathbf{1}_A$. By 14.2,

$$\infty > \int f \geq \int n\mathbf{1}_A = \mathcal{S}(n\mathbf{1}_A) = n\mu(A)\,.$$

But this shows that $\mu(A) \leq \inf\{\int f/n : n \in \mathbb{N}\} = 0$, and so $\mu(A) = 0$. □

Lemma 14.6. *Let f be a non-negative measurable function. Then $\int f = 0$ if and only if $\mu(Q(f)) = 0$.*

Proof. Let (f_n) be an isnsf tending to f. If $\mu(Q(f)) = 0$, then evidently $\mu(Q(f_n)) = 0$ for each n, since $Q(f_n) \subseteq Q(f)$, and thus $\mathcal{S}(f_n) = 0$. Hence $\int f = \lim \mathcal{S}(f_n) = 0$.

Conversely, suppose $\mu(Q(f)) > 0$. Define $A(n) \coloneqq \{x \in \Omega : f(x) > 1/n\} \in \Sigma$, for each $n \in \mathbb{N}$. Then $A(n) \subseteq A(n+1)$ for each n, and $Q(f) = \bigcup_{n=1}^{\infty} A(n)$ as f is non-negative. By 8.5*(a)*, $\mu(Q(f)) = \lim \mu(A(n))$, so that, for sufficiently large n, $\mu(A(n)) > 0$. Since $f_n \geq n^{-1}\mathbf{1}_{A(n)}$, 14.2 shows that then $\int f \geq \int f_n \geq n^{-1}\mu(A(n)) > 0$. This establishes the desired result. □

Remark 14.7. The above lemma has a number of consequences. For a start, it is important to realize that *a non-negative measurable function may have zero integral without vanishing identically*. If f is the "Dirichlet function" on $[0,1]$ mentioned in §1A,

$$f(t) = 1 \quad \text{when } t \in \mathbb{Q} \cap [0,1]\,, \qquad f(t) = 0 \quad \text{otherwise,}$$

then f is in fact the indicator function of $\mathbb{Q} \cap [0,1]$ (and so a simple function). $\mathbb{Q} \cap [0,1]$ is

countable, so its Lebesgue measure is 0, and $\int f\, d\lambda = 0$, λ denoting Lebesgue measure. This is the more surprising in that f is not Riemann-integrable at all.

In elementary courses, one often meets assumptions of the kind that the (Riemann) integral of a non-negative function is zero only if the function is zero; generally speaking, this is because the functions are continuous. Indeed, it is amazingly difficult to prove, on the basis of the Riemann integral alone, even that the Riemann integral of a function that is Riemann-integrable and *everywhere* positive on $[0, 1]$ must be positive. Any attempt to prove it carries you a long way towards the Lebesgue integral. Yet our proof above of a more general fact for the Lebesgue integral was not hard, and, generally speaking, we shall find that many results which arise fairly naturally for Lebesgue integration and ought, therefore, to be true for the Riemann integral (in suitable formulations) are either false or far less easy to prove.

In these remarks I am taking for granted that Lebesgue integration with respect to Lebesgue measure generalizes Riemann integration, in the sense that a Riemann-integrable function on a multi-interval is necessarily Lebesgue-integrable with the same value for the integral. It is rather probable that I shall not give a proof of this — although it is not very difficult, it is rather time-consuming — but I shall assume it occasionally.

The fact that certain functions are not "detected" by the integral has led to a whole list of phrases expressing the idea.

Definition 14.8. Let P be a property of points of Ω [for example, if (f_n) is a sequence of functions $\Omega \longrightarrow \mathbb{R}$, $P(x)$ might mean that $f_n(x) \to f(x)$; this may be true for some $x \in \Omega$ and false for some others]. We say that P holds (Σ, μ)-*almost everywhere in* Ω if there is a set $E \in \Sigma$ such that $\mu(E) = 0$ and $P(x)$ is true whenever $x \notin E$. That is to say, the set of points x for which $P(x)$ is *false* is a subset of a measurable set of measure 0.

It is of course "usually" the case that the set of points x for which $P(x)$ is false (often called, especially in informal discussion, the *exceptional set*) is itself measurable; in §10, I explained that there are reasons why this should be so in $\mathbb{R}^n$, although we cannot assume it, for all "practical" properties P. However, the case may be altered for peculiar spaces or measures or properties.

One often says "P holds a.e." In older books in English, "p.p." (for *presque partout*; Lebesgue was French) is sometimes used instead of a.e., and of course corresponding abbreviations are (sometimes) used in other languages. There are also other forms of words, such as "$P(x)$ for almost all x in Ω".

Lemma 14.9. *Suppose that $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ are non-negative measurable functions and α, β are non-negative extended real numbers. Then*

$$\int (\alpha f + \beta g) = \alpha \int f \, + \, \beta \int g \, .$$

Proof. Suppose $\alpha = \infty$. If $\int f > 0$, then by 14.6 $\mu(Q(f)) > 0$, and since

$$\{x \in \Omega : \alpha f(x) = \infty\} = Q(f) \, ,$$

14.5*(b)* shows αf is non-integrable. So $\int \alpha f = \infty = \alpha \int f$. If $\alpha = \infty$ and $\int f = 0$, then,

from 14.6, $\mu(Q(f)) = 0$; hence, $\mu(Q(\alpha f)) = 0$; and 14.6 shows in turn that

$$\int \alpha f = 0 = \infty.0 = \alpha \int f \,.$$

Now suppose $0 \le \alpha < \infty$, and let (f_n) be an isnsf with $f_n \uparrow f$. Then (αf_n) is also an isnsf, $\alpha f_n \uparrow \alpha f$, and $\mathcal{S}(\alpha f_n) = \alpha \mathcal{S}(f_n)$ for each n, by 13.4*(c)*. So

$$\int \alpha f = \lim \mathcal{S}(\alpha f_n) = \lim \alpha \mathcal{S}(f_n) = \alpha \lim \mathcal{S}(f_n) = \alpha \int f \,.$$

It will now suffice to show that $\int (f+g) = \int f + \int g$. Take isnsfs $f_n \uparrow f$, $g_n \uparrow g$; then $(f_n + g_n)$ is an isnsf and $(f_n + g_n) \uparrow (f+g)$. So, again using 13.4*(c)*,

$$\begin{aligned} \int (f+g) &= \lim \mathcal{S}(f_n + g_n) = \lim(\mathcal{S}(f_n) + \mathcal{S}(g_n)) \\ &= \lim \mathcal{S}(f_n) + \lim \mathcal{S}(g_n) = \int f + \int g \,. \end{aligned}$$ □

Definition 14.10. Let $f : \Omega \longrightarrow \overline{\mathbb{R}}$. Define

$$f^+ := \max(f, 0) \,, \quad f^- := -\min(f, 0) = \max(-f, 0) \,.$$

These are, of course, the *pointwise* maximum and minimum (see 7.15): for each $x \in \Omega$, $f^+(x) = \max(f(x), 0)$ and $f^-(x) = -\min(f(x), 0) = \max(-f(x), 0)$.

Lemma 14.11. (a) f^+, f^- *are non-negative.*
(b) *If* $f : \Omega \longrightarrow \overline{\mathbb{R}}$ *is* Σ*-measurable, both* f^+ *and* f^- *are* Σ*-measurable.*
(c) $f = f^+ - f^-$.

Proof. Only *(b)* is not obvious, and it follows from 12.6*(b), (c)*. For *(c)*, notice that f^+ and f^- cannot take opposite infinite values at the same point. □

Briefly, any function f can be expressed as the difference of two non-negative functions, both measurable if f is. However, f can usually be so expressed in many ways. The form $f^+ - f^-$ is the most 'economical', because f^+ and f^- are the smallest possible functions that can appear in such a decomposition. Another way of putting the same idea is that, for each point $x \in \Omega$, either $f^+(x) = 0$ or $f^-(x) = 0$. In graphical terms, f^+ correspond to the "part of the graph that is above the axis", and so on.

Definition 14.12. Let $f : \Omega \longrightarrow \overline{\mathbb{R}}$ be Σ-measurable. The *integral of* f *is defined* when $\int f^+$ and $\int f^-$ are not both infinite. f is *integrable* if both $\int f^+$ and $\int f^-$ are finite. In either case, the *integral of* f is to be

$$\int f := \int f^+ - \int f^- \,.$$

Thus, f is integrable precisely when its integral *is defined* and *is finite*. Also, notice that, by specifying f^+ and f^- , we are ensuring that the class of integrable functions is as large as

possible for a definition that depends on expressing f as the difference of two non-negative measurable functions. The most important elementary observation is this:

Lemma 14.13. *The measurable function f is integrable if and only if $|f|$ (which is measurable and non-negative) is integrable.*

Proof. $|f| = f^+ + f^-$, by considering the cases $f(x) \leq 0$ and $f(x) > 0$ separately. So, if $\int f^+ < \infty$ and $\int f^- < \infty$, $\int |f| = \int f^+ + \int f^- < \infty$ by 14.9. Similarly, from 14.2, $\int f^+ \leq \int |f|$ and $\int f^- \leq \int |f|$, so that, if $|f|$ is integrable, f^+ and f^- must be too. □

It is unfortunate that Definition 14.12 does not immediately lead to the central properties of the integral, which I shall postpone to the next section. You pay a price for the simplicity of the definition. For the moment, it should be observed that we have only defined the integral over the whole of Ω.

Lemma 14.14. *Let $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ be Σ-measurable functions, and suppose that $f = g$ a.e. Then $\int f$ exists if and only if $\int g$ exists, and $\int f = \int g$. (In particular, f is integrable if and only if g is integrable.)* □

Lemma 14.15. (a) *Let $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ be measurable, suppose $|f| \leq g$ a.e. (which implies that g is a.e. non-negative), and let g be integrable. Then f is integrable.*

(b) *If the integral of $h : \Omega \longrightarrow \overline{\mathbb{R}}$ exists, then $|\int h| \leq \int |h|$ (which is defined, as h is non-negative measurable).*

Proof. *(a)* follows from 14.2, 14.8, and 14.13. As for *(b)*, by 14.9

$$\left| \int h \right| = \left| \int h^+ - \int h^- \right| \leq \int h^+ + \int h^- = \int |h| \,. \qquad \square$$

Remark 14.16. Suppose that $M \in \Sigma$. Define

$$\Sigma^M := \{E \cap M : E \in \Sigma\} \,.$$

It is easily seen that Σ^M is a σ-field of subsets of M, and, of course, $\Sigma^M \subseteq \Sigma$. Consequently, the restriction of μ to Σ^M is a measure on Σ^M, and our definitions will yield the ideas of Σ^M-measurability or "relative measurability" (both of subsets of M and of functions $M \longrightarrow \overline{\mathbb{R}}$), of Σ^M-simplicity of functions, and of the integral with respect to Σ^M and $\mu|\Sigma^M$.

In particular, a Σ-measurable function $f : \Omega \longrightarrow \overline{\mathbb{R}}$ restricts to a Σ^M-measurable function on M, whose integral, when it exists, may be denoted $\int_M f$. That is,

$$\int_M f := \int_{M, \Sigma^M, \mu|\Sigma^M} f|M \,.$$

Then $\int_M f$ exists if and only if $\int_\Omega f\mathbf{1}_M$ exists, in which case they are equal.

(a) Furthermore, if $\int_\Omega f$ exists, then so does $\int_M f$.

(b) If $\int_\Omega f$ exists and is finite (that is, if f is integrable on Ω), then so is $\int_M f$ (f is integrable on M).

(c) If $\int_\Omega f$ exists and $\int_M f$ has an infinite value, then $\int_\Omega f$ must have the same infinite value.

(d) If f is non-negative (on Ω), then $\int_\Omega f \geq \int_M f$.

On the other hand, if $g : M \longrightarrow \overline{\mathbb{R}}$ is Σ^M-measurable, define $\widehat{g} : \Omega \longrightarrow \overline{\mathbb{R}}$ by

$$\widehat{g}(x) := g(x) \quad \text{when } x \in M\,, \qquad \widehat{g}(x) := 0 \quad \text{when } x \notin M\,.$$

Then $\widehat{g}$ is Σ-measurable, and $\int_\Omega \widehat{g}$ is defined if and only if $\int_{M,\,\Sigma^M,\,\mu|\Sigma^M} g$ is defined, and in that case they are equal.

These facts explain why I have defined the integral only on the whole of Ω. They are straightforward consequences of the definitions, and I leave the proofs as exercises.

§15. Properties of the integral.

It will be easiest to begin with one of the great theorems about the Lebesgue integral, which on our definition (unlike Lebesgue's) becomes rather easy. As before, the measure space (Ω, Σ, μ) is fixed. First, a lemma:

Lemma 15.1. *Suppose that, for each* $n \in \mathbb{N}$, $f_n : \Omega \longrightarrow \overline{\mathbb{R}}$, *and that* $f_n \leq f_{n+1}$ *a.e. for each* n. *Then the sequence* (f_n) *is increasing a.e.*

Proof. By hypothesis, there is, for each n, a set $Z_n \in \Sigma$ such that $\mu(Z_n) = 0$ and $f_n(x) \leq f_{n+1}(x)$ when $x \notin Z_n$. Take $Z := \bigcup_{k=1}^\infty Z_k$; then $Z \in \Sigma$, and 8.3 shows that

$$0 \leq \mu(Z) \leq \sum\nolimits_{k=1}^\infty \mu(Z_k) = 0\,,$$

whilst, if $x \notin Z$, then $x \notin Z_n$ for all n, so that $f_n(x) \leq f_{n+1}(x)$ for all n simultaneously. That is, the sequence (f_n) is increasing a.e. □

Remark 15.2. More generally, the argument shows that the conjunction of any countable class of properties, each of which holds a.e., also holds a.e. This general fact saves authors the effort of distinguishing clearly between the various meanings of "a.e." when applied to properties of sequences — as, for instance, "increasing a.e." ought to mean, as I used it above, that the property "for all n, $f_n(x) \leq f_{n+1}(x)$" holds a.e., but it could easily be misunderstood to mean that, for each individual n, $f_n(x) \leq f_{n+1}(x)$ a.e. The Lemma reassures us that the distinction is unimportant in practice.

Now the big theorem. Lately it is usually called in English the Monotone Convergence Theorem; but it is still sometimes known as Beppo Levi's theorem. (The use of his forename is, I suppose, because there was another well-known Italian mathematician of the same surname at the time, E. E. Levi.) Because of our approach to the integral, which was influenced by this theorem to begin with, the proof becomes mere "bookkeeping", albeit not completely obvious; all the hard work is in 13.6; but deducing it directly from Lebesgue's definition as B. Levi did is less easy.

Theorem 15.3. Let (f_n) be an a.e. increasing sequence of a.e. non-negative measurable functions $\Omega \longrightarrow \overline{\mathbb{R}}$. Suppose $f : \Omega \longrightarrow \overline{\mathbb{R}}$ is measurable and $f_n \uparrow f$ a.e. Then

$$\int f = \lim \int f_n \,.$$

Proof. By 15.2, we may remove an exceptional set Z of measure zero. It will suffice, by 14.14, to prove the result on the assumption that $0 \le f_n \le f_{n+1} \uparrow f$ at every point of Ω, and then consider $\Omega \setminus Z$ (see 14.16) instead of Ω.

For each n, let $(g_k^{(n)})_{k=1}^{\infty}$ be an isnsf such that $g_k^{(n)} \underset{k}{\uparrow} f_n$. (The non-standard symbol $\underset{k}{\uparrow}$ means "tends as $k \to \infty$", with the implication that n is fixed.) Set, for given k,

$$h_k := \max\{g_k^{(n)} : n \le k\} \,.$$

By 12.12, h_k is also simple. But also, $g_{k+1}^{(n)} \ge g_k^{(n)}$ for each n and k, so that

$$\begin{aligned} h_{k+1} &= \max\{g_{k+1}^{(n)} : n \le k+1\} \ge \max\{g_k^{(n)} : n \le k+1\} \\ &\ge \max\{g_k^{(n)} : n \le k\} = h_k \,, \end{aligned}$$

so that (h_k) is increasing. Fix n. When $k \ge n$, $h_k \ge g_k^{(n)}$ by definition, so that

$$\lim_k h_k \ge \lim_k g_k^{(n)} = f_n \,.$$

Since this is true for each n, $\lim_k h_k \ge \sup_n f_n = \lim_n f_n = f$.

On the other hand, for every n and l, $g_l^{(n)} \le f_n$, so that $h_l \le f_l$ and $\lim_l h_l \le f$. The conclusion must be that $\lim_k h_k = f$. Thus (h_k) is an isnsf, $h_k \uparrow f$, and by 14.1

$$\int f = \lim \mathcal{S}(h_k) \,.$$

But, from 13.6, as $h_k \le f_k = \lim_l g_l^{(k)}$, $\mathcal{S}(h_k) \le \lim_l \mathcal{S}(g_l^{(k)}) = \int f_k \le \int f$ by 14.2. So

$$\int f \le \lim \int f_k \le \int f \,.$$

This is the result. □

Notice the advantage of allowing infinite values here. The corresponding result for Riemann integrals would include the hypothesis that f, as well as each f_n, is Riemann-integrable; and it would not be easy to prove by purely "Riemann-integral" methods.

Corollary 15.4. If (f_n) is a sequence of a.e. non-negative measurable functions on Ω,

$$\int \left(\sum_{n=1}^{\infty} f_n \right) = \sum_{n=1}^{\infty} \left(\int f_n \right) .$$

Proof. Apply 15.3 to the sequence of partial sums, recalling 14.9. □

This result is more striking than useful. The next one, however, is different; although superficially rather unmemorable, it turns out to be by far the easiest method of deducing some non-obvious facts later on. Because of this, it has a name: *Fatou's lemma*.

Lemma 15.5. *Let* (f_n) *be* any *sequence of non-negative measurable functions. Then*

$$\int(\liminf f_n) \leq \liminf \int f_n\,.$$

Proof. Set $g_k := \inf_{n\geq k} f_n$, for each $k \in \mathbb{N}$. Then g_k is also measurable and non-negative, and $g_k \underset{k}{\uparrow} \underline{\lim}\, f_n$. By the monotone convergence theorem 15.3,

$$\int \underline{\lim}_n f_n = \lim_k \int g_k\,. \tag{37}$$

However, for each $n \geq k$, $f_n \geq g_k$ and so $\int g_k \leq \int f_n$ (by 14.2). This being so for all $n \geq k$, in fact $\int g_k \leq \inf_{n\geq k} \int f_n$. Take the limit as $k \to \infty$: $\lim_k \int g_k \leq \underline{\lim}_n \int f_n$. The result follows by putting this together with (37). □

Although we cannot at present fill in the details, Fatou's lemma may be understood in terms of an alternative definition of the integral. It is possible to define, from the given measure space (Ω, Σ, μ) and from $\mathbb{R}$ with Lebesgue measure, a "product measure space" consisting of a "product σ-algebra" $\Sigma^{\otimes}$ in $\Omega \times \mathbb{R}$ and a "product measure" $\mu^{\otimes}$ thereon, satisfying various appropriate properties. A non-negative function f on Ω is Σ-measurable if and only if the ordinate sets of 7.14 are $\Sigma^{\otimes}$-measurable; in that case the integral of f is the product measure of either ordinate set. (This is obviously the measure-theoretic interpretation of Leibniz's definition of the integral.) Then Fatou's lemma results from the first assertion of 7.16 and from 8.6. The interest of this way of looking at it is that it reinforces the need to consider lower limits, which may at first seem rather strange.

We now return to the properties of the integral.

Proposition 15.6. *Let (M_n) be a sequence of measurable sets such that $\mu(M_i \cap M_j) = 0$ when $i \neq j$; let $M := \bigcup_{n=1}^{\infty} M_n$, and suppose that $f : M \longrightarrow \overline{\mathbb{R}}$ is relatively measurable on M and a.e. non-negative. Then*

$$\int_M f = \sum\nolimits_{n=1}^{\infty} \left(\int_{M_n} f \right).$$

Proof. Let $f_n := f\mathbf{1}_{M_n}$. Then f_n is relatively measurable and a.e. non-negative on M, and

$$f = \sum\nolimits_{n=1}^{\infty} f_n \quad \text{a.e. on } M$$

(the points where equality fails form a subset of $\bigcup_{i \neq j} (M_i \cap M_j)$, which, as a countable union of measurable sets of measure zero, is itself measurable of measure zero). The result follows from 14.14 and 15.4. □

The above proposition concerns a.e. non-negative functions, integrable or not. The next one is about general measurable functions.

Lemma 15.7. Suppose $M_1, M_2, \ldots, M_n$ are measurable sets and $M := \bigcup_{i=1}^{n} M_i$; let $f : M \longrightarrow \overline{\mathbb{R}}$ be relatively measurable on M. Then f is integrable on M if and only if it is integrable on M_i for each i. If $\mu(M_i \cap M_j) = 0$ whenever $i \neq j$, then

$$\int_M f = \sum\nolimits_{i=1}^{n} \left(\int_{M_i} f \right).$$

Proof. In one direction, the result follows from 14.16*(b)*. So, suppose that f is integrable on each M_i. 'Disjunctify' in the usual way, 4.7, so that M is expressed as a disjoint union of measurable sets M_i' for which $M_i' \subseteq M_i$ for each i. By 14.16*(b)*, f is integrable on each M_i'. Now apply 15.6, taking $M_{n+1}' = M_{n+2}' = \cdots = \emptyset$, to obtain

$$\int_M f^+ = \sum\nolimits_{i=1}^{n} \left(\int_{M_i'} f^+ \right), \quad \int_M f^- = \sum\nolimits_{i=1}^{n} \left(\int_{M_i'} f^- \right),$$

which are both finite; so f is integrable on M.

If $\mu(M_i \cap M_j) = 0$ when $i \neq j$, then 15.6 applies without disjunctification:

$$\int_M f^+ = \sum\nolimits_{i=1}^{n} \left(\int_{M_i} f^+ \right), \quad \int_M f^- = \sum\nolimits_{i=1}^{n} \left(\int_{M_i} f^- \right),$$

and the final assertion follows. □

It is a consequence of our definition of the integral that very often statements are true in two cases: when everything in sight is non-negative (although some values may be infinite), and when the functions are integrable. This is the reason why some very basic properties have been postponed until now.

Since an integrable function is finite a.e. by 14.5*(b)* and 14.13, one may usually assume, by omitting an exceptional set of measure 0, that functions are *everywhere* finite-valued; this

avoids the difficulty of 12.5. As a reminder, I state the next result for finite-valued measurable functions only.

Proposition 15.8. *Let* $f, g : \Omega \longrightarrow \mathbb{R}$ *be integrable, and* $\alpha, \beta \in \mathbb{R}$. *Then* $\alpha f + \beta g$ *is also integrable, and*

$$\int (\alpha f + \beta g) = \alpha \int f + \beta \int g .$$

Proof. If $\alpha \geq 0$, $(\alpha f)^+ = \alpha f^+$ and $\alpha f^- = \alpha f^-$; if $\alpha \leq 0$, $(\alpha f)^+ = -\alpha f^-$ and $(\alpha f)^- = (-\alpha) f^+$. Applying 14.9, one obtains in the first case

$$\int (\alpha f)^+ = \alpha \int f^+ , \quad \int (\alpha f)^- = \alpha \int f^-$$

(which are both finite), and in the second case

$$\int (\alpha f)^+ = (-\alpha) \int f^- , \quad \int (\alpha f)^- = (-\alpha) \int f^+$$

(which again are both finite). So, in either case, αf is integrable, and $\int(\alpha f) = \alpha \int f$.

To complete the proof of the Proposition, it remains to deal with addition. Given the integrable functions f and g, there are in principle eight cases to consider: set

$$\begin{aligned} E_+ &:= \{x \in \Omega : \ f(x) \geq 0\} , & E_- &:= \{x \in \Omega : f(x) < 0\} , \\ F_+ &:= \{x \in \Omega : \ g(x) \geq 0\} , & F_- &:= \{x \in \Omega : g(x) < 0\} , \\ G_+ &:= \{x \in \Omega : \ f(x) + g(x) \geq 0\} , & G_- &:= \{x \in \Omega : \ f(x) + g(x) < 0\} , \\ & H(\epsilon, \zeta, \eta) := E_\epsilon \cap F_\zeta \cap G_\theta , \quad \text{for } \epsilon, \zeta, \theta = \pm . \end{aligned}$$

The eight sets $H(\epsilon, \zeta, \eta)$ are measurable, disjoint, and cover Ω. Some of them, specifically $H(+,+,-)$ and $H(-,-,+)$, are always empty, whatever the functions f and g. By 15.7, it will suffice to prove the result on each $H(\epsilon, \zeta, \eta)$ separately, and in each case this is straightforward from 14.9. For instance, on $H(-,+,-)$, $-f$ and $-f-g$ and g are non-negative and $g + (-f-g) = -f$, so that 14.9 gives

$$\int g + \int (-f-g) = \int (-f) ;$$

as $\int g$ and $\int(-f) = -\int f$ are finite, we deduce that $\int(-f-g)$ must also be finite, and so

$$\int (f+g) = -\int (-f-g) = \int g - \int (-f) = \int f + \int g .$$

The same sort of argument works on any $H(\epsilon, \zeta, \eta)$. □

The conclusion of the proposition is true in some other cases. It is sufficient, for instance, for α to be finite (i.e. in $\mathbb{R}$), f to be integrable, and $\int g$ to exist. I leave this as an exercise.

Lemma 15.9. *Suppose* $f, g : \Omega \longrightarrow \overline{\mathbb{R}}$ *are measurable and their integrals are defined, and* $f \leq g$ *a.e. Then* $\int f \leq \int g$. *Furthermore, if one of* f, g *is integrable and* $f \leq g$ *a.e., then the integral of the other function exists, and* $\int f = \int g$ *only if* $f = g$ *a.e.*

Proof. If $\int f^- = \infty$ (so that $\int f = -\infty$) or if $\int g^+ = \infty$ (so that $\int g = \infty$) then $\int f \leq \int g$ immediately. If $\int f^- < \infty$ and $\int g^+ < \infty$, then (as, rather trivially, $f^+ \leq g^+$ and $g^- \leq f^-$) both f and g are integrable, and finite a.e., so that 15.8 applies. But

$$\int g - \int f = \int (g - f)$$

(to repeat the point of the remark before 15.8: $g - f$ is defined except on the set of measure zero where f and g have opposite infinite values, so it makes no difference how we define it there — provided it is measurable). But $\int (g - f)$ is the integral of a a.e. non-negative function, so is non-negative. This proves the first assertion.

For the second, suppose f is integrable. Then, as $g^- \leq f^-$, $\int g^- < \infty$, and $\int g$ exists. If $\int f = \int g$, then g is integrable, and by 15.8, $\int (g - f) = 0$. But $g - f \geq 0$ a.e.; omitting the exceptional set, we may apply 14.6 to deduce $g = f$ a.e. □

We may now generalize the monotone convergence theorem.

Lemma 15.10. *Let (f_n) be an a.e. increasing sequence of measurable functions such that $\int f_1$ is defined and is not $-\infty$. If $f = \lim f_n$ a.e., then $\int f = \lim \int f_n$.*

Proof. Apply the monotone convergence theorem to the sequence $(f_n - f_1)$. □

The details of the above proof are left as an exercise.

It is a long time since signed measures were mentioned. They are very common.

Proposition 15.11. *Let (F_n) be a disjoint sequence of measurable sets, $F \coloneqq \bigcup_{n=1}^{\infty} F_n$. Suppose $f : F \longrightarrow \overline{\mathbb{R}}$ is relatively measurable and $\int_F f$ is defined. Then*

$$\int_F f = \sum\nolimits_{n=1}^{\infty} \left(\int_{F_n} f \right).$$

(All these integrals are defined; see 14.16*).*

Proof. Apply 15.6 to f^+ and to f^- separately. □

Notice that the existence of $\int_F f$ must be assumed; it is quite possible for $\sum_{n=1}^{\infty} \left(\int_{F_n} f \right)$ to be defined, and even to be zero, although $\int_F f$ is undefined. (There are easy examples).

One may interpret this in an interesting way. If $\int_\Omega f$ is defined, then the "indefinite integral of f" is the function $\sigma : \Sigma \longrightarrow \overline{\mathbb{R}}$ defined by

$$(\forall E \in \Sigma) \quad \sigma(E) \coloneqq \int_E f\,,$$

and the Proposition states that σ is a signed measure on Σ.

Beppo Levi's theorem is, from our point of view, the first of the great theorems of the Lebesgue theory. It has the obvious disadvantage that monotone sequences of functions are rather rare. The second great theorem, which is free from that disadvantage, is the *dominated convergence theorem*. Apart from these two theorems, Lebesgue's extraordinary contributions

to integration theory had to do with differentiation, and we, like many other recent authors, shall not discuss them seriously; this is regrettable, for they are profound and illuminating, and are of great significance in harmonic analysis — but they have had relatively little influence in wider mathematics, and my aim is to treat the most essential topics.

Theorem 15.12. *Let (f_n) be a sequence of measurable functions $\Omega \longrightarrow \overline{\mathbb{R}}$, and suppose g is a non-negative integrable function on Ω such that $|f_n| \leq g$ a.e. for all n. Then $\overline{\lim} f_n$ and $\underline{\lim} f_n$ are both integrable, and*

$$\int \overline{\lim} f_n \geq \overline{\lim} \int f_n \,, \quad \int \underline{\lim} f_n \leq \underline{\lim} \int f_n \,.$$

In particular, if $f_n \to f$ a.e. (where f is measurable), then f is integrable and

$$\int f = \lim \int f_n \,.$$

Proof. Evidently $|\limsup f_n| \leq g \geq |\liminf f_n|$ a.e. The first assertion follows from 12.7 and 14.15. As g is a.e. finite by 14.12 and 14.5*(b)*, we may remove an exceptional set of measure 0 and assume that all the functions g and f_n are finite-valued (everywhere) and $g \geq f_n \geq -g$ everywhere. As $g - f_n$ and $g + f_n$ are now non-negative measurable functions, we can apply Fatou's lemma 15.5. In view of the linearity of the integral, 15.8, we find

$$\int g - \int \overline{\lim} f_n = \int (g - \overline{\lim} f_n) = \int \underline{\lim} (g - f_n) \leq \underline{\lim} \int (g - f_n) = \int g - \overline{\lim} \int f_n \,,$$

and, since all these numbers are finite, it follows that $\overline{\lim} \int f_n \leq \int \overline{\lim} f_n$. Similarly,

$$\int g + \int \underline{\lim} f_n = \int (g + \underline{\lim} f_n) = \int \underline{\lim} (g + f_n) \leq \underline{\lim} \int (g + f_n) = \int g + \underline{\lim} \int f_n \,,$$

so that $\int \underline{\lim} f_n \leq \underline{\lim} \int f_n$.

To complete the proof, notice that, if $f_n \to f$, then $\underline{\lim} f_n = \overline{\lim} f_n = f$ a.e., and the chain of inequalities (the middle one comes from 7.6*(c)*)

$$\int \underline{\lim} f_n \leq \underline{\lim} \int f_n \leq \overline{\lim} \int f_n \leq \int \overline{\lim} f_n$$

has ends equal to each other. So all the inequalities must be equalities. □

The name of the theorem arises from the statement that the sequence (f_n) is "dominated" by the integrable function g. A particular case is the *bounded convergence theorem*: if f_n is measurable for all n, and so is f, and $f_n \to f$ a.e., and $\mu(\Omega) < \infty$, and there exists some constant K such that $|f_n| \leq K$ a.e. for all n, then $\int f_n \to \int f$ as $n \to \infty$. (This results from taking "g" in 15.0 to be the constant function with value K.)

If we assume that the functions f_n, their limit f, and the dominating function g are all Riemann-integrable, we obtain a theorem for Riemann integrals. Again, it may be proved (very painfully) by using the theory of the Riemann integral, but there is no point in doing so.

The proof above obviously relies on the possibility of integrating such functions as $\inf_{n\geq k} f_n$.

It is very tempting to suppose that the two theorems above about interchanging limits and integrals — the monotone and dominated convergence theorems — altogether supersede the theorems one meets in undergraduate courses. For the simplest such theorem (uniform convergence on a bounded domain), of course, we now have a much better result; *uniform* convergence has been weakened to *dominated* convergence. However, the more advanced and specialized theorems tend to deal with improper integrals of various sorts, and recall that Lebesgue integrals must be "absolutely convergent". Hence, the Lebesgue theorems are sometimes inapplicable because the Lebesgue integral is unavailable. This is irritating, but if you define "integrals" in unusual ways for special purposes you cannot also preserve all the properties of the standard definition and must expect to need special proofs of some results.

Remark 15.13. The dominated convergence theorem only gives a *sufficient* condition that $\int \lim f_n = \lim \int f_n$. For instance, take

$$f_n(x) := \begin{cases} 0 & \text{when } x \leq (n+1)^{-2} \text{ and when } x \geq n^{-2}, \\ n^{5/2} & \text{when } (n+1)^{-2} < x < n^{-2}. \end{cases}$$

Then $f_n \to 0$ pointwise (indeed, there is at most one n such that $f_n(x) \neq 0$), and

$$\int f_n = n^{5/2}(n^{-2} - (n+1)^{-2}) = \frac{(2n+1)n^{5/2}}{n^2(n+1)^2} \to 0,$$

so the limit of the integrals is the integral of the limit. However, if $f \geq f_n$ for all n, then

$$f(x) \geq n^{5/2} \quad \text{for } (n+1)^{-2} < x < n^{-2}, \quad \text{and so}$$
$$\int f \geq \sum n^{5/2}(n^{-2} - (n+1)^{-2}),$$

which diverges by comparison with $\sum n^{-1/2}$. Thus (f_n) is not dominated.

It follows that even the dominated convergence theorem is not the last word.

§16. Introductory remarks on probability theory.

This section could be inserted almost anywhere, but is perhaps most appropriate here.

I commented long ago (in §1A) that the modern approach to probability theory is to say "probability is a measure on the space of events". In fact, the question "what *is* probability?" had been a serious philosophical puzzle for a long time, and I imagine we have all felt some unease when told, for instance, that if you toss a coin randomly (whatever that means), the proportion of heads will settle down in the long run to some number, near $\frac{1}{2}$, and that the limiting proportion is the probability of heads. Why should there be a limiting proportion at all? To be sure, experiments may be, and have been, carried out, but they can neither establish the genuine existence of a limit nor be genuinely "random". I suspect that a good deal of the training one gets in statistics courses is intended to put this unease to sleep. However, it must also be admitted that statistical theory was quite highly developed long before the logical foundations of probability were decently established, and that many famous probabilists did brilliant work on the basis of what one supposes to be intuition.

To get the philosophical question out of the way first: I don't think the situation is really different, except for its complexity, to what one has in other branches of science. We say a body has a "mass", which we suppose to be in principle a precise number in whatever units we are using, even though experiments to measure this mass give imprecise and somewhat variable results. Similarly, we say that an event has a "probability", despite the lack of any direct method of measuring it. In both cases, the assumption that there really is a quantity called "mass" or "probability" is suggested to us by intuition — possibly supported by experience — and is justified by later deductions. To be a little more explicit, we construct a mathematical model of the phenomena that interest us, involving relations between various quantities (for instance $force = mass \times acceleration$); experimental observations then tell us whether these relations are plausible and what values should be assigned to the quantities that are not directly observable, let us say the mass of the proton. Our intention is that the "laws of nature" we propose should be exact relations between exact quantities; observed discrepancies from the laws should be ascribed to "experimental error" unless there is evidence that it is an inadequate or untenable explanation. The greater unease we often feel about probability as a quantity is perhaps due to the rather abstract entities ("events" consisting of "outcomes" of "experiments") to which it is attached, which means that it has no immediate appeal to our senses. By contrast, we experience mass or electric charge, say, rather directly, and are accustomed to regard the phenomena of ordinary life as deterministic. (One would expect even the spin of a roulette wheel to be completely determined by the conditions; there is an intellectual problem in assuming that its results are genuinely "random".) When we come to theories that do not impinge on everyday perception, such as the structure of the atom, probabilistic interpretations are perhaps less offensive to our imagination, partly because we have fewer preconceptions about the laws which should operate; and they do seem to be validated by the predictions they lead to.

The real philosophical problem, although it is hard to pose it exactly and it could be described as "psychological", seems to me to be with our intuition. Where do we get the idea that events *have* probabilities, or that successive throws of the die are "independent"? No-one who considers the questions doubts the answers, whatever bizarre misconceptions may also be common. The same applies to many other mathematical (and not only mathematical) concepts. But this is a digression.

The modern approach to probability, therefore, is to set up a model without worrying overmuch about the values of the quantities that appear. Putting it very crudely, it is the domain of statistics to discuss the values that should be inserted. The reason for taking a model based on measure theory is that *you want probabilities to be countably additive* (and in general to behave well under "countable" operations). Without this, one could not discuss the probabilities associated with a sequence of experiments.

Definition 16.1. A *probability space* is a triple $(\Omega, \mathcal{F}, P)$, where Ω is a set (the *event space*), $\mathcal{F}$ is a σ-field in Ω, and P is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$. In general, a measure μ on a measurable space (Ω, Σ) is called a *probability measure*, or a *probability*, if $\mu(\Omega) = 1$.

My impression is that probabilists prefer to call their σ-algebras $\mathcal{F}$, although analysts tend to use Σ. Where analysts say "almost everywhere", or "for almost all $x \in \Omega$", probabilists say "almost surely" (a.s.), or "almost always".

One should think of Ω as the set of possible outcomes of an experiment. An idealized example might be the following. A gun at the centre O of a sphere S shoots bullets in random directions; then Ω might be the set of all points of S, an "outcome" being the point you hit. $\mathcal{F}$ would be the class of subsets of S to which a probability might be assigned, which we call "events". We expect, this being roughly what we mean by "random", that the probability of hitting a point in a set A — of the "event" A — will be proportional to the solid angle A subtends at O; more precisely, it ought to be $\lambda(A)/\lambda(S)$, λ being Lebesgue measure (I slide discreetly over the question of defining "Lebesgue measure" on a sphere; it is possible, of course.) There are three aspects to this.

In the first place, we *assume* or *postulate* the values of the probability. They are *not* based on experiment. A probability is in most cases a very complicated entity, which could not be found even approximately by purely experimental results without assumptions on its general character. Our postulated values may need subsequent modification if experiments suggest they are wrong; that is the province of so-called Bayesian statistics.

When, as so often, people say that some observations have only a 1 in 10,000 probability of having arisen by pure chance, the "pure chance" refers, at root, to a probability they have themselves defined. I am not suggesting dishonesty or stupidity here, but only that, very often, entirely credible assumptions are involved that cannot be seriously or even superficially tested. An example is when expert witnesses give odds on DNA matches. It is obvious that these odds are *not* based on convincing statistical sampling — the odds quoted are usually such that no large enough sample could ever have been tested —, but on *a priori* assumptions about the random behaviour of DNA sequences. A rather similar but older example is the crude assertion that "no two people have the same fingerprints", which I heard in primary school; it must have meant that the standard characteristics used to analyze fingerprints are in principle sufficient to distinguish any two people in the world, and was based on the unspoken assumptions that these characteristics vary independently and so on. It is inconceivable that the assertion has ever been checked by any scrupulous survey of fingerprints. (A final "proof" would require fingerprinting everybody without exception, and analyzing all the fingerprints.)

To put much the same idea another way, what constitutes "random behaviour" depends on your point of view and the information you have available. If the gun in my example were known to be able only to fire in a plane, we should want to take the probability in this case to be proportional to Lebesgue measure on the *circle* of possible hits; "random" behaviour would be random relative to the added information. (Conversely, if we found in practice that it seemed over many firings to hit points near to a fixed plane more often than others, we

should be inclined to suspect some asymmetry in the mounting and to modify our assumptions about the probability accordingly; this is the Bayesian idea.)

Secondly, the event space Ω is not, in real life, as obvious as all that. Each firing of the gun is associated with many phenomena, not just with a bullet-hole in the sphere, and so it might be more realistic to consider an Ω far "larger" than S. We might, for example, have reason to expect that the bullets' trajectories will for some reason be very irregular. In that case we might want to take Ω to be the whole "space of possible trajectories" starting at the gun and ending on the sphere. Then the position of the hits, which was all we looked at before, would give the function $\Omega \longrightarrow S$ that assigns to each trajectory its end-point on the sphere. Similarly, there might be other "random" phenomena (the mass of the bullets or their initial speed or whatever) that we want to take into account, and the space Ω may be given more structure, more dimensions as it were, to allow for that.

This being so, probabilists tend to avoid explicit mention of Ω. Where an analyst might write $\mu(\{x \in \Omega : f(x) > 3\})$, for instance, a probabilist might write instead $P(f > 3)$. In general, probabilists use rather abbreviated notation, because of the intuitive approach that was typical of the subject. It is usually easy enough to see what is meant, and I may on occasion be similarly casual; but it is important to appreciate that an argument in which an analyst might mention Ω on every line may be phrased by a probabilist in such a way that Ω is neither named nor mentioned, although the mathematical substance is identical.

The third remark is the oddest. We have seen that, on the Axiom of Choice, there must be subsets of the sphere (actually, we argued for the interval $(0, 1]$, and some slight modifications are needed for the sphere) that are *not* Lebesgue-measurable. So there are possible events — that the bullet should hit such a non-measurable set — to which no probability can be assigned. This is a little disquieting. The idea of probability arose from gambling, and we might expect that you can lay a bet on anything. The suggestion that some events are unavailable for wagers is at first surprising; but, of course, *the events in question* (the non-measurable sets) *cannot even be described explicitly*, and laying a bet on an event that you cannot specify in any realistic way is unimaginable; you would never know whether you have won. In short, our intuitive feeling that all events have probabilities should not be interpreted too generously — it is only "specifiable" events that it can apply to, and other events may *exist* by logical necessity.

Several other probabilistic notations may be mentioned here.

A measurable function is called a *random variable*, and commonly denoted by upper-case italics like $Z, Y, X, \dots$, without mention of Ω or Σ. The integral is called the *expectation* or *expected value* (on the "frequency" interpretation of probability, it would be the average value of the random variable over an infinite sequence of repetitions of the experiment, if such a thing were possible):

$$EX = E(X) \coloneqq \int_\Omega X(\omega)\, dP(\omega)\,.$$

In cases where P, or Σ or Ω, is in doubt, they may be incorporated in the notation: $E(X; P)$ or $E(X; \Omega)$, and so on.

The standard joke is that probability theory is measure theory plus the notion of independence. It should not be taken too seriously, but there is no doubt that a theorem in probability theory in which the idea of independence has no part at all is likely to be a theorem (though possibly an uninteresting one) in pure analysis.

Definition 16.2. Let (Ω, Σ, P) be a probability space. The events $A, B \in \Sigma$ are *independent* if $P(A \cap B) = P(A)P(B)$.

Once again, the mathematical theory of probability is not interested in the practical question when the "actual probabilities" of events in the real world make them "independent"; that, in two successive throws of a die, the outcome of the first and the outcome of the second exactly satisfy the law just stated is, in the first instance, a *belief* rather than an experimental observation. To be specific, in this case

$$\Omega := \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} = M \times M\,, \text{ say,}$$

and we have an event $A := K \times M$ which consists all of the outcomes for which the first throw is in $K \subseteq M$, and an event $B := M \times L$ of the outcomes for which the second throw is in L. Then we *expect* $P(A \cap B) = P(K \times L) = P(K \times M)P(M \times L) = P(A)P(B)$ by our intuitive feeling that the first and second throws are "independent" in some real-world sense. Of course, our belief would have to be abandoned if it appeared to be substantially contrary to experience, which it isn't; and, equally, the theoretical concept is introduced to model similar situations to this one. But, from the theoretical point of view, P is given, and independence is defined relative to P. The physical circumstances in which we suppose the concept of independence to be instantiated are irrelevant to the theory.

Definition 16.2 has various extensions, which I omit here.

§17. Types of convergence.

To simplify the statements of this section, let us establish the conventions that (Ω, Σ, μ) is a fixed measure space; that $n \in \mathbb{N}$; and that the functions f, g, h, f_n, g_n, h_n are Σ-measurable functions $\Omega \longrightarrow \mathbb{R}$. α, β, γ will be real numbers. Several of the definitions and results do not require these conventions (for instance, the definitions 17.1 do not require the functions to be measurable or the existence of a measure). The restriction to finite-valued functions is not in practice very significant — see the remark after 15.7; but it avoids irrelevancies.

Definition 17.1. *(a)* $f_n \to f$ *pointwise on* Ω (or just *pointwise*, when there is no ambiguity) means that, for each $x \in \Omega$, the numerical sequence $(f_n(x))$ converges to $f(x)$:

$$(\forall x \in \Omega)(\forall \epsilon > 0)(\exists N) \quad n \geq N \Longrightarrow |f_n(x) - f(x)| < \epsilon\,,$$

where N is specific to the particular x and ϵ under consideration.

We have already used this idea many times, writing just $f_n \to f$ or $f_n \uparrow f$.

(b) (f_n) *is pointwise Cauchy on* Ω means that, for each $x \in \Omega$, the numerical sequence $(f_n(x))$ is Cauchy:

$$(\forall x \in \Omega)(\forall \epsilon > 0)(\exists N) \quad n, m \geq N \Longrightarrow |f_n(x) - f_m(x)| < \epsilon\,.$$

Lemma 17.2. *If* $f_n \to f$ *pointwise, then* (f_n) *is pointwise Cauchy. If* (f_n) *is pointwise Cauchy, then there exists* f *such that* $f_n \to f$ *pointwise.* □

Definition 17.3. *(a)* $f_n \to f$ *uniformly on* Ω means that

$$(\forall \epsilon > 0)(\exists N)(\forall x \in \Omega) \quad n \geq N \Longrightarrow |f_n(x) - f(x)| < \epsilon\,. \tag{38}$$

In other words, N is now no longer specific to a single x; for the given ϵ, N has to 'work' for all x. (38) implies that

$$(\forall \epsilon > 0)(\exists N) \quad n \geq N \Longrightarrow \sup\{|f_n(x) - f(x)| : x \in \Omega\} \leq \epsilon\,, \tag{39}$$

which could be taken as an alternative definition. A still more abstract version is

$$\sup\{|f_n(x) - f(x)| : x \in \Omega\} \to 0 \quad \text{as} \quad n \to \infty\,. \tag{40}$$

The equivalence of (38), (39), and (40) is a trivial exercise, but notice that the passage from (38) to (39) involves a change, for the chosen ϵ, from $<$ to $\leq$; to return from (39) to (38) it is necessary to consider different values for ϵ.

(b) (f_n) *is uniformly Cauchy on* Ω means that

$$(\forall \epsilon > 0)(\exists N)(\forall x \in \Omega) \quad n, m \geq N \Longrightarrow |f_n(x) - f_m(x)| < \epsilon\,. \tag{41}$$

This condition also may be formulated in other ways, which I leave to you.

Lemma 17.4. *If* $f_n \to f$ *uniformly, then* (f_n) *is uniformly Cauchy. If* (f_n) *is uniformly Cauchy, then there exists* f *such that* $f_n \to f$ *uniformly.* □

In this lemma, as in the previous one and others to come, the assertion is of a kind of "completeness"; it is not always the case that there is a metric involved, so Definition 0.14 may not apply in the form I gave, but, nevertheless, the main statement is that a Cauchy sequence must necessarily converge (the converse is usually trifling). The method is always the same. To find the putative limit of the sequence, you consider some other kind of convergence — in spaces of functions often pointwise convergence. Having found a candidate for the limit in this weaker sense, you must check, firstly, that it, too, belongs to the space under consideration, and then that it is the limit in the sense desired, not just in the sense used to construct it. (My usual joke here is that it is like the procedure for electing an American president. The parties look, by any means available, for suitable *candidates*. Having found one, they must check that he or she supports the right *party*. There have been instances in living memory where both parties wanted to field the same candidate, having no idea which he favoured. And finally, he must be elected.) One or more of these steps may be redundant. In the lemma above, "uniformly Cauchy" implies "pointwise Cauchy", which implies "pointwise convergent" by 17.2; the 'hard' part, not very hard in this case, is to show that convergence pointwise to the pointwise limit is also uniform convergence. Indeed, given $\epsilon > 0$, there exists N such that $m, n \geq N \Longrightarrow (\forall x \in \Omega)\ |f_n(x) - f_m(x)| < \epsilon$; as this is so for any $m \geq N$ (when $n \geq N$ is kept fixed), one has in the limit $|f_n(x) - f(x)| \leq \epsilon$. This is so for any $n \geq N$ and any $x \in \Omega$, so $f_n \to f$ uniformly on Ω.

So far Ω might have been any set.

Definition 17.5. $f_n \to f$ *a.e. on* Ω if the set $\{x \in \Omega : f_n(x) \not\to f(x)\}$, which must belong to Σ, is of μ-measure zero.

(If I had not demanded at the start that f_n, f were measurable, I could still have defined "$f_n \to f$ a.e." to mean that there is a measurable set A of measure 0 such that, for any $x \notin A$, $f_n(x) \to f(x)$. See 14.8. That, on our assumptions, the "exceptional set" is in Σ, follows from the equality

$$\{x \in \Omega : f_n(x) \not\to f(x)\} = \bigcup_{m=1}^{\infty} \left(\bigcap_{N=1}^{\infty} \left(\bigcup_{n \geq N}^{\infty} \{x \in \Omega : |f_n(x) - f(x)| \geq m^{-1}\} \right) \right),$$

which it is an amusing exercise to prove.)

Definition 17.6. (f_n) *is a.e. Cauchy on* Ω means that $\{x \in \Omega : (f_n(x))$ is not Cauchy$\}$, which must belong to Σ, is of measure 0.

A similar remark applies to this definition. Notice that

$$\{x \in \Omega : (f_n(x)) \text{ not Cauchy}\} = \bigcup_{m=1}^{\infty} \left(\bigcap_{N=1}^{\infty} \left(\bigcup_{k,l \geq N}^{\infty} \{x : |f_k(x) - f_l(x)| \geq m^{-1}\} \right) \right).$$

Lemma 17.7. *If* (f_n) *converges a.e. to* f*, then* (f_n) *is a.e. Cauchy. If* (f_n) *is a.e. Cauchy on* Ω*, then there is some* f *such that* $f_n \to f$ *a.e.*

Proof. Remove the exceptional sets and apply 17.2. □

Definition 17.8. $f_n \to f$ *a.e. uniformly* (or *uniformly a.e.*) if there exists a set $Z \in \Sigma$ such that $\mu(Z) = 0$ and $f_n \to f$ uniformly on $\Omega \setminus Z$.

Definition 17.9. (f_n) is *a.e. uniformly Cauchy on* Ω, or *uniformly Cauchy a.e.*, if there exists $Y \in \Sigma$ such that $\mu(Y) = 0$ and (f_n) is uniformly Cauchy on $\Omega \setminus Y$.

Lemma 17.10. *If (f_n) is a.e. uniformly Cauchy, then there exists f such that $f_n \to f$ uniformly a.e.* □

Although the definitions 17.8 and 17.9 are natural ones to make, there is something odd about them. It may be seen from the alternative formulation of 17.8:

$$(\forall m \in \mathbb{N})(\exists N \in \mathbb{N}) \quad \mu\Big(\bigcup\nolimits_{n\geq N} \{x \in \Omega : |f_n(x) - f(x)| \geq m^{-1}\}\Big) = 0\,.$$

It is rather strange to demand that, for some N that need not be chosen independently of m, the measure of the set in question be *exactly* 0. (To deduce 17.8, take the union of these sets over all m.) Indeed, all the definitions so far given rely only on the structure of the sets of measure zero — the non-zero values of μ have no influence on them.

This remark brings us to the first kind of convergence that is strikingly new and really exploits the measure. Although I state it for measure spaces, its simplest interpretation is in terms of probabilities.

Definition 17.11. $f_n \to f$ *in measure on* Ω (or, when μ is a probability, $f_n \to f$ *in probability*) if, for each $\epsilon > 0$, $\mu(\{x \in \Omega : |f_n(x) - f(x)| \geq \epsilon\}) \to 0$ as $n \to \infty$. That is,

$$(\forall \epsilon > 0)(\forall \eta > 0)(\exists N \in \mathbb{N})(\forall n \geq N) \quad \mu(\{x \in \Omega : |f_n(x) - f(x)| \geq \epsilon\}) < \eta\,. \quad (42)$$

The general idea behind this definition is, obviously enough, that the *probability* that f_n and f will differ by more than ϵ becomes very small for large n; in probabilist-speak,

$$P(|f_n - f| \geq \epsilon) \to 0\,.$$

However, the definition is quite different from the previous ones, since there is no fixed exceptional set off which the convergence occurs; rather, there is an exceptional set that depends on ϵ and n, and μ is used to measure the "size" of this set, so that, at least in principle, the whole structure of μ may be called on. (It turns out that this is not entirely true in practice, at least in many useful cases.)

The statement (42) is often cast in a different form:

$$(\forall \epsilon > 0)(\exists N \in \mathbb{N})(\forall n \geq N) \quad \mu(\{x \in \Omega : |f_n(x) - f(x)| \geq \epsilon\}) < \epsilon\,.$$

It is easy and instructive to show that this implies (42); the converse implication is trivial.

Example 17.12. Let Ω be $\mathbb{R}$ or $[0,1]$, with Lebesgue measure λ. Define a sequence (E_n) of sets in Ω as follows. $E_1 := [0,1]$, $E_2 := [0, \frac{1}{2})$, $E_3 := [\frac{1}{2}, 1]$, $E_4 := [0, \frac{1}{3})$, $E_5 := [\frac{1}{3}, \frac{2}{3})$, $E_6 := [\frac{2}{3}, 1]$, $E_7 := [0, \frac{1}{4})$, ... The rule is that, if $\frac{1}{2}r(r+1) < n < \frac{1}{2}(r+1)(r+2)$ and $i := n - \frac{1}{2}r(r+1)$, $E_n := [\frac{i-1}{r+1}, \frac{i}{r+1})$, whilst $E_n := [\frac{r}{r+1}, 1]$ when $n = \frac{1}{2}(r+1)(r+2)$.

Take $f_n := 1_{E_n}$. Then, if $\epsilon > 1$, $\{x \in \Omega : |f_n(x)| \geq \epsilon\} = \emptyset$, whilst, if $0 < \epsilon \leq 1$,

$$\{x \in \Omega : |f_n(x)| \geq \epsilon\} = E_n\,.$$

However, $\lambda(E_n) = 1/(r+1)$, where r is the largest non-negative integer such that $r(r+1) < 2n$. It follows that $\lambda(E_n) \to 0$ as $n \to \infty$, and, therefore, that $f_n \to 0$ in measure (or in probability).

On the other hand, if $x \in [0,1]$, $x \in E_n$ for infinitely many indices n (exactly once in the range $\frac{1}{2}r(r+1) < n \le \frac{1}{2}(r+1)(r+2)$ for each r). So the numerical sequence $(f_n(x))$ consists of 0s and 1s, and there are infinitely many of both; the relative frequency of 1s diminishes as n increases, but they never die out. Thus, for any fixed $x \in [0,1]$, the numerical sequence $(f_n(x))$ does not converge. One might say that the "exceptional set" wanders across the whole of $[0,1]$ again and again, and that it is only "on the whole" (that is, in probability!) that f_n tends to 0. This example, and similar examples, should be kept in mind as the theory is developed.

For the previous kinds of convergence (pointwise, uniform, a.e. pointwise, a.e. uniform) we had a corresponding Cauchy condition and "completeness theorem", and it was obvious that the limit was "linear": that is, if $f_n \to f$ and $g_n \to g$, then $\alpha f_n + \beta g_n \to \alpha f + \beta g$ for any $\alpha, \beta \in \mathbb{R}$. None of these ideas is so straightforward for convergence in measure. However, the clauses that $\cdots (\exists N \in \mathbb{N})(\forall n \ge N) \cdots$ ensure the following Lemma is true.

Lemma 17.13. *Let $f_n \to f$ in any of the senses listed above (pointwise, uniform, a.e. pointwise, a.e. uniform, in measure). Then any subsequence of (f_n) also converges to f in the same sense.* □

Definition 17.14. The sequence (f_n) is *Cauchy in measure* (or *in probability*) if, for each $\epsilon > 0$, $\mu(\{x \in \Omega : |f_n(x) - f_m(x)| \ge \epsilon\}) \to 0$ as $m, n \to \infty$; that is,

$$(\forall \epsilon > 0)(\forall \eta > 0)(\exists N \in \mathbb{N})(\forall m, n \ge N) \quad \mu(\{x \in \Omega : |f_n(x) - f_m(x)| \ge \epsilon\}) < \eta \,. \tag{43}$$

As before, (43) is equivalent to the somewhat simpler condition

$$(\forall \epsilon > 0)(\exists N \in \mathbb{N})(\forall m, n \ge N) \quad \mu(\{x \in \Omega : |f_n(x) - f_m(x)| \ge \epsilon\}) < \epsilon \,. \tag{44}$$

Lemma 17.15. *If $f_n \to f$ in measure, then (f_n) is Cauchy in measure.*

Proof. Given $\epsilon, \eta > 0$, take $N \in \mathbb{N}$ such that

$$(\forall n \ge N) \quad \mu(\{x \in \Omega : |f_n(x) - f(x)| \ge \tfrac{1}{2}\epsilon\} < \tfrac{1}{2}\eta \,.$$

Then, if $m, n \ge N$,

$$\{x : |f_n(x) - f_m(x)| \ge \epsilon\} \subseteq \{x : |f_n(x) - f(x)| \ge \tfrac{1}{2}\epsilon\} \cup \{x : |f_m(x) - f(x)| \ge \tfrac{1}{2}\epsilon\} \,.$$

To prove this, suppose x is not in the right-hand side; then $|f_n(x) - f(x)| < \frac{1}{2}\epsilon$ and $|f_m(x) - f(x)| < \frac{1}{2}\epsilon$, so that $|f_n(x) - f_m(x)| < \epsilon$ and x is not in the left-hand side. Now

$$\begin{aligned} &\mu(\{x : |f_n(x) - f_m(x)| \ge \epsilon\}) \\ &\qquad \le \mu(\{x : |f_n(x) - f(x)| \ge \tfrac{1}{2}\epsilon\}) + \mu(\{x : |f_m(x) - f(x)| \ge \tfrac{1}{2}\epsilon\}) \\ &\qquad < \tfrac{1}{2}\eta + \tfrac{1}{2}\eta = \eta \,. \end{aligned}$$

□

Lemma 17.16. *Suppose $f_n \to f$ in measure and $g_n \to g$ in measure, and $\alpha, \beta \in \mathbb{R}$. Then $\alpha f_n + \beta g_n \to \alpha f + \beta g$ in measure.* □

It is clear that if $f_n \to f$ and $g_n \to g$ pointwise or pointwise a.e., then $f_n g_n \to fg$ pointwise or pointwise a.e. The analogous statements for uniform or a.e. uniform convergence are not true without additional hypotheses to do with the boundedness of the functions considered. For instance, let $f_n(x) = x$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$, and let $g_n(x) = n^{-1}$ for all x. Then $f_n g_n$ does *not* converge uniformly to 0. For convergence in measure, the situation is discussed later, at.

I shall now introduce an idea which constitutes a sort of amalgam of convergence a.e. and convergence in measure. As far as I know, it was first named by Munroe, although the concept had been used before.

Definition 17.17. $f_n \to f$ *almost uniformly on* Ω if, for any $\eta > 0$, there is a set $E \in \Sigma$ such that $\mu(E) < \eta$ and $f_n \to f$ uniformly on $\Omega \setminus E$.

It is important to grasp that the *rate* of uniform convergence on $\Omega \setminus E$ will (in principle) depend on E. By this I mean that, for a given ϵ, different Es may require different Ns to validate (39). A very simple and familiar example, which should be kept in mind as the theory proceeds, is when $\Omega := [0,1)$ with Lebesgue measure and $f_n(t) = t^n$. Then $f_n \to 0$ pointwise on Ω, and, for any $\eta \in (0,1)$, $f_n \to 0$ uniformly on $[0, 1-\eta]$. However, N in (39) must be taken to be greater than $\log \epsilon / \log(1-\eta)$, which (if $\epsilon \in (0,1)$) may be made as large as you wish by taking η sufficiently small.

Definition 17.18. (f_n) is *almost uniformly Cauchy on* Ω if, for any $\eta > 0$, there is a set $E \in \Sigma$ such that $\mu(E) < \eta$ and (f_n) is uniformly Cauchy on $\Omega \setminus E$.

Lemma 17.19. (a) *Suppose that $f_n \to f$ almost uniformly. Then $f_n \to f$ a.e. and $f_n \to f$ in measure.*

(b) *If (f_n) is almost uniformly Cauchy, then (f_n) is a.e. Cauchy and Cauchy in measure.*

Proof. *(a)* Given $m \in \mathbb{N}$, let $E_m \in \Sigma$ be such that $\mu(E_m) < 1/m$ and $f_n \to f$ uniformly on $\Omega \setminus E_m$. Then $f_n \to f$ pointwise on $\Omega \setminus E_m$. As $f_n(x) \to f(x)$ for each $x \notin E_m$ for each m, $f_n \to f$ for $x \notin \bigcap_{m=1}^{\infty} E_m$; but $\mu(\bigcap_{m=1}^{\infty} E_m) = 0$. So $f_n \to f$ a.e. on Ω. Similarly, given $\epsilon, \eta > 0$, there exists some $E \in \Sigma$ with $\mu(E) < \eta$ such that $f_n \to f$ uniformly on $\Omega \setminus E$; in particular, there is some $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ whenever $x \notin E$ and $n \geq N$. This proves convergence in measure.

The argument for *(b)* is analogous. □

Lemma 17.20. (f_n) *is almost uniformly Cauchy if and only if there exists f such that $f_n \to f$ almost uniformly.*

Proof. "If" is easy. Suppose that (f_n) is almost uniformly Cauchy. Then, by 17.19*(b)*, it is a.e. Cauchy; by 17.7, it converges a.e. to some function f. Take $\eta > 0$. By hypothesis, there is some $E \in \Sigma$ such that $\mu(E) < \eta$ and (f_n) is uniformly Cauchy on $\Omega \setminus E$. Hence $f_n \to g_E$ uniformly on $\Omega \setminus E$, by 17.4, where the limit function g_E may depend on E. Since $f_n \to g_E$ pointwise, necessarily $g_E = f$ a.e. on $\Omega \setminus E$. Letting Z be the exceptional set, we deduce that $f_n \to f$ uniformly on $\Omega \setminus (E \cup Z)$, and $\mu(E \cup Z) = \mu(E) < \eta$. Thus $f_n \to f$ almost uniformly on Ω. □

Before we begin to discuss the major theorems, it should be noted that the definitions given above are sometimes stated in different forms. A rather striking instance is the following.

Lemma 17.21. *Let* $\mu(\Omega) < \infty$. *Then* $f_n \to f$ *a.e. if and only if either*

(a) $(\forall \epsilon, \eta > 0)(\exists N \in \mathbb{N}) \quad \mu(\{x \in \Omega : (\exists n \geq N) \ |f_n(x) - f(x)| \geq \epsilon\}) < \eta$ *or*
(b) $g_n \to 0$ *in measure, where* $g_n := \sup\{|f_m - f| : m \geq n\}$.

Proof. Let $f_n \to f$ a.e., $\epsilon, \eta > 0$. Set $E_N^\epsilon := \{x \in \Omega : (\forall n \geq N) \ |f_n(x) - f(x)| < \epsilon\}$. Evidently $E_N^\epsilon \uparrow$ with N, and

$$\lim_N (\Omega \setminus E_N^\epsilon) = \Omega \setminus \left(\bigcup_{N=1}^\infty E_N^\epsilon\right) = \{x : |f_n(x) - f(x)| \geq \epsilon \text{ i.o.}\}$$

(where "i.o." denotes "infinitely often", i.e. for infinitely many values of n) must be of measure 0 as it is a set of points at which (f_n) does not tend to f. However, since $\mu(\Omega) < \infty$, 8.5*(b)* implies that $\mu(\Omega \setminus E_N^\epsilon) \downarrow 0$. Consequently, given ϵ and η, there must exist some N such that $\mu(\Omega \setminus E_N^\epsilon) < \eta$. This is the assertion of *(a)*.

For the reverse implication, it is unnecessary to assume $\mu(\Omega) < \infty$. Indeed, define

$$E^\epsilon := \bigcup_{N=1}^\infty E_N^\epsilon,$$

and then $\mu(\Omega \setminus E^\epsilon) \leq \mu(\Omega \setminus E_N^\epsilon)$ for each N; therefore, by *(a)*, $\mu(\Omega \setminus E^\epsilon) < \eta$. Ergo, $\mu(\Omega \setminus E^\epsilon) = 0$. Now take $Z := \Omega \setminus \left(\bigcap_{m=1}^\infty E^{1/m}\right)$. Certainly

$$\mu(Z) \leq \sum_{m=1}^\infty \mu(\Omega \setminus E^{1/m}) = 0,$$

and if $x \notin Z$ and $m \in \mathbb{N}$, then $x \in E^{1/m}$. Hence, for some N, $|f_n(x) - f(x)| < \epsilon$ whenever $n \geq N$. That is, $f_n \to f$ except on Z.

Now consider *(b)*. The condition that $g_n \to 0$ in measure is

$$(\forall \epsilon, \eta > 0)(\exists N \in \mathbb{N}) \quad n \geq N \Longrightarrow \mu(\{x \in \Omega : |g_n(x)| \geq \epsilon\}) < \eta. \tag{45}$$

As $g_n \downarrow$, the conclusion says merely that $\mu(\{x : g_N(x) \geq \epsilon\}) < \eta$. On the other hand,

$$\{x : (\exists n \geq N) \ |f_n(x) - f(x)| \geq \epsilon\} \subseteq \{x : g_N(x) \geq \epsilon\},$$

so that *(a)* holds if $g_n \to 0$ in measure. Conversely, for $n \geq N$,

$$\{x : |g_n(x)| \geq \epsilon\} \subseteq \{x : g_N(x) \geq \epsilon\} \subseteq \{x : (\exists n \geq N) \ |f_n(x) - f(x)| \geq \tfrac{1}{2}\epsilon\},$$

so that, if *(a)* holds, and N is chosen to correspond to $\frac{1}{2}\epsilon$ and η, (45) follows. This proves that *(a)* implies *(b)*. □

Corollary 17.22. *A monotone sequence of functions on a space of finite measure converges in measure to a given limit if and only if it converges a.e. to the same limit.* □

The next theorem can be used as a step in the proofs of many other results, but in my exposition it seems to have few consequences.

Theorem 17.23. (Egorov[4]'s theorem). *Suppose that* $\mu(\Omega) < \infty$ *and* $f_n \to f$ *a.e. Then* $f_n \to f$ *almost uniformly.*

Proof. Let $E_{nk} := \{x \in \Omega : (\exists l \geq n)\ |f_l(x) - f(x)| > 1/k\}$ for $n, k \in \mathbb{N}$. This set is measurable for each choice of n and k. For fixed k, E_{nk} decreases as n increases, and, if $f_n(x) \to f(x)$, $x \notin \bigcap_{n=1}^{\infty} E_{nk}$; that means $\mu(\bigcap_{n=1}^{\infty} E_{nk}) = 0$ for each k, as $f_n \to f$ a.e. By 8.5*(b)*, $\mu(E_{nk}) \underset{n}{\downarrow} 0$. Given $\epsilon > 0$, there exists $n(k, \epsilon)$ such that

$$\mu(E_{n(k,\epsilon),k}) < 2^{-k}\epsilon\,.$$

Set $A := \bigcup_{k=1}^{\infty} E_{n(k,\epsilon),k}$. Then

$$\mu(A) = \mu\Big(\bigcup\nolimits_{k=1}^{\infty} E_{n(k,\epsilon),k}\Big) \leq \sum\nolimits_{k=1}^{\infty} 2^{-k}\epsilon = \epsilon\,.$$

If $x \notin A$, then, for any given $k \in \mathbb{N}$, $x \notin E_{n(k,\epsilon),k}$, so that, by the definition of $E_{n(k,\epsilon),k}$,

$$(\forall l \geq n(k,\epsilon)) \quad |f_l(x) - f(x)| \leq 1/k\,.$$

As a consequence, $f_n \to f$ uniformly on $\Omega \setminus A$. □

The "rate of uniform convergence", represented by the sequence $(n(k,\epsilon))_{k=1}^{\infty}$, depends on ϵ (cf. the remark after 17.17). There is a similar result that a sequence Cauchy a.e. on a space of finite measure is almost uniformly Cauchy, but it is unnecessary to prove it separately, in view of 17.7, 17.23, and 17.20. The moral is that convergence a.e. on a space of finite measure is *equivalent* to almost uniform convergence, which at first glance is more demanding (and also implies convergence in measure). For a probability space, almost sure convergence of random variables implies convergence in probability.

The question arises whether Egorov's theorem can be extended to spaces of infinite measure. The example of $\Omega := \mathbb{R}$, with Lebesgue measure, and $f_n := \mathbf{1}_{[n,n+1)}$, which tends everywhere to 0 but is not almost uniformly convergent, shows that the theorem definitely fails unless extra restrictions are imposed. One possibility is the following.

Theorem 17.24. *Let* $\phi : [0,\infty) \longrightarrow [0,\infty)$ *be an increasing function such that* $\phi^{-1}(\{0\}) = \{0\}$. *Suppose that* $f_n \to f$ *a.e. on* Ω, *and that there exists an integrable function* g *on* Ω *such that, for all* $x \in \Omega$ *and all* n, $\phi(|f(x)|) \leq g(x) \geq \phi(|f_n(x)|)$. *Then* $f_n \to f$ *almost uniformly.*

Proof. Define $E_{nk} := \{x \in \Omega : (\exists l \geq n)\ |f_l(x) - f(x)| > 1/k\}$ for $n, k \in \mathbb{N}$, exactly as in 17.23. Then $E_{nk} \underset{n}{\downarrow}$, and, as $f_n \to f$ a.e., $\mu(\bigcap_{n=1}^{\infty} E_{nk}) = 0$ for any k. Now

$$\begin{aligned} E_{nk} &\subseteq \{x \in \Omega : |f(x)| > 1/(2k)\} \cup \Big(\bigcup\nolimits_{l \geq n} \{x \in \Omega : |f_l(x)| > 1/(2k)\}\Big) \\ &\subseteq F_k := \{x \in \Omega : g(x) \geq \phi(1/(2k))\}\ , \end{aligned}$$

and, as $\phi(1/(2k)) > 0$ and g is integrable, it follows that $\mu(F_k) < \infty$. By 8.5*(b)*, there exists $n(k,\epsilon)$ such that $\mu(E_{n(k,\epsilon),k}) < 2^{-k}\epsilon$, and the argument proceeds as before. □

[4] The name is transliterated from Cyrillics, so, as usual, appears in several forms. You may see Egoroff, Jegorow, Yegorov, and presumably mixtures of all three.

This result implies Egorov's theorem; take $\phi(0) := 0$, $\phi(\xi) := 1$ for all $\xi > 0$. It may also be used to prove the dominated convergence theorem (and some related results to be mentioned later).

Theorem 17.25. *Suppose (f_n) is Cauchy in measure. There exists a subsequence $(f_{n(k)})$ which is almost uniformly Cauchy.*

Proof. Take $n(0) := 0$ by convention. Now, if $k \geq 1$, suppose $n(k-1)$ has been chosen. By hypothesis (see (44)), there exists some $N \in \mathbb{N}$ such that

$$(\forall m, n \geq N) \quad \mu(\{x \in \Omega : |f_m(x) - f_n(x)| \geq 2^{-k-1}\}) < 2^{-k-1}.$$

Choose $n(k)$ be the least such natural number which exceeds $n(k-1)$. Thus $k \leq n(k)$ for all k, so that $(f_{n(k)})$ is an infinite subsequence of (f_n).

Given $\epsilon > 0$, take p so that $2^{-p} < \epsilon$, and let

$$E := \bigcup_{k=p}^{\infty} \{x \in \Omega : \left|f_{n(k)}(x) - f_{n(k+1)}(x)\right| \geq 2^{-k-1}\}.$$

Then $\mu(E) \leq \sum_{k=p}^{\infty} 2^{-k-1} = 2^{-p} < \epsilon$, and, if $x \notin E$ and $j > i \geq q \geq p$,

$$\left|f_{n(i)}(x) - f_{n(j)}(x)\right| \leq \sum_{k=i}^{j-1} \left|f_{n(k)}(x) - f_{n(k+1)}(x)\right| < \sum_{k=i}^{j-1} 2^{-k-1} < 2^{-q}.$$

This shows that $(f_{n(k)})$ is uniformly Cauchy on $\Omega \setminus E$, and so is almost uniformly Cauchy on the whole space Ω. □

We can now fill the obvious gap in the story so far.

Proposition 17.26. *If (f_n) is Cauchy in measure, then there exists f such that $f_n \to f$ in measure.*

Proof. By 17.25, there is a subsequence $(f_{n(k)})$ which is almost uniformly Cauchy. By 17.20, there is f such that $f_{n(k)} \to f$ almost uniformly. By 17.19*(a)*, $f_{n(k)} \to f$ in measure. This is sufficient to ensure that $f_n \to f$ in measure. Indeed, given $\epsilon > 0$,

$$\begin{aligned}
&(\exists N_1 \in \mathbb{N})(\forall m, n \geq N_1) \quad \mu(\{x \in \Omega : |f_m(x) - f_n(x)| \geq \tfrac{1}{2}\epsilon\}) < \tfrac{1}{2}\epsilon \qquad \text{and} \\
&(\exists N_2 \in \mathbb{N})(\forall k \in \mathbb{N}) \quad n(k) \geq N_2 \Longrightarrow \mu(\{x \in \Omega : \left|f_{n(k)}(x) - f(x)\right| \geq \tfrac{1}{2}\epsilon\}) < \tfrac{1}{2}\epsilon.
\end{aligned}$$

Take $N := \max(N_1, N_2)$. Then, if $n \geq N$, as in the proof of 17.15

$$\begin{aligned}
\mu(\{x \in \Omega : |f_n(x) - f(x)| \geq \epsilon\}) &\leq \mu(\{x \in \Omega : \left|f_{n(N)}(x) - f_n(x)\right| \geq \tfrac{1}{2}\epsilon\}) \\
&\quad + \mu(\{x \in \Omega : \left|f_{n(N)}(x) - f(x)\right| \geq \tfrac{1}{2}\epsilon\}) \\
&\leq \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon.
\end{aligned}$$

□

Behind the above result lies the idea formulated after 7.13, that a Cauchy sequence must converge if it has a convergent subsequence. I cannot state this as a theorem, because I have not given a general definition of a Cauchy sequence except in a metric space.

I insert here a result which is often quoted; its proof is essentially the same as in 17.25.

Proposition 17.27. *Let (Ω, Σ, μ) be a measure space. If the sequence (f_n) converges in measure to f on Ω, it has a subsequence which converges a.e. to f.* □

The restriction $\mu(\Omega) < \infty$ is not needed. But here is a probabilist's proof:

Proposition 17.28. *Let (Ω, Σ, P) be a probability space. Let a sequence (X_n) of random variables converge in probability to X. Then there is a subsequence $(X_{n(k)})$ which converges to X almost surely.*

Proof. Take any sequence $\epsilon_k \downarrow 0$; choose $n(k)$ such that $P(\left|X_{n(k)} - X\right| \geq \epsilon_k) < 2^{-k}$. Thus $\sum_{k=1}^{\infty} P(\left|X_{n(k)} - X\right| \geq \epsilon_k) < 1$, and by the first Borel-Cantelli lemma 8.9

$$P(\left|X_{n(k)} - X\right| \geq \epsilon_k \text{ i.o.}) = 0$$

(recall that "i.o." means "infinitely often"). Thus, in fact, $X_{n(k)} \to X$ almost surely. (For any point not in the exceptional set, there are only finitely many indices k for which $\left|X_{n(k)} - X\right| \geq \epsilon_k$.) □

If you study this proof carefully, you will see that it remains fundamentally the same as 17.25, although 17.25 gets more out of the ideas because it assumes less.

Definition 17.29. Let $\mathcal{N}$ denote the set of measurable functions on Ω such that

$$\mu(\{x \in \Omega : |f(x)| \geq K\}) \to 0 \quad \text{as} \quad K \to \infty\,.$$

If $\mu(\Omega) < \infty$, all measurable functions that are a.e. finite belong to $\mathcal{N}$, by 8.5*(b)*. It is easily checked that $\mathcal{N}$ is a vector space of measurable functions on Ω.

Lemma 17.30. *Suppose that $f_n \to f$ and $g_n \to g$ in measure, where $f, g \in \mathcal{N}$. Then $f_n g_n \to fg$ in measure. A similar statement holds for almost uniform convergence.*

Proof. Suppose first that $f_n \to 0$ and $g_n \to 0$ in measure. Then, for any $\epsilon > 0$,

$$\{x : |f_n(x)g_n(x)| \geq \epsilon\} \subseteq \{x : |f_n(x)| \geq \sqrt{\epsilon}\} \cup \{x : |g_n(x)| \geq \sqrt{\epsilon}\}\text{, so}$$
$$\mu(\{x : |f_n(x)g_n(x)| \geq \epsilon\}) \leq \mu(\{x : |f_n(x)| \geq \sqrt{\epsilon}\}) + \mu(\{x : |f_n(x)| \geq \sqrt{\epsilon}\})\,,$$

and it follows that $f_n g_n \to 0$ in measure.

Suppose now that $f_n \to 0$ in measure. Take $\epsilon > 0$. As $g \in \mathcal{N}$, there exists $q \in \mathbb{N}$ such that $\mu(\{x \in \Omega : |g(x)| \geq q\}) < \frac{1}{2}\epsilon$. Next, there exists $N \in \mathbb{N}$ such that

$$(\forall n \geq N) \quad \mu(\{x \in \Omega : |f_n(x)| \geq \epsilon/q\}) < \tfrac{1}{2}\epsilon\,.$$

Consequently, if $n \geq N$

$$\mu(\{x : |f_n(x)g(x)| \geq \epsilon\}) \leq \mu(\{x : |f_n(x)| \geq \epsilon/q\}) + \mu(\{x : |g(x)| \geq q\}) < \epsilon\,.$$

This shows that $f_n g \to 0$ in measure.

Finally, suppose $f_n \to f$ and $g_n \to g$ in measure, where $f, g \in \mathcal{N}$. Then, by 17.16, $f_n - f \to 0$ and $g_n - g \to 0$ in measure, and so, by the results just proved and 17.16,

$$f_n g_n - fg = (f_n - f)(g_n - g) + (f_n - f)g + f(g_n - g) \to 0$$

and $f_n g_n \to fg$.

The argument for almost uniform convergence is very similar. □

By the remark after 17.29, the Lemma holds if $\mu(\Omega) < \infty$ without the restriction on f and g (provided they are a.e. finite).

Remark 17.31. In this section several notions of convergence or of "Cauchyness" of a sequence of measurable functions have been introduced. All the definitions and results after 17.5 really concern only equivalence classes of functions under equality a.e. (for instance, if $f_n \to f$ in measure and $g_n = f_n$ a.e. for each n and $f = g$ a.e., then $g_n \to g$ in measure). In the first place, then, our assumption at the beginning that all functions were finite-valued could have been relaxed later to finiteness a.e. But, in the second place, one could formulate the theory in terms of equivalence classes; take $\mathcal{M}$ to be the vector space of all measurable a.e. finite functions on Ω, let $\mathcal{Z}$ be the vector subspace of functions that are a.e. zero, and define a.e. convergence, a.e. uniform convergence, almost uniform convergence, and convergence in measure, and the corresponding Cauchy properties, for sequences in the quotient space $\mathcal{M}/\mathcal{Z}$. This will be equally possible for the kinds of convergence introduced later. However, to avoid notational difficulties it is customary in some of these contexts to blur the distinction between functions and their equivalence classes.

It is a curiosity of these definitions that they concern *sequences*. Certain sequences are described as having limits, which satisfy various desirable properties. It does not necessarily follow that there is a topology that determines which sequences converge and to what limits, or that, if there is a topology, that it is unique.

Pointwise convergence (everywhere) can be described by treating $\mathcal{M}$ as a subset (with the subspace topology) of $\mathbb{R}^\Omega$ with the product topology. Uniform convergence (everywhere) is derivable from a metric, $d(f,g) := \sup\{|f(x) - g(x)| : x \in \Omega\}$, provided that one restricts attention to bounded functions. Convergence a.e. and uniform convergence a.e. are convergence in the corresponding quotient topologies. Almost uniform and a.e. convergence coincide in spaces of finite measure (or if the functions are suitably restricted).

The really interesting case from this point of view is convergence in measure. There is a topology of convergence in measure on the space $\mathcal{N}$ of 17.29, in which a base of neighbourhoods at 0 is furnished by the sets

$$\{f \in \mathcal{N} : \mu(\{x \in \Omega : |f(x)| \geq \epsilon\}) < \epsilon\}.$$

This topology is metrizable; when Ω is of finite measure (so that $\mathcal{N} = \mathcal{M}$) one can exhibit simple formulæ for suitable metrics.

Theorem 17.32. *Let $\mu(\Omega) < \infty$. Define, for a.e. finite measurable functions f, g on Ω,*

$$d_1(f,g) := \int \frac{|f(x) - g(x)|}{1 + |f(x) - g(x)|}\, \mu(dx) \quad \textit{and}$$
$$d_2(f,g) := \int \min(f(x), 1)\, \mu(dx).$$

Then d_1, d_2 are metrics on the space $\mathcal{M}/\mathcal{Z}$ of measurable functions modulo a.e. equality such that the sequences that are metrically convergent or metrically Cauchy are precisely the sequences that are convergent or Cauchy in measure.

Proof. This is a straightforward exercise. □

§18. Some inequalities.

We need to know some standard inequalities for real numbers; like §0, these are matters that really come *before* the course, and these notes are intended to give appropriate "revision", in a setting that is as general as is reasonable. The principal inequality that is already familiar, and could in fact be treated as the basic result, is the relation between arithmetic and geometric means:

$$\sqrt{\alpha\beta} \leq \tfrac{1}{2}(\alpha+\beta) \qquad \text{for non-negative } \alpha, \beta\,. \tag{46}$$

Definition 18.1. Let E be a real vector space. A subset A of E is *convex* if

$$(\forall a, b \in A)(\forall \alpha \in [0,1]) \quad \alpha a + (1-\alpha)b \in A\,. \tag{47}$$

If A is such a convex set, and $f : A \longrightarrow \mathbb{R}$, we say that the *function* f is *convex* if

$$(\forall a, b \in A)(\forall \alpha \in [0,1]) \quad f(\alpha a + (1-\alpha)b) \leq \alpha f(a) + (1-\alpha) f(b)\,. \tag{48}$$

A set A is convex if, for any two points a and b of A, A also includes the whole straight-line segment joining a and b. In intuitive geometrical terms this means that A does not have "gaps, holes or slits" in its "interior", or "bays, bights, or re-entrants" on its "boundary" — I shall not attempt to clarify these vague expressions. In the case we shall be dealing with, where $E = \mathbb{R}$, the convex sets are precisely the intervals. It is necessary to assume that the set A is convex for the definition (48) to make sense.

If the inequality in (48) went the other way, we should say f is concave. Obviously f is concave if and only if $-f$ is convex, so concave functions are of little independent interest.

Should one know in advance that f is continuous (where $E = \mathbb{R}$, for instance; more generally, if E is a topological vector space), the condition (48) might be weakened. For a continuous function f, convexity is ensured if

$$(\forall a, b \in A) \quad f\left(\frac{a+b}{2}\right) \leq \tfrac{1}{2}(f(a)+f(b))\,.$$

From this it may be deduced by induction that (48) holds whenever α is of the form $2^{-n}k$, for $0 \leq k \leq 2^n$, and it therefore holds in general by continuity.

Lemma 18.2. *Let J be an interval in $\mathbb{R}$, and $f : J \longrightarrow \mathbb{R}$. Then f is convex if and only if any one of the following equivalent conditions holds:*

(i) *for any* $a, b, c \in J$ *with* $a < b < c$, $\dfrac{f(b)-f(a)}{b-a} \leq \dfrac{f(c)-f(b)}{c-b}$,

(ii) *for any* $a, b, c \in J$ *with* $a < b < c$, $\dfrac{f(b)-f(a)}{b-a} \leq \dfrac{f(c)-f(a)}{c-a}$,

(iii) *for any* $a, b, c \in J$ *with* $a < b < c$, $\dfrac{f(c)-f(a)}{c-a} \leq \dfrac{f(c)-f(b)}{c-b}$,

(iv) *if* $a, b, c, d \in J$ *and* $a < b \leq c < d$, $\dfrac{f(c)-f(a)}{c-a} \leq \dfrac{f(d)-f(b)}{d-b}$.

Proof. If a, b, c are given, write $\alpha = \frac{c-b}{c-a}$, and then $b = \alpha a + (1-\alpha)c$. Convexity of f implies that $f(b) \leq \alpha f(a) + (1-\alpha)f(c)$, or, on rearrangement,

$$\alpha(f(b) - f(a)) \leq (1-\alpha)(f(c) - f(b)),$$

which (after multiplying by $(c-a)/\{(c-b)(b-a)\}$) is the required conclusion for (i). Conversely, if a, α and c are given, set $b = \alpha a + (1-\alpha)c$, and the argument reverses to give the inequality (48). The equivalence of *(i)* with *(ii)* and with *(iii)* is simple algebra; *(iv)* follows (if $b = c$, it is just *(i)*; otherwise it results from combining *(ii)* and *(iii)*). □

Geometrically, this result may be vaguely paraphrased as saying that the chords of the graph of f have increasing slope as one moves to the right. (Since a chord has two ends, some care is needed to explain what is included in "moving to the right").

Lemma 18.3. *Suppose $f : J \longrightarrow \mathbb{R}$ is convex, $a_1, a_2, \ldots, a_n \in J$, and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are nonnegative numbers for which $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$. Then*

$$f(\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n) \leq \alpha_1 f(a_1) + \alpha_2 f(a_2) + \ldots + \alpha_n f(a_n). \qquad (49)$$

Proof. Assume $a_1 \leq a_2 \leq \cdots \leq a_n$. For $n = 1$, it is obvious; when $n = 2$, the result is just (48). Proceed by induction on n. Let $n > 2$. If $\alpha_n = 1$, all other α's vanish, and there is nothing to prove. If $\alpha_n < 1$, let $b_1 = (\alpha_1 a_1 + \alpha_2 a_2 + \ldots \alpha_{n-1} a_{n-1})/(1-\alpha_n) \leq a_{n-1}$ and $b_2 = a_n$. Then

$$\begin{aligned} f(\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n) &= f((1-\alpha_n)b_1 + \alpha_n b_2) \\ &\leq (1-\alpha_n)f(b_1) + \alpha_n f(b_2) \quad \text{by virtue of (48);} \end{aligned} \qquad (50)$$

however, since $\alpha_1(1-\alpha_n)^{-1} + \ldots + \alpha_{n-1}(1-\alpha_n)^{-1} = 1$, one has by the previous step

$$f(b_1) \leq \frac{\alpha_1 f(a_1) + \alpha_2 f(a_n) + \ldots + \alpha_{n-1} f(a_{n-1})}{1-\alpha_n}.$$

Substituting this in (50), we obtain the result. □

This result is sometimes called Jensen's theorem, and should not be confused with the "Jensen's formula" which is important in complex analysis, or with the theorem probabilists perversely call Jensen's inequality (presented below as 19.1).

Lemma 18.4. *A nonnegative multiple of a convex function is convex; the sum of convex functions is convex; linear functions of the form $f(x) = \beta x + \gamma$ are convex.* □

Although I shall not present the proofs (which depend on 18.2), it is worth noting that a convex function must be continuous, differentiable except at countably many points, and twice differentiable almost everywhere. Thus, the arguments which follow are not based on grossly excessive hypotheses; and, whilst they are not the best possible, they suffice in nearly all applications. We start with a weak version of Lagrange's form of Taylor's theorem (you have probably seen a much fuller form of the same argument before, since it can be used to obtain several forms of the remainder after n terms of the Taylor series). Notice that a function $f : [a, b] \longrightarrow \mathbb{R}$ is said to be differentiable on $[a, b]$ if it has a right derivative (written $f'(a)$) at a, a left derivative $f'(b)$ at b, and a two-sided derivative $f'(x)$ at every

point $x \in (a,b)$. It is said to be C^1 on $[a,b]$ if it is differentiable on $[a,b]$ and f', defined as just described, is continuous on $[a,b]$.

Lemma 18.5. *Suppose* $f : [a,b] \longrightarrow \mathbb{R}$ *is* C^1, *and* f'' *is defined at each point of* (a,b). *Then there exists* $c \in (a,b)$ *such that*

$$f(b) = f(a) + (b-a)f'(a) + \tfrac{1}{2}(b-a)^2 f''(c). \tag{51}$$

If $b < a$, *then, on corresponding hypotheses, there again exists* $c \in (b,a)$ *satisfying* (51).

Proof. Set, for $x \in [a,b]$,

$$F(x) := f(x) + (b-x)f'(x) + Q(b-x)^2, \tag{52}$$

where $Q := (b-a)^{-2}(f(b) - f(a) - (b-a)f'(a))$, so that $F(b) = f(b) = F(a)$. Now F is continuous on $[a,b]$, and differentiable on (a,b), so that the hypotheses of Rolle's theorem are fulfilled, and there exists $c \in (a,b)$ such that $F'(c) = 0$. However,

$$F'(x) = f'(x) - f'(x) + (b-x)f''(x) - 2Q(b-x),$$

so that $2Q(b-c) = (b-c)f''(c)$ or $Q = \frac{1}{2}f''(c)$. Substitute this in (52). □

Theorem 18.6. *Suppose in* 18.5 *that* f'' *is nonnegative on* (a,b). *Then* f *is convex on* $[a,b]$.

Proof. Take $a_1 < a_2$ in $[a,b]$, and $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, such that $\alpha_1 + \alpha_2 = 1$. Apply 18.5 to the intervals $[a_1, q]$ and $[q, a_2]$ separately, where $q = \alpha_1 a_1 + \alpha_2 a_2 \in [a_1, a_2]$. Thus there exist points $s \in [q, a_2]$ and $r \in [a_1, q]$ such that

$$f(a_2) = f(q) + (a_2 - q)f'(q) + \tfrac{1}{2}(a_2 - q)^2 f''(s),$$
$$f(a_1) = f(q) + (a_1 - q)f'(q) + \tfrac{1}{2}(a_1 - q)^2 f''(r).$$

Combining these equalities,

$$\alpha_1 f(a_1) + \alpha_2 f(a_2) = f(q) + (\alpha_1 a_1 + \alpha_2 a_2 - q)f'(q) + \tfrac{1}{2}\{\alpha_1(a_1 - q)^2 f''(r) + \alpha_2(a_2 - q)^2 f''(s)\}.$$

The expression in braces is nonnegative; the previous term vanishes by the definition of q. Hence $\alpha_1 f(a_1) + \alpha_2 f(a_2) \geq f(q) = f(\alpha_1 a_1 + \alpha_2 a_2)$, as required. □

Definition 18.7. f is *strictly convex* on $[a,b]$ if the inequality

$$f(\alpha_1 a_1 + \alpha_2 a_2) < \alpha_1 f(a_1) + \alpha_2 f(a_2)$$

holds whenever $a_1, a_2 \in [a,b]$, $a_1 \neq a_2$, $\alpha_1 > 0$, $\alpha_2 > 0$, and $\alpha_1 + \alpha_2 = 1$.

That is, strict convexity means that the inequality (48) is an equality only if the "point in the middle" is in fact an end-point. The proofs of 18.6 and of 18.3, with slight modifications, show in addition that

Corollary 18.8. *If* f'' *in* 18.5 *is strictly positive on* (a,b), *then* f *is strictly convex on* (a,b). *If* $a_1 \le a_2 \le \cdots \le a_n$, *equality in* (49) *occurs only if, for any index* k *such that* $a_k < a_{k+1}$, *either* $\alpha_1 + \cdots + \alpha_k = 0$ *or* $\alpha_1 + \cdots + \alpha_k = 1$. □

These results can be used to deduce most of the standard elementary inequalities of mathematics, a fact first pointed out by Jensen[5] in 1906 when he introduced the notion of convex functions. Here are two examples.

Lemma 18.9. *If* $0 \le x \le \frac{1}{2}\pi$, *then* $\sin x \ge 2x/\pi$, *with equality only when* $x = 0, \frac{1}{2}\pi$.

Proof. Let $f(x) := -\sin x$. Then $f''(x) = \sin x > 0$ for $0 < x < \frac{1}{2}\pi$. Hence, by 18.3,

$$-\sin((1-\beta)0 + \beta\pi/2) \le -(1-\beta)\sin 0 - \beta\sin(\pi/2)$$

for any $\beta \in [0,1]$, with equality only when $\beta = 0, 1$. The result follows by taking $\beta := 2x/\pi$. □

Lemma 18.10. *For any* $n \in \mathbb{N}$ *and positive numbers* $b_1, b_2, \ldots, b_n$,

$$(b_1 b_2 \ldots b_n)^{1/n} \le \frac{b_1 + b_2 + \cdots + b_n}{n}, \tag{53}$$

with equality only when $b_1 = b_2 = \cdots = b_n$.

Proof. Take $f(x) := \exp x$, and then $f''(x) = \exp x > 0$ for all x. Let $a_i := \log b_i$ and $\alpha_i := 1/n$ for all n, and then (49) gives

$$\exp\left(\frac{\log b_1 + \log b_2 + \cdots + \log b_n}{n}\right) \le \frac{\exp(\log b_1) + \cdots + \exp(\log b_n)}{n},$$

exactly as required. Furthermore, equality occurs only if $b_1 = b_2 = \cdots = b_n$. □

This is perhaps the easiest proof of the *general* arithmetic-geometric mean inequality. (The "harmonic mean inequality" results by taking a_i^{-1} instead of a_i). However, it is not purely algebraic, and some authors expend considerable ingenuity on more "elementary" proofs which begin with $\sqrt{a_1 a_2} \le \frac{1}{2}(a_1 + a_2)$ and work up to (53). The inequality we shall need later, which follows, is obtainable from (53) by taking limits, but it is simpler to prove it directly. The case $\alpha = \frac{1}{2}$ is just (46).

Lemma 18.11. *Suppose* $a > 0$, $b > 0$, $0 \le \alpha \le 1$. *Then*

$$a^\alpha b^{1-\alpha} \le \alpha a + (1-\alpha)b,$$

with equality only if $a = b$ *or if* $\alpha = 0, 1$.

[5] Acta Math. 30 (1906), 175-193.

Proof. Take $f(x) = -\log x$, for $x > 0$. Then $f''(x) = x^{-2} > 0$. By 18.8,

$$f(\alpha a + (1-\alpha)b) \leq \alpha f(a) + (1-\alpha) f(b), \quad \text{or}$$
$$\alpha \log a + (1-\alpha)\log b \leq \log(\alpha a + (1-\alpha)b),$$

with equality in the cases stated. Exponentiate to obtain the stated result. □

Our use for this inequality will be as a step in proving the Hölder and Minkowski inequalities, which lead to the "Lebesgue spaces" L^p we shall be discussing. The argument involves some rather puzzling manipulations with 'conjugate indices', and there is a more general inequality due to W. H. Young (only one of several famous inequalities he discovered) in which more general functions are substituted for exponentiations and the manipulations appear more naturally. This leads to "Orlicz spaces" instead of "Lebesgue spaces". However, the Lebesgue spaces are vastly more important.

§19. "Jensen's inequality".

The quotation marks indicate that this is what the probabilists call it, rather confusingly for the rest of us. In complex analysis, a quite different result bears much the same name.

Theorem 19.1. *Let* (Ω, Σ, P) *be a probability space, and* $f : \Omega \longrightarrow \mathbb{R}$ *an integrable random variable. Suppose* $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ *is convex and continuous. Then* $\int \phi \circ f$ *exists (though* $\phi \circ f$ *need not be integrable), and*

$$\phi\left(\int f\right) \leq \int \phi \circ f\,.$$

The assumption that ϕ be continuous is redundant; see the remark after 18.4. The proof that follows is complicated by some cases that are not particularly useful in practice — the fundamental idea is rather simple.

Proof. Since ϕ is continuous, $\phi \circ f$ is measurable (by 12.8).

Let $\phi(0) = \alpha$, $\phi(0) - \phi(-1) = \beta$. According to 18.4, the function

$$\psi(\xi) := \phi(\xi) - \beta\xi - \alpha \tag{54}$$

is also convex, and $\psi(0) = \psi(-1) = 0$. Now, by 18.2,

$$\text{if } \xi > 0\,, \qquad (\psi(\xi) - \psi(0))/(\xi - 0) \geq 0\,, \quad \text{so } \psi(\xi) \geq 0\,; \tag{55}$$

$$\text{if } \xi < -1\,, \quad (\psi(-1) - \psi(\xi))/(-1 - \xi) \leq 0\,, \quad \text{so } \psi(\xi) \geq 0\,. \tag{56}$$

Of course $\psi(\xi) \leq 0$ when $-1 \leq \xi \leq 0$, by convexity. Since ψ is continuous, however, it is bounded below on $[-1,\, 0]$. Hence, ψ is bounded below for all ξ, say by γ; $\psi \circ f$ is also bounded below by γ, and $\int \psi \circ f$ exists (for $P(\Omega) = 1$, and so $\int (\psi \circ f)^- < \infty$). I shall first prove the theorem for ψ instead of ϕ.

Construct a sequence (s_n) of simple functions such that

$$\text{when } f(x) \geq 0\,,\ s_n(x) \uparrow f(x)\,;\ \text{when } f(x) < 0\,,\ s_n(x) \downarrow f(x)\,, \tag{57}$$

and, in addition, $s_1(x) = -1$ when $f(x) \leq -1$. (One may take $s_n = s_n^+ - s_n^-$, where (s_n^+) is a misnsf tending pointwise to f^+, and (s_n^-) is a misnsf tending to f^-; s_1^+ may be chosen to be 0, and s_1^- to be the characteristic function of $\{x : f(x) \leq -1\}$). Let

$$s_n = \sum\nolimits_{i=1}^{m(n)} a_{ni}\, \mathbf{1}_{E_{ni}}$$

be a standard form for s_n (see 12.13 et seqq.). Then

$$\psi \circ s_n = \sum\nolimits_{i=1}^{m(n)} \psi(a_{ni})\, 1_{E_{ni}} \qquad \text{and, applying 18.3,}$$

$$\psi\left(\int s_n\right) = \psi(\mathcal{S}(s_n)) = \psi\left(\sum\nolimits_i a_{ni}\, P(E_{ni})\right) \leq \sum\nolimits_i \psi(a_{ni})\, P(E_{ni}) = \int \psi \circ s_n\,, \tag{58}$$

since, for each i, $P(E_{ni}) \geq 0$ and $\sum_{i=1}^{m(n)} P(E_{ni}) = P(\Omega) = 1$.

Now, $\int s_n^{\pm} \longrightarrow \int f^{\pm}$ (by the definition 14.1), so that $\int s_n \longrightarrow \int f$. But ψ is continuous, and so $\psi(\int s_n) \longrightarrow \psi(\int f)$. Thus, the theorem for ψ will follow from (58) as $n \longrightarrow \infty$, if we can prove that $\int \psi \circ s_n \longrightarrow \int \psi \circ f$.

Partition Ω into three disjoint measurable subsets:

$$\Omega_1 := \{x : f(x) \le -1\}, \quad \Omega_2 := \{x : -1 < f(x) < 0\}, \quad \Omega_3 := \{x : f(x) \ge 0\}.$$

On Ω_1, (57) ensures that $-1 \ge s_n \downarrow f$. However, if $\eta < \xi \le -1$, by 18.2

$$\frac{\psi(\xi) - \psi(\eta)}{\xi - \eta} \le \frac{\psi(0) - \psi(-1)}{0 - (-1)} = 0, \quad \text{or } \psi(\xi) \le \psi(\eta). \tag{59}$$

That is, ψ is decreasing on $(-\infty, -1]$. Consequently, $\psi \circ s_n$ is increasing on Ω_1. By monotone convergence, $\int_{\Omega_1} \psi \circ s_n \uparrow \int_{\Omega_1} \lim(\psi \circ s_n) = \int \psi \circ f$ (by continuity of ψ; the value of the integral may be ∞).

On Ω_3, $0 \le s_n \uparrow f$, and, as at (59), 18.2 shows that ψ is increasing on $[0, \infty)$. Hence, on Ω_3, $\psi \circ s_n \uparrow \psi \circ f$, and, by monotone convergence, $\int_{\Omega_3} \psi \circ s_n \uparrow \int \psi \circ f$ (which may be ∞).

On Ω_2, we have $-1 < s_n \le 0$ for all n. Recall that ψ is bounded below by γ; hence, for all n, $\gamma \le \psi \circ s_n \le 0$, and $P(\Omega) = 1$. Thus $\int_{\Omega_2} \psi \circ s_n \longrightarrow \int_{\Omega_2} \psi \circ f$ by bounded convergence (in this case the limit is finite, though it may be negative).

Adding the three integrals, we deduce that $\int_\Omega \psi \circ s_n \longrightarrow \int_\Omega \psi \circ f$, and, as already remarked, this suffices, with (58), to establish the theorem for ψ. But, by (54),

$$\begin{aligned}
\phi\left(\int f\right) &= \psi\left(\int f\right) + \beta \int f + \alpha \qquad \text{(by hypothesis, } \int f \text{ is finite)} \\
&\le \int (\psi \circ f + \beta f + \alpha) \qquad \text{(as } P(\Omega) = 1\text{)} \\
&= \int \phi \circ f \qquad \text{as required.}
\end{aligned}$$

□

Notice how often I had to invoke the fact that Ω is a probability measure.

The most obvious example of Jensen's inequality is obtained by taking $\phi(\xi) = |\xi|$. This is a convex function of ξ. The result is that $\left|\int f\right| \le \int |f|$, which of course we already knew (and without restricting ourselves to probability spaces). But there are easy non-trivial examples; for instance, it is far from clear at first glance that, for any function f integrable on $(0,1)$ with Lebesgue measure dt,

$$\exp\left(\int_{(0,1)} f(t)\, dt\right) \le \left(\int_{(0,1)} \exp(f(t))\, dt\right),$$

which follows from taking $\phi(\xi) = e^{\xi}$ in Jensen's theorem.

§20. The Hölder and Minkowski inequalities.

Definition 20.1. Let p and q be positive real numbers. They are described as *conjugate* or as *conjugate indices* or *conjugate exponents* (p is conjugate to q, q is conjugate to p) if

$$\frac{1}{p}+\frac{1}{q}=1\,. \tag{60}$$

The symbolic exponent ∞ is also considered as conjugate to 1, and vice versa.

Note 20.2. If p and q are conjugate indices, then $1\le\min(p,q)\le 2\le\max(p,q)\le\infty$. If both are finite, then both are greater than 1. 2 is the only exponent conjugate to itself, and the only natural number whose conjugate is also an integer; and the relation (60) (for finite p,q) is equivalent to each of the equalities

$$p+q=pq\,,\quad (p-1)q=p\,,\quad \frac{1}{p-1}=q-1\,, \tag{61}$$

all of which are used from time to time.

In what follows, (Ω,Σ,μ) is a fixed measure space.

Theorem 20.3. (Hölder's inequality for integrals). *Let p,q be finite conjugate indices, and let $f,g:\Omega\longrightarrow\mathbb{R}$ be measurable functions such that $|f|^p$ and $|g|^q$ are integrable. Then fg is integrable, and*

$$\int|fg|\le\left(\int|f|^p\right)^{1/p}\left(\int|g|^q\right)^{1/q}.$$

Proof. Let $M:=\{x\in\Omega: f(x)g(x)\ne 0\}\in\Sigma$. If $\mu(M)=0$, $\int|fg|=0$, so there is nothing to prove; assume $\mu(M)>0$, and then $0<\int|f|^p<\infty$ and $0<\int|g|^q<\infty$. For any $x\in M$, take in 18.11 $\alpha:=1/p$, $1-\alpha:=1/q$, and $a:=\dfrac{|f(x)|^p}{\int_M|f|^p}$, $b:=\dfrac{|g(x)|^q}{\int_M|g(x)|^q}$:

$$\frac{|f(x)g(x)|}{\left(\int_M|f|^p\right)^{1/p}\left(\int_M|g|^q\right)^{1/q}}\le\frac{|f(x)|^p}{p\int_M|f|^p}+\frac{|g(x)|^q}{q\int_M|g|^q}\,.$$

Integrate this inequality over M:

$$\frac{\int_M|fg|}{\left(\int_M|f|^p\right)^{1/p}\left(\int_M|g|^q\right)^{1/q}}\le\frac{1}{p}+\frac{1}{q}=1\,.$$

Thus

$$\begin{aligned}\int_\Omega|fg|=\int_M|fg|\ &\le\left(\int_M|f|^p\right)^{1/p}\left(\int_M|g|^q\right)^{1/q}\\ &\le\left(\int_\Omega|f|^p\right)^{1/p}\left(\int_\Omega|g|^q\right)^{1/q}.\end{aligned}$$

□

As so often in mathematics, the proof renders prosaic what at first sight might seem rather unexpected; one might initially suppose, although for no good reason, that fg is not constrained by the integrals of $|f|^p$ and $|g|^q$. The case $p = q = 2$ is the familiar Cauchy-Schwarz inequality, to which the same comment applies.

Hölder's inequality was originally stated — by Rogers in 1888, a year before Hölder — for finite sums of numbers, more or less in the form:

$$\left|\sum_{k=1}^n a_k b_k\right| \leq \left(\sum_{k=1}^n |a_k|^p\right)^{1/p} \left(\sum_{k=1}^n |b_k|^q\right)^{1/q}. \tag{62}$$

This is an easy corollary of the Theorem; simply take $\Omega \coloneqq \{1, 2, \ldots, n\}$, and define $\Sigma = \mathcal{P}(\Omega)$, $\mu(E) = \#(E)$ ("counting measure in Ω"). It is possible to prove the Theorem itself by gradually working up from this algebraic version, but there is no advantage in doing so. Hardy and his students and collaborators[6] tended to avoid 18.11 in expositions of Hölder's and similar inequalities, because they felt that it was a disproportionately advanced result for the purpose. There are interesting proofs both of 18.11 and of (62) which are "elementary" in the sense of avoiding calculus entirely[7]. They deal directly, by algebraic methods, with the case when p and q are both rational, and then pass to a limit. However, there is no circularity involved in our proof, which is far shorter.

Lemma 20.4. *Let $p > 0$. Let f and g be measurable functions such that $|f|^p$ and $|g|^p$ are integrable. Then $|f + g|^p$ is integrable, and in fact*

$$\int |f + g|^p \leq 2^p \int |f|^p + 2^p \int |g|^p.$$

Proof. Let $A_1 \coloneqq \{x \in \Omega : |f(x)| \geq |g(x)|\}$, $A_2 \coloneqq \Omega \setminus A_1$. These are measurable subsets of Ω. If $x \in A_1$, then $|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq 2|f(x)|$, whilst, if $x \in A_2$, similarly $|f(x) + g(x)| \leq |f(x)| + |g(x)| < 2|g(x)|$. Thus, for all $x \in \Omega$,

$$|f(x) + g(x)|^p \leq 2^p |f(x)|^p + 2^p |g(x)|^p,$$

and the result follows by integration over Ω. □

Theorem 20.5. (Minkowski's inequality for integrals.) *Let (Ω, Σ, μ) be any measure space, and suppose that $1 < p < \infty$. Let f and g be measurable functions such that $|f|^p$ and $|g|^p$ are integrable. Then $|f + g|^p$ is integrable, and*

$$\left(\int_\Omega |f + g|^p\right)^{1/p} \leq \left(\int_\Omega |f|^p\right)^{1/p} + \left(\int_\Omega |g|^p\right)^{1/p}.$$

Proof. 20.4 proves integrability. Now let q be the index conjugate to p. Recall from (61) that $p = (p-1)q$. Ergo, $|f|^p$, $|g|^p$ and $|f + g|^{(p-1)q}$ are integrable, and by Hölder's inequality

---

[6] See for instance Hardy, Littlewood and Pólya, *Inequalities*, O.U.P. 1934 and 1952; the relevant section is 2.7, although 2.5 and 2.6 may clarify what is going on. A similar eccentricity is found in books by Phillips and Titchmarsh, who were also under Hardy's influence.

[7] Hardy, Littlewood, and Pólya, *loc. cit.*

$$\int |f||f+g|^{p-1} \le \left(\int |f|^p\right)^{1/p}\left(\int |f+g|^{(p-1)q}\right)^{1/q},$$
$$\int |g||f+g|^{p-1} \le \left(\int |g|^p\right)^{1/p}\left(\int |f+g|^{(p-1)q}\right)^{1/q},$$

and, therefore,

$$\begin{aligned}\int |f+g|^p \le & \int |f||f+g|^{p-1} + \int |g||f+g|^{p-1} \\ & \le \left(\int |f|^p\right)^{1/p}\left(\int |f+g|^{(p-1)q}\right)^{1/q} + \left(\int |g|^p\right)^{1/p}\left(\int |f+g|^{(p-1)q}\right)^{1/q} \\ & \le \left\{\left(\int |f|^p\right)^{1/p} + \left(\int |g|^p\right)^{1/p}\right\}\left(\int |f+g|^p\right)^{1/q}.\end{aligned}$$

If $\int |f+g|^p = 0$, there is nothing to prove. But, if $\int |f+g|^p > 0$, the last inequality may be divided by $\left(\int |f+g|^p\right)^{1/q}$ on both sides:

$$\left(\int |f+g|^p\right)^{1/p} = \left(\int |f+g|^p\right)^{1-\frac{1}{q}} \le \left(\int |f|^p\right)^{1/p} + \left(\int |g|^p\right)^{1/p}. \qquad \square$$

Lemma 20.6. *Suppose that f and g are integrable functions $\Omega \longrightarrow \mathbb{R}$. Then $f+g$ is also integrable, and $\int |f+g| \le \int |f| + \int |g|$.*

Proof. Indeed, the inequality $|f(x)+g(x)| \le |f(x)| + |g(x)|$ holds for every x. $\square$

This is Minkowski's inequality when $p = 1$. It remains to consider its extension to the case where $p = \infty$, and "Hölder's inequality" when $p = 1$, $q = \infty$.

Definition 20.7. A function $f : \Omega \longrightarrow \mathbb{R}$ is *essentially bounded*, or "a.e. bounded", if there are a measurable set Z of zero measure and a number $K_Z \ge 0$ such that

$$(\forall x \in \Omega \setminus Z) \quad |f(x)| \le K_Z .$$

This definition is quite consistent with 14.8, but it has a peculiarity: the exceptional set Z may depend on the choice of K_Z (or vice versa). For instance, the function $g : \mathbb{R} \longrightarrow \mathbb{R}$ which is 0 except at the non-zero rationals, and takes value q at the rational which is p/q (where $q \in \mathbb{N}$) in lowest terms, is essentially bounded; if Z consists of all the rationals, K_Z may be 0, but if Z consists of the rationals with denominators exceeding 10 in lowest terms, then K_Z must be at least 9, and so on. There is no possible concept of "a.e. supremum", although I have come across authors who carelessly adopt the phrase.

Definition 20.8. Let $f : \Omega \longrightarrow \overline{\mathbb{R}}$ be measurable. Define the *essential supremum* and *essential infimum* of f, essup f and essinf f, by

$$\text{essup } f = \sup\{\alpha \in \overline{\mathbb{R}} : \mu\{x \in \Omega : f(x) \geq \alpha\} > 0\},$$
$$\text{essinf } f = \inf\{\alpha \in \overline{\mathbb{R}} : \mu\{x : f(x) \leq \alpha\} > 0\}.$$

There are several definitions of the essential supremum and the essential infimum that are for practical purposes equivalent. It is easily proved from the above definition that f is essentially bounded if and only if essup $f < \infty$ and essinf $f > -\infty$. (If $\mu(\Omega) = 0$, then essup $f = -\infty$, essinf $f = \infty$, and f is still essentially bounded.)

Lemma 20.9. essinf $f \leq f \leq$ essup f *a.e. on* Ω.

Proof. Suppose that $\beta := \text{essup } f < \infty$. Then, for any $n \in \mathbb{N}$,

$$E_n := \{x \in \Omega : f(x) \geq \beta + \tfrac{1}{n}\}$$

is measurable of measure 0. Hence $E := \{x \in \Omega : f(x) > \beta\} = \bigcup_{n=1}^{\infty} E_n$ is also measurable of measure 0, and $f(x) \leq \text{essup } f$ for $x \notin E$. If essup $f = \infty$, there is nothing to prove. There is a similar argument for the essential infimum. □

It follows that if $\mu(\Omega) > 0$, then essinf $f \leq$ essup f.

Theorem 20.10. *Let $f, g : \Omega \longrightarrow \mathbb{R}$ be measurable, and suppose that both $|f|$ and $|g|$ are essentially bounded. Then so is $|f + g|$, and*

$$\text{essup}|f + g| \leq \text{essup}|f| + \text{essup}|g|. \qquad \square$$

This result is the analogue of Minkowski's inequality for the case $p = \infty$, in which essup$|f|$ takes the place of $(\int |f|^p)^{1/p}$.

Theorem 20.11. *Suppose that f and g are measurable functions $\Omega \longrightarrow \mathbb{R}$, and let f be integrable and $|g|$ be essentially bounded. Then $|fg|$ is integrable, and*

$$\int_\Omega |fg| \leq (\text{essup}|g|) \int_\Omega |f|.$$

Proof. By 20.9, there is a measurable set E of zero measure such that $|g| \leq \text{essup}|g|$ on $\Omega \setminus E$. Thus

$$\int_\Omega |fg| = \int_{\Omega\setminus E} |fg| \leq (\text{essup}|g|) \int_{\Omega\setminus E} |f| = (\text{essup}|g|) \int_\Omega |f|. \qquad \square$$

When $0 < p < 1$, Minkowski's inequality definitely fails. For instance, let $\Omega := \{1, 2\}$, with counting measure; then $\left(\int \mathbf{1}_{\{1\}}^p\right)^{1/p} = \left(\int \mathbf{1}_{\{2\}}^p\right)^{1/p} = 1$, but

$$\left(\int (\mathbf{1}_{\{1\}} + \mathbf{1}_{\{2\}})^p\right)^{1/p} = 2^{1/p} > 2\,.$$

Theorem 20.12. *Let $f, g : \Omega \longrightarrow \mathbb{R}$ be measurable. If $0 < p < 1$, then*

$$\int |f + g|^p \leq \int |f|^p + \int |g|^p\,.$$

Proof. It is easily seen that, if $\xi, \eta \geq 0$, then $(\xi + \eta)^p \leq \xi^p + \eta^p$. The result follows by integration of this pointwise inequality. □

§21. Complexification.

My definition of the integral assumed that the functions we consider are all real-valued (or extended-real-valued; although I have, for simplicity, stated most of the theorems on the assumption that all values are finite, it is just a matter of excluding a set of measure zero to obtain the more general versions). However, one often wishes to have a theory of integration for complex-valued functions, and even for complex-valued measures. This is easily accomplished.

As usual, let (Ω, Σ, μ) be a measure space. A function $f : \Omega \longrightarrow \mathbb{C}$ is defined to be Σ-measurable if both its real part and its imaginary part are Σ-measurable; in that case $|f|$ is measurable as a real-valued function. f is defined to be integrable with respect to Σ and μ if both its real and imaginary parts are integrable with respect to Σ and μ. Since

$$\max(|\Re f|, |\Im f|) \leq |f| \leq |\Re f| + |\Im f|\,,$$

it follows that a measurable complex-valued function f is integrable in the complex sense if and only if $|f|$ is integrable in the real sense. The analogues of the results of §17 are readily proved by looking at real and imaginary parts.

Hölder's inequality involves only the integrals of moduli, and so remains true in the complex case without any change in the proof. The same applies to Minkowski's inequality (with some use of the triangle inequality for moduli of complex numbers) and to the extended inequalities 20.10 (only the essential supremum of the modulus is taken), 20.11, and 20.12.

It is natural to wonder whether the integral of a complex-valued function — or, more generally still, of a vector-valued function — might be defined directly, and indeed it is possible. However, such definitions are less straightforward and natural than ours.

§22. The Lebesgue spaces.

I am not sure exactly why they are called Lebesgue spaces. As before, (Ω, Σ, μ) is a fixed measure space.

Definition 22.1. Let $0 < p < \infty$. The class of real or complex measurable functions f on Ω such that $|f|^p$ is integrable on Ω is described as the *space of pth power integrable functions* on Ω, and is denoted by $\mathcal{L}^p(\Omega, \Sigma, \mu)$. The class of all real or complex measurable functions on Ω is denoted by $\mathcal{M}(\Omega, \Sigma, \mu)$. The class of all essentially bounded real or complex measurable functions on Ω is denoted by $\mathcal{L}^\infty(\Omega, \Sigma, \mu)$.

The notation varies according to the context; for instance, it may be desirable to indicate whether the values are to be real or complex, and there may be no need to mention Ω, Σ, or μ explicitly. The affix p may be superscript or subscript. My impression is that subscripts are commoner nowadays, but superscripts are preferred by the better mathematicians.

These classes of functions are vector spaces (over the field $\mathbb{K}$, which is $\mathbb{R}$ or $\mathbb{C}$ as the case may be) under pointwise operations. For $\mathcal{L}^p$, where $0 < p < \infty$, it is trivial that $f \in \mathcal{L}^p \wedge \lambda \in \mathbb{K} \Longrightarrow \lambda f \in \mathcal{L}^p$, and 20.4 shows that $f, g \in \mathcal{L}^p \Longrightarrow f + g \in \mathcal{L}^p$. For $\mathcal{L}^\infty$, 20.10 takes the place of 20.4. For $\mathcal{M}$, 12.6 *(a)* and *(c)* settle the matter at once.

Definition 22.2. *(a)* For $f \in \mathcal{L}^p$, where $1 \le p < \infty$, $\|f\|_p^\wedge := \left(\int |f|^p \right)^{1/p}$.

(b) If $p = \infty$, $\|f\|_\infty^\wedge := \operatorname{essup}|f|$.

It follows that, if $1 \le p \le \infty$, then, for all $\lambda \in \mathbb{K}$ and $f, g \in \mathcal{L}^p$,

$$\|\lambda f\|_p^\wedge = |\lambda| \|f\|_p^\wedge, \quad \|f + g\|_p^\wedge \le \|f\|_p^\wedge + \|g\|_p^\wedge .$$

(The first statement is trivial; the second is 20.5, 20.6, or 20.10.) These are two of the requirements for $\|\|_p^\wedge$ to be a *norm* in $\mathcal{L}^p$. It is clear that $\|f\|_p^\wedge \ge 0$ for any $f \in \mathcal{L}^p$. The remaining axiom of a norm $\|\|$ in a real or complex vector space V is that $\|x\| = 0$ only when $x = 0$ in V; for each of the functions $\|\|_p^\wedge$, this would clearly amount to saying that a measurable function that is zero a.e. is zero everywhere. This is definitely false in $\mathcal{L}^p(\Omega, \Sigma, \mu)$ unless the only set of measure zero is $\emptyset$. Thus, $\|\|_p^\wedge$ is usually a *seminorm* (or *pseudonorm*) in $\mathcal{L}^p$, rather than a norm.

However, we may introduce an equivalence relation $\sim$ in $\mathcal{L}^p$, setting $f \sim g$ whenever f and g are equal a.e.

Definition 22.3. The set of equivalence classes of elements of $\mathcal{L}^p(\Omega, \Sigma, \mu)$, for $0 < p \le \infty$, under the relation $\sim$ of almost everywhere equality constitutes the *Lebesgue space of exponent p on the measure space* (Ω, Σ, μ), denoted $L^p(\Omega, \Sigma, \mu)$.

If $[f]$ denotes the $\sim$-equivalence class in $\mathcal{L}^p$ of the function f, one defines the vector space operations in L^p, for $0 < p \le \infty$, by

$$(\forall \alpha, \beta \in \mathbb{K})(\forall f, g \in \mathcal{L}^p) \quad \alpha[f] + \beta[g] := [\alpha f + \beta g],$$

and, for $1 \le p \le \infty$, one defines a norm in L^p by

$$(\forall f \in \mathcal{L}^p) \quad \|[f]\|_p := \|f\|_p^\wedge .$$

It is easily checked that the left-hand sides are well-defined by these prescriptions (that is: if $f \sim f'$ and $g \sim g'$, then $\alpha f + \beta g \sim \alpha f' + \beta g'$ and $\|f\|_p\hat{} = \|f'\|_p\hat{}$), and that L^p becomes a vector space over $\mathbb{K}$ and $\|\|_p$ a norm in L^p.

In this way, the Lebesgue spaces L^p for $1 \leq p \leq \infty$ become normed vector spaces over $\mathbb{K}$. In practice, the conceptual distinction between L^p and $\mathcal{L}^p$ tends to be blurred; people write of a *function* f that $f \in L^p$, rather than $f \in \mathcal{L}^p$. This does occasionally involve some confusion, but it is rarely serious.

The procedure which constructs L^p from $\mathcal{L}^p$ is quite general. If $\|\|\hat{}$ is a seminorm in a vector space V, let $Z := \{v \in V : \|v\|\hat{} = 0\}$. Then Z is a vector subspace of V, and $\|\|\hat{}$ induces a norm $\|\|$ on the quotient vector space V/Z. In our construction, Z is the set of $\mathbb{K}$-valued measurable functions that are a.e. equal to 0, and $L^p := \mathcal{L}^p/Z$. It is worth noting, however, that Z is the same for all $p \geq 1$.

Definition 22.4. In the vector space L^p for $0 < p < 1$, define a metric d_p by

$$(\forall f, g \in \mathcal{L}^p) \quad d_p([f],[g]) := \int_\Omega |f-g|^p .$$

It is easily checked that d_p is well-defined; that it is a metric follows from 20.12. (The right-hand integral defines a *pseudometric* or *semimetric* on $\mathcal{L}^p$.) This metric cannot be derived from a norm (except in a trivial case!), since $d_p(\alpha[f], \alpha[g]) = |\alpha|^p d_p([f],[g])$, and if it came from a norm we should have $|\alpha|$ rather than $|\alpha|^p$. In fact the spaces $L^p([0,1])$ (with respect to Lebesgue measure, for $0 < p < 1$) are standard examples of topological vector spaces that are not locally convex.

Definition 22.5. Let Ω be any set, and take μ to be counting measure on $\Sigma := \mathcal{P}(\Omega)$. In this case, $L^p(\Omega, \Sigma, \mu)$ is naturally identified with $\mathcal{L}^p(\Omega, \Sigma, \mu)$ (the $\sim$-equivalence classes are singletons); each element is a function $f : \Omega \longrightarrow \mathbb{K}$ which is "countably non-zero", which means that it takes the value 0 except on a countable set (see 14.5), and such that

$$\sum\nolimits_{x\in\Omega} |f(x)|^p < \infty \quad (\text{when } 0 < p < \infty), \qquad \sup\nolimits_{x\in\Omega} |f(x)| < \infty \quad (\text{for } p = \infty).$$

The space $L^p(\Omega, \Sigma, \mu)$ is customarily denoted $l^p(\Omega)$.

When $\Omega := \mathbb{N}$, one customarily writes l^p, and describes it as "the sequence space l^p"; its elements are usually written as sequences $(x_n)_{n=1}^\infty$ with terms $x_n \in \mathbb{K}$. There is also a "bilateral sequence space" $l^p(\mathbb{Z})$ when $\Omega = \mathbb{Z}$.

Definition 22.6. The sequence space c_0 consists of those sequences $(x_n)_{n=1}^\infty$, with terms in $\mathbb{K}$, such that $x_n \to 0$ as $n \to \infty$. It is a vector space under termwise operations, and is normed by $\|(x_n)\| := \sup\limits_n |x_n|$.

In fact, c_0 is a subset of l^∞, it is a vector subspace of l^∞, and the norm on c_0 is the restriction of $\|\|_\infty$. It is a worthwhile exercise to prove the completeness of these various sequence spaces directly, i.e. without using any facts from the theory of the integral.

Definition 22.7. Let (Ω, Σ, μ) be any measure space, $0 < p < \infty$. A sequence (f_n) in $\mathcal{L}^p$ is *Cauchy in p-mean*, or *Cauchy in* $\mathcal{L}^p$, if $\int |f_m - f_n|^p \to 0$ as $m, n \to \infty$. I recall that this means (cf. 0.11)

$$(\forall \epsilon > 0)(\exists N \in \mathbb{N}) \quad m, n \geq N \Longrightarrow \int |f_m - f_n|^p < \epsilon \,.$$

Similarly, if $f \in \mathcal{L}^p$, $f_n \to f$ *in p-mean*, or *in* $\mathcal{L}^p$, if $\int |f_n - f|^p \to 0$ as $n \to \infty$.

When $p = 1$, the phrases *Cauchy in mean* and *convergent in mean* are sometimes used; for $p = 2$, *Cauchy* or *convergent in mean square*. The definition above has the advantage of covering both the case $p \geq 1$, when the concepts are those appropriate to the seminorm $\| \|_p$ on $\mathcal{L}^p$, and the case $0 < p < 1$, when they may be derived from the pseudometric d_p. As before, L^p is often written for $\mathcal{L}^p$; one has Cauchy and convergent sequences in L^p.

Lemma 22.8. *If (f_n) is Cauchy in $\mathcal{L}^p$, where $0 < p < \infty$, it is Cauchy in measure. If $f_n \to f$ in $\mathcal{L}^p$, then $f_n \to f$ in measure.*

Proof. Given $\epsilon > 0$, choose N so that $m, n \geq N \Longrightarrow \int |f_m - f_n|^p < \epsilon^{p+1}$. Then, for $m, n \geq N$, $\mu(\{x \in \Omega : |f_m(x) - f_n(x)| \geq \epsilon\}) < \epsilon$. Similarly for the second assertion. □

Theorem 22.9. *Let (f_n) is a Cauchy sequence in $\mathcal{L}^p$, where $0 < p < \infty$, then there is a function $f \in \mathcal{L}^p$ such that $f_n \to f$ in $\mathcal{L}^p$.*

Proof. As (f_n) is Cauchy in $\mathcal{L}^p$, there exists N such that $n \geq N \Longrightarrow \int |f_n - f_N|^p \leq 1$. By 20.4, $n \geq N \Longrightarrow \int |f_n|^p \leq 2^p \int |f_N|^p + 2^p \int |f_n - f_N|^p \leq 2^p \int |f_N|^p + 2^p$, and so

$$(\forall n \in \mathbb{N}) \quad \int |f_n|^p \leq K := \max\left\{ \int |f_1|^p, \int |f_2|^p, \ldots, \int |f_{N-1}|^p, 2^p \int |f_N|^p + 2^p \right\}.$$

(Using 20.4, we need not distinguish the cases $0 < p < 1$ and $1 \leq p < \infty$. But the bound K could be improved by treating them separately and using 20.5 or 20.12 as appropriate.)

By 22.8, (f_n) is Cauchy in measure. By 17.25, there is a subsequence $(f_{n(k)})_{k=1}^{\infty}$ which is almost uniformly Cauchy. By 17.20, there is a measurable function f such that $f_{n(k)} \to f$ as $n \to \infty$. By Fatou's lemma, 15.5,

$$\int |f|^p = \int \liminf_{k \to \infty} |f_{n(k)}|^p \leq \liminf_{k \to \infty} \int |f_{n(k)}|^p \leq K \,.$$

Hence $f \in \mathcal{L}^p$. If $\epsilon > 0$, there exists N such that $m, n \geq N \Longrightarrow \int |f_m - f_n|^p < \infty$, and, if $n \geq N$, it follows that

$$\int |f - f_n|^p = \int \liminf_{k \to \infty} |f_{n(k)} - f_n|^p \leq \liminf_{k \to \infty} \int |f_{n(k)} - f_n|^p \leq \epsilon \,.$$

So $f_n \to f$ in $\mathcal{L}^p$. □

Remark 22.10. The general form of the above proof follows the lines suggested in 17.4; we construct f as a weaker kind of limit (an almost uniform limit), and only subsequently show that it is both in the right space and a limit in the right sense. It is not necessary to use almost uniform convergence, and most authors prefer a.e. convergence, which is weaker still. Of

course, in the complex case one must establish the almost uniform convergence $f_{n(k)} \to f$ by considering the real and imaginary parts separately.

Fatou's lemma appears here as a sort of deus ex machina, descending suddenly from the heavens to resolve an otherwise refractory problem (the estimate of the integral of the limit). I know at least one book, by Lusternik and Sobolev, in which — for no very clear reason — they avoid Fatou's lemma at this point. It is possible, but the result is a much more complicated proof.

It is clear that 22.9 amounts to a proof that L^p is a complete normed space, i.e. a Banach space, when $1 \leq p < \infty$, and a complete metric space when $0 < p < 1$. These completeness assertions are sometimes called the Riesz-Fischer theorem, but the name is also applied to some equivalent results, more particularly in L^2, which they show to be a Hilbert space.

Theorem 22.11. *The normed space L^∞ is complete.*

Proof. Let (f_n) be a Cauchy sequence in $\mathcal{L}^\infty$. It is a.e. uniformly Cauchy, by definition, and thus converges uniformly a.e. to some f, which must be in $\mathcal{L}^\infty$ too. (See 17.10; but, of course, one must consider real and imaginary parts separately in the complex case.) □

These results on completeness of the Lebesgue spaces are the beginning of *functional analysis*, the idea of which is to establish theorems about, for instance, the solutions of differential equations by considering properties of functions in the mass, rather than by constructing them individually. They are the most substantial practical justification for the Lebesgue integral. The completeness of L^p (and, especially, of L^2) is fundamental for many purposes of applied mathematics, for instance for Fourier series or least-squares regression, and *is definitely false if only the Riemann theory is assumed*. An L^p-Cauchy sequence of Riemann-integrable functions need not converge to a Riemann-integrable function.

Theorem 22.12. *Suppose $0 < p < \infty$, $g \in \mathcal{L}^p$, and (f_n) is a sequence of measurable functions converging* either *a.e.* or *in measure to the measurable function f, such that $|f_n| \leq g$ a.e. for all n. Then $f_n, f \in \mathcal{L}^p$ for all n, and $f_n \to f$ in $\mathcal{L}^p$.*

Proof. That $f_n \in \mathcal{L}^p$ for all n is obvious, and clearly $|f| \leq g$ a.e., so $f \in \mathcal{L}^p$ too (why?) Suppose that $\int |f_n - f|^p$ does *not* tend to 0. Then there exists $\epsilon > 0$ and a subsequence $(f_{n(k)})_{k=1}^\infty$ such that $\int |f_{n(k)} - f|^p \geq \epsilon$ for all k. As $f_{n(k)} \to f$ a.e. or in measure, there is a further subsequence (h_i) such that $h_i \to f$ a.e., by 17.25 and 17.19.

Now, however, $|h_i - f|^p \to 0$ a.e. and $|h_i - f|^p \leq 2^p g^p$, which is integrable. By the dominated convergence theorem 15.12, $\int |h_i - f|^p \to 0$; but this is a contradiction, since, by construction, $\int |h_i - f|^p \geq \epsilon$ for all i. The result follows. □

The Theorem is an extension to $\mathcal{L}^p$ of the dominated convergence theorem, to which it reduces when $p = 1$.

To conclude, I give a Lemma which was used by Halmos as his definition of the integral. Its *advantage* is that it can be put in a form equally applicable to functions with values in $\mathbb{C}$ or even in a Banach space; but it has several evident disadvantages, perhaps principally that the definition of the integral must be postponed until the "kinds of convergence" have been sorted out. As I commented at the start, there are many possible ways of defining the integral, and my aim was to present the one that seemed most "natural" in a certain elementary sense, whilst also introducing the idea of "measure".

Lemma 22.13. *Let* $f : \Omega \longrightarrow \mathbb{R}$ *be a* Σ*-measurable function. Then* f *is integrable*

(a) *if and only if there is a sequence* (f_n) *of integrable simple functions such that* $f_n \to f$ *in measure and* (f_n) *is Cauchy in mean, or*

(b) *if and only if there is a sequence* (f_n) *of integrable simple functions such that* $f_n \to f$ *a.e. and* (f_n) *is Cauchy in mean.*

If either (a) *or* (b) *is satisfied, then* $\int f = \lim \int f_n$.

Proof. By definition (14.1 and 14.12), f is integrable if and only if there are misnsfs (f_n^+) and (f_n^-) such that $f_n^+ \uparrow f^+$, $f_n^- \uparrow f^-$ a.e., and $\lim \int f_n^+$, $\lim \int f_n^-$ are finite. Then $f_n^+ \uparrow f^+$, $f_n^- \uparrow f^-$ in measure (I leave this as an exercise, but compare 17.22), so that, by 17.16, $f_n \coloneqq f_n^+ - f_n^- \to f$ both a.e. and in measure. Certainly

$$\int |f_m - f_n| \le \int (f_m^+ - f_n^+) + \int (f_m^- - f_n^-) \quad \text{for } m \ge n\,,$$

which tends to 0 as $m, n \to \infty$, since $\int f_m^+ \to \int f^+$ and so on. So (f_m) is Cauchy in mean.

Conversely, if (f_n) is Cauchy in $\mathcal{L}^1$ and $f_n \to f$ a.e. or in measure, by 22.9 there exists $g \in \mathcal{L}^1$ with $\int |f_n - g| \to 0$. Hence $f_n \to g$ in measure (by 22.8), and some subsequence tends to g a.e., by 17.25 and 17.19. Thus $f = g$ a.e., and $f_n \to f$ in mean. Since

$$\left| \int f_n - \int f \right| \le \int |f_n - f|\,,$$

the final assertion follows. □