1

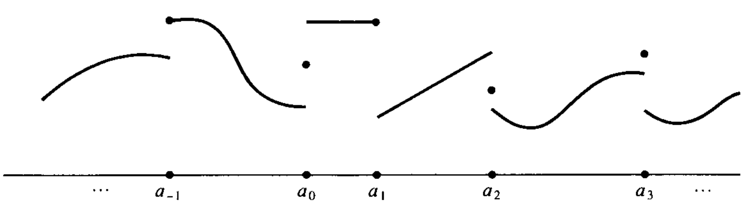
Introduction

A 'distribution' is a kind of 'generalized function'. That is, every reasonable function $f(x)$ corresponds to a distribution, but there are distributions which do not correspond to functions. Here we must define what we mean by a 'reasonable' function.

By a reasonable function (or ordinary function or *bona fide* function), we mean a piecewise continuous function $f(x)$ of one real variable x (see Figure 0). This class of functions is sufficient as a starting point. The values of f may be either real or complex numbers, but the variable x is real. In a later chapter we shall extend the theory to several variables.

Now the class of distributions or 'generalized functions' includes many objects which are not functions at all. Why do we study these? The reason is not mere generality. Rather, the theory of distributions has a coherence and power that the classical theory of functions lacks. There are many aspects of this conceptual power which we hope to demonstrate later on, but one

Figure 0. A typical piecewise continuous function $f(x)$. The function is continuous except at the points a_i. For each a_i, the function $f(x)$ has finite left and right hand limits as $x \to a_i$, but these limits may differ (producing the 'jumps'). The value of $f(x)$ at $x = a_i$ itself is immaterial. The partition points a_i are either finite in number, or else approach infinity as $i \to \infty$ and approach minus infinity as $i \to -\infty$.



of them can be mentioned right away: the operation of taking a derivative applies without restriction to distributions. That is, the derivative of a distribution always exists and is another distribution. By contrast, there are many continuous functions which have no derivative in the classical sense. This defect in the classical approach is eliminated by distribution theory, for, since continuous functions are distributions, the derivative will exist as a distribution.

In this introductory chapter, we intend to lead up to the theory of distributions by a series of intuitive steps based on physical or phenomenological considerations. We interrupt the flow briefly in the next section to give a rigorous definition of the term 'test function'. Since the whole theory is based on test functions, it seemed worthwhile to pin this idea down before proceeding further.

Technical note. For readers familiar with measure theory, we remark that we could replace the class of piecewise continuous functions introduced above (Figure 0) by functions which are locally integrable. One indication of the power of distribution theory is that this extension adds no generality. Every locally integrable function is the derivative of a continuous function!

1 Test functions

This section is slightly technical because of the need to carry out a construction. The reader may prefer to skim it, glance at Figures 1 and 2, and pass on to the next section. As mentioned above, in these early chapters we shall concentrate mainly on functions of one real variable.

A *test function* is a C^∞ function with compact support: 'C^∞' means that the function has continuous derivatives of all orders, and 'compact support' describes functions which vanish outside of some bounded set. The motivations underlying the term 'test function' will be developed as we proceed. But the essential idea is this: curves having certain crude geometrical shapes (like pulses or mesas – see Figures 1 and 2) can be constructed in an infinitely differentiable manner.

We now give the construction. As our point of departure we take

$$h(x) = \begin{cases} e^{-1/x} & \text{for } x > 0, \\ 0 & \text{for } x \leqslant 0. \end{cases}$$

(*The function* $h(x)$ is often presented to students of elementary calculus as a counterexample. It is curious that this erstwhile counterexample has since become the foundation of a major theory.) The crucial property of $h(x)$ is that, at the transition point $x = 0$, all of its derivatives exist and are zero. This is easily checked. We simply apply the standard rules of calculus to $e^{-1/x}$ and observe that the nth derivative must have the general form

$(d/dx)^n(e^{-1/x}) = (a_N x^{-N} + \cdots + a_0)\cdot e^{-1/x}$. No matter what the values of $a_N, \ldots, a_0$ are, the factor $e^{-1/x}$ decreases so rapidly as $x \to 0^+$ that it annihilates all of the other terms.

Now to construct our first basic test function $\varphi(x)$ we set (see Figure 1a):

$$\varphi(x) = h(x)h(1-x).$$

From this function many others can be constructed. Recall that the *support* of a function f is the closure of the set of x-values for which $f(x) \neq 0$. Thus the function φ in Figure 1a has support on $[0, 1]$. To build a test function with support on any preassigned interval $[a, b]$, we simply set (see Figure 1b)

$$\varphi_{a,b}(x) = \varphi\left(\frac{x-a}{b-a}\right).$$

Figure 1a. A pulse function supported on [0, 1].

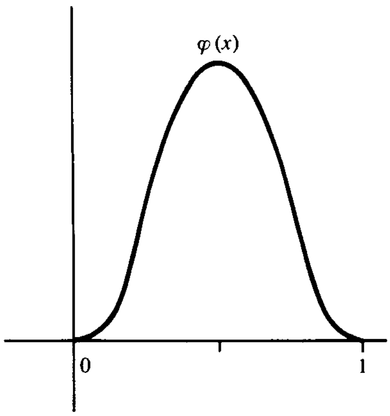
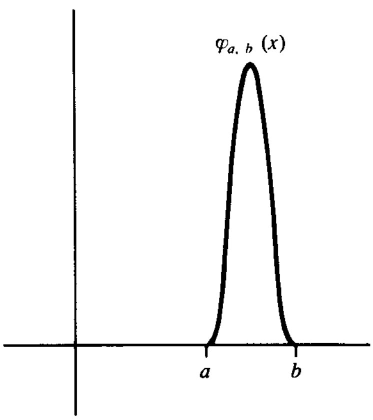


Figure 1b. The same pulse shifted and contracted so that its support is $[a, b]$.



A useful extension of this idea is to integrate the pulse $\varphi_{a,b}$, i.e. to let

$$\psi_{a,b}(x) = \int_{-\infty}^{x} \varphi_{a,b}(u)\,du.$$

Then $\psi_{a,b}$ is a C^∞ function which climbs up from zerto to a positive value and then levels off; the climbing takes place on the interval $[a, b]$. By glueing two such curves together we create a 'mesa function' (Figure 2).

Although we are mainly interested in functions of one variable, we may as well show how to extend these constructions from one to several dimensions. This is easily accomplished by multiplication. For example, to build a pulse on the unit square $\{0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ in $\mathbb{R}^2$, we would take the function $\varphi(x)$ above and then multiply $\varphi(x)\varphi(y)$. Mesas can be extended in a like manner.

Before proceeding further, we should clarify the role that test functions play in the theory. From the viewpoint of applied mathematics, the test functions are not important. In fact, the authors can think of no test function which is of the slightest interest, in and of itself. *The test functions serve as tools in studying other functions.*

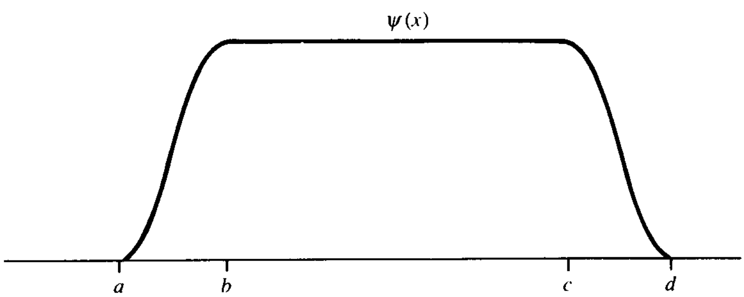
2 Three modes for the evaluation of functions

The first mode is the standard pointwise evaluation of $y = f(x)$ at each particular point x. The second mode involves averaging over intervals $\{a \leqslant x \leqslant b\}$, and the third mode uses test functions. We will argue that the third mode is the most powerful, and in a sense the most natural.

(As promised in the preface, we are going to concentrate for the time being on functions of one variable.)

We begin with the first mode. Pointwise evaluation has the advantage of logical simplicity. This may explain why it came first historically, and why it still comes first in everyone's mathematical education. Its defects are a certain rigidity and the fact that it does not correspond to physical reality.

Figure 2. A mesa function.



Consider, for example, a chemist studying some property of a substance at temperature T_o. He or she cannot achieve the temperature T_o exactly, or even achieve a uniform temperature: almost certainly, the temperatures throughout the system will vary over some range $a \leqslant T \leqslant b$. This suggests that the correct mathematical model might involve averaging over $a \leqslant T \leqslant b$ (the second mode above). However, even that is not an accurate description of reality. If a and b are the extreme temperatures, then temperatures close to these extremes will probably occur rarely, and the bulk of the temperatures will be bunched somewhere in the middle. A possible temperature density is indicated by the function $\varphi(T)$ in Figure 3. If the function $\varphi(T)$ is infinitely differentiable, a reasonable hypothesis, then it is a test function.

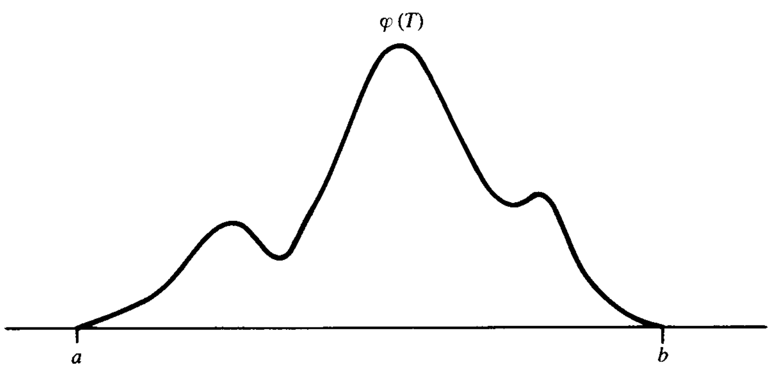
At this stage we can say a little more about the origin of the term 'test function'. The chemist wants to test the properties of some substance at temperature T_o. So he brings together a batch of the stuff with temperatures distributed in a pulse $\varphi(T)$ near $T = T_o$. The objective is to find some scientific law, and the batch of chemicals (the pulse $\varphi(T)$) tests the law. In a mathematical model, the scientific law will normally be represented by a function, say $f(T)$. Now, just like the chemist, we are interested in f, not φ! We use φ to test f. Let us see how this works out mathematically.

Return briefly to the second mode, averaging. The average value of a function $f(T)$ over an interval $[a, b]$ is $[1/(b-a)] \int_a^b f(T)\,\mathrm{d}T$. The division by $(b-a)$ can be carried out arithmetically, so we may as well consider

$$\int_a^b f(T)\,\mathrm{d}T.$$

Of course, this corresponds in our model to a uniform temperature distribution over $a \leqslant T \leqslant b$. In the true situation, the temperatures have a

Figure 3. A typical pulse $\varphi(T)$ representing a possible temperature density on the interval $a \leqslant T \leqslant b$. The values a and b are the extreme temperatures, so that $\varphi(T)$ vanishes outside of the interval $[a, b]$.



density $\varphi(T)$ supported on $a \leqslant T \leqslant b$. From this we get a 'weighted average',

$$\int_a^b f(T)\varphi(T)\,dT,$$

in which different temperatures T are weighted according to their frequency $\varphi(T)$ of occurrence. This, of course, is the third mode. It forms the basis for distribution theory.

(From now on we shall drop the second, or averaging, mode. If pushed to its logical conclusion it leads to measure theory. That development is not our objective in this book.)

To summarize our conclusions so far: we are measuring f, and using φ as a tester or probe to do it. Different φ's can probe different temperature ranges. By applying enough probes φ to the same function f, we eventually obtain a knowledge of the structure of f. (Or, as the chemist might say, by enough experiments φ we learn the 'law' described by f.)

There is one technical change we must make in the last formula above, and then we will have our basic definition. Because a and b are the extreme temperature values, there is no harm in extending the integration beyond a or b:

$$\int_a^b f(T)\varphi(T)\,dT = \int_{-\infty}^{\infty} f(T)\varphi(T)\,dT,$$

since $\varphi(T) = 0$ for T not in $[a, b]$. It is convenient to use $\int_{-\infty}^{\infty}$ for reasons of homogeneity. But the reader should understand that the integration is really over a finite interval $[a, b]$ determined by the support of φ.

Definition. Let f be a real or complex valued function of one real variable. Suppose that $\int_a^b |f(x)|\,dx$ exists and is finite for any finite interval $[a, b]$. Then the action of an arbitrary test function φ upon f is defined to be

$$\int_{-\infty}^{\infty} f(x)\varphi(x)\,dx.$$

(Sometimes for brevity we write $\int f\varphi$ in place of $\int_{-\infty}^{\infty} f(x)\varphi(x)\,dx$.)

To underline the fact that many test functions φ are used to probe a particular function f, we prove our first theorem. (It is curious that one book on distribution theory omits this theorem entirely. The book stresses – quite correctly – that the simplicity of the theory is based on the severe restrictions imposed on the test functions. It is left for the reader to observe that the theory would be even simpler if there were no test functions at all! Every theorem in that book would be true. Only the following would fail.)

Theorem 1.1. Let f and g be continuous real valued functions of one real variable. Suppose that every test function φ has the same action on f and g, that is $\int f\varphi = \int g\varphi$ for all φ. Then $f = g$.

Proof. The idea of the proof is to construct a narrow 'pulse' $\varphi(x)$ supported on a very short interval $[a, b]$, where $f(x) \neq g(x)$ (see Figure 4). This is what the chemist does when making an accurate experiment, so that the extreme temperatures a and b are very close together. Incidentally, we shall here bid farewell to our hypothetical chemist; the reader has probably grown quite tired of him. Now for the details of the proof.

Suppose $f \neq g$. Then there is some point x_0 where $f(x_0) \neq g(x_0)$, and without loss of generality we can assume that $f(x_0) > g(x_0)$. By continuity, there is some $\varepsilon > 0$ and some $\delta > 0$ such that

$$f(x) \geqslant g(x) + \varepsilon \qquad \text{for } x_0 - \delta \leqslant x \leqslant x_0 + \delta.$$

(Thus for our interval $[a, b]$ we take $[x_0 - \delta, x_0 + \delta]$.) Let $\varphi(x) \geqslant 0$ be a test function (not identically zero) supported on the interval $[x_0 - \delta, x_0 + \delta]$; the existence of such a φ was proved in the previous section. Then,

$$\int_{-\infty}^{\infty} f(x)\varphi(x)\,dx - \int_{-\infty}^{\infty} g(x)\varphi(x)\,dx = \int_{x_0-\delta}^{x_0+\delta} [f(x) - g(x)]\varphi(x)\,dx$$

(because the support of φ is on $[x_0 - \delta, x_0 + \delta]$)

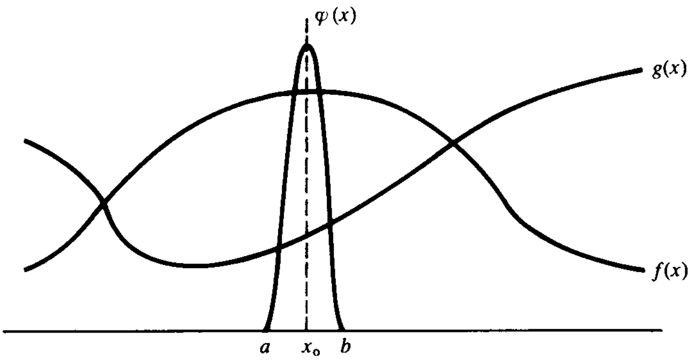
$$\geqslant \int_{x_0-\delta}^{x_0+\delta} \varepsilon \cdot \varphi(x)\,dx > 0,$$

since $f - g \geqslant \varepsilon$ on this interval, $\varphi \geqslant 0$, and φ does not vanish identically.

Hence $\int f\varphi > \int g\varphi$; that is, φ distinguishes between f and g, as desired.

Remark. Of course, the continuity of f and g is essential in this theorem. If, for example, we changed the definition of g at one point, then the operation $\int g\varphi$ – which depends on integration – would not be altered.

Figure 4. Schematic picture of the proof of Theorem 1.1. We select a small interval $[a, b] = [x_0 - \delta, x_0 + \delta]$, where $f(x) > g(x)$. Then we take a test function $\varphi(x)$ supported on $[a, b]$ and verify that $\int f\varphi > \int g\varphi$.



3 Distributions

The time has come to define what we mean by a 'distribution'. The idea is already implicit in the preceding section. We have seen that we can probe a function $f(x)$ (which is the object of study) with test functions $\varphi(x)$, and that each $\varphi(x)$ is mapped into the number $\int f\varphi = \int_{-\infty}^{\infty} f(x)\varphi(x)\,dx$. Thus we decide that:

A distribution is a mapping from test functions to numbers.

The above statement leaves something out. Not all mappings qualify as distributions. There are two side conditions. One of these is slightly technical: it involves the idea of a sequence of test functions $\varphi_1(x), \varphi_2(x), \varphi_3(x), \ldots$ 'converging to zero'. This is spelled out in Chapter 2. Accepting for now that there is such a notion, we make the following tentative definition.

Definition (provisional). A distribution T is a mapping from test functions to numbers with the following properties. (For a distribution T and a test function φ, we write the action of T on φ as $\langle T, \varphi\rangle$.)

(a) (Linearity.) $\langle T, a\varphi(x) + b\psi(x)\rangle = a\cdot\langle T, \varphi(x)\rangle + b\cdot\langle T, \psi(x)\rangle$ for all test functions φ, ψ and all constants a, b.

(b) (Continuity.) If a sequence of test functions $\varphi_1(x), \varphi_2(x), \ldots$ 'converges to zero' (definition in Chapter 2), then

$$\langle T, \varphi_n(x)\rangle \to 0.$$

Remarks on notation. The 'inner product' notation $\langle T, \varphi\rangle$ is standard in distribution theory. It reminds us that the transition from functions to distributions is given by an integral $\int f\varphi = \int_{-\infty}^{\infty} f(x)\varphi(x)\,dx$. Incidentally, this is a 'real inner product'; i.e. even if the functions f or φ are complex valued, the integral involves no complex conjugate.

The notation $\langle T, \varphi\rangle$ also brings out the 'duality' which is implicit in our definition: just as different test functions φ can act on a single distribution T, different distributions can act on the same test function. Now we make the further notational conventions:

(1) We use small Latin letters like $f, g, h, \ldots$ for functions and the capital letters $S, T, U, \ldots$ for distributions. (Test functions are denoted by φ, ψ, etc.)

(2) Every piecewise continuous function f gives rise to a distribution, but (as we shall see) not every distribution comes from a function. When a distribution does come from a function, and we want to be careful about logical distinctions, we write:

$f =$ the function evaluated pointwise ($x \mapsto f(x)$);

$T_f =$ the distribution corresponding to f,

$$\langle T_f, \varphi\rangle = \int_{-\infty}^{\infty} f(x)\varphi(x)\,dx.$$

As an example of how distribution theory works, we will now see how to take the 'derivative' of a non-differentiable function. We break the discussion into three steps.

Step 1. As our starting point, we ask what the operation would look like if the function f *were* continuously differentiable. This question has already been answered: the interaction between a function $f'(x)$ and a test function $\varphi(x)$ is given by

$$\int_{-\infty}^{\infty} f'(x)\varphi(x)\,dx.$$

Step 2. The next step forms a model for every construct in the theory of distributions. We try to 'get rid of' the derivative operation on f by 'doing something' to the test function φ. This is achieved using integration by parts. Remembering that $\varphi(x)$ (which has compact support) vanishes as $x \mapsto \pm\infty$, we find that

$$\int_{-\infty}^{\infty} f'(x)\varphi(x)\,dx = -\int_{-\infty}^{\infty} f(x)\varphi'(x)\,dx.$$

On the right hand side we have $f(x)$ (without derivative) – the operation of differentiation has been carried over to φ. The derivative of φ exists since φ is C^∞.

Step 3. Now, for the last step we consider a function $f(x)$ which may not be differentiable. For simplicity, we assume that f is at least piecewise continuous (see Figure 0). We have seen that if f *were* differentiable, then

$$\int f'\varphi = -\int f\varphi',$$

and we have noted that the right hand side involves only f and not f'. So we *define* the distribution f' to be the operation which carries every test function φ into the number $-\int_{-\infty}^{\infty} f(x)\varphi'(x)\,dx$.

The same idea extends to arbitrary distributions T. Recall that there we write $\langle T, \varphi\rangle$ in place of $\int T\varphi$. Then we define T' by the formula

$$\langle T', \varphi\rangle = -\langle T, \varphi'\rangle.$$

We will now give our first example of a distribution that is not an ordinary function. This is the famous 'Dirac Delta Function' $\delta(x)$ (pictured crudely in Figure 5b). Strictly speaking, $\delta(x)$ is not a function of x at all.

Remark A. (See, however, the antithesis in Remark B below.) In elementary texts and lectures, the delta function is sometimes treated as though it were an ordinary function, taking a value 'infinity' at $x = 0$. The pulse is assumed to have 'infinite height and width zero, but total area = 1'. This means that we are led to a number system with an '∞' so that $\infty \cdot 0 = 1$. Logically this leads to difficulties. For example, if we suppose that $\infty \cdot 0 = 1$, then there must be a number '$2 \cdot \infty$' so that $(2 \cdot \infty) \cdot 0 = 2$. The trouble is that the laws of algebra fail in this system: $(2 \cdot \infty) \cdot 0 = 2$ but $\infty \cdot (2 \cdot 0) = \infty \cdot 0 = 1$.

The theory of distributions provides an easy way out of this impasse. We start with the discontinuous function $F(x)$ shown in Figure 5a:

$$F(x) = \begin{cases} 1 & \text{for } x \geqslant 0 \\ 0 & \text{for } x < 0. \end{cases}$$

Classically the derivative $F'(x) = 0$, for $x \neq 0$, and $F'(0)$ does not exist. Intuitively, however, we could think of $F'(x)$ as being a unit 'impulse' located at $x = 0$. Now, in distribution theory, this intuitive perception becomes a rigorous fact. Since $F(x)$ (not $F'(x)$) is a *bona fide* function, we have, from our distribution–theoretic definition of F',

$$\begin{aligned} \langle F', \varphi \rangle &= -\langle F, \varphi' \rangle \\ &= -\int_{-\infty}^{\infty} F(x)\varphi'(x)\, dx \\ &= -\int_{0}^{\infty} \varphi'(x)\, dx \quad \text{(by definition of } F\text{)} \\ &= -\varphi(x)\big|_0^{\infty} \quad \text{(by elementary calculus)} \\ &= \varphi(0) \quad \text{(since } \varphi(x) \text{ vanishes as } x \to \infty\text{)}. \end{aligned}$$

Figure 5. (a) The 'step' function $F(x)$ whose derivative is $\delta(x)$. (b) A schematic picture of the delta function $\delta(x)$. The pulse is to be viewed as 'very thin and very high', with its total area equal to 1. It is centered above the point $x = 0$.