OUTLINE OF LEBESGUE THEORY: A HEURISTIC INTRODUCTION

ROBERT E. WERNIKOFF

TECHNICAL REPORT 310

JANUARY 11, 1957

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U. S. Army (Signal Corps), the U. S. Navy (Office of Naval Research), and the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-64637, Department of the Army Task 3-99-06-108 and Project 3-99-00-100.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 310 January 11, 1957

OUTLINE OF LEBESGUE THEORY: A HEURISTIC INTRODUCTION

Robert E. Wernikoff

Preface

This report is an attempt to give a heuristic exposition of measure theory and of the theory of integration that derives from it. Its purpose is to acquaint communication engineers with a language that has been found most useful in probability theory, statistics, ergodic theory, the theory of linear operators in function spaces – in fact, the language in which much of the mathematical foundation of communication theory is most frequently and most naturally expressed. Hopefully, this descriptive introduction will serve as a source of motivated and meaningful definitions of the principal terms and concepts, and perhaps as an aid in interpreting the rather concise rigorous expositions of the theory.

The author wishes to thank Professor Y. W. Lee for his friendly support and encouragement during this work and to acknowledge gratefully the benefits derived from many stimulating discussions with K. H. Powers, J. Y. Hayase, and A. H. Nuttall, and their help in checking and proofreading the manuscript.

Table of Contents

1. INTRODUCTION: THE RIEMANN INTEGRAL

In this section we shall establish the basic motivation for the results to be described in the report. We start with a review of the Riemann definition of the integral and a derivation of some of the basic properties of the integral implied by this definition. We then show, by means of an example, that there are definite limitations to the class of functions to which Riemann's definition of integration is applicable. It is natural to attempt to isolate that feature of Riemann's definition which is responsible for its limitations, and in so doing we shall find that an apparently simple change of point of view leads us to an approach in which these limitations are irrelevant. The new point of view leads to what is, in effect, a rudimentary form of Lebesgue's definition of the integral. We shall use this imprecise form to determine the new concepts that must be studied in order to make possible a rigorous new definition of integration.

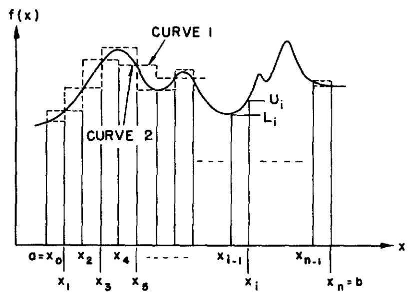

Fig. 1.1. Pertinent to the definition of the Riemann integral.

The Riemann integral, which is the ordinary integral discussed in elementary calculus, can be defined by proceeding as follows. [For a more complete exposition, see, for example, Whittaker and Watson (6), p. 61 et seq., or Rudin (4), p. 87 et seq.] We consider a bounded function f(x) defined on the interval (a, b), as in Fig. 1.1. We subdivide the interval into n parts at the points

$$a_o = x_o \leqslant x_1 \leqslant x_2 \leqslant \dots \leqslant x_{n-1} \leqslant x_n = b$$

and for each subinterval (x_{i-1}, x_i) we define

$$U_i = \text{upper bound of } f(x) \text{ in } (x_{i-1}, x_i)$$

$$L_i = \text{lower bound of } f(x) \text{ in } (x_{i-1}, x_i)$$

and we form the sums

$$S_n = U_1(x_1 - a) + U_2(x_2 - x_1) + \dots + U_n(b - x_{n-1})$$

$$s_n = L_1(x_1 - a) + L_2(x_2 - x_1) + \dots + L_n(b - x_{n-1})$$

We shall call S_n the upper sum; s_n, the lower sum. Since S_n is the area under stepped curve 1, and s_n is the area under curve 2, it is clear from Fig. 1.1 (and can easily be proved analytically) that

$$S_n \geqslant s_n$$

As the number n of subdivisions is increased, S_n can only decrease in value, since curve 1 approaches f(x) more closely. For the analogous reason, as n increases, s_n can only increase in value. Now we consider all possible ways of subdividing (a, b), and let n approach infinity in such a way that the length of the longest subdivision tends to zero. For each way of subdividing (a, b) we obtain a different set of values for S_n and s_n. Let S be the smallest value taken by S_n, and let s be the largest value taken by s_n. It is still true that

$$S \geqslant s$$

If it should happen that S = s, the common value is called the <u>Riemann</u> <u>integral</u> of f(x) between the limits a and b, and is denoted by

$$\int_a^b f(x)\, dx$$

As a result of this definition, the Riemann integral exists (or, f(x) is integrable in the sense of Riemann) if and only if S_n and s_n have a common limit as n approaches infinity in such a way that the longest interval tends to zero, the limit being independent of the mode of subdivision of (a, b). Therefore, if f(x) is Riemann-integrable, given any $\epsilon > 0$ however small, there must exist a $\delta > 0$ which is such that, if the length of the longest subinterval is less than δ, then

$$S_n - s_n < \epsilon$$

We have all of the tools that are necessary to prove that all continuous bounded functions are Riemann-integrable (on a finite interval). For, by definition, f(x) is continuous if, given any $\epsilon > 0$, we can find a $\delta > 0$ such that

$$|f(x) - f(x')| < \epsilon \quad \text{whenever } |x - x'| < \delta$$

If we pick a δ small enough so that

$$|f(x) - f(x')| < \frac{\epsilon}{(b-a)} \quad \text{whenever } |x - x'| < \delta$$

and if we subdivide (a, b) in such a way that all intervals are shorter than δ, then the upper and lower bounds of f(x) in any subinterval (x_i, x_{i+1}) must differ by less than $\epsilon/(b-a)$. That is, we have

$$U_i - L_i < \frac{\epsilon}{b - a}$$

Therefore

$$S_n - s_n = \sum_i U_i(x_i - x_{i-1}) - \sum_i L_i(x_i - x_{i-1}) = \sum_i (U_i - L_i)(x_i - x_{i-1})$$

$$< \sum_i \frac{\epsilon}{b-a}(x_i - x_{i-1}) = \frac{\epsilon}{b-a}\sum_i (x_i - x_{i-1}) = \epsilon$$

That is, we have

$$S_n - s_n < \epsilon$$

Thus we have proved that S_n and s_n approach a common limit, and therefore any continuous bounded function is Riemann-integrable.

It is just as easy to prove that if a bounded function has a finite number of (finite) discontinuities, and is continuous elsewhere on (a, b), it is still integrable. In fact, it is integrable even if it has a countable number of finite discontinuities. [For a definition of "countable," see Appendix I, p. 67.] Heuristically, we can see this simply by thinking of the original interval (a, b) as split up into smaller intervals (a_i, b_i) in such a way that f(x) is continuous within each (a_i, b_i), so that all discontinuities occur at the boundaries between contiguous intervals. Then we simply integrate f(x) in each (a_i, b_i) and add up the results to obtain the integral of f(x) over (a, b).

We can prove just as easily that not all bounded functions are integrable. Consider, for example, the function f(x) defined on (a, b) by

$$\begin{aligned} f(x) &= 1 \quad \text{if } x \text{ is irrational and } a \leq x \leq b \\ &= 0 \quad \text{if } x \text{ is rational and } a \leq x \leq b \end{aligned}$$

Then f(x) is bounded, and discontinuous at every point of (a, b). Following the procedure of page 1, we subdivide (a, b) into n parts. No matter how this is done, within each subinterval (x_{i-1}, x_i) there are both rational and irrational numbers, so that the upper value of f(x) is 1, and the lower value is zero. Therefore, for all n,

$$S_n = \sum_i U_i(x_i - x_{i-1}) = 1\,(b-a) = b - a$$

$$s_n = \sum_i L_i(x_i - x_{i-1}) = 0\,(b-a) = 0$$

The upper sum never approaches the lower sum, and therefore the Riemann integral of f(x) does not exist; that is, it simply is not defined.

The example has shown that there are bounded functions that are not Riemann-integrable. In particular, it would seem that in order to be integrable, a function must not be "too" discontinuous. Let us see why this is so.

To simplify the argument, we shall consider a slightly different definition of the integral, one which may be more familiar than the one given above. Let f(x) be a function defined on (a, b) and divide (a, b) in the same way as above (see Fig. 1.1). Then we define the Riemann integral as the limit of the sum

$$\sigma_n = \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \quad \text{with } x_{i-1} \leqslant \xi_i \leqslant x_i$$

as n approaches infinity in such a way that the longest subdivision tends to zero. The present definition is derivable from the previous one. To see this, note that ξ_i is an arbitrary point in the subinterval (x_{i-1}, x_i), which implies that, for all i,

$$L_i \leqslant f(\xi_i) \leqslant U_i$$

since L_i and U_i are the extreme values of the function in the subinterval. Therefore it must be true that

$$s_n \leqslant \sigma_n \leqslant S_n$$

for all values of n. But if s_n and S_n tend to a common limit, then σ_n, being squeezed in between s_n and S_n, must tend to the same limit. So we have that, independently of the mode of subdivision or the choice of points ξ_i, it must be true that

$$\sum_i f(\xi_i)(x_i - x_{i-1}) \rightarrow \int_a^b f(x)\,dx \tag{1.1}$$

if f(x) is Riemann-integrable on (a, b). This result is an immediate consequence of the definition of Riemann integrability. In fact, it is precisely the way in which Riemann originally defined the integral.

Clearly it is very important, if "integral" is to be a useful concept, that the value of the limit shall be independent of the detailed way in which the sum on the left-hand side of Eq. 1.1 was formed. In constructing the sum, we must be able to choose any ξ_i in the interval (x_{i-1}, x_i), and changing our choice must not produce any sensible changes in the value of the limit of the sum. Now, under what conditions can it be true that the limit in Eq. 1.1 will be independent of the choice of the ξ_i?

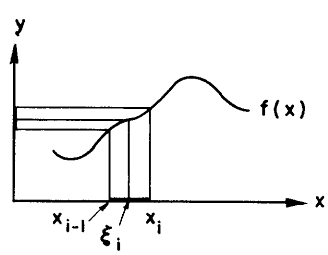

Fig. 1.2. Pertinent to the existence of Riemann integrals.

Following Natanson (2), we see that the limit (Eq. 1.1) can be independent of the choice of ξ_i only if changing ξ_i within the interval (x_{i-1}, x_i) changes the value of $f(\xi_i)$ only imperceptibly. But since any point in (x_{i-1}, x_i) may be chosen as ξ_i, we might ask, What property is common to all points of the subinterval?

The answer is simple: They are close together, since as n is increased, the length of the subinterval tends to zero. Therefore, as shown in Fig. 1.2, our requirements will be satisfied if we stipulate that small intervals along the x-axis correspond to small intervals along the y-axis. If this condition obtains, then it does not matter very much how we pick ξ_i, since any one particular choice may be thought of as yielding a value $f(\xi_i)$ which will be representative of all of the values of the function within the subinterval.

The condition that small x-intervals correspond to small y-intervals is satisfied by continuous functions and is obviously not satisfied by discontinuous functions. And yet, as we have seen, this condition is fundamental in making possible the existence of Riemann integrals. Thus we see why there is an intimate connection between Riemann integrability and continuity, and why there are functions for which the Riemann definition of integration is meaningless.

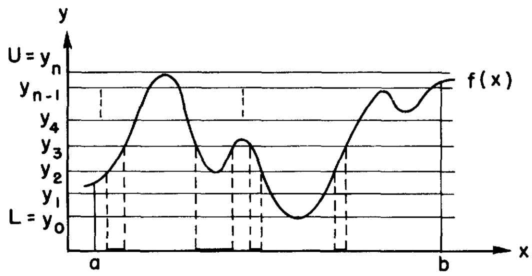

Fig. 1.3. Pertinent to the Lebesgue definition of the integral.

We wish to generalize our concept of "integral" so that it will be meaningful for a class of functions larger than the class of continuous or piecewise continuous functions. As before, we want the net area that is included between a curve and the abscissa, but this time, instead of subdividing the x-axis we shall subdivide the y-axis. As a result of following up this apparently simple change in point of view, we shall end up with a new and very general definition of integration, the Lebesgue integral.

Consider a bounded function f(x) defined on the interval (a, b). Let U be the upper bound of f(x) and L be its lower bound. As in Fig. 1.3, we divide the y-interval (L, U) into n parts:

$$L = y_o \leqslant y_1 \leqslant y_2 \leqslant \dots \leqslant y_{n-1} \leqslant y_n = U$$

Now let E_k be the set of values of x for which

$$y_k \leqslant f(x) \leqslant y_{k+1}$$

One such E_k, for example E_2, is the set formed by all points marked in black along the x-axis of Fig. 1.3. Let $\ell(E_k)$ stand for the total length of the set E_k (that is, the sum of the lengths of the x-intervals for which $y_k \leqslant f(x) \leqslant y_{k+1}$). Then, if we form the sum

$$\sum_{k=1}^{n} \eta_k \,\ell(E_k) \quad \text{where } y_k \leqslant \eta_k \leqslant y_{k+1} \tag{1.2}$$

we again have a quantity which approximates the area under f(x). Thus, if an appropriate limit process is performed, the sum (Eq. 1.2) may serve to define the integral of f(x) between a and b.

It is instructive to compare the sum of Eq. 1.2 to that of Eq. 1.1. In fact, the comparison is more striking if we rewrite Eq. 1.1 slightly: Let the set of points contained in the interval (x_{i-1}, x_i) be denoted by E_i, and write $(x_i - x_{i-1})$ as $\ell(E_i)$. This notation is then the same as that used in Eq. 1.2. With the new notation, the sum in Eq. 1.1 becomes

$$\sum_{i=1}^{n} f(\xi_i)\, \ell(E_i) \tag{1.3}$$

where ξ_i is a point contained in the set E_i. Now we see that Eq. 1.2 and Eq. 1.3 have exactly the same form. But there is one important difference: in forming Eq. 1.3, the points of E_i were chosen by the rather arbitrary criterion that they shall be close together. We then found that we had to place severe restrictions on f(x) to make sure that the various possible $f(\xi_i)$ would also be close together. But in forming Eq. 1.2, the points of E_k were chosen, not because the points themselves happened to be close to each other, but because the <u>values of the function</u> on those points are close to each other. Thus we have no difficulty in picking a representative value of the function on E_k, and we can do this without saying anything at all about the continuity properties of f(x)! And so our, or rather, Lebesgue's, simple change of approach has pointed out a way of defining integration that will be meaningful for a class of functions far larger than the class to which Riemann's definition is appropriate.

Of course, the new point of view is only simple on the surface; before we can use it, we must find a precise concept which will correspond to our vague idea of "length of a set." This problem forms the subject matter of the theory of measure. In the next three sections, we shall discuss the parts of measure theory that are relevant to the theory of integration. Then, having acquired the necessary tools, we shall give a precise definition of Lebesgue integration, and devote the rest of the paper to a study of some of its properties.

2. SIMPLE FUNCTIONS

Measure theory can be, and usually is, developed as an abstract, independent discipline completely divorced from such applications as integration theory. While this approach lends great unity and elegance to the theory, it makes it almost impossible to provide motivation and intuition for the numerous seemingly arbitrary definitions and theorems that arise in the development of the theory. Such an approach is too formal for our present purposes. We are interested in integration, and we shall study measure theory only to obtain answers to questions that arise as we develop the theory of integration.

Before starting the discussion of this section we need to enlarge our catalogue of symbols.

If E is a set (of numbers or objects) we write

$$x \in E$$

to indicate that the number (or object) x belongs to E or is a member of E. For example, if E is the set of points of the real line between zero and one, and we wish to talk about those values of x which satisfy the inequality $0 < x < 1$, we simply say $x \in E$. If x does not belong to E, we write $x \notin E$.

The set of all elements x which have a given property P will be denoted by

$$\{x : P\}$$

For example, the symbol $\{x : 0 \leq x \leq 1\}$ stands for the unit interval; and the symbol $\{x : f(x) < \alpha\}$ stands for the set of values of x for which some given function f(x) has a value less than some given α. The symbol $\{x : f(x) < \alpha\}$ may be read: "the set of all values of x for which $f(x) < \alpha$."

By <u>function</u> we mean any rule for associating a number f(x) with each element x of a given set E. That is, we say "a function f is defined on E" if, with every $x \in E$, there is associated a number f(x). E is called the <u>domain of definition</u> of f; the numbers f(x) are called the <u>values</u> of f. The set of values of f is called the <u>range</u> of f. If the set E is an interval of the real line, this definition reduces to the elementary concept of function.

We shall call E a subset of F if every element of E is also an element of F, and we shall write

$$E \subset F$$

which may be read: "E is contained (or is included) in F." For example, if F is the set of all positive integers and E is the set of all positive even numbers, then $E \subset F$. It is true for any set E that $E \subset E$. If $E \subset F$, and at the same time $F \subset E$, then we write $E = F$. That is, two sets are equal if all the elements of one are contained in the other, and vice-versa. Notice the difference between the meanings of symbols $\in$ and $\subset$: $\subset$ denotes a relation between two things of the same kind (like two sets);

ϵ denotes a relation between an element of one kind and an element of the next higher kind (like a relation between a point and a set of points).

We shall use the notation $\{x_n\}$ to denote the sequence $x_1, x_2, x_3, \ldots$. For example, the sequence of numbers

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{n}, \ldots$$

is represented by the symbol $\left\{\frac{1}{n}\right\}$.

We are now ready to turn to the main topic of this section, and we begin, typically enough, with a definition. Let s be a function defined on a set X. If s has only a finite number of different values, then s is called a simple function. For example, if X is the real line, then the function shown in Fig. 2.2 is a simple function, since it takes on only five different values: $0, a_1, a_2, a_3, a_4$. A particularly useful simple function is the characteristic function, K_E, defined for any set $E \subset X$ by

$$\begin{aligned} K_E(x) &= 1 \quad \text{if } x \in E \\ &= 0 \quad \text{if } x \notin E \end{aligned}$$

For example, if X is the real line, and E is the unit interval, $E = \{x : 0 \leqslant x \leqslant 1\}$, then $K_E(x)$ is the unit pulse shown in Fig. 2.1.

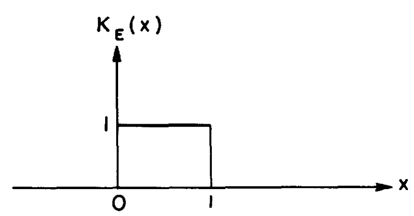

Fig. 2.1. The characteristic function of the unit interval.

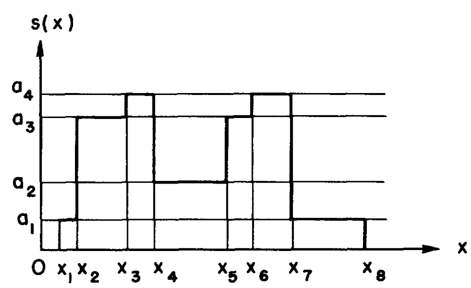

Fig. 2.2. A simple function.

Any simple function can be expressed as a finite linear combination of characteristic functions. For example, if s takes on the values $c_1, c_2, \ldots, c_n$ on the sets $E_1, E_2, \ldots, E_n$, respectively, we can write

$$s(x) = \sum_{i=1}^{n} c_i \, K_{E_i}(x)$$

To verify this, let us evaluate the sum at some particular point x. If x is in E_j, then $K_{E_j} = 1$ and $K_{E_i} = 0$, for all $i \neq j$, so that the sum reduces to the single term c_j, which is precisely the value of $s(x)$ on E_j.

It is easy to see how we should go about defining the integral of a simple function. Consider, for example, the simple function shown in Fig. 2.2. As usual, the Riemann integral of $s(x)$ is given by

$$\int_{x_1}^{x_8} s(x)\,dx = a_1(x_2 - x_1) + a_3(x_3 - x_2) + a_4(x_4 - x_3) + \ldots + a_1(x_8 - x_7)$$

However, an equally natural way of determining the integral would be to add the total length of the abscissa on which $s(x) = a_1$, the total length on which $s(x) = a_2$, and so on, then multiply the lengths by the corresponding heights and add. Suppose $s(x) = a_i$ over a set E_i of length $\ell(E_i)$. Then

$$\begin{aligned}\int_{x_1}^{x_8} s(x)\,dx &= a_1\left[(x_2 - x_1) + (x_8 - x_7)\right] + a_2\left[(x_5 - x_4)\right] + \ldots \\ &= \sum_{i=1}^{4} a_i\, \ell(E_i)\end{aligned} \tag{2.1}$$

Evidently, both methods of integration yield the same answer, since both yield the area between $s(x)$ and the abscissa. But it is worth noting that the result in Eq. 2.1 has the same form as the sum in Eq. 1.2, and that the way of arriving at Eq. 2.1 begins to give expression to our thoughts of Section 1. By analogy with Eq. 2.1, we formulate the following provisional definition of the integral of a simple function: If s is a simple function defined on X and given by

$$s(x) = \sum_{i=1}^{n} c_i\, K_{E_i}(x)$$

then its integral over the whole space X is defined as

$$\int s(x)\,dx = \sum_{i=1}^{n} c_i\, \ell(E_i) \tag{2.2}$$

This definition is provisional because, among other things, we still do not have a clear meaning for the symbol $\ell(E)$. If in the preceding examples its meaning appeared evident, that is only because the examples were so chosen that all sets E turned out to be intervals, so that we had a natural feeling for what their length should be. However, it is very easy to find an example in which our intuition becomes helpless. We need only consider again the function

$$f(x) = 1 \quad \text{if } x \text{ is irrational and } a \leqslant x \leqslant b$$
$$\quad = 0 \quad \text{if } x \text{ is rational and } a \leqslant x \leqslant b$$

In our present terminology, f is nothing more than the characteristic function of the set of irrationals of the interval (a, b). That is, if I is the set of irrationals of (a, b) and R is the set of rationals of the same interval, then f is just the simple function

$$f = 1\, K_I + 0\, K_R = K_I$$

Therefore, by Eq. 2.2, the integral of f over (a, b) is

$$\int f\, dx = 1 \cdot \ell(I) = \ell(I)$$

But now, what is the "length" of the set of irrationals?

The question will be taken up in Section 4. Meanwhile, let us continue our informal discussion and see how the definition of Eq. 2.2 might be extended from simple functions to more arbitrary functions. To achieve this extension, we shall use an important property of simple functions – that any other function can be expressed as the limit of a sequence of simple functions. That is, simple functions can be used to approximate, as closely as desired, the behavior of any given function.

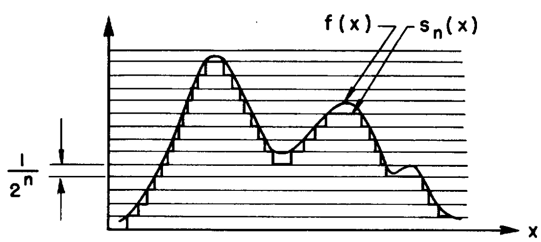

Fig. 2.3. Approximation of f(x) by simple functions.

Suppose that we have an arbitrary function f defined on X. For the moment, we assume that f is nonnegative and bounded (these restrictions will be removed presently). We divide each unit interval of the axis of ordinates into 2^n parts, as is done in Fig. 2.3, where X is taken to be the real line. Let us call the total number of divisions N. Then, over each subset E_{nk} of X such that

$$\frac{k-1}{2^n} \leqslant f(x) < \frac{k}{2^n} \qquad (k = 1, 2, \ldots, N)$$

we define

$$s_n(x) = \frac{k-1}{2^n}$$

the lower value of f on that subset. In other words, we define

$$E_{nk} = \left\{ x : \frac{k-1}{2^n} \leqslant f(x) < \frac{k}{2^n} \right\}$$

and

$$s_n(x) = \sum_{k=1}^{N} \left(\frac{k-1}{2^n}\right) K_{E_{nk}} \tag{2.3}$$

These expressions have a complicated appearance, but their content is simple: they are nothing but an analytical way of representing stepped curves like $s_n(x)$ of Fig. 2.3.

From the way $s_n(x)$ was defined (and as can be seen from Fig. 2.3), it is always true that $s_n(x) \leq f(x)$, that $s_{n+1}(x) \geq s_n(x)$, and that as the number of subdivisions increases ($n \to \infty$), $s_n(x) \to f(x)$. In fact, since for bounded functions it is true by definition that

$$|f(x) - s_n(x)| \leq \frac{1}{2^n} \quad \text{for all } x$$

the sequence $\{s_n(x)\}$ converges to $f(x)$ uniformly as n approaches ∞. (The inequality follows, as is seen in Fig. 2.3, from the fact that $f(x)$ and $s_n(x)$ never are farther apart then the width of one subdivision, which is $1/2^n$.)

We shall now remove the restrictions that $f(x)$ be positive and bounded. If $f(x)$ is positive but not bounded, we modify the definition of Eq. 2.3 as follows. For each n, let F_n be the set of values of x on which $f(x) \geq n$, that is,

$$F_n = \{x : f(x) \geq n\}$$

Then, on those points for which $f(x) < n$, we form the sum (Eq. 2.3) as before, and for $x \in F_n$ we simply set $s_n(x) = n$. Since there are 2^n divisions in each unit of the ordinate axis, in n units there will be $n2^n$ divisions, and Eq. 2.3 becomes

$$s_n(x) = \sum_{k=1}^{n2^n} \left(\frac{k-1}{2^n}\right) K_{E_{nk}} + nK_{F_n} \tag{2.4}$$

As n approaches infinity the sequence $\{s_n(x)\}$ defined in Eq. 2.4 still converges to $f(x)$, although now the convergence is no longer uniform.

If $f(x)$ assumes both positive and negative values, then we form the positive and negative parts of $f(x)$, defined as

$$f^+(x) = \begin{cases} f(x) & \text{when } f(x) \geq 0 \\ 0 & \text{when } f(x) < 0 \end{cases}$$
$$f^-(x) = \begin{cases} -f(x) & \text{when } f(x) \leq 0 \\ 0 & \text{when } f(x) > 0 \end{cases} \tag{2.5}$$

Then $f = f^+ - f^-$, where both f^+ and f^- are always positive, and we can apply the construction (Eq. 2.4) to f^+ and f^- separately.

So we have the result that any function can be represented as the limit of a sequence of simple functions. We shall use this property later in arriving at a precise definition of integration, but we use it now in continuing our informal discussion of how we should go about defining integration.

We have already seen (in Eq. 2.2) how we can define the integral of a simple function. We have also seen that we can use simple functions to approximate any other function as closely as we wish. It seems natural, then, to define the integral of an arbitrary function f in terms of the integrals of a sequence of simple functions tending to f. And so we formulate the following (provisional) definition:

If f is a function defined on X, and if $\{s_n\}$ is a sequence of simple functions defined on X and such that

$$\lim_{n\to\infty} s_n(x) = f(x)$$

then the integral of f over the whole space X is defined to be

$$\int f(x)\,dx = \lim_{n\to\infty} \int s_n(x)\,dx \tag{2.6}$$

Note that we already know the meaning of the symbol on the right-hand side of Eq. 2.6; it is merely the integral of a simple function. We have succeeded in expressing the integral of an arbitrary function in terms of integrals of simple functions.

It is instructive to write Eq. 2.6 in greater detail. If f is any function defined on X, then, according to the procedure of Eq. 2.5 for approximating it with simple functions, we define

$$E_{nk} = \left\{ x : \frac{k-1}{2^n} \leqslant f(x) < \frac{k}{2^n} \right\}$$

and

$$s_n(x) = \sum_{k=1}^{N} \left(\frac{k-1}{2^n}\right) K_{E_{nk}}$$

The quantity $(k-1)/2^n$ is just the lower value of f(x) on the set E_{nk}. Let us abbreviate it by writing

$$\eta_{nk} = \left(\frac{k-1}{2^n}\right)$$

Then we have, as in Eq. 2.2,

$$\int s_n(x)\,dx = \sum_{k=1}^{N} \eta_{nk}\, \ell(E_{nk})$$

and so the defining Eq. 2.6 becomes

$$\int f(x)\, dx = \lim_{n \to \infty} \sum_{k=1}^{N} \eta_{nk}\, \ell(E_{nk}) \tag{2.7}$$

In the latter part of Section 1 we discussed how an integral should be defined so that it would be more generally useful than the Riemann integral. Equation 1.2 was our first tentative formulation of the results of that discussion. Now if we compare Eqs. 2.7 and 1.2, we find that they are identical in form and content. Thus we find that our work in this section has resulted in an analytical embodiment of our ideas of Section 1.

In Section 5 we shall give the precise form of definition 2.7. Meanwhile, now that we know how to proceed in order to arrive at a definition of the integral, it is time that we go back and give consideration to some fundamental problems which form the essential basis of our new point of view. More specifically, definition 2.7 makes it quite clear that the success of the method depends almost exclusively on our being able to assign some definite meaning to the symbol $\ell(E)$, the "length" of a set. This is the measure-theoretic problem, which we treat in Sections 3 and 4.

In our discussion of measure theory we shall need some of the elements of set algebra; they are summarized in Appendix I.

3. MEASURE THEORY

We begin this section with a discussion of terminology. In what follows, the words "class," "collection," and "aggregate," which are all synonymous with "set," will be used interchangeably to denote a set of objects. For example, we shall speak of a "class $\mathscr{E}$ of sets E"; this means "a set $\mathscr{E}$ each element of which is itself a set E." A set of sets will be denoted by a script capital letter (such as $\mathscr{E}$), and we shall continue to use Roman capitals for sets of points and Roman lower-case letters for points. The only object in using synonyms for the word set is to avoid or mitigate such syntactical obstacle courses as "the set of all sets which are subsets of a set E." In particular, we shall always call a set of sets a class.

To be consistent with the definitions given in the last section, the inclusion relations $\in$ and $\subset$ must be used as follows: A point may belong to a set ($x \in E$), a set may be contained in another set ($E \subset F$) or it may belong to a class of sets ($E \in \mathscr{E}$); or a class of sets may be contained in another class of sets ($\mathscr{E} \subset \mathscr{F}$). However, the symbols $x \subset E$ or $E \subset \mathscr{E}$ are meaningless.

The <u>empty set</u> (or <u>vacuous set</u>) is the set that contains no element. It will be denoted by 0. If a set has at least one element, it is called nonempty (or nonvacuous).

In this section we are interested in finding a meaning for the symbol $\ell(E)$ or, more generally, in finding a way of associating a number with each set of a given collection. Therefore, the first new idea we shall need is that of a set function. Suppose that we have a collection $\mathscr{E}$ of sets E. We say that ϕ is a <u>set function</u> defined on $\mathscr{E}$ if ϕ assigns to each $E \in \mathscr{E}$ a real number $\phi(E)$. While an ordinary function (a point function) has a set of points as its domain of definition, the domain of a set function is a collection of sets. To emphasize the difference between point functions and set functions, we shall denote point functions by lower case Roman letters (f, g, etc.), and shall denote set functions by lower case Greek letters (μ, ν, etc.).

A large part of measure theory consists of answering or elaborating two questions:

I. With what properties should we endow a set function so that it corresponds to a useful $\ell(E)$ function?

II. Over what domain can or should such a function be defined?

The first question can be answered rather simply. We want a set function $\phi(E)$ which is as unrestricted as possible and yet one which, for the special case in which the sets E are subsets of the real line, can have the ordinary properties that we associate with length. An obvious requirement is that $\phi(E) \geqslant 0$, since lengths are always nonnegative. Another requirement is that the length of "nothing" shall be zero; i.e., that for the empty set 0, $\phi(0) = 0$. A third requirement is that the length of the "sum" of two line segments shall equal the sum of their individual lengths. This last statement is formalized by saying that if ϕ is defined on the class $\mathscr{E}$, if A and B are any two sets in $\mathscr{E}$ ($A, B \in \mathscr{E}$), if A and B do not overlap ($A \cap B = 0$) and if the "sum" or union of A and B is also contained in $\mathscr{E}$ ($(A \cup B) \in \mathscr{E}$), then

$$\phi(A \cup B) = \phi(A) + \phi(B)$$

As will be shown later, any set function that satisfies this equation also satisfies a similar equation in which sets A and B are replaced by any finite number of sets. Such set functions are said to be finitely additive. Our ordinary conception of the length of intervals of the real line is a finitely additive set function; in fact, it is more than that, because length is additive even when the number of sets is countably infinite, as can be seen from the following example.

Let $\{E_n\}$ be the sequence of disjoint intervals of the real line defined by

$$E_n = \left\{x : \frac{1}{2^n} \leqslant x < \frac{1}{2^{n-1}}\right\} \qquad (n = 1, 2, 3, \ldots)$$

The union of all the E_n forms the unit interval. That is

$$\bigcup_{n=1}^{\infty} E_n = \{x : 0 \leqslant x < 1\}$$

and the length of the unit interval is one. On the other hand, the length of each interval E_n is $\frac{1}{2^{n-1}} - \frac{1}{2^n} = \frac{1}{2^n}$, and therefore the total length of all the E_n is

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$$

Therefore, for this example we have shown that

$$\text{Length of}\left[\bigcup_{n=1}^{\infty} E_n\right] = \sum_{n=1}^{\infty} [\text{length of } E_n] = 1$$

Since, as we said before, we want our $\ell(E)$ functions to have the ordinary properties of length for the special case in which E is a subset of the real line, we must require that $\ell(E)$ have the property illustrated in the example. Stated formally, to make ϕ a suitable "length" function we require that, if ϕ is a set function defined on a class $\mathscr{E}$, and if $\{E_n\}$ is any sequence of pairwise disjoint sets belonging to $\mathscr{E}$ and is such that

$$\left(\bigcup_{n=1}^{\infty} E_n\right) \in \mathscr{E}$$

then

$$\phi\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \phi(E_n)$$

A set function that satisfies this last requirement is said to be countably additive (or completely additive).

Thus far we have been considering some of the properties that a set function ϕ must have in order to be useful as an $\ell(E)$ function. The conclusion of this discussion will be an exact definition of the concept of measure, but first we must consider in a little more detail the classes of sets that are the domains of our set functions.

In defining finite and countable additivity, we made statements of the form "... if $A, B \in \mathscr{E}$ and if $(A \cup B) \in \mathscr{E}$, then ... ." The reason for specifying that $(A \cup B) \in \mathscr{E}$ is simple: The set function ϕ is defined only for sets belonging to class $\mathscr{E}$. Therefore, we can talk about $\phi(A \cup B)$ only if we know that $(A \cup B) \in \mathscr{E}$. Otherwise, the symbol $\phi(A \cup B)$ is meaningless. We had to stipulate explicitly that $(A \cup B) \in \mathscr{E}$ because it is not obvious that the statement "A and B belong to $\mathscr{E}$" necessarily implies "$A \cup B$ also belongs to $\mathscr{E}$." In fact, the following example shows that this is not true in general.

EXAMPLE 3.1. Let $\mathscr{E}$ be the class of intervals of the real line of the form $a \leqslant x \leqslant b$. Let

$$A = \{x : 0 \leqslant x \leqslant 1\}, \qquad B = \{x : 9 \leqslant x \leqslant 10\}$$

Then $A \in \mathscr{E}$, $B \in \mathscr{E}$, but

$$A \cup B = \{x : 0 \leqslant x \leqslant 1 \quad \text{or} \quad 9 \leqslant x \leqslant 10\}$$

is not an interval (since it cannot be expressed in the form $a \leqslant x \leqslant b$) and so $A \cup B$ does not belong to class $\mathscr{E}$.

Since most of our interest in set functions is centered on "length" functions, for which combinations of sets like

$$A \cup B, \ A - B, \ \bigcup_{n=1}^{N} E_n, \ \text{etc.}$$

are important, let us restrict ourselves to classes of sets for which such combinations automatically belong to the class. It is convenient to give such classes definite names, and we proceed to do this in the following definitions.

A class $\mathscr{R}$ of sets is called a ring if $A \in \mathscr{R}$ and $B \in \mathscr{R}$ implies $(A \cup B) \in \mathscr{R}$ and $(A - B) \in \mathscr{R}$. From the definition, we can immediately derive some simple properties of rings.

1. The empty set 0 belongs to every ring. Proof: Let $\mathscr{R}$ be a ring of sets and A be a set belonging to $\mathscr{R}$. Then $A - A = 0$ also belongs to $\mathscr{R}$.
2. The intersection of two sets of a ring belongs to the ring. Proof: if $\mathscr{R}$ is a ring and $A \in \mathscr{R}$, $B \in \mathscr{R}$, then $(A - B) \in \mathscr{R}$ and from the identity $A \cap B = A - (A - B)$ it follows that $(A \cap B) \in \mathscr{R}$. (See Appendix I for proof of the identity.)
3. If $E_1, E_2, \ldots, E_N$ are any N sets belonging to a ring $\mathscr{R}$, then their union

$$\bigcup_{n=1}^{N} E_n$$

also belongs to $\mathscr{R}$. Proof: we simply apply the definition of ring to two sets at a time:

since $E_1 \in \mathscr{R}$ and $E_2 \in \mathscr{R}$, it follows that $E_1 \cup E_2 \in \mathscr{R}$

since $E_1 \cup E_2 \in \mathscr{R}$ and $E_3 \in \mathscr{R}$, it follows that $E_1 \cup E_2 \cup E_3 \in \mathscr{R}$

and so on up to E_n.

The two following examples of classes that are rings may help to clarify the definitions.

EXAMPLE 3.2. Let us call a set <u>finite</u> if the set is empty or contains a finite number of points. Then, given any set X, the class $\mathscr{F}$ of all finite subsets of X is a ring.

PROOF. If A and B are any two sets of the class $\mathscr{F}$, A and B are finite by definition, and their union and difference are certainly finite also. Therefore, $A \cup B$ and $A - B$ belong to $\mathscr{F}$, which proves that $\mathscr{F}$ is a ring.

EXAMPLE 3.3. We have already seen from Example 3.1 that the class of intervals of the real line of the form $a \leqslant x \leqslant b$ is not a ring, since the union of two sets does not necessarily belong to the class. The present example will show how a ring may be built up out of intervals of the real line. First, let us change our definition of interval to mean a set of the form

$$I = \{x : a < x < b\}$$

where a and b are any two numbers and where either one or both of the $<$ signs may be replaced with $\leqslant$ signs. [In particular, a and b may be the same number, in which case $I = \{x : a < x < a\}$ = empty set = $\{x : a \leqslant x < a\} = \{x : a < x \leqslant a\}$; and $\{x : a \leqslant x \leqslant a\}$ is the one-point set containing just the point a.] Now let $\mathscr{E}$ be the class of all sets which are finite unions of intervals; that is, $\mathscr{E}$ is the class of all sets E which can be expressed in the form,

$$E = \bigcup_{n=1}^{N} I_n$$

Then $\mathscr{E}$ is a ring.

PROOF.

1. The union of two sets of $\mathscr{E}$ belongs to $\mathscr{E}$: Let A and B be two sets belonging to $\mathscr{E}$. Then, by definition of the class $\mathscr{E}$, A and B can be written

$$A = \cup_i I_i \qquad B = \cup_j I_j$$

Therefore $A \cup B$ can be written as a finite union of intervals, whence $A \cup B \in \mathscr{E}$.

2. The difference of two sets of $\mathscr{E}$ belongs to $\mathscr{E}$: Let A and B be defined as above. Then $(A - B)$ can have three possible forms; if A and B have no points in common, $A - B = A$, which belongs to $\mathscr{E}$; if A is contained in B, $A - B = 0$, which belongs to $\mathscr{E}$; if A and B overlap partially, then $A - B$ is a finite union of sets of the form $I_i - I_j$, and this difference yields either one or two intervals. Therefore $A - B$ is still a finite union of intervals and thus belongs to $\mathscr{E}$.

Since unions and differences of sets of $\mathscr{E}$ belong to $\mathscr{E}$, we have proved that $\mathscr{E}$ is a ring.

The reader will have noticed that nothing was said about countable unions in our discussion of rings. We talked only about finite unions. If $E_1, E_2, E_3, \ldots$ is a countable sequence of sets in the ring $\mathscr{R}$, it is <u>not</u> true in general that

$$\bigcup_{n=1}^{\infty} E_n$$

is also in $\mathscr{R}$, as the following example shows.

EXAMPLE 3.4. We can see very easily that the ring $\mathscr{F}$ defined in Example 3.2 does not, in general, contain infinite unions. Suppose that the set X consists of an infinite number of points. Then, if we pick $E_1, E_2, \ldots,$ to be one-point sets all different from each other, the set defined by

$$\bigcup_{n=1}^{\infty} E_n$$

contains an infinite number of points and so does not belong to $\mathscr{F}$.

We can demonstrate the same thing for the ring $\mathscr{E}$ defined in Example 3.3. There, we can pick the one-point sets

$$E_n = \left\{ x : \frac{1}{n} \leqslant x \leqslant \frac{1}{n} \right\} \qquad (n = 1, 2, \ldots)$$

Each E_n belongs to $\mathscr{E}$, and the set E defined by

$$E = \bigcup_{n=1}^{\infty} E_n$$

consists of the points $1, 1/2, 1/3, \ldots$. But it is impossible to construct a finite union of intervals which will equal E. Therefore E does not belong to the ring.

We now define a special type of ring for which it will be true that a countable union of sets of the class will also belong to the class. A class $\mathcal{S}$ of sets is called a <u>σ-ring</u> if $\mathcal{S}$ is a ring and if for any infinite sequence $E_1, E_2, \ldots$ of sets of $\mathcal{S}$ it is true that

$$\bigcup_{n=1}^{\infty} E_n$$

also belongs to $\mathcal{S}$. The σ-rings are important because, as we shall see, they are the natural domains for our "length" functions.

A simple example of a σ-ring is the class of all sets which are subsets of some given set A. To prove that such a class is a σ-ring, we need only note that 1) any union (finite or countable) of subsets of A will still be a subset of A and hence will belong to the class, and 2) the difference of any two subsets of A is still a subset of A and thus belongs to the class. Since the class contains countable unions and differences of its members, it is a σ-ring, as was asserted.

As a second illustration, we shall show how the class $\mathcal{F}$ defined in Example 3.2 can be made into a σ-ring. Using the same notation as in that example, we see that if the set X is itself a finite set, then $\mathcal{F}$ is a σ-ring. Proof: We have already seen that $\mathcal{F}$ is a ring, no matter how many points X has. Therefore, we need only show that when X is finite, any countable union of sets of $\mathcal{F}$ is finite. But this follows immediately from the fact that such a union will be a subset of X and the greatest number of different points that any subset of X can have is finite. Therefore every such union belongs to $\mathcal{F}$, whence $\mathcal{F}$ is a σ-ring.

In Section 4 we shall discuss σ-rings further and become more familiar with them. For the present, we shall list two important properties of σ-rings:

1. It follows from the definition that every σ-ring is a ring and therefore has all the properties of a ring. But Example 3.4 shows that the converse is not true: a ring need not be a σ-ring.
2. If $\mathcal{S}$ is a σ-ring, and $E_1, E_2, \ldots$ is a countable sequence of sets belonging to $\mathcal{S}$, then

   $$\bigcap_{n=1}^{\infty} E_n$$

   also belongs to $\mathcal{S}$. To prove this, we need the identity

   $$\bigcap_{n=1}^{\infty} E_n = E_1 - \bigcup_{n=1}^{\infty} (E_1 - E_n)$$

   which is proved in Appendix I. Since $E_1 \in \mathcal{S}$ and $E_n \in \mathcal{S}$ for all n, $E_1 - E_n \in \mathcal{S}$, therefore the union belongs to $\mathcal{S}$, therefore E_1 minus the union belongs to $\mathcal{S}$, therefore the countable intersection belongs to $\mathcal{S}$.

We can now return to our discussion of "length" functions. In the first few pages of this section we arrived at some properties that a set function must have if it is to be useful as a "length" function. The properties listed there turn out to be sufficient, and we can proceed immediately to the definition of the rigorous counterpart of our $\ell(E)$ functions.

A measure is an extended real-valued, nonnegative and countably additive set function μ, defined on a ring $\mathscr{R}$, and such that $\mu(0) = 0$.

The following are some examples of measure functions:

EXAMPLE 3.5.

a) On the ring $\mathscr{F}$ defined in Example 3.2, let $\mu(E)$ be the number of points contained in the set E. That is, if $E \in \mathscr{F}$ and there are n points in E, we define $\mu(E) = n$.

b) On the ring $\mathscr{E}$ defined in Example 3.3, let $\mu(I)$ be defined as follows: if $I = \{x : a \leqslant x \leqslant b\}$, where, as before, either one or both of the end points of the interval may be missing, let $\mu(I) = (b-a)$. Then, by definition of the class $\mathscr{E}$, any set $E \in \mathscr{E}$ may be written

$$E = \bigcup_{n=1}^{N} I_n$$

where the I_n are chosen to be pairwise disjoint. To define the measure of E, we set

$$\mu(E) = \sum_{n=1}^{N} \mu(I_n) = \sum_{n=1}^{N} (b_n - a_n)$$

The set function μ defined in this example is the most important practical example of a measure. It is called the Lebesgue measure on the real line, and corresponds exactly to the ordinary idea of length. If we define a ring analogous to $\mathscr{E}$ for two-dimensional intervals, or three-dimensional intervals, we can, in an entirely similar way, define the Lebesgue measure in two dimensions, corresponding to area, in three dimensions, corresponding to volume, and so on.

c) Suppose f is any continuous, monotonically nondecreasing function defined on the real line. For any interval I of the real line, instead of writing $\mu(I) = b - a$ as in b), we can define

$$\mu_f(I) = f(b) - f(a)$$

and similarly, for any set E belonging to the class $\mathscr{E}$ used in b), we can define

$$\mu_f(E) = \sum_{n=1}^{N} \mu_f(I_n) = \sum_{n=1}^{N} [f(b_n) - f(a_n)]$$

The set function defined in this example is the most natural generalization of the Lebesgue measure. It is called the Lebesgue-Stieltjes measure induced by f.

Considering the wide range of possible choices of f, this measure provides great flexibility in adapting measure theory to a physical problem. For example, suppose that we have a nonuniform mass distribution along the real line. Let $f(x)$ be the total mass on the line between $-\infty$ and x. Then the Lebesgue-Stieltjes measure induced by f "weights" the measure of an interval I in accordance with how much mass lies on I, since if $I = \{x : a < x < b\}$, then $\mu_f(I) = f(b) - f(a)$ = mass between a and b. In particular, if f increases linearly with x (the mass is distributed uniformly along the real line), then μ_f reduces to the ordinary Lebesgue measure of the line.

d) A particularly interesting weighting of the measure of sets is the one which assigns to every set E of some suitable ring $\mathscr{R}$, the probability that a given physical experiment results in a number x belonging to E. Here every set $E \in \mathscr{R}$ represents an event (the points of E representing the various possible ways in which the event may occur) and we choose $\mu(E)$ so that it equals our physical idea of the probability of occurrence of the event represented by E. In effect, the probability measure is a special case of the Lebesgue-Stieltjes measure in which the inducing function f (see part c of this example) is the probability distribution of the process. [For a simple and complete description of the generation of probability measure, see Kolmogoroff (7), Chap. I, or Halmos (1), Chap. IX.]

The reader will perhaps have seen from these examples how general and flexible is the concept of measure. We conclude this section by deriving some simple and useful properties common to all measure functions.

1. By definition, all measures are countably additive set functions. This implies immediately that all measures are also finitely additive. As a result, if $\mathscr{R}$ is a ring, μ a measure defined on $\mathscr{R}$, A and B two sets belonging to $\mathscr{R}$, and if $A \cap B = 0$, then

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

2. A measure is monotone. That is, if a set A is contained in a set B, the measure of A is less than or equal to the measure of B. Stated formally, if $\mathscr{R}$ is a ring, μ a measure on $\mathscr{R}$, A and B two sets belonging to $\mathscr{R}$, and if $A \subset B$, then

$$\mu(A) \leqslant \mu(B)$$

Proof: Since $A \in \mathscr{R}$ and $B \in \mathscr{R}$, $B - A \in \mathscr{R}$. Also, since $A \subset B$, $B = A \cup (B-A)$. This last identity expresses B as the union of two disjoint sets belonging to $\mathscr{R}$. Therefore, by property 1,

$$\mu(B) = \mu(A) + \mu(B-A) \tag{3.1}$$

But since all measures are nonnegative by definition, $\mu(B-A) \geqslant 0$. Therefore $\mu(B) \geqslant \mu(A)$, as was asserted.

3. We notice from Eq. 3.1 that if $\mu(A)$ is finite, it may be subtracted from both sides of the equation, thus proving that if $\mathscr{R}$ is a ring, μ a measure on $\mathscr{R}$, $A \in \mathscr{R}$, $B \in \mathscr{R}$, $A \subset B$ and $\mu(A)$ is finite, then

$$\mu(B-A) = \mu(B) - \mu(A)$$

This last property is described by saying that a measure is <u>subtractive</u>. (We had to stipulate that $\mu(A)$ be finite because otherwise our proof might have involved a symbol of the form $\infty - \infty$, which is meaningless.)

4. If $\mathscr{R}$ is a ring, μ a measure on $\mathscr{R}$, and A and B are two sets of $\mathscr{R}$, then

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

(In the special case when A and B are disjoint, so that $A \cap B = 0$, this relation reduces to property 1, since $\mu(0) = 0$, be definition.) Proof: We can write $A \cup B$ as a union of disjoint sets belonging to $\mathscr{R}$.

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

By using property 1 we obtain

$$\mu(A \cup B) = \mu(A-B) + \mu(B) \tag{3.2}$$

Now we can write $A - B = A - (A \cap B)$, and we note that $A \cap B \in \mathscr{R}$ and that $(A \cap B) \subset A$. Therefore, since by property 3 measures are subtractive [provided $\mu(A \cap B) < \infty$],

$$\mu(A-B) = \mu(A) - \mu(A \cap B)$$

Substituting this result in Eq. 3.2 yields

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \tag{3.3}$$

as asserted. The truth of the identities used in the proof is made evident by the use of the circle diagram described in Appendix I. In fact, from Fig. 3.1 it is clear that if we write $\mu(A \cup B) = \mu(A) + \mu(B)$, then we have counted twice the points which are common to A and B. Therefore, to get the right answer we have to subtract once the measure of the points common to A and B, that is, $\mu(A \cap B)$. This is just what is stated in Eq. 3.3.

This concludes our discussion of the concept of measure per se. In the next section we shall consider the domains on which measures are defined and the relation of these ideas to our original problem, which, as the reader may remember, was integration.

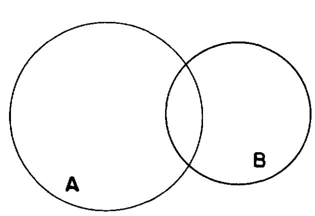

Fig. 3.1. Pertinent to deriving the measure of the union of two sets.

4. MORE MEASURE THEORY

We started Section 3 by asking two questions:

I - What properties should a set function have in order to be a suitable measure function?

II - On what domain should a measure be defined?

Question I was answered in the last section; we now proceed to answer question II. Despite appearances, the reason for considering this question is not idle mathematical pedantry. As will be shown presently, the domain of a measure is the most important single factor that determines the range of applicability of the new integral that we wish to define.

In the definition of measure given in Section 3, the domain of the measure was specified to be a ring. Let us forget this requirement for the moment, and just think of a measure as being defined on some undescribed class $\mathcal{S}$ of sets. We shall let the present discussion discover for us what sort of a class $\mathcal{S}$ should be. As a matter of terminology, a set E will be said to be measurable if it belongs to the domain of definition of a measure. In other words, E is measurable if and only if $E \in \mathcal{S}$. Our question is, What is a useful class of measurable sets?

Whether or not a given domain will be useful depends on what measures will be used for. This brings us back to our discussion of integration, since integration provided the original motive for our interest in measure theory. In Section 2 it was shown that any function f can be approximated by a convergent sequence $\{s_n(x)\}$ of simple functions, where

$$s_n(x) = \sum_{k=1}^{N} \left(\frac{k-1}{2^n}\right) K_{E_{nk}} \tag{4.1}$$

and

$$E_{nk} = \left\{ x : \frac{k-1}{2^n} \leqslant f(x) < \frac{k}{2^n} \right\} \tag{4.2}$$

Tentatively, the integral of f was defined in Eq. 2.7 to be

$$\int f\,dx = \lim_{n\to\infty} \int s_n\,dx = \lim_{n\to\infty} \sum_{k=1}^{N} \left(\frac{k-1}{2^n}\right) \mu(E_{nk}) \tag{4.3}$$

but we left suspended the question of whether or not a meaning can be given to $\mu(E_{nk})$. In the language of the present section, the question can be restated: For what functions f are the sets E_{nk} included in the domain of the measure μ? The importance of this question is clearly visible in Eq. 4.3: if the sets E_{nk} corresponding to some given function f are not within the domain of μ, then the symbol $\mu(E_{nk})$ is meaningless, and the whole definition (Eq. 4.3) becomes meaningless. For such functions, the integral (as we define it above) does not exist. Thus the usefulness of the definition of integration given in Eq. 4.3 depends on our being able to so choose the domain of μ that a large class of functions f will have their sets E_{nk} in the domain of μ. Since the measurability of the sets E_{nk} is the only factor that determines whether or not the integral will have a definite meaning, we see why the choice of a domain for μ is such an important step.

The problem motivates an important new definition: A function f is said to be a measurable function if the set

$$\{x : f(x) < \alpha\} \tag{4.4}$$

is measurable for every real α. The reader is reminded that a set is measurable if and only if it belongs to the domain of a measure.* It can be shown (see, e.g., Rudin (4), p. 200) that measurable functions can also be defined by using any one of the sets

$$\{x : f(x) \geq \alpha\}, \quad \{x : f(x) > \alpha\}, \quad \{x : f(x) \leq \alpha\} \tag{4.5}$$

instead of the set of Eq. 4.4, without changing the meaning of the definition in any way.

We can determine immediately the relationship between the measurability of f and the measurability of the sets E_{nk} of Eq. 4.3. Suppose that we have a measure μ defined on a class $\mathcal{S}$ of sets. While we have thus far placed no restrictions on the nature of $\mathcal{S}$, let us now require that the intersection of any two sets of $\mathcal{S}$ be a set of $\mathcal{S}$. That is, we stipulate that if $A \in \mathcal{S}$, and $B \in \mathcal{S}$, then

$$(A \cap B) \in \mathcal{S} \tag{4.6}$$

Then we can show that if f is a measurable function, the set

$$E_{nk} = \left\{x : \frac{k-1}{2^n} \leq f < \frac{k}{2^n}\right\}$$

is a measurable set. In fact, this follows from the simple identity

$$E_{nk} = \left\{x : \frac{k-1}{2^n} \leq f < \frac{k}{2^n}\right\} = \left\{x : f \geq \frac{k-1}{2^n}\right\} \cap \left\{x : f < \frac{k}{2^n}\right\} \tag{4.7}$$

By definition of measurability of f, both sets in this intersection are measurable, whence, from Eq. 4.6, their intersection is measurable, as was asserted.

To summarize: Equation 4.3 shows that the existence of the integral** of f depends

---

*It should be emphasized that measurability of a set (or of a function) has nothing whatever to do with the functional nature of the measure being used in a particular problem. Measurability merely describes whether or not the set in question belongs to the domain of the measure being used. It has nothing to do with the measure itself.

**The definition of integration (Eq. 4.3) is not the final, rigorous one; it will be given in Section 5. However, Eq. 4.3 has all of the essential features of the final definition, and, while it needs some qualification (which will be added in Sec. 5), it is sufficiently good for our present purposes.

only on the measurability of the sets E_{nk}; Eq. 4.7 shows that the sets E_{nk} are measurable if f is a measurable function. In effect, therefore, f is integrable if f is a measurable function.

Our argument has demonstrated two things: 1) That the idea of measurable function is a central concept in the theory of integration, and 2) that in order to make a large class of functions integrable, we must make a large class of functions measurable. Thus our criterion in choosing an appropriate domain for measure functions must be that a large class of functions shall be measurable with respect to the chosen domain.

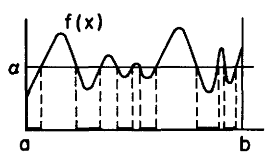

Fig. 4.1. The points belonging to the set $\{x : f(x) < \alpha\}$ are shown in black.

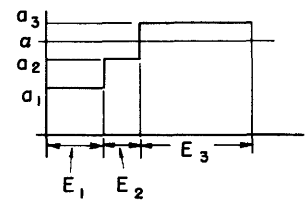

Fig. 4.2. Illustration of the test of measurability of a simple function.

To gain some intuitive feeling for what is involved in measurability, consider Fig. 4.1. As is shown there, the set $\{x : f(x) < \alpha\}$, which tests the measurability of f, will, in general, consist of a conglomeration of subsets. It follows that the more types of conglomerations that can be included in the domain of μ, the more varied will be the class of μ-measurable functions. In other words, there is an intimate connection between the richness of variety of the types of sets included in the domain of a measure, and the variety of functions f which are measurable with respect to that domain.

There is no unique procedure for deciding what to use as a domain of μ. We shall arrive at our answer by considering two examples at the beginnings of which we make certain requirements, and then ask what sort of a domain will satisfy our requirements. The derivation of Eq. 4.7 has shown that it is very useful to define measures on a class $\mathcal{S}$ of sets that contains the intersections of sets belonging to $\mathcal{S}$. Our first example shows the same thing with respect to finite unions, and thus, in effect, indicates that $\mathcal{S}$ should be at least a ring.

EXAMPLE 4.1. It seems reasonable to require that a simple function defined on measurable sets be itself a measurable function. Let us see what this implies. Suppose that s(x) is a simple function that takes on n values $a_1, a_2, \ldots, a_n$ on the sets $E_1, E_2, \ldots, E_n$, respectively. With suitable relabeling, we can arrange the values so that $a_1 < a_2 < a_3 < \ldots < a_n$. Let $\mathcal{S}$ be a class on which is defined a measure μ, and let the sets $E_1, E_2, \ldots, E_n$ be measurable. (For a concrete example, see Fig. 4.2.) The question is, Is s(x) a measurable function?

For an affirmative answer, it is necessary that the sets $\{x : s(x) < \alpha\}$ be measurable for every α. Let us list these sets:

$$\begin{array}{lll} \text{For} & \alpha < a_1 & \{x : s(x) < \alpha\} = 0 \\ \text{for} & a_1 < \alpha \leqslant a_2 & \{x : s(x) < \alpha\} = E_1 \\ \text{for} & a_2 < \alpha \leqslant a_3 & \{x : s(x) < \alpha\} = E_1 \cup E_2 \\ \text{for} & a_3 < \alpha \leqslant a_4 & \{x : s(x) < \alpha\} = E_1 \cup E_2 \cup E_3 \\ \cdots & \cdots & \cdots \\ \text{for} & \alpha > a_n & \{x : s(x) < \alpha\} = \bigcup_{i=1}^{n} E_i \end{array}$$

Remembering that by hypothesis the sets $E_1, E_2, \ldots, E_n$ belong to $\mathcal{S}$, the above relations indicate that, to make $s(x)$ measurable, our class must include the empty set and all finite unions of sets belonging to $\mathcal{S}$. Thus the derivation of Eq. 4.7 and this example, taken together, have shown that we want the class $\mathcal{S}$ of measurable sets to include finite unions and intersections of sets belonging to $\mathcal{S}$, and that we want $\mathcal{S}$ to include the empty set. All of these requirements are met if $\mathcal{S}$ is chosen to be a ring.

EXAMPLE 4.2. In this example we attempt to make plausible the notion that the class $\mathcal{S}$ should not only be a ring, but in fact should be a σ-ring. It happens frequently in practical problems that we are given a sequence of integrable functions that converge to a limit, and we wish to know whether or not the limit will also be integrable. This sort of problem will be treated more fully later on. For the present, we need only observe that in order to be integrable the limit function must first be measurable. This leads us to ask, Given a sequence $\{f_n(x)\}$ of measurable functions and given $\lim f_n(x) = f(x)$, will $f(x)$ also be measurable? Clearly, it would be very desirable to be able to give an affirmative answer. Let us see what such an answer implies for the following special case.

Let $\{f_n(x)\}$ be a convergent sequence of measurable functions defined on a set E. Let $\mathcal{S}$ be a class of sets (which contains E) on which is defined a measure μ. Let the sequence $\{f_n\}$ be monotone nondecreasing, that is, let $f_1(x) \leqslant f_2(x) \leqslant f_3(x) \leqslant \ldots$, and define $f(x) = \lim f_n(x)$. Our question is, Is $f(x)$ measurable? For an affirmative answer, it must be true for every real α that $\{x : f(x) > \alpha\}$ is a measurable set. Consider the relation

$$\{x : f(x) > \alpha\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) > \alpha\} \tag{4.8}$$

which is proved in Appendix I. Since $f_n(x)$ is measurable for every n (by hypothesis), each of the sets in the countable union is measurable. That is, for all n,

$$\{x : f_n(x) > a\} \in \mathcal{S}$$

Now if the class $\mathcal{S}$ is such that every countable union of sets of $\mathcal{S}$ also belongs to $\mathcal{S}$, then $\{x : f(x) > a\}$ belongs to $\mathcal{S}$, and our limit function is measurable. We have already seen in the previous example that $\mathcal{S}$ should be at least a ring. Our present requirement will be satisfied if $\mathcal{S}$ is a σ-ring.

To illustrate what we have accomplished by making our class of measurable sets a σ-ring, we proceed to list some properties of measurable functions the proofs of which are given in Halmos (1), Rudin (4), or Munroe (3). Let $\mathcal{S}$, the class of measurable sets, be a σ-ring, and let f be a measurable function. Then the following functions are all measurable:

1. $af + b$ (a, b are any real numbers)
2. $|f|$
3. f^{α}, $|f|^{\alpha}$ (α is any positive real number)
4. f^{+}, f^{-}
5. $F(f(x))$ (F is any continuous function)

Furthermore, if g is also a measurable function, the following functions are all measurable:

6. $f + g$, fg, $\frac{f}{g}$ ($g \neq 0$)
7.* $\max(f, g)$, $\min(f, g)$

In addition, we have the extremely important property:

8. the limit function of any convergent sequence of measurable functions is measurable.

In other words, if the class of measurable functions is a σ-ring, then measurability is preserved under practically all ordinary processes of analysis. As a result of this circumstance, any function we are likely to meet will almost certainly be measurable. [An exception: It is not true, in general, that a measurable function of a measurable function is measurable. For proof and comment, see Halmos (1), p. 83, or McShane (9), p. 241. This case is rather unusual and we shall ignore it here.]

Can any significant advantage be gained by continuing the process of including increasingly more complicated conglomerations of sets in the class $\mathcal{S}$? The question is imprecise and so does not have a definite answer, but Halmos (ref. 1, Sec. 16) answers it indirectly by showing that it is quite difficult to find nonmeasurable sets when the class of measurable sets is a σ-ring. It follows that, under these conditions, it is equally difficult to find nonmeasurable functions. Thus it appears that a σ-ring

*The expression $\max(f, g)$ means the function $h(x)$ which has the property that, for every value of x, $h(x) = f(x)$ if $f(x) \geqslant g(x)$, $h(x) = g(x)$ if $f(x) < g(x)$. Similarly, for $\min(f, g)$.

is a class of measurable sets that is large enough for all known purposes; that is why σ-rings are usually chosen as the domains of measure functions.

The only problem now is that, while we have just concluded that we want σ-rings as domains for measure functions, in Section 3 measures were defined on rings, not on σ-rings. How do we reconcile the definition of Section 3 with our present requirements? The process that is needed here is somewhat reminiscent of the theory of analytic continuation, whereby the domain of a function of a complex variable is extended (in a nonarbitrary way) from one region of the complex plane to a larger region. For our purposes we shall simply take it for granted that a measure function (defined on a ring) can be extended uniquely and meaningfully to a measure function defined on a σ-ring which contains the ring. In effect, then, we shall proceed as though measure functions had been defined on σ-rings to start with.

However, there are some difficult and important problems associated with the extension of measures from rings to σ-rings. The example that follows illustrates how these problems arise, and at the same time presents some terminology and ideas that occur frequently.

EXAMPLE 4.3. This example is meant to provide an extremely brief glimpse of the problem of extending a measure from a ring to a σ-ring for the special case of the Lebesgue measure of the real line. We shall employ the definitions and notations of Examples 3.3 and 3.5b. There we had the ring $\mathscr{E}$ of all finite unions of intervals, and on $\mathscr{E}$ we defined the measure μ which to every interval assigns its length.

Our first problem is to construct a σ-ring out of the ring $\mathscr{E}$. We achieve this by performing all possible finite or countable unions, intersections, and differences of sets in $\mathscr{E}$. The class of all sets which are reached by performing these operations a finite or countable number of times is called the class $\mathscr{B}$ of Borel sets of the line. It can be verified from the definition that $\mathscr{B}$ is a σ-ring. Suppose, now, that the set A belongs to $\mathscr{B}$ but not to $\mathscr{E}$. Then the symbol $\mu(A)$ is meaningless, since μ is defined for sets of $\mathscr{E}$ only. And yet, it would be desirable to associate a measure with every set of $\mathscr{B}$. Our problem, therefore, is how to extend the measure from μ-defined-on-$\mathscr{E}$ to μ-defined-on-$\mathscr{B}$. To see that we cannot do this in an arbitrary manner, consider two disjoint sets A and B belonging to $\mathscr{B}$ but not to $\mathscr{E}$, but with the property that $A \cup B$ does belong* to $\mathscr{E}$. If μ' is a measure which is the extension of μ to $\mathscr{B}$, we want $\mu'(E)$ to equal $\mu(E)$ for those special sets which belong both to $\mathscr{B}$ and to $\mathscr{E}$. Therefore, in our example, we want $\mu'(A \cup B) = \mu'(A) + \mu'(B) = \mu(A \cup B)$, which will certainly not be generally true if $\mu'(A)$ and $\mu'(B)$ are assigned arbitrarily. How should we proceed? We note first that any set in $\mathscr{B}$ can

*Example to show that this is possible: Let A be the set of rationals of $(0, 1)$, and B the set of irrationals. Then $A \cap B = 0$, $A \cup B = (0, 1)$. A and B belong to $\mathscr{B}$ but not to $\mathscr{E}$; on the other hand, $A \cup B$, being an interval, belongs to $\mathscr{E}$.

be covered by a countable union of sets of $\mathscr{E}$; that is, if $A \in \mathscr{B}$, there exists a sequence $E_1, E_2, E_3, \ldots$ that is such that $A \subset \bigcup_{i=1}^{\infty} E_i$ and $E_i \in \mathscr{E}$ for $i = 1, 2, \ldots$. We define the <u>outer measure</u> of A, $\mu^*(A)$, to be the set function

$$\mu^*(A) = \inf \left[\sum_{i=1}^{\infty} \mu(E_i) \right] \tag{4.9}$$ [†]

with $A \subset \bigcup_{i=1}^{\infty} E_i$ and $E_i \in \mathscr{E}$. That is, we fit the set A with the tightest possible cover of sets of $\mathscr{E}$, where tightest is interpreted to mean, having the least total measure. This least total measure of the covering of A is then the outer measure of A. We have, then, that μ^* is defined on the σ-ring $\mathscr{B}$, and has the property that if $E \in \mathscr{E}$, $\mu^*(E) = \mu(E)$, since the tightest possible cover of E is just E itself. It can be shown that μ^*, as defined, is unique and that it is a countably additive set function. Since it follows from Eq. 4.9 that μ^* is nonnegative, and that $\mu^*(0) = \mu(0) = 0$ (since $0 \in \mathscr{E}$), we have that μ^* is indeed a measure. Thus μ^* is an extension of μ to the class $\mathscr{B}$, and is our desired measure defined on a σ-ring.

As an illustration of Eq. 4.9, we shall determine the measures of the set R of rationals of the unit interval $0 \leq x \leq 1$, and then the measure of the set I of irrationals of the same interval. Our derivation for the rationals is made possible by the fact that there is a countable number of rationals in the unit interval (for proof, see e.g., Rudin (4), p. 23). Let E_n be the interval of length $\epsilon/2^n$ which covers the n^{th} rational number. Then $R \subset \bigcup_{n=1}^{\infty} E_n$, and

$$\mu^*(R) = \inf \left[\sum_{n=1}^{\infty} \mu(E_n) \right] \leq \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon$$

Since this result is true for any value of ϵ, and since μ^* is nonnegative, it follows that $\mu^*(R) = 0$. Thus the set of rationals (or, for that matter, any other countable set) has Lebesgue measure zero. If now we denote the unit interval by E (and note that $E \in \mathscr{E}$), then $E = R \cup I$, the union of two disjoint sets. Thus

$$\mu^*(E) = \mu^*(R) + \mu^*(I) = \mu(E) = 1$$

Since $\mu^*(R) = 0$, $\mu^*(I) = 1$. Thus the set of irrationals of the unit interval has measure one.

[†] The abbreviation "inf" stands for "infimum" (least). Given a set of numbers $\{a_i\}$, the <u>infimum</u> (or greatest lower bound) of the set is the largest number a for which it is true that $a \leq a_i$ for all i. If the set $\{a_i\}$ consists of a finite collection of numbers, the infimum is just the smallest number in the set.

Having extended the measure from $\mathscr{E}$ to $\mathscr{B}$, we denote both μ and μ^* simply by μ, and call μ the <u>Lebesgue</u> <u>measure</u> defined on the σ-ring $\mathscr{B}$ of all Borel sets of the real line. This concludes our example.

We have now collected all of the necessary tools for the study of integration. In the remainder of the work, we shall always start out by assuming (either explicitly or implicitly) that we are given a space X, a σ-ring $\mathcal{S}$ of subsets of X, and a measure μ defined on $\mathcal{S}$. We shall summarize these data with the symbol $(X, \mathcal{S}, \mu)$.

5. INTEGRATION

In the first two sections of this paper we showed, in an intuitive way, how integration should be defined. In Sections 3 and 4 we developed the language and concepts necessary to lend precision to the desired definition. Now we shall bring together the ideas of all the previous sections and proceed to the rigorous definition of the integral and the derivation of some of its properties.

The procedure is the following:

1. Define integration for simple functions.
2. Consider any arbitrary, nonnegative function f as the limit of a sequence of simple functions, and define the integral of f as the limit of the integrals of the simple functions which approximate f.
3. Extend step 2 to functions which are both positive and negative.

In effect, then, the integral of any function is obtained in terms of integrals of simple functions.

We start with a space X, and a σ-ring $\mathcal{S}$ of subsets of X, this σ-ring being the domain on which is defined the set function μ, our measure.

Let s(x) be a measurable simple function defined on X, s(x) having the values $\alpha_1, \alpha_2, \ldots, \alpha_N$ on the sets $E_1, E_2, \ldots, E_N$, respectively. As usual, we represent s(x) by

$$s(x) = \sum_{i=1}^{N} \alpha_i K_{E_i}$$

where K_{E_i} is the characteristic function of the set E_i. The <u>integral</u> of s(x) with respect to μ, over the whole space X, is defined to be

$$\int_X s \, d\mu = \sum_{i=1}^{N} \alpha_i \, \mu(E_i) \tag{5.1}$$

This definition corresponds to our tentative definition 2.2. In particular, if the simple function s is the characteristic function of the measurable set E, $s = K_E$, definition 5.1 yields

$$\int_X s \, d\mu = \int_X K_E \, d\mu = \mu(E)$$

The integral in Eq. 5.1 extends over the whole space X. If we wish it to extend only over the measurable subset E of X, we can achieve this by defining

$$\int_E s \, d\mu = \int_X K_E \, s \, d\mu \tag{5.2}$$

Thus we achieve our purpose by making the integrand zero on all points outside the set E. Since K_E is a measurable simple function, so is sK_E, and therefore we can apply definition 5.1 to determine the right-hand side of Eq. 5.2. It is instructive to work this out a little further. The product sK_E can be written

$$sK_E = \sum_{i=1}^{N} \alpha_i K_{E_i} K_E$$

Noting that for any two sets A and B the product $K_A K_B$ is nonzero only on points belonging to both A and B, we have

$$K_A K_B = K_{A \cap B}$$

so that

$$sK_E = \sum_{i=1}^{N} \alpha_i K_{E \cap E_i}$$

Therefore, definition 5.1 applied to Eq. 5.2 yields

$$\int_E s\, d\mu = \int_X K_E\, s\, d\mu = \sum_{i=1}^{N} \alpha_i\, \mu(E_i \cap E) \tag{5.3}$$

That is, to integrate over a set E we use only those portions of the sets E_i which are contained in E.

Now consider any nonnegative measurable function f defined on X. As was shown in Sections 2 and 4, any such function may be considered as the limit of a sequence $\{s_n\}$ of measurable simple functions. We define the integral of f with respect to μ over the space X by

$$\int_X f\, d\mu = \lim_{n\to\infty} \int_X s_n\, d\mu \tag{5.4}$$

This definition is meaningful because all of the members of the right-hand side can be evaluated by using Eq. 5.1. Note that definition 5.4 is the rigorous counterpart of our earlier tentative definition 2.6.

To extend the definition to an arbitrary measurable function f defined on X, we split the function into positive and negative parts f^+ and f^-, as in Eq. 2.5. Then both f^+ and f^- are nonnegative measurable functions, and

$$f = f^+ - f^-$$

We now define the integral of f with respect to μ to be

$$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu \tag{5.5}$$

Both terms on the right of Eq. 5.5 fit the requirements of definition 5.4.

As before, if E is a measurable subset of X, we define the integral of f with respect to μ over E by

$$\int_E f\,d\mu = \int_X K_E\,f\,d\mu \tag{5.6}$$

[The requirement that E be measurable is necessary because if it should happen that $f = c$, a constant, then $\int_E f\,d\mu = c\int_X K_E\,d\mu = c\,\mu(E)$ which would be meaningless unless E is measurable.]

For the special case when μ is the Lebesgue measure on the real line, the integral of f over E is sometimes denoted by

$$\int_E f\,dx \qquad \text{or by} \qquad \int_a^b f\,dx \tag{5.7}$$

if E happens to be the interval (a, b). The integral is then called the Lebesgue integral of f on E. [This nomenclature is not uniform; some authors call the general integral (Eq. 5.5) the Lebesgue integral with respect to μ and call Eq. 5.7 the Lebesgue integral with respect to Lebesgue measure.]

As a matter of terminology, a measurable function f is said to be integrable (or summable) if $\int f\,d\mu$ is finite. If $\int f\,d\mu$ is infinite, the integral is still defined, but f is not integrable. Only when both $\int f^+\,d\mu$ and $\int f^-\,d\mu$ are infinite is the integral not defined and then because $\int f\,d\mu = \infty - \infty$, which is meaningless. If

$$\int_E f\,d\mu$$

is finite, then we say f is integrable with respect to μ on E, and we abbreviate this by writing the symbol $f \in \mathscr{L}(\mu)$ on E. (The symbol says that f belongs to the class of functions which are integrable with respect to μ on E.)

To illustrate these ideas, we shall use the fundamental definitions of Eqs. 5.4 and 5.1 to determine the integral of a function.

EXAMPLE 5.1. We shall determine the integral with respect to the Lebesgue measure of the function $f(x) = Ax$, on the interval (0, 1). The function is shown in Fig. 5.1.

The first step is to find a sequence $\{s_n\}$ of simple functions whose limit is f(x). To determine the n^{th} member of the sequence, the axis of ordinates is subdivided in strips A/n wide, as shown in Fig. 5.1. Then $s_n(x)$ is given by

$$s_n(x) = \begin{cases} 0 & \text{for all } x \in \left\{x : 0 \leq f(x) < \frac{A}{n}\right\} = \left\{x : 0 \leq x < \frac{1}{n}\right\} = E_{n0} \\ \frac{A}{n} & \text{for all } x \in \left\{x : \frac{A}{n} \leq f(x) < \frac{2A}{n}\right\} = \left\{x : \frac{1}{n} \leq x < \frac{2}{n}\right\} = E_{n1} \\ \frac{2A}{n} & \text{for all } x \in \left\{x : \frac{2A}{n} \leq f(x) < \frac{3A}{n}\right\} = \left\{x : \frac{2}{n} \leq x < \frac{3}{n}\right\} = E_{n2} \\ \cdots & \cdots \\ \frac{mA}{n} & \text{for all } x \in \left\{x : \frac{mA}{n} \leq f(x) < \frac{(m+1)A}{n}\right\} = \left\{x : \frac{m}{n} \leq x < \frac{m+1}{n}\right\} = E_{nm} \\ \cdots & \cdots \\ & (0 \leq m \leq n-1) \end{cases}$$

Since the measure of every set E_{nm} is just $1/n$, a constant, we have, using definition 5.1, that the integral of the n^{th} simple function is

$$\int_0^1 s_n(x)\,dx = \sum_{m=0}^{n-1} \frac{mA}{n}\,\mu(E_{nm}) = \sum_{m=0}^{n-1} \frac{mA}{n} \cdot \frac{1}{n}$$

$$= \frac{A}{n^2} \sum_{m=0}^{n-1} m = \frac{A}{n^2}\left[\frac{(n-1)(n)}{2}\right] = \frac{A}{2}\left(\frac{n-1}{n}\right)$$

And now, from definition 5.4, the Lebesgue integral of $f(x)$ is

$$\int_0^1 f(x)\,dx = \lim_{n\to\infty} \int_0^1 s_n\,dx = \lim_{n\to\infty} \frac{A}{2}\left(\frac{n-1}{n}\right) = \frac{A}{2}$$

which is the expected answer.

Example 5.1 has shown two things. The first is that the result of Lebesgue integration is the same that we would have obtained with a Riemann integral. This is

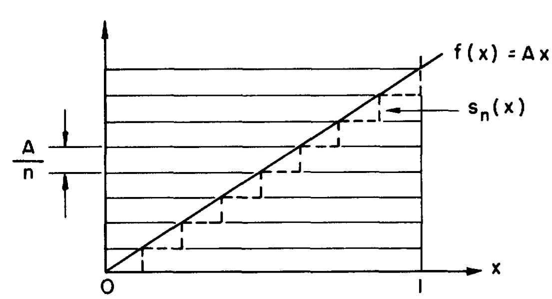

Fig. 5.1. Construction involved in determining the integral of f(x).

reasonable when we recall how Lebesgue integrals were constructed in Sections 1 and 2, and it leads us to suspect that the following statement might be (as in fact it is) generally true: Whenever a function is both Riemann integrable and Lebesgue integrable, the two integrals are equal. The second thing that we can see from the example is that the use of the fundamental definitions to determine Lebesgue integrals is not very practical, any more than it is practical to determine a Riemann integral by using its fundamental definition. [For a convincing example of this, see Whittaker and Watson (6), p. 64.]

How do we evaluate Riemann integrals? Except in cases in which numerical integration is unavoidable, we proceed in the following manner: Using their fundamental definition, we derive various properties of the integrals, and in particular we find that the following theorem is true: If

$$F(x) = F(a) + \int_a^x f(t)\, dt$$

then, at every point of continuity of $f(x)$,

$$\frac{dF}{dx} = f(x)$$

Knowing this fact, a table of integrals is constructed by the simple expedient of constructing a table of derivatives. Then, when we wish to determine the indefinite Riemann integral of some function $f(x)$, we look in the table for a function $F(x)$ which, when differentiated, yields $f(x)$. Thus we avoid completely the cumbersome fundamental definition. The question is, What can we do to evaluate our more general integrals? For the general case, we cannot do very much; but for the special case of Lebesgue integrals on the real line (the case most often encountered in practice) if the integrand also happens to be Riemann integrable, the solution is simple. We make use of the fact that the Lebesgue and Riemann integrals will be equal, and look up the integrand in a table of Riemann integrals, thus completely sidestepping the problem of integration.

If the integrand is not Riemann integrable, or if the function is to be integrated over a general set, or with respect to a general measure, then other means must be found to determine the value of the integral. However, from a theoretical point of view, the entire question of determining values of integrals is not particularly important. The criterion in judging the value of a given definition of integration is whether or not the integral so defined will be generally useful in analysis. Flexibility, ease of manipulation, general applicability, and so forth, are far more important as criteria than the more or less arithmetical question of how to associate a number with the symbol $\int_E f\, d\mu$. Therefore we abandon the question of the determination of integrals, and proceed, in the following sections and the remainder of this one, to derive some of the properties that make our general integrals more useful, and simpler, than the Riemann integral.

In the following theorems we shall give complete proofs only when they are either very brief or else very instructive.

THEOREM 5.1. If f and g are two functions defined on X, and if both are integrable with respect to μ on E, so that f, g ϵ $\mathscr{L}(\mu)$ on E, and if α and β are any two real numbers, then $(\alpha f + \beta g) \epsilon \mathscr{L}(\mu)$ on E, and

$$\int_E (\alpha f + \beta g)\, d\mu = \alpha \int_E f\, d\mu + \beta \int_E g\, d\mu$$

THEOREM 5.2. If f is defined on X, f ϵ $\mathscr{L}(\mu)$ on E, and $f \geqslant 0$, then

$$\int_E f\, d\mu \geqslant 0$$

THEOREM 5.3. If f and g are two functions defined on X, f, g ϵ $\mathscr{L}(\mu)$ on E, and $f \geqslant g$, then

$$\int_E f\, d\mu \geqslant \int_E g\, d\mu$$

PROOF: Since $f \geqslant g$, $(f-g) \geqslant 0$. Therefore, from Theorem 5.2,

$$\int_E (f-g)\, d\mu \geqslant 0 \qquad \text{and} \qquad \int_E f\, d\mu \geqslant \int_E g\, d\mu$$

THEOREM 5.4. If f ϵ $\mathscr{L}(\mu)$ on E, then $|f|$ ϵ $\mathscr{L}(\mu)$ on E.

PROOF: In terms of positive and negative parts,

$$f = f^+ - f^- \qquad \text{and} \qquad |f| = f^+ + f^-$$

and since f ϵ $\mathscr{L}(\mu)$ on E, both

$$\int_E f^+\, d\mu \qquad \text{and} \qquad \int_E f^-\, d\mu$$

are finite. Therefore

$$\int_E |f|\, d\mu = \int_E f^+\, d\mu + \int_E f^-\, d\mu$$

is also finite.

It is easy to show that the converse of this theorem does not hold. That is, $|f| \in \mathscr{L}$ does not, in general, imply that $f \in \mathscr{L}$. The following counterexample shows why. Suppose that the measurable set E can be written as the union of two disjoint nonmeasurable sets A_1 and A_2, and let $f(x) = 1$ for $x \in A_1$, and $f(x) = -1$ for $x \in A_2$. Then $|f| = 1$ on E, and therefore $|f|$ is measurable and integrable, and $\int_E |f|\, d\mu = \mu(E)$. On the other hand, f is not measurable, and therefore its integral is not even defined.

The difficulty in the counterexample is obviated if we assume explicitly that f is measurable. With this additional hypothesis, Theorem 5.4 does have a converse. The revised statement reads:

THEOREM 5.4a. If f is measurable, then $f \in \mathscr{L}(\mu)$ on E if and only if $|f| \in \mathscr{L}(\mu)$ on E.

The fact that the integrability of f implies that of $|f|$ is sometimes described by saying that our integrals are <u>absolutely</u> <u>convergent</u>. This property makes possible, among other things, the existence of functions which are Riemann (or, more properly, Riemann-Cauchy) integrable and not Lebesgue integrable. [Riemann-Cauchy integrals are the generalization of Riemann integrals to unbounded sets and unbounded functions.] For example, $\frac{\sin x}{x}$ is measurable, and Riemann-Cauchy integrable on $(-\infty, \infty)$. However, while

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\, dx$$

is finite,

$$\int_{-\infty}^{\infty} \left| \frac{\sin x}{x} \right| dx$$

is not. Therefore, according to Theorem 5.4a, $\frac{\sin x}{x}$ is not Lebesgue-integrable over the interval $(-\infty, \infty)$. As another example, consider functions whose integrals exist only as Cauchy Principal Values, such as $1/x$ on the interval $(-1, 1)$. Since

$$\int_{-1}^{1} \left| \frac{1}{x} \right| dx$$

is infinite, $1/x$ is not Lebesgue integrable on $(-1, 1)$. These difficulties have led to the definition of Perron and Denjoy integrals, which have many of the properties of Lebesgue integrals but are not absolutely convergent. For further discussion of this problem, the reader is referred to Munroe (3), page 189.

THEOREM 5.5. If $f \in \mathscr{L}(\mu)$ on E (whence $|f|$ is also integrable) then

$$\int_E |f| \, d\mu \geq \left| \int_E f \, d\mu \right|$$

PROOF: Since $|f| \geq f$, $\int |f| \, d\mu \geq \int f \, d\mu$ (Theorem 5.3). Similarly, since $|f| \geq -f$, $\int |f| \, d\mu \geq -\int f \, d\mu$. Therefore

$$\int_E |f| \, d\mu \geq \left| \int_E f \, d\mu \right|$$

THEOREM 5.6. If E is a measurable subset of X and $\mu(E) = 0$, then every function is integrable on E, and

$$\int_E f \, d\mu = 0$$

PROOF: If f is bounded, so that for some finite K, $|f| \leq K$, then, by using Theorem 5.5, we obtain

$$\left| \int_E f \, d\mu \right| \leq \int_E |f| \, d\mu \leq K \, \mu(E) = 0$$

whence $\int_E f \, d\mu = 0$.

This theorem can also be proved for f not bounded.

For Riemann integrals it is well known that if $a < c < b$, then

$$\int_a^b f(x) \, dx = \int_a^c f(x) \, dx + \int_c^b f(x) \, dx$$

The following theorem is a generalization of this property to general sets. The proof is given as an illustration of the use of our definitions.

THEOREM 5.7. If A and B are two measurable subsets of X, and if $E = A \cup B$, and $A \cap B = 0$, then for every function f integrable on E,

$$\int_E f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu$$

PROOF: For any measurable simple function s, given by

$$s = \sum_i \alpha_i K_{E_i}$$

we have

$$\int_E s\, d\mu = \sum_i \alpha_i\, \mu(E_i \cap E) = \sum_i \alpha_i\, \mu[E_i \cap (A \cup B)]$$

$$= \sum_i \alpha_i\, \mu[(E_i \cap A) \cup (E_i \cap B)]$$

and because of the finite additivity of μ, and the disjointness of A and B, we have

$$\mu[(E_i \cap A) \cup (E_i \cap B)] = \mu(E_i \cap A) + \mu(E_i \cap B)$$

Therefore

$$\int_E s\, d\mu = \sum_i \alpha_i\, \mu(E_i \cap A) + \sum_i \alpha_i\, \mu(E_i \cap B) = \int_A s\, d\mu + \int_B s\, d\mu$$

Thus the theorem holds for simple functions. For any integrable function f,

$$\int_E f\, d\mu = \lim_{n\to\infty} \int_E s_n\, d\mu = \lim_{n\to\infty} \left[\int_A s_n\, d\mu + \int_B s_n\, d\mu\right]$$

and since limits and finite sums may always be interchanged,

$$\int_E f\, d\mu = \lim_{n\to\infty} \int_A s_n\, d\mu + \lim_{n\to\infty} \int_B s_n\, d\mu = \int_A f\, d\mu + \int_B f\, d\mu$$

This proof* may be extended immediately to any finite number of sets.

Suppose that, using the symbols of the preceding theorem, $\mu(B) = 0$. Then according

*A much simpler but less instructive proof of the theorem: Using the same notation as above, let K_A and K_B be the characteristic functions of A and B, respectively. Then, for every $x \in E$, it is certainly true that $f = K_A f + K_B f$, whence, from Theorem 5.1, we obtain

$$\int_E f\, d\mu = \int_E K_A f\, d\mu + \int_E K_B f\, d\mu = \int_A f\, d\mu + \int_B f\, d\mu$$

to Theorem 5.6 we have

$$\int_E f\,d\mu = \int_A f\,d\mu + \int_B f\,d\mu = \int_A f\,d\mu$$

This equation shows that sets of measure zero do not contribute anything to an integral and may therefore be ignored completely. Thus, for example, if a function is bounded everywhere except at most on a set of measure zero, we may just as well assume, in problems involving integrals, that it is bounded everywhere, without thereby losing generality or affecting the results. The possibility of ignoring the behavior of functions on sets of measure zero is so useful so often that the following abbreviated terminology is in common use. If a property P holds for every $x \in E$ except at most on a subset A of E of measure zero, we say that the property P holds for <u>almost</u> <u>all</u> $x \in E$, or holds <u>almost</u> <u>everywhere</u> on E. [Sometimes the phrase is further abbreviated to "P holds a.e. on E," or, in some British and French texts, to "P holds p.p. on E," where p.p. stands for presque partout (the French for almost everywhere).] Of course, whether or not a set has measure zero depends on what is the measure under consideration. Therefore, if more than one measure enters into a particular discussion and we wish to say, for example, that $f = g$ except at most on a set of μ_1-measure zero, we write

$$f = g \quad \text{a.e.} \quad (\mu_1) \qquad \text{or} \qquad f = g \quad [\mu_1]$$

Now, suppose that we have two integrable functions f and g defined on a set E, and that $f = g$ a.e. on E. Let A be the subset of E on which $f = g$, and let B be the subset of E on which $f \neq g$. That is,

$$A = \{x : f = g\} \cap E, \qquad B = \{x : f \neq g\} \cap E$$

Clearly, $A \cup B = E$, and since A and B have no points in common, $A \cap B = 0$. Also, by definition of almost everywhere equality, $\mu(B) = 0$. Then, using Theorem 5.7, we have

$$\int_E (f-g)\,d\mu = \int_A (f-g)\,d\mu + \int_B (f-g)\,d\mu$$

Of the two integrals on the right, the first one is zero because the integrand is identically zero, and the second one is zero because $\mu(B) = 0$. Therefore

$$\int_E (f-g)\,d\mu = 0, \qquad \text{so that} \int_E f\,d\mu = \int_E g\,d\mu$$

Thus we have proved that if two functions are equal almost everywhere, their integrals are equal. Since the set E is arbitrary, the integrals in question are, in effect, indefinite integrals. Now, when the indefinite integrals of two functions are equal, the integrands themselves may be thought of as being equal in some sense or other, since from the point of view of the integral they are indistinguishable. (This feeling is strengthened by the known fact that if the indefinite Riemann integrals of two continuous functions are equal, the functions themselves are exactly equal.) It is desirable to have a notation to express this effective equality, and so we write

$$f \sim g \quad \text{on } E$$

if $\{x : f \neq g\} \cap E$ has measure zero, and we say that f is equivalent to g on E.

Now let us suppose that we have a class of functions defined on a set E, any two functions in the class differing from each other at most on a set of μ-measure zero. We call that class of functions an equivalence class. If we have several equivalence classes, the indefinite integral cannot distinguish between members of any one equivalence class — it can only distinguish between any two equivalence classes. For example, one says that under suitable restrictions the Fourier spectrum of a time function is uniquely related to the time function in the sense that given one, the other is uniquely derivable from it. This cannot possibly be true, since altering the time function at a finite or even countably infinite number of points does not affect the value of the integral which defines the spectrum, and vice-versa. The Fourier transform pair can only relate uniquely equivalence classes of time functions with equivalence classes of spectra.

We might note, by the way, that the definition of such concepts as "almost everywhere" and "equivalence class" allows, in effect, the introduction of a controlled amount of imprecision into a perfectly rigorous discussion. As a result, we are no longer forced, in the course of an argument, to say more than we need or can say about the behavior of a function on unimportant sets. While this is hardly a fundamental point in favor of the Lebesgue theory (something like this could have been developed within the framework of the Riemann theory), it is nevertheless a worthwhile improvement in the language in which we think. It allows us to satisfy our desire for rigor while allowing us to circumvent the need for increasingly restrictive hypotheses at every turn of a discussion. The net effect of this increased flexibility in language is that we are enabled to make statements about less restricted classes of functions, hence to increase the generality of our results and to simplify our arguments.

Having derived some of the basic properties of the integral, we proceed in the following sections to the study of some related topics in which the properties of the integral are applied. In Section 6 we use the ideas of the present section to arrive at various possible concepts of convergence for a sequence of functions. In Section 7 we shall be concerned with the interchange of limits and integration (or summation and integration), and we shall present some important results which illustrate the advantages of the general integral over the Riemann integral. In Section 8, the integral is considered as a set function, a point of view which leads in a natural way to the concept of the Radon-Nikodym derivative.

6. CONVERGENCE

A sequence of numbers $\alpha_1, \alpha_2, \alpha_3, \ldots$ is said to converge to a number α if, given any $\epsilon > 0$, there exists a number N with the property that

$$|\alpha_n - \alpha| < \epsilon$$

for all $n \geq N$. Then α is called the limit of the sequence. A necessary and sufficient condition for the existence of the limit is that, given any $\epsilon > 0$, there exists a number N with the property that

$$|\alpha_m - \alpha_n| < \epsilon$$

for all $m \geq N$ and $n \geq N$. Any sequence that satisfies this condition is called a Cauchy sequence.

We now ask, How should convergence and limit be defined if, instead of a sequence of numbers, we are given a sequence of functions? The most obvious first thought is to consider pointwise convergence of functions. That is, suppose that we have a sequence $f_1, f_2, f_3, \ldots$ of functions defined on a set E. Let x be some fixed point of E. Then the sequence $f_1(x), f_2(x), f_3(x), \ldots$ is just a sequence of numbers, as before, and we say that it converges to a number f(x) if, given any $\epsilon > 0$, there exists an N such that

$$|f_n(x) - f(x)| < \epsilon \qquad \text{for all } n \geq N \tag{6.1}$$

The set of numbers f(x) which are the limits of the sequence $\{f_n(x)\}$ at each value of x defines a function f on the set E, and we may think of this function f as the limit of the sequence of functions $\{f_n\}$. Note that, in general, the N in Eq. 6.1 is a function both of ϵ and x. If N is independent of x (that is, if one single value of N will satisfy Eq. 6.1 for all points $x \in E$) the convergence is said to be uniform on E.

While our definition of convergence for a sequence of functions is acceptable and useful, we might ask whether it is not excessively demanding for some purposes. After all, suppose that there are some points x of E at which the sequence $\{f_n\}$ fails to converge, so that at those points the limit function f is undefined. If the set of points on which the sequence fails to converge is small enough, this does not seem to be a sufficient reason for throwing out the whole sequence, considering, for example, that for some problems the behavior of a function on a set of measure zero may be ignored completely. Again, as we saw in the preceding section, for some purposes we are not really interested in specifying functions but only equivalence classes of functions. Since pointwise convergence defines a particular function, it might be useful to find some other, softened, form of convergence which only defines an equivalence class of functions, and does not pinpoint any one function exactly. In still other problems, we might not even be interested in knowing whether the sequence $\{f_n\}$ has an equivalence class as a limit, but only whether some expression like

$$\int_E |f - f_n|^p \, d\mu \qquad (p > 0)$$

approaches zero with increasing n.

All of these possibilities, and others which will not be discussed, have in common the general idea that a sequence of functions assumes a definite character (in some sense or other) for n sufficiently large, and therefore are included within the general intuitive meaning of convergence. In this section we shall define and present some of the properties of three types of convergence (other than pointwise convergence) which occur frequently and which we shall need in later sections.

The first definition of convergence embodies our willingness to ignore failure of pointwise convergence on sufficiently small sets. Let $\{f_n\}$ be a sequence of functions defined on a set E. $\{f_n\}$ is said to converge almost everywhere to a limit f if there exists a subset A of E such that $\mu(A) = 0$, and $\{f_n(x)\}$ converges pointwise to f(x) for all $x \in (E-A)$. That is, $\{f_n(x)\}$ converges pointwise to f(x) everywhere on E except at most on a set of measure zero. Symbolically, convergence almost everywhere is denoted by

$$\lim_{n\to\infty} f_n = f \quad (\text{a.e.}) \quad \text{or } f_n \to f \quad (\text{a.e.})$$

Since the set A on which pointwise convergence fails may be empty, in which case we have ordinary convergence, it is clear that if a sequence converges in the ordinary way (i.e., at every point of the set) then it also converges almost everywhere.

EXAMPLE 6.1. In this example we consider a sequence $\{f_n\}$ of functions defined on the real line which converges almost everywhere but not everywhere. For this example, the measure used is the Lebesgue measure of the real line, so that the phrase "almost everywhere" means everywhere except at most on a set of Lebesgue measure zero. Let

$$f_n(x) = \left(\frac{\cos n\pi}{2n+1}\right) \frac{\sin(2n+1)\,\pi x}{\sin \pi x} \qquad (n = 1, 2, 3, \ldots)$$

with value $f_n = \cos n\pi$ at the points $x = 0, \pm 1, \pm 2, \ldots$ Then for a fixed, noninteger value of x, we have

$$|f_n(x)| = \left| \frac{\cos n\pi}{2n+1} \cdot \frac{\sin(2n+1)\pi x}{\sin \pi x} \right| \leq \frac{1}{2n+1} \cdot \left| \frac{1}{\sin \pi x} \right| < \epsilon$$

That is, for any $\epsilon > 0$, $|f_n(x)|$ can be made smaller than ϵ by choosing a sufficiently large n. Therefore, for noninteger values of x, the sequence converges to zero:

$$f_n(x) \to 0 \qquad \text{for } x \neq 0, \pm 1, \pm 2, \ldots$$

But if x has any integer value k, then

$$f_n(k) = \frac{\cos n\pi}{2n + 1}(2n + 1) = \cos n\pi$$

Therefore, for $x = k$, the sequence, written out, is

$$-1, 1, -1, 1, -1, 1, -1, 1, \ldots$$

which obviously does not converge to any number. So we find that $\{f_n\}$ converges for all values of x except the integer values. Since the integers form a countable set, the set of all integers has Lebesgue measure zero (cf. Ex. 4.3). Thus $\{f_n\}$ converges everywhere on the real line except on a set of measure zero. In symbols, this is expressed as

$$f_n \to 0 \quad \text{a.e. (Lebesgue measure)}$$

This concludes the example.

With ordinary (pointwise) convergence, if a sequence converges to a limit function, that function is defined exactly and uniquely. In contrast to this, convergence a.e. defines uniquely only an equivalence class of functions, as is shown in the two following theorems.

THEOREM 6.1. If, for the sequence $\{f_n\}$, $f_n \to f$ (a.e.), and if there is a function g such that $f \sim g$ [read: f is equivalent to g], then it is also true that $f_n \to g$ a.e.

PROOF: $f_n \to f$ a.e. means that for n large enough, $|f_n - f| < \epsilon$ a.e.
$f \sim g$ means $f = g$ a.e., or $|f - g| = 0$ a.e.

But now $|f_n - g| = |(f_n - f) - (g - f)| \leq |f_n - f| + |g - f| = |f_n - f|$ a.e.

Therefore, for n large enough

$$|f_n - g| < \epsilon \quad \text{a.e.} \quad \text{or } f_n \to g \quad \text{a.e.}$$

THEOREM 6.2. If, for the sequence $\{f_n\}$, $f_n \to f$ a.e. and at the same time $f_n \to g$ a.e., then $f \sim g$. (The limit is unique up to an equivalence.)

PROOF: For any $\epsilon > 0$, choose N sufficiently large so that, for all $n \geq N$,

$$|f_n - f| < \frac{\epsilon}{2} \quad \text{a.e.}$$

$$|f_n - g| < \frac{\epsilon}{2} \quad \text{a.e.}$$

This is possible by hypothesis. Then

$$|f - g| = |(f - f_n) - (g - f_n)| \leq |f - f_n| + |g - f_n| < \epsilon \quad \text{a.e.}$$

Therefore

$$|f - g| < \epsilon \quad \text{a.e.}$$

and since ϵ is arbitrary

$$|f - g| = 0 \quad \text{a.e.} \quad \text{or } f \sim g$$

A second type of convergence is convergence in measure, which is defined as follows. A sequence $\{f_n\}$ of a.e. finite-valued, measurable functions f_n is said to <u>converge in measure</u> to the measurable function f if, given any $\epsilon > 0$, the measure of the set on which

$$|f_n - f| \geq \epsilon$$

approaches zero with increasing n. That is, $f_n \to f$ in measure if, given any $\epsilon > 0$,

$$\lim_{n\to\infty} \mu\,[\{x : |f_n(x) - f(x)| \geq \epsilon\}] = 0$$

Symbolically, convergence in measure is denoted by

$$\lim_{n\to\infty} f_n = f \quad (\text{meas.}) \quad \text{or } f_n \to f \quad (\text{meas.})$$

When put into words, the definitions of convergence in measure and convergence a.e. sound almost indistinguishable. This is misleading, because while it is true that if the sequence $\{f_n\}$ converges a.e. on a set of finite measure it also converges in measure, the converse is not true. The truth of these statements is not obvious, but the proof is too involved to give here.* Those readers who have developed an intuitive understanding of the difference between the Strong Law of large numbers and the Weak Law of large numbers in probability theory, can apply that understanding to the present problem since, when probability is regarded as a measure, the Strong Law corresponds exactly to convergence a.e., while the Weak Law corresponds to convergence in measure. [For a discussion of the problem from a probability point of view, see Feller (10), p. 191 or Munroe (3), p. 226 and p. 227, Exercises a and b.]

The following example will illustrate the use of the definition in testing a given sequence for convergence in measure.

EXAMPLE 6.2. Let X be the real line and μ the Lebesgue measure of the line, and $\mathcal{S}$ the class of Borel sets of the line. Let A_n be the interval $\left(0, \frac{1}{n}\right)$, $(n = 1, 2, 3, \ldots)$, and let K_{A_n} be the characteristic function of A_n, so that $K_{A_n}(x) = 1$ if $x \in A_n$, and is zero otherwise. Since $A_n \in \mathcal{S}$, the functions K_{A_n} are measurable, and they certainly are finite. Consider the sequence $\{K_{A_n}\}$, some members of which are shown in Fig. 6.1. We shall show that the sequence converges in measure to zero. In fact, the n^{th} member of the sequence, K_{A_n}, is greater than zero only for those points x contained in A_n, and so for any $\epsilon > 0$ (but less than 1, of course),

$$\mu\left[\left\{x : (K_{A_n} - 0) \geq \epsilon\right\}\right] = \frac{1}{n}$$

Therefore, $\lim_{n\to\infty} \mu\left[\left\{x : (K_{A_n} - 0) \geq \epsilon\right\}\right] = 0$, which proves our assertion.

---

*For a proof that convergence a.e. on a finite set implies convergence in measure, see Natanson (2), p. 95, or Munroe (3), p. 224. Counterexamples that prove that the converse is not true are given in the same places.

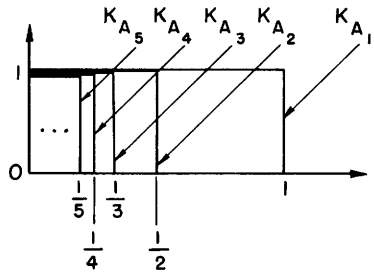

Fig. 6.1. Example of convergence in measure.

Incidentally, the sequence $\{K_{A_n}\}$ also converges pointwise (and therefore also a.e.) to zero. To see this, note that for each x contained in (0, 1), it is possible to find a number N so large that for all $n \geq N$, $K_{A_n}(x) = 0$. Therefore

$$K_{A_n}(x) \to 0 \quad \text{for all } x \in \left\{x : 0 < x < 1\right\} .$$

Note, however, that the convergence is not uniform. We shall digress briefly to illustrate the difference between ordinary (pointwise) convergence and uniform convergence. By definition, if the sequence $\{K_{A_n}\}$ converges to zero uniformly, given any $\epsilon > 0$ we must be able to find an N large enough so that, for all $n \geq N$, $|K_{A_n}(x) - 0| < \epsilon$ for all values of $x \in (0, 1)$. In this example, this is impossible. For, suppose that N_0 is the appropriate number. Then it would have to be true that $|K_{A_n}| < \epsilon$ for all $n \geq N_0$, and for all values of x in the unit interval. The inequality is certainly true for $x > \frac{1}{N_0}$; but for $0 < x < \frac{1}{N_0}$, say $x = \frac{1}{2N_0}$, $f_n\left(\frac{1}{2N_0}\right) = 1$ for all $N_0 < n < 2N_0$, so that the inequality is not satisfied and the convergence is not uniform. Note that in proving ordinary convergence we had to choose a point x first, and then find an appropriate N. This would not be necessary if the sequence converged uniformly.

A sequence that converges in measure, like one that converges almost everywhere, does not define a unique limit function. Both types of convergence define only an equivalence class of functions. As Example 6.2 has indicated, and as can be shown in general, if a sequence converges both in measure and almost everywhere, the two limit functions are equivalent.

Yet another type of convergence which is frequently useful, especially in the study of series of orthonormal functions and in the theory of Fourier integrals, is convergence in the mean. Let $\{f_n\}$ be a sequence of measurable functions defined on a set E, and with the additional property that $|f_n|^p$ $(p \geq 1)$ is integrable. If there exists a measurable function f such that $|f|^p$ is integrable on E and such that

$$\lim_{n\to\infty} \int_E |f_n - f|^p \, d\mu = 0$$

we say that $\{f_n\}$ <u>converges in the mean</u> of order p to f, and write

$$\lim_{n\to\infty} f_n = f \quad (\text{mean}^p) \quad \text{or} \quad f_n \to f \quad (\text{mean}^p)$$

The special case $p = 2$ arises in a natural way in work involving orthonormal series and Fourier integrals. For this case, it has become customary simply to say that $\{f_n\}$ converges in the mean to f, and to write

$$\underset{n\to\infty}{\text{l.i.m.}}\, f_n = f, \quad \text{or} \quad f_n \to f \quad (\text{mean})$$

Instances of sequences which converge in the mean are plentiful, for example, in the study of convergence of Fourier integrals. Here we shall give a very simple example of a mean convergent sequence, principally for the purpose of illustrating the use of the definition.

EXAMPLE 6.3. Consider the sequence of functions defined in Example 6.2. We had

$$\begin{aligned} f_n(x) &= 1 \quad \text{for all } x \in \left\{x : 0 < x < \frac{1}{n}\right\} \\ &= 0 \quad \text{elsewhere} \end{aligned}$$

We shall prove that this sequence converges in the mean to zero. In fact

$$\int_E |f_n - f|^2 \, d\mu = \int_0^{1/n} (1 - 0)^2 \, dx = \frac{1}{n}.$$

Therefore

$$\lim_{n\to\infty} \int_E |f_n - f|^2 \, d\mu = \lim_{n\to\infty} \frac{1}{n} = 0$$

whence

$$\underset{n\to\infty}{\text{l.i.m.}}\, f_n = 0$$

as asserted.

For simplicity, all of the following statements refer only to mean convergence of order two, although most of them are true for arbitrary $p \geq 1$.

As is the case for the other types of convergence, the limit in the mean is unique only up to an equivalence, as the next theorem shows.

THEOREM 6.3. Suppose that the sequence $\{f_n\}$ converges in the mean to the functions f and g. Then $f \sim g$.

PROOF: The fact that

$$f_n \to f \quad (\text{mean}) \quad \text{and} \quad f_n \to g \quad (\text{mean})$$

means that, for any $\epsilon > 0$, there exists an N so large that for all $n \geq N$ we can make

$$\int_E (f_n - f)^2 \, d\mu < \frac{\epsilon}{4}$$

$$\int_E (f_n - g)^2 \, d\mu < \frac{\epsilon}{4}$$

Now consider the relation

$$\int_E (f - g)^2 \, d\mu = \int_E [(f - f_n) - (g - f_n)]^2 \, d\mu$$

$$= \int_E (f - f_n)^2 \, d\mu + \int_E (g - f_n)^2 \, d\mu - 2 \int_E (f - f_n)(g - f_n) \, d\mu$$

We already have bounds for the first two integrals. To find a bound for the third, we use the Schwartz inequality, and obtain

$$\left| \int_E (f - f_n)(g - f_n) \, d\mu \right| \leq \sqrt{\left[\int_E (f - f_n)^2 \, d\mu\right]\left[\int_E (g - f_n)^2 \, d\mu\right]} < \sqrt{\frac{\epsilon}{4} \cdot \frac{\epsilon}{4}} = \frac{\epsilon}{4}$$

Therefore,

$$\int_E (f - g)^2 \, d\mu < \frac{\epsilon}{4} + \frac{\epsilon}{4} + 2\frac{\epsilon}{4} = \epsilon$$

Since ϵ is arbitrary and the left-hand side is nonnegative, we have

$$\int_E (f - g)^2 \, d\mu = 0$$

and since the integrand is always positive, this implies that $f = g$ a.e., or $f \sim g$, as asserted.

How is mean convergence related to the other types of convergence? One simple relationship is expressed in the following theorem.

THEOREM 6.4. If a sequence $\{f_n\}$ converges in the mean to a function f, then it also converges to f in measure.

PROOF: The statement

$$f_n \to f \text{ (mean)}$$

says that, for any given $\epsilon > 0$, there exists an N with the property that, for all

$n \geqslant N$,

$$\int_E (f_n - f)^2 \, d\mu < \epsilon$$

Consider the subset A_n of E defined by

$$A_n = \{x : |f_n(x) - f(x)| \geqslant \alpha\}$$

where α is some given positive number which remains fixed. Then, since the integrand is always positive,

$$\int_E (f_n - f)^2 \, d\mu \geqslant \int_{A_n} (f_n - f)^2 \, d\mu \geqslant \alpha^2 \mu(A_n)$$

Therefore, $\alpha^2 \mu(A_n) \leqslant \int_E (f_n - f)^2 \, d\mu < \epsilon$, which is the same as saying that as n increases, $\mu(A_n) \to 0$, which, by definition, means that $f_n \to f$ (meas.).

It is true, and perhaps plausible, that convergence in the mean does not imply pointwise convergence or even convergence a.e. [For an example which proves this statement, see Wiener (8), p. 29.] What is more unexpected is that pointwise convergence does not imply mean convergence either, as the following counterexample shows.

EXAMPLE 6.4. Let our space be the interval $[0, 1]$, and our measure the Lebesgue measure of the line. Let the sequence $\{f_n\}$ be defined by

$f_n(x) = 0$ at $x = 0$

$f_n(x) = n$ for $0 < x < 1/n$

$f_n(x) = 0$ everywhere else

Then, for any $x \in [0, 1]$, it is possible to find an N sufficiently large so that for all $n \geqslant N$, $f_n(x) = 0$. Therefore $f_n(x) \to 0$ everywhere as $n \to \infty$ (although not uniformly). On the other hand,

$$\int_0^1 (f_n(x) - 0)^2 \, dx = \int_0^{1/n} n^2 \, dx = n$$

which grows without limit as n tends to infinity. Therefore the sequence $\{f_n\}$ does not converge in the mean.

Since convergence everywhere implies convergence a.e. and convergence in measure, the example has also shown that these latter two types of convergence do not, in general, imply convergence in the mean.

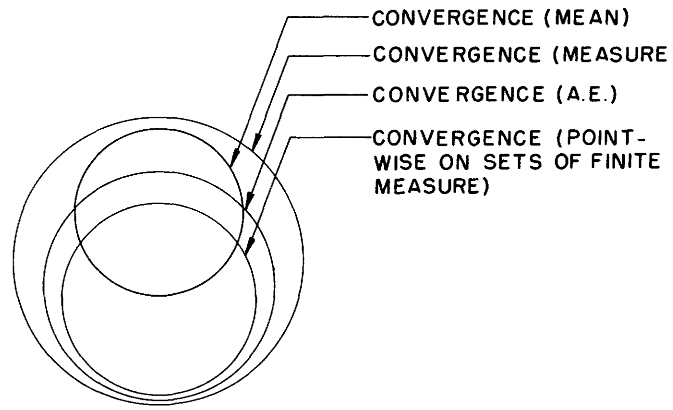

Fig. 6.2. Summary of interrelationships for various types of convergence.

In this section we have spoken of convergence pointwise, a.e., in measure, and in the mean. Figure 6.2 is an attempt to summarize graphically the rather involved interrelationships that obtain among these various types of convergence. Each circle in the figure represents a class of sequences that converge in the mode specified by the label on the circle. The diagram shows, for example, that all listed types of convergence imply convergence in measure (since all the points of the smaller circles are also points of the large circle) but that there are sequences that converge in measure but not, say, in the mean.

In addition to the modes of convergence that were discussed above, several others have been found useful. A more complete discussion of convergence is given in Halmos (1) or Munroe (3), and an excellent graphical summary of interrelationships can be found in Munroe (3), p. 237. We shall leave this subject here, and proceed to consider briefly one of the important features of Lebesgue integration — its properties with respect to limit processes.

7. INTEGRALS AND LIMIT PROCESSES

Suppose that we have a sequence $\{f_n\}$ of integrable functions defined on a set E, and suppose that the sequence converges, in some sense or other, to a limit function f. In this section we shall consider under what conditions it is true that

$$\lim_n \int_E f_n \, d\mu = \int_E \left[\lim_n f_n \, d\mu\right] \equiv \int_E f \, d\mu \tag{7.1}$$

That is, our problem is to determine the conditions under which the order of a limit process and integration may be interchanged.

The answer to this problem has frequent practical application, for example, in determining when an infinite series may be integrated term by term. This comes about as follows: Suppose that we are given a convergent series of integrable functions $f_k(x)$, so that

$$\sigma(x) = \sum_{k=1}^{\infty} f_k(x)$$

We define the n^{th} partial sum, $\sigma_n(x)$, by

$$\sigma_n(x) = \sum_{k=1}^{n} f_k(x) \tag{7.2}$$

If the infinite series is convergent, then the set of partial sums

$$\sigma_1(x), \sigma_2(x), \sigma_3(x), \ldots, \sigma_n(x), \ldots$$

forms a convergent sequence, and

$$\lim_n \sigma_n(x) = \sigma(x) \tag{7.3}$$

In terms of partial sums, asking whether or not it is true that

$$\int_E \sum_{k=1}^{\infty} f_k \, d\mu = \sum_{k=1}^{\infty} \int_E f_k \, d\mu \tag{7.4}$$

is the same as asking whether or not

$$\int_E \left[\lim_n \sigma_n\right] d\mu = \lim_n \int_E \sigma_n \, d\mu \tag{7.5}$$

To see this, note first that the left-hand members of Eqs. 7.4 and 7.5 are identical by definition. Next, using the definition of σ_n, we have

$$\int_E \sigma_n \, d\mu = \int_E \sum_{k=1}^{n} f_k \, d\mu = \sum_{k=1}^{n} \int_E f_k \, d\mu$$

the last equality following from Theorem 5.1, which states that the order of finite sums and integrals may always be interchanged. Therefore,

$$\lim_n \int_E \sigma_n \, d\mu \equiv \lim_n \sum_{k=1}^{n} \int_E f_k \, d\mu = \sum_{k=1}^{\infty} \int_E f_k \, d\mu$$

so that the right-hand members of Eqs. 7.4 and 7.5 are also identical. Therefore, the problem of determining when the order of limit and integration may be interchanged for sequences contains, as a special case, the problem of when an infinite series may be integrated term by term. These considerations provide us with a practical motivation for our interest in the topic of this section.

With Riemann integration, the standard theorem used in connection with problems of interchanging limits and integration is the following (see, for example, Rudin (4), p. 121):

THEOREM 7.1. Suppose that f_n is Riemann integrable on $[a, b]$ for every n, and suppose that $\{f_n\}$ converges uniformly to f on $[a, b]$. Then

1. f is Riemann integrable on $[a, b]$, and

2. $$\int_a^b f(x)\, dx \equiv \int_a^b \left[\lim_n f_n(x)\, dx\right] = \lim_n \int_a^b f_n(x)\, dx \qquad (7.6)$$

The fact that this is the standard theorem used with Riemann integrals reveals some of the fundamental limitations of the Riemann definition. We shall study in detail the reasons for making the hypotheses of Theorem 7.1 and show, by means of examples, why some other apparently possible sets of hypotheses do not work. In order not to interrupt the discussion with the examples, they will all be presented in Appendix II, and reference to them will be made at appropriate points in the development.

We notice first that the conditions specified by the theorem are sufficient, but not necessary. That is, there are sequences that do not converge uniformly, but for which the conclusions of the theorem are nevertheless true, as in Example AII-1. On the other hand, the conclusions of the theorem are definitely false, in general, if the adverb "uniformly" is omitted in the statement of the theorem. A simple instance of this failure is given in Example AII-2, in which omitting uniformity leads to an unbounded sequence. On the evidence from these examples, we might ask, Why is ordinary (pointwise) convergence, together with boundedness ($|f_n| \leqslant K$ for all n, which implies $|f| \leqslant K$) not enough? Why is uniformity of convergence involved in the problem?

The relation that we want to justify is given by Eq. 7.6. There we notice that two things must be proved: 1) that the limit function f is integrable, since otherwise the symbols in Eq. 7.6 are meaningless, and 2) that the interchange itself is possible. With Riemann integrals, most of the difficulty of interchange theorems arises in making conclusion 1 possible. In fact, once the first conclusion is established, the second follows under very mild conditions. To see this, let us assume the truth of conclusion 1 as a separate hypothesis. Now the interchange theorem becomes simply:

THEOREM 7.2. (Arzelà-Osgood Theorem). Let $\{f_n\}$ be a sequence of functions Riemann integrable on $[a, b]$, and converging pointwise to f, and let f be Riemann integrable on $[a, b]$. Then, if there exists a finite constant K such that $|f_n(x)| < K$ for all n and all $x \in [a, b]$,

$$\lim_n \int_a^b f_n \, dx = \int_a^b \left[\lim_n f_n\right] dx \equiv \int_a^b f \, dx$$

Thus we find that ordinary convergence and boundedness can replace uniform convergence provided that the Riemann integrability of the limit is eliminated as a problem by postulating it separately. (The proviso is necessary: As is shown in Example AII-3, the limit of a convergent, bounded sequence of Riemann integrable functions need not be Riemann integrable.)

Theorem 7.2 begs the question: it solves the simple problem, the interchange problem, but leaves us with the difficult one. To use the theorem in practice, we must be able to guarantee that the limit of a sequence will be Riemann integrable without (usually) knowing what that limit is. But how are we to know ahead of time whether or not the limit of a sequence is Riemann integrable? This is where uniformity comes in.

As we have seen before, a bounded function is Riemann integrable (on a finite interval) only if it is not "too" discontinuous. The precise statement reads:

THEOREM 7.3. Let f be bounded on $[a, b]$. Then f is Riemann integrable on $[a, b]$ if and only if f is continuous almost everywhere on $[a, b]$.

That is, the set of points of $[a, b]$ on which f is discontinuous must have Lebesgue measure zero. If a function is discontinuous on a set of positive measure, as in the example in Section 1, its Riemann integral does not exist because the lower and upper sums defined in Section 1 do not approach a common value. Our problem is, Given a sequence $\{f_n\}$ of functions (which might even be continuous for every finite n), how can we know ahead of time that the limit function will also be continuous enough to be Riemann integrable? On the one hand, we know that limit processes in general do not preserve continuity (Example AII-4); on the other hand, some form of continuity is the most basic requirement in making possible the existence of the Riemann integral.

Now we can see why uniformity appears as a requirement of Theorem 7.1: It is there mainly to guarantee the integrability of the limit, because with uniform convergence we have an a priori guarantee that continuity will be preserved. We can also see why the requirements of the theorem are sufficient but not necessary: As is shown in Example AII-5, uniform convergence is a sufficient, but not a necessary, condition for the preservation of continuity. Besides, according to Theorem 7.3, we do not really require the limit to be continuous everywhere – it need only be bounded and continuous almost everywhere. Thus the requirement of uniform convergence is too stringent and that of pointwise, bounded convergence is too lax. Unfortunately, classical analysis has not developed anything useful in between.

In an imprecise way, then, this discussion shows why Riemann integrals become inflexible in connection with limit processes: The integral is based on continuity, and continuity is an awkward property when considered in connection with limit processes.

In this respect, the more general integral is far superior to the Riemann integral. It does not depend on evanescent properties like continuity; instead, it is based on measurability, a dependable quality which, as we saw in Section 4, is always preserved in the course of limit processes. This circumstance virtually eliminates, right from the start, the possibility of ending up with limit functions for which an integral cannot be defined.* Thus it is possible to find relatively undemanding conditions under which the order of limits and integrals may be interchanged.

We shall start with a special case of the Lebesgue Bounded Convergence Theorem:

THEOREM 7.4. Let $\{f_n\}$ be a sequence of measurable functions, defined on the measurable set E of finite measure, and converging pointwise to a function f. If there exists a finite constant K such that $|f_n(x)| < K$ for all n and all $x \in E$, then

$$\lim_n \int_E f_n \, d\mu = \int_E \left[\lim_n f_n\right] d\mu \equiv \int_E f \, d\mu$$

The interesting thing about this theorem is that it is almost identical in form with Theorem 7.2. But there is one important difference: the integrability of the limit does not have to be postulated separately; instead, it is a direct consequence of the other hypotheses. The gain in ease of application, and hence in usefulness, is obvious – we have all of the simplicity of the Arzelà-Osgood theorem without its major drawback. Of course, we also automatically gain all of the generality inherent in the Lebesgue language: the functions in the sequence need not be Riemann integrable, they need only be measurable; their domains can be general sets, instead of being limited to intervals of the real line; and the integrals can be taken with respect to any measure, instead of

*The reader is reminded that, from the definitions given in Section 5, the integral of f (with respect to μ on E) exists (i.e., is defined) if f is measurable and either $\int_E f^+ \, d\mu$ or $\int_E f^- \, d\mu$ is integrable. A function f is integrable (with respect to μ on E) if it is measurable and if $\int_E f \, d\mu$ is finite. Since the limit of a sequence of measurable functions is always measurable, to insure the integrability of the limit we need only place just enough constraints on the sequence to insure that the integral of the limit will be finite.

just the length of intervals.

Is it possible to eliminate the boundedness requirement ($|f_n| < K$) in Theorem 7.4, and thus justify interchange without any special hypotheses? No, as is demonstrated by Example AII-6. But we can generalize the theorem tremendously, and do this in two directions. 1) Since what happens on sets of measure zero does not affect the values of integrals, we might suspect that all we really need is that the sequence shall converge almost everywhere, not everywhere. As a matter of fact, we need even less than that: The theorem is even true for convergence in measure. 2) The sequence does not have to be bounded by a constant; it is sufficient if it is dominated by an integrable function, and, of course, the inequality need only apply almost everywhere. And so we arrive at Theorem 7.5.

THEOREM 7.5. (Lebesgue Dominated Convergence Theorem). Let $\{f_n\}$ be a sequence of measurable functions defined on a measurable set E. Let $\{f_n\}$ converge to the limit f in measure. If there exists an integrable function g that is such that, for all n and for almost all $x \in E$,

$$|f_n(x)| \leq g(x)$$

then

$$\lim_n \int_E f_n \, d\mu = \int_E \left[\lim_n f_n\right] d\mu \equiv \int_E f \, d\mu \ .$$

[The existence and finiteness of the integral of the limit is part of the conclusion.]

The requirements of Theorem 7.5, mild as they are (compare with uniform convergence of a sequence of Riemann integrable functions), are still more stringent than is necessary. More general conditions are given in Natanson* (2), p. 153 et seq., but they are expressed in terms of concepts which have not been presented here, and which it would be too complicated to introduce at this point. However, there is a theorem whose hypotheses are simpler than those of Theorem 7.5, this simplicity being gained by restricting the type of sequences to which it applies. This is the Lebesgue Monotone Convergence Theorem, which applies only to monotonic sequences. We shall present it here in the form it takes when it is to be applied to the problem of interchanging summation and integration of infinite series.

THEOREM 7.6. Given an infinite series of nonnegative functions, each integrable on E. If for almost all $x \in E$

*Other variations and examples of interchange theorems may be found in: Munroe (3), p. 186 and p. 233 et seq.; Riesz-Nagy (11), p. 33 et seq.; Graves (12), p. 190 et seq.

$$\sum_{n=1}^{\infty} f_n(x) = f(x)$$

then

1. f is integrable on E if and only if

$$\sum_{n=1}^{\infty} \int_E f_n \, d\mu < \infty$$

and if this is the case, then

2. $$\int_E \left[\sum_{n=1}^{\infty} f_n \right] d\mu = \sum_{n=1}^{\infty} \int_E f_n \, d\mu = \int_E f \, d\mu$$

One of the most interesting and suggestive features of this theorem is the form in which its conclusions are stated. As in the familiar engineering rule of thumb, the theorem says, essentially, that a series of nonnegative functions may be integrated term by term, provided the result does not blow up. In its use, therefore, the theorem is indistinguishable from the usual pragmatic attitude toward the interchange of summation and integration. But there is, of course, this difference: Despite its informal appearance, the theorem is a precise statement whose validity can be established with full rigor, so that its use is always completely justified.

However, the similarity between the theorem and the rule of thumb does suggest an interesting thought: despite its abstract form and basis, the Lebesgue theory may, in its results, be considerably closer to engineering thought and practice than the Riemann theory. We remarked on this once before, in connection with the structure and flexibility of the Lebesgue language, and we see it again in this section, in the comparative simplicity of the Lebesgue theorems and the ease and naturalness with which they may be employed.

While in practical analysis there is no real difference (in the value of the result) between justifying an interchange by means of a theorem or proceeding by the usual combination of physical insight and casualness, it is interesting to realize that, often, an interchange performed in a purely heuristic spirit is actually rigorously defensible just as it stands, without any changes. Without any changes, that is, except for the insertion of the qualification that the integrals are to be understood in the sense of Lebesgue. This is not true with the Riemann theory. As long as Theorem 7.1 is the main available guide, so that something in the way of uniform convergence is required, formal interchanges of limit and integration (or summation and integration) are frequently indefensible. And this is so despite the fact that we find it practically impossible to think of a single example, especially a physically meaningful example, of a bounded , pointwise convergent sequence that ends up in a limit that fails to be Riemann integrable.

The difference is not in the integrals – if we just consider integrals defined on intervals, there is no practical difference between a Riemann integral and a Lebesgue integral – the difference is in the method of constructing the integrals. The new construction allows us to establish properties of the integral which, while they seemed intuitively evident, could not be proved within the framework of the Riemann theory. The abstract formulation studies more fundamental, more cogent properties of functions than does the Riemann theory, and the results are simpler, more powerful theorems. Thus it seems that although the Riemann theory is more commonsensical in its construction, it is the Lebesgue theory that, in its actual results, parallels and justifies the procedures of practical analysis.

We conclude this section with an example that illustrates the use of Theorem 7.5 and at the same time proves a property of integrals that we shall require in the next section.

EXAMPLE 7.1. In Theorem 5.7 we showed that integrals are finitely additive. That is, given any integral $\int_E f\, d\mu$, if E is the union of N measurable, disjoint sets $A_1, A_2, \ldots, A_N$, then the integral of f over E is equal to the sum of the integrals of f over each of the component sets:

$$\int_E f\, d\mu = \sum_{i=1}^{N} \int_{A_i} f\, d\mu$$

We shall extend this result to show that integrals are countably additive, so that Theorem 5.6 remains true even when the number of component sets is countably infinite. Our proof will proceed along the lines of that in the footnote to Theorem 5.7, and will serve to illustrate the use of the Lebesgue Dominated Convergence Theorem.

Let $\{A_i\}$ be a collection of pairwise disjoint measurable sets whose union is a given measurable set E, so that

$$E = \bigcup_{i=1}^{\infty} A_i, \quad A_i \cap A_j = 0 \text{ if } i \neq j, \text{ and } A_i \in \mathcal{S} \text{ for } i = 1, 2, \ldots$$

Let $\mu(E)$ be finite. We wish to prove that, for any measurable function f which is integrable on E,

$$\int_E f\, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f\, d\mu \tag{7.7}$$

PROOF: Let K_{A_i} be the characteristic function of the set A_i. Then, since every $x \in E$ belongs to exactly one A_i,

$$f = K_{A_1} f + K_{A_2} f + K_{A_3} f + \dots$$

$$= \sum_{i=1}^{\infty} K_{A_i} f \tag{7.8}$$

We notice that if it were possible to integrate the infinite series (Eq. 7.8) term by term, we could get the result of Eq. 7.7 immediately, since then

$$\int_E f \, d\mu = \int_E \left[\sum_{i=1}^{\infty} K_{A_i} f \right] d\mu = \sum_{i=1}^{\infty} \int_E K_{A_i} f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu$$

The object of the whole proof is the justification of the interchange of infinite summation and integration. We shall proceed in a roundabout way in order to make use of some of the ideas and definitions presented in previous sections.*

Let σ_n denote the n^{th} partial sum of the series Eq. 7.8,

$$\sigma_n = \sum_{i=1}^{n} K_{A_i} f \tag{7.9}$$

Then, by definition of the characteristic function, we have immediately

$$\sigma_n(x) = f(x) \quad \text{for } x \in \bigcup_{i=1}^{n} A_i$$

$$= 0 \qquad \text{for } x \in \left[E - \bigcup_{i=1}^{n} A_i \right] = \bigcup_{i=n+1}^{\infty} A_i \tag{7.10}$$

and since the A_i are measurable sets, the σ_n are measurable functions. Our proof will be divided in two parts: First, we shall show that $\sigma_n \to f$ (meas.); and then use this fact to prove countable additivity.

Let Q_n be the set of values of x for which σ_n does not equal f:

$$Q_n = \left\{ x : \sigma_n(x) \neq f(x) \right\}$$

---

*For the reader in a hurry: notice, from the definitions (Eqs. 7.9 and 7.10), that $|\sigma_n(x)| \leq f(x)$ for every $x \in E$, that f is integrable, and that $\sigma_n \to f$ pointwise. Therefore, the conditions of Theorem 7.5 are satisfied (since pointwise convergence implies convergence in measure) so that the desired interchange of summation and integration is justified.

Our problem is, essentially, to prove that $\mu(Q_n) \to 0$. From Eq. 7.10 we see that

$$Q_n \subset \left[E - \bigcup_{i=1}^{n} A_i \right] = \bigcup_{i=n+1}^{\infty} A_i$$

Therefore, because of the monotonic nature of measures (property 2, Section 3),

$$\mu(Q_n) \leqslant \mu\left[\bigcup_{i=n+1}^{\infty} A_i \right] = \sum_{i=n+1}^{\infty} \mu(A_i) \tag{7.11}$$

The last equality in Eq. 7.11 is made possible by the countable additivity of measures and the fact that the sets A_i are disjoint. We recall that the set E is a countable union of the A_i, and that E has finite measure, that is,

$$\mu(E) = \mu\left[\bigcup_{i=1}^{\infty} A_i \right] = \sum_{i=1}^{\infty} \mu(A_i) < \infty \tag{7.12}$$

Therefore it must be true that, for any $\epsilon > 0$, there exists an N sufficiently large so that for all $n \geqslant N$

$$\sum_{i=n+1}^{\infty} \mu(A_i) < \epsilon$$

Substituting this result in Eq. 7.11, we have that, for all $n \geqslant N$,

$$\mu(Q_n) < \epsilon \tag{7.13}$$

Now take any number $\delta > 0$. The set of values of x for which $|f(x) - \sigma_n(x)| \geqslant \delta$ is certainly a subset of Q_n, since Q_n is the set where $f(x) \neq \sigma_n(x)$. That is,

$$\left\{ x : |f(x) - \sigma_n(x)| \geqslant \delta \right\} \subset Q_n$$

so that, for $n \geqslant N$,

$$\mu\left[\left\{ x : |f(x) - \sigma_n(x)| \geqslant \delta \right\}\right] \leqslant \mu(Q_n) < \epsilon \tag{7.14}$$

Since this is true for any $\delta > 0$, it corresponds precisely to the definition of convergence in measure given in Section 6. Therefore, we have proved that

$$\sigma_n \to f \ \text{(meas.)}$$

Now we note that from Eq. 7.10 it follows that, for every $x \in E$,

$$|\sigma_n(x)| \leqslant |f(x)|$$

and since f was assumed integrable on E, so is $|f|$. Therefore, $\{\sigma_n\}$ is a sequence of measurable functions which converges in measure to f and is dominated by the integrable function $|f|$. Thus $\{\sigma_n\}$ fits all of the specifications of Theorem 7.5, and we have immediately

$$\lim_n \int_E \sigma_n \, d\mu = \int_E (\lim_n \sigma_n) \, d\mu \tag{7.15}$$

To identify the terms in Eq. 7.15, note that the right side is just $\int_E f \, d\mu$, and that on the left side,

$$\int_E \sigma_n \, d\mu = \int_E \sum_{i=1}^{n} K_{A_i} f \, d\mu = \sum_{i=1}^{n} \int_E K_{A_i} f \, d\mu = \sum_{i=1}^{n} \int_{A_i} f \, d\mu$$

Since

$$\lim_n \sum_{i=1}^{n} \int_{A_i} f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu$$

Eq. 7.15 becomes

$$\int_E f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu$$

as we wished to show. Thus the integral is countably additive.

8. SET FUNCTIONS, ABSOLUTE CONTINUITY, AND DIFFERENTIATION

If f is a measurable function defined on a set X, and if it is integrable on X, then its indefinite integral

$$\nu(E) = \int_E f \, d\mu \tag{8.1}$$

is defined for every measurable subset E of X. If $\mathcal{S}$ is the class of measurable subsets of X, then $\nu(E)$ is a set function defined on $\mathcal{S}$ since ν assigns a number to every $E \in \mathcal{S}$.

Using the properties of integrals derived in previous sections, we can easily determine many of the properties of the set functions defined as in Eq. 8.1. A very interesting special case occurs when $f \geqslant 0$ a.e. In this case, the set function ν is a measure function defined on the same domain as μ. To see this, we need only recall that, by definition (see Sec. 3), a measure is an extended real-valued, nonnegative, and countably additive set function, defined on a ring of sets, and such that its value for the empty set is zero. But when $f \geqslant 0$ a.e.,

$$\int_E f \, d\mu \geqslant 0$$

for any E, and

$$\int_E f \, d\mu = 0$$

if $\mu(E) = 0$. Furthermore, as we showed in Example 7.1, $\int_E f \, d\mu$ is countably additive. Thus we have proved that when $f \geqslant 0$ a.e., ν is a measure. Therefore we can generate new measures ν from a given measure μ by means of any nonnegative integrable function f. A measure defined as in Eq. 8.1 is sometimes called the <u>Lebesgue-Stieltjes measure induced by f</u>. It is a generalization of the measure defined in Example 3.5c.

It is interesting to ask, Can any set function defined on the class of measurable sets be represented as in Eq. 8.1? That is, are all set functions indefinite integrals? Quite clearly, the answer to this question is "no," because all integrals are countably additive, so that a set function must at least be countably additive, in order to qualify. But besides this, what characterizes set functions for which an f exists which makes possible the representation given in Eq. 8.1?

The search for the property that distinguishes indefinite integrals from other set functions led to the concept of absolute continuity, which is defined as follows. If μ is a measure defined on a σ-ring $\mathcal{S}$ of sets, and ν is a set function also defined on $\mathcal{S}$, then ν is absolutely continuous with respect to μ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|\nu(E)| < \epsilon$ for every measurable set E for which $\mu(E) < \delta$. In an imprecise way, absolute continuity of ν with respect to μ requires that $|\nu(E)|$ be small whenever $\mu(E)$ is small. (Absolute value signs are used around ν because it is just an arbitrary set function, not necessarily a measure, and therefore not necessarily positive.) There is no standard notation to denote absolute continuity. We shall say "ν is μ-continuous" as an abbreviation for the statement "ν is absolutely continuous with respect to μ." The relationship is not symmetric; that is, the fact that ν is μ-continuous does not, in general, imply that μ is ν-continuous.

It is easy to give an example of a set function ν which is not absolutely continuous with respect to μ. Let μ be the Lebesgue measure of the line, and let ν be the measure which to every set assigns the number of points contained in the set. Then, for a countable set E, $\nu(E) = \infty$, while $\mu(E) = 0$. Therefore there is no δ with the property that $\nu(E) < \epsilon$ whenever $\mu(E) < \delta$. The following example will help to clarify the notion of absolute continuity.

EXAMPLE 8.1. We shall be concerned with absolute continuity with respect to Lebesgue measure of the real line, and for this special case it is convenient to restate the definition of absolute continuity in a somewhat different form. A bounded function $f(x)$ defined on the interval $[a, b]$ is said to be absolutely continuous if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for every finite set of subintervals $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$ of total length less than δ

$$\sum_{k=1}^{n} (b_k - a_k) < \delta \tag{8.2}$$

it is true that

$$\sum_{k=1}^{n} |f(b_k) - f(a_k)| < \epsilon \tag{8.3}$$

This definition is derivable from our original definition, as is shown in Halmos (1), page 181.

We can use our new definition to find a large class of functions that are absolutely continuous. In fact, any function that satisfies, for every choice of x and x', the inequality

$$|f(x) - f(x')| \leq K\,|x - x'| \qquad \text{(K finite)} \tag{8.4}$$

will do. To see this, we need only form the sum Eq. 8.3 and obtain

$$\sum_{k=1}^{n} |f(b_k) - f(a_k)| \leq K \sum_{k=1}^{n} |b_k - a_k| \tag{8.5}$$

Then, given $\epsilon > 0$, we can make expression 8.5 less than ϵ simply by choosing $\delta < \epsilon/K$ because then

$$\sum_{k=1}^{n} |f(b_k) - f(a_k)| \leq K \sum_{k=1}^{n} (b_k - a_k) < K\delta < \epsilon$$

Now we shall show that any function that has a bounded derivative at every point of (a, b) satisfies the inequality of Eq. 8.4. This follows immediately from the mean value theorem, which states that if f(x) has bounded derivatives at every point of (a, b), and if x and x', $x \geq x'$, are any two points of (a, b), there exists a point ξ with the property that

$$|f(x') - f(x)| = |f'(\xi)| \cdot |x' - x| \qquad (x' \leq \xi \leq x) \tag{8.6}$$

Since the derivatives f'(x) are assumed bounded, there exists a finite K large enough so that $|f'(x)| \leq K$ for all $x \in (a, b)$. Substituting in Eq. 8.6 we have

$$|f(x') - f(x)| \leq K|x' - x|$$

which was to be shown. Thus every function that has bounded derivatives at every point is absolutely continuous. It follows immediately from Eq. 8.3, by setting n = 1, that all absolutely continuous functions are also continuous. That the converse is not true is somewhat awkward to prove. We shall simply refer the interested reader to the counterexample given in Munroe (3), page 193, or to the simpler one in Natanson (2), pages 248-9 (footnotes).

Why is absolute continuity important? Because it isolates precisely that quality which distinguishes set functions that are integrals from set functions that are not. Its importance is shown in the following two imprecise but suggestive statements:

1. All integrals are absolutely continuous: if $\nu(E) = \int_E f\, d\mu$, then ν is μ-continuous.
2. All (finite,* countably additive) absolutely continuous set functions are intetrals: if ν has the stated properties, there exists an integrable f such that $\nu(E) = \int_E f\, d\mu$.

*The finiteness of ν is not necessary, since there are perfectly good functions f that have finite integrals on some sets but not on others. Thus the condition "ν finite everywhere" can actually be relaxed to something less stringent. However, doing this involves introducing new concepts which will not add anything to our understanding of the main ideas. Therefore, in all that follows we shall always require our set functions to be finite, with the understanding that this is not necessary and that more general conditions can be found in Halmos (1), Chap. VI. Our restriction will not diminish in any way the meaningfulness of our results, and it will simplify them.

The next two theorems will render our statements more precise. We assume as usual that we are given a space X, a σ-ring $\mathcal{S}$ of subsets of X, and a finite measure μ defined on $\mathcal{S}$.

THEOREM 8.1. If f is an integrable function defined on X, and if for every measurable set E,

$$\nu(E) = \int_E f \, d\mu$$

then ν is μ-continuous.

PROOF: We shall prove the theorem only for functions f which are bounded, that is, for which $|f| \leq K$, where K is a finite number. A proof can also be found for unbounded (but integrable) f. In our case,

$$|\nu(E)| = \left| \int_E f \, d\mu \right| \leq \int_E |f| \, d\mu \leq K \mu(E)$$

so that $|\nu(E)| < \epsilon$ if $\mu(E) < \delta = \epsilon/K$. We have thus shown that there exists a δ such that $|\nu(E)| < \epsilon$ whenever $\mu(E) < \delta$. Therefore ν is μ-continuous.

The second proposition is more difficult. It is essentially the Radon-Nikodym theorem which, stated precisely, reads as follows.

THEOREM 8.2. (Radon-Nikodym) Let μ be a finite measure defined on $\mathcal{S}$, and let ν be a finite, countably additive set function also defined on $\mathcal{S}$. If ν is absolutely continuous with respect to μ, then there exists a finite-valued integrable function f defined on X with the property that

$$\nu(E) = \int_E f \, d\mu$$

for every measurable set E. The function f is unique up to an equivalence $[\mu]$. That is, if it is also true that

$$\nu(E) = \int_E g \, d\mu$$

then $f = g$ a.e. $[\mu]$

One of the most interesting things about this theorem is that it gives us one possible approach to the problem of defining differentiation for set functions. The indefinite integral

$$\nu(E) = \int_E f \, d\mu$$

which holds for any measurable set E, suggests that the Radon-Nikodym integrand f might be regarded as the derivative of ν with respect to μ, in the sense that f, when integrated with respect to μ, yields ν. The suggestion is strengthened by defining a new notation for Radon-Nikodym integrands. We shall write

$$f = \frac{d\nu}{d\mu} \qquad \text{or} \qquad f \, d\mu = d\nu$$

and instead of calling f a Radon-Nikodym integrand, we shall call it a Radon-Nikodym derivative.*

The value of this notation is enhanced by the fact that all of the properties of $d\nu/d\mu$ that are suggested by the ordinary differential formalism turn out to correspond to true theorems, with the qualification that they do not hold everywhere, only almost everywhere.

EXAMPLE 8.2. If μ, ν, λ are set functions that satisfy the requirements of the Radon-Nikodym theorem, then we have:

a. if ν is μ-continuous and λ is μ-continuous,

$$\frac{d}{d\mu}(\nu + \lambda) = \frac{d\nu}{d\mu} + \frac{d\lambda}{d\mu} \qquad \text{a.e. } [\mu]$$

b. if ν is μ-continuous and μ is λ-continuous,

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \qquad \text{a.e. } [\lambda]$$

c. if ν is μ-continuous and at the same time μ is ν-continuous,

$$\frac{d\nu}{d\mu} = 1 \Big/ \left(\frac{d\mu}{d\nu} \right) \qquad \text{a.e. } [\mu]$$

d. if μ is λ-continuous and if f is integrable with respect to μ,

$$\int f \, d\mu = \int f \frac{d\mu}{d\lambda} \, d\lambda$$

Formally, therefore, the Radon-Nikodym derivative bears a very close resemblance

*The use of differential notation for Radon-Nikodym integrands does not in any way imply that a connection can be shown to exist, in general, between ordinary derivatives and the symbols written above, nor that f can be obtained from ν by the ordinary processes of differentiation. Regardless of their usefulness in suggesting new ideas, the symbols themselves only mean what they are defined to mean.

to the ordinary derivative.

Unfortunately, so far as the specific purpose of defining differentiation is concerned, there is something rather sterile about the approach used above, because, while the Radon-Nikodym theorem asserts the existence of the corresponding integrand, it does not give the slightest indication of how to determine it. What is wanted in that case is a constructive definition of differentiation, that is, one that shows how to determine the derivative of a given function. Such an approach must start from something that resembles the limit process that is used in defining ordinary differentiation. It then becomes necessary to show that the two definitions are consistent, and to prove the truth, for example, of a statement like: A function is equal to the integral of its derivative. These problems become rather complex, and will not be considered here. A general treatment is given in Munroe (3), Chap. VII. The special case of differentiation on the real line is treated in Natanson (2), Chap. IX, and Burkill (5), Chap. IV.

APPENDIX I

In this section we shall present some useful definitions and results from elementary set algebra.

If A and B are two sets, the union of A and B, written $A \cup B$, is the set of all points that belong either to A, or to B, or to both. In other words, $A \cup B$ is the set of all points that belong to at least one of the sets A and B, For example, if sets A and B are thought of as circular regions, as shown in Fig. AI-1, then $A \cup B$ is the set of points contained within the dotted line. The union of n sets $E_1, E_2, \ldots, E_n$ is, as before, the set of all points that belong to at least one of the sets E_i $(i = 1, 2, \ldots, n)$. It is written

$$E_1 \cup E_2 \cup \ldots \cup E_n$$

or, more conveniently,

$$\bigcup_{i=1}^{n} E_i$$

For a countable* sequence of sets $E_1, E_2, E_3, \ldots$ the union is defined in the same way and is denoted by

$$\bigcup_{i=1}^{\infty} E_i$$

If A and B are two sets, the intersection of A and B, written $A \cap B$, is the set of all points common to A and B. If sets A and B are thought of as circular regions, as shown in Fig. AI-2, then $A \cap B$ is the set of points shown shaded. For n sets $E_1, E_2, \ldots, E_n$, the set of all points that belong simultaneously to all n sets is denoted by

$$\bigcap_{i=1}^{n} E_i$$

*A sequence, or a set, is said to be countable (or denumerable) if its members can be put into a one-to-one correspondence with the members of the set of all positive integers. Thus a countable set is infinite, but, speaking loosely, it is the smallest type of infinity. As an example, the set E of all positive even integers is countable since, if $e \in E$, we can establish the required correspondence with the integers by letting e correspond to the positive integer $e/2$. It has been shown (see, for example, Rudin (4), p. 23) that the set of all rational numbers is countable, but that the set of all irrational numbers is uncountable (or nondenumerable). Speaking loosely again, this means that there are many more irrational numbers than there are rationals.

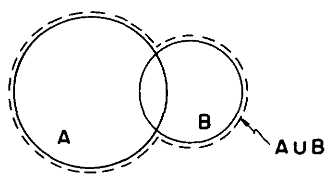

Fig. AI-1. Union of two sets.

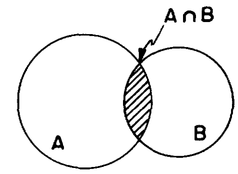

Fig. AI-2. Intersection of two sets.

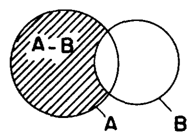

Fig. AI-3. Difference of two sets.

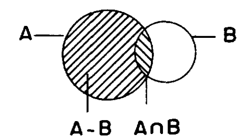

Fig. AI-4. Illustrating Identity I.

and the intersection of a countable number of sets is written

$$\bigcap_{i=1}^{\infty} E_i$$

The operations of union and intersection of sets are very similar to addition and multiplication of numbers. As in arithmetic, the commutative and associative laws are true for unions:

$$A \cup B = B \cup A$$

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

and similarily for intersections. The distributive law also holds:

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

If A and B are two sets, the <u>difference</u> A - B is the set of all points that belong to A and not to B. Fig. AI-3, where A - B is shown shaded, illustrates the relationship.

The <u>empty</u> (or <u>vacuous</u>) set is the set that has no members. It is denoted by 0. If the intersection of two sets A and B contains no points (i.e., if A and B have no points in common) we write $A \cap B = 0$, and say that A and B are disjoint.

As examples of our definitions, we shall prove three identities which, while not particularly important in themselves, are useful in the discussions in Sections 3 and 4.

IDENTITY I. $A \cap B = A - (A-B)$ (AI. 1)

That this relation is plausible can be seen immediately from Fig. AI-4. The points that belong to the intersection are shown crisscrossed; the set (A–B), the points that belong to A but not to B, are shown shaded. From the figure it follows immediately that the points that belong to A but not to (A–B) are just the points common to A and

B, which is what is asserted in Eq. AI-1. For brevity, let $E = A \cap B$, $F = A - (A-B)$. Then, for an analytical proof of the equality of E and F, we must show (by definition of equality of two sets, Section 2) that all the points of E belong to F $(E \subset F)$ and at the same time all the points of F belong to E $(F \subset E)$. If these two requirements are satisfied, then $E = F$. We show this in two steps:

1. If $x \in E$, then, by definition of intersection, $x \in A$ and $x \in B$. Therefore, $x \notin (A-B)$. Therefore, since x is a point of A which is not in $(A-B)$, $x \in [A - (A-B)]$. Thus $x \in E$ implies $x \in F$, which means that $E \subset F$.

2. Now suppose that $x \in F$. Then it must be true that $x \in A$ but not in $(A-B)$. Therefore $x \in B$. Thus x is a point common to A and B, so that $x \in A \cap B = E$. Therefore $x \in F$ implies $x \in E$, which means $F \subset E$.

Results 1 and 2 establish that $E = F$, which was to be shown.

IDENTITY II. $$\bigcap_{i=1}^{\infty} E_i = E_1 - \bigcup_{i=1}^{\infty} (E_1 - E_i)$$

This relation is a generalization of Identity I to the case in which instead of two sets, we have a countable number of sets. The interpretation and proof are exactly analogous to those for Identity I. Let

$$E = \bigcap_{i=1}^{\infty} E_i, \qquad F = E_1 - \bigcup_{i=1}^{\infty} (E_1 - E_i)$$

Then,

1. If $x \in E$, it follows that $x \in E_i$ for all i. Therefore $x \notin (E_1 - E_i)$ $(i = 1, 2, \ldots)$, and so $x \notin \bigcup_{i=1}^{\infty} (E_1 - E_i)$. Thus $x \in F$, which shows that $x \in E$ implies $x \in F$. Therefore $E \subset F$.

2. If, on the other hand, $x \in F$, then $x \in E_1$ but $x \notin \bigcup_{i=1}^{\infty} (E_1 - E_i)$. Therefore $x \notin (E_1 - E_i)$ for any value of i. But since $x \in E_1$, this implies that $x \in E_i$ for all values of i. Therefore x is common to all E_i $(i = 1,2,\ldots)$, and so $x \in E$. Thus $F \subset E$.

Results 1 and 2 show that $E = F$, which establishes the desired identity.

IDENTITY III. The usefulness and meaning of this identity will become apparent to the reader when he reaches Section 4. Here we limit ourselves to its statement and proof.

Let $\{f_n(x)\}$ be a convergent sequence of functions, and let the sequence be monotone nondecreasing, that is,

$$f_1(x) \leq f_2(x) \leq f_3(x) \leq \ldots \tag{AI. 2}$$

Define $f(x) = \lim_{n \to \infty} f_n(x)$. Then

$$\{x : f(x) > \alpha\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) > \alpha\} \tag{AI. 3}$$

Proof. Let

$$A = \{x : f(x) > \alpha\}, \qquad B = \bigcup_{n=1}^{\infty} \{x : f_n(x) > \alpha\}$$

To prove that Eq. AI. 3 is correct, we must show that $A \subset B$ and $B \subset A$, which means that $A = B$.

1. Consider any $x \in B$. Since $x \in B$, it must be true for some value of n, say N, that $f_N(x) > \alpha$. But then, from Eq. AI. 2 it follows that $f_n(x) > \alpha$ for all $n \geq N$. Thus it must also be true in the limit that $f(x) > \alpha$, so that $x \in A$. Therefore $x \in B$ implies $x \in A$, which means that $B \subset A$.

2. To prove the inverse relation, choose any $x \in A$. For that value of x, $f(x) > \alpha$. Since the sequence $\{f_n(x)\}$ converges to f(x), given any $\epsilon > 0$, there must exist an N such that for all $n \geq N$, $f(x) - f_n(x) < \epsilon$, so that $f_n(x) > f(x) - \epsilon$. Choose $\epsilon = (f(x) - \alpha)/2$ (ϵ is greater than zero, since $f(x) > \alpha$). Then

$$f_n(x) > f(x) - \left[\frac{f(x) - \alpha}{2}\right] = f(x) - \frac{f(x)}{2} + \frac{\alpha}{2} = \frac{f(x)}{2} + \frac{\alpha}{2} > \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

We have proved that if $x \in A$, then there exists an n with the property that, for the chosen value of x, $f_n(x) > \alpha$. For that n, then, the set $\{x : f_n(x) > \alpha\}$ is nonempty, and our chosen value of x belongs to the set. Since the set is a subset of B, it follows that $x \in B$. Therefore $x \in A$ implies $x \in B$, which means that $A \subset B$.

The results of 1 and 2 show that $A = B$.

APPENDIX II

We present here some examples that illustrate the discussion of Section 7.

EXAMPLE AII-1. Consider the sequence of functions $\left\{K_{A_n}\right\}$ where K_{A_n} is the characteristic function of the interval $\left(0, \frac{1}{n}\right)$. It was shown in Example 6.2 that $K_{A_n} \to 0$ for all $x \in (0, 1)$ but that the convergence is not uniform. We have

$$\int_0^1 K_{A_n}(x)\, dx = \int_0^{1/n} 1\, dx = \frac{1}{n}$$

so that, in spite of nonuniformity of convergence,

$$\lim_n \int_0^1 K_{A_n}\, dx = \int_0^1 \left[\lim_n K_{A_n}\right] dx = 0$$

EXAMPLE AII-2. Consider the sequence $\{f_n\}$ defined in Example 6.4. It was shown there that $f_n(x) \to 0$ for all $x \in [0, 1]$, but that the convergence is not uniform. Clearly,

$$\int_0^1 \left[\lim_n f_n\right] dx = 0$$

On the other hand,

$$\int_0^1 f_n(x)\, dx = \int_0^{1/n} n\, dx = 1$$

so that

$$\lim_n \int_0^1 f_n\, dx = 1 \neq \int_0^1 \left[\lim_n f_n\right] dx$$

We notice that the functions f_n are rectangles of height n and width $1/n$, so that they become taller and narrower as the limit is approached, while their area remains constant. There is a δ-function lurking here, and one way to exorcise it is to require that the sequence be bounded.

EXAMPLE AII-3. Consider the sequence of functions $\{f_n(x)\}$ defined on the unit interval $[0, 1]$ by

$$f_n(x) = \mathrm{sgn}\left[\sin^2 n!\, \pi x\right] \qquad (n = 1, 2, 3, \ldots)$$

where sgn(y) is defined by

$$\operatorname{sgn}(y) = \begin{cases} -1 & \text{if } y < 0 \\ 0 & \text{if } y = 0 \\ 1 & \text{if } y > 0 \end{cases}$$

Then, for rational values of x, say $x = p/q$, $\sin^2 n!\,\pi(p/q) = 0$ for all $n \geqslant q$, so that, for rational values of x, $f_n(x) \to 0$ as $n \to \infty$. On the other hand, for irrational values of x, $n!\,\pi x$ will never be an integer multiple of π, so that $\sin^2 n!\,\pi x > 0$. Therefore, for irrational values of x, $f_n(x) = 1$ for all n, so that, for these values of x,

$$\lim_n f_n(x) = 1$$

Thus the limit of the sequence $\{f_n(x)\}$ is the characteristic function of the set of irrationals of the unit interval. This function, as we saw in Section 1, is discontinuous at every point of $[0, 1]$, so that a Riemann integral cannot be defined for it.

On the other hand, for every finite n, $f_n(x)$ is a bounded function (its values being either zero or one) and $f_n(x)$ is certainly Riemann integrable, since it equals one everywhere except at the points $x = \frac{k}{n!}$ $(k = 0, 1, \ldots, n!)$, so that it has only a finite number of discontinuities. Thus $\{f_n\}$ is a convergent, bounded sequence of Riemann integrable functions whose limit is not Riemann integrable.

EXAMPLE AII-4. Let $f_n(x) = x^n$ on $[0, 1]$. Then each $f_n(x)$ is continuous. On the other hand,

$$\lim_n x^n = 0 \qquad \text{for } 0 \leqslant x < 1$$

$$\lim_n x^n = 1 \qquad \text{for } x = 1$$

so that the limit is discontinuous.

EXAMPLE AII-5. Consider the sequence $\{f_n(x)\}$, where

$$f_n(x) = \frac{1}{nx} \qquad \text{for } 0 < x \leqslant 1$$

(Note that the point $x = 0$ is not included in the interval of definition, so that every $f_n(x)$ is continuous and finite.) Clearly, $f_n(x) \to 0$ as $n \to \infty$, for all $0 < x \leqslant 1$, so that the limit, being constant, is continuous. On the other hand, the convergence is not uniform, since, given $\epsilon > 0$, it is impossible to find one single N such that, for all $n \geqslant N$, $1/(nx) < \epsilon$ for <u>every</u> value of x. The appropriate N is necessarily a function of x.

EXAMPLE AII-6. If we disregard the boundedness requirement of Theorem 7.4, the sequence of Example AII-2 fits all the requirements of the theorem (with the measure μ taken as the Lebesgue measure of the line). Thus, from Example AII-2, we see that the conclusion of the theorem is false if the boundedness condition is removed.

BIBLIOGRAPHY

1. P. R. Halmos, Measure Theory (D. Van Nostrand Co., Inc., New York, 1956).
2. I. P. Natanson, Theorie der Funktionen Einer Reellen Veränderlichen (Akademie-Verlag, Berlin, 1954).
3. M. E. Munroe, Introduction to Measure and Integration (Addison-Wesley, Cambridge, Mass., 1953).
4. W. Rudin, Principles of Mathematical Analysis (McGraw-Hill Book Co., Inc., New York, 1953).
5. J. C. Burkill, The Lebesgue Integral (Cambridge University Press, London, 1951).
6. E. T. Whittaker and G. N. Watson, Modern Analysis (Cambridge University Press, London, 1952).
7. A. N. Kolmogoroff, Foundations of the Theory of Probability (Chelsea Publishing Co., New York, 1950).
8. N. Wiener, The Fourier Integral (Dover Publications, New York, 1951).
9. E. J. McShane, Integration (Princeton University Press, Princeton, 1944).
10. W. Feller, Probability Theory and Its Applications (John Wiley and Sons, Inc., New York, 1951).
11. F. Riesz and B. Sz.-Nagy, Functional Analysis (Frederick Ungar Publishing Co., New York, 1955).
12. L. M. Graves, Theory of Functions of Real Variables (McGraw-Hill Book Co., Inc., New York, 1946), 1st edition.

Index of Definitions