JOSÉ SEBASTIÃO E SILVA

TEXTOS DIDÁCTICOS

Volume III

SERVIÇO DE EDUCAÇÃO E BOLSAS
FUNDAÇÃO CALOUSTE GULBENKIAN | LISBOA

Edição da
FUNDAÇÃO CALOUSTE GULBENKIAN
Av. de Berna — Lisboa
1999

ISBN 972-31-0971-9
Depósito Legal n.º 148 805/00

III.1

THEORY OF DISTRIBUTIONS*

* Este texto tem por base apontamentos coligidos por diversos alunos de José Sebastião e Silva na sequência de um curso que realizou em 1958 na Universidade de Maryland, e que posteriormente foram utilizados, e por ele revistos, na Faculdade de Ciências de Lisboa.

CHAPTER III

SPECIAL TYPES OF DISTRIBUTIONS

3.1. Locally summable functions

A function f is said to be locally summable on an open set Ω in $I\!R$ iff f is summable on every compact interval contained in Ω. For example, the function $x^{-1}(x-1)^{-1/3}$ is locally summable on the interval $]0, +\infty[$ or even on the set Ω of all points $x \neq 0$; but it is not locally summable on $I\!R$, for it is not summable on any interval containing 0. On the contrary, *$\log|x|$ is locally summable (thought not summable) on $I\!R$.*

Instead of an open set, we may consider more generally any set which results from an open set by adding to it one or more of its boundary points.

If f is a function locally summable on an interval I, we shall call a **primitive** of f any function F of the form:

$$F(x) = K + \int_c^x f(\xi)\,d\xi\,, \qquad \forall x \in I$$

where c is an arbitrary point of I and K an arbitrary complex number.

From the properties of the Lebesgue integral, the following theorem is deduced:

3.1.1. *If F is a primitive of a locally summable function f on I, then F is continuous on I and has a derivative a.e. (almost everywhere) in the ordinary sense such that:* $F'(x) = f(x)$, *almost everywhere on I.*

It should be observed that the converse of theorem 3.1.1. is not true. There are examples of continuous functions which have a derivative a.e. in the ordinary sense on an interval *I and which are primitives of no locally summable functions on I.*

The functions which are primitives of locally summable functions are said to be **absolutely continuous**. A direct characterization of such functions was given by Vitali.

Theorem 3.1.1. suggests calling the function f **a derivative** of its primitive F. But then F would have, of course, *infinitely many derivatives* (of 1st order).

3.1.2. *Two locally summable functions* f_1 *and* f_2 *on I have the same primitive F if and only if* $f_1(x) = f_2(x)$ *almost everywhere on I.*

In such a case the functions are said to be **equivalent** and it is written $f_1 \sim f_2$ (on I).

It is readily seen that this is actually an equivalence relation. The class of all functions which are equivalent, in this sense, to a given function f, locally summable on I, will be denoted by $[f]$. That being so, if F is any primitive of f, it will be natural to call the class $[f]$ the **derivative of** F and to write:

$$DF = [f].$$

So the derivative of F is uniquely defined as *one class* of functions instead of a single function. On the other hand it is natural to define the sum of two such classes $[f]$ and $[g]$ and the product of $[f]$ by a complex number α according to the formulas:

$$[f] + [g] = [f + g], \quad \alpha[f] = [\alpha f]$$

It is readily seen that with these definitions the set of all such classes $[f]$ becomes a complex vector space. Finally, if f is a continuous function on I, it is natural to identify $[f]$ with f, so that $C(I)$ becomes a subspace of the preceding vector space.

However, it will be troublesome to have to speak throughout of classes of functions. To avoid this we shall use a simple device.

Let f be any locally summable function on I and let us place:

$$\widetilde{f}(x) = \frac{d}{dx}\int_c^x f(\xi)\,d\xi \quad (\text{with } c \in I).$$

Then $\widetilde{f}$ is defined *only* at the points x of I for which the preceding derivative exists in the ordinary sense. On the other hand, we have, of course:

$$\widetilde{f} \sim f \quad \text{and} \quad \widetilde{\widetilde{f}} = \widetilde{f}.$$

We shall call the operation $f \to \widetilde{f}$, **standardization**, and the functions f such that $\widetilde{f} = f$, **standard functions**. For example, the Heaviside function:

$$H(x) = \begin{cases} 1, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases}$$

is not a standard function. By standardization of H we obtain the standardized Heaviside function:

$$\widetilde{H}(x) = \begin{cases} 1, & \text{for } x > 0 \\ 0, & \text{for } x < 0 \end{cases}$$

which is not defined at $x = 0$.

In particular, all continuous functions are standard functions. It is natural to replace any equivalence class $[f]$ by the standard function $\widetilde{f}$ belonging to this class.

From now on, when we speak of locally summable functions, it will be understood that they are standard functions. We shall denote by $\mathring{L}(I)$, or simply $\mathring{L}$, the set of all (standard) locally summable functions on I. *According to the preceding remarks,* $\mathring{L}$ *is a complex vector space and* C *is a linear subspace of* $\mathring{L}$.

3.2. Locally summable functions as distributions

Let f be any (standard) locally summable function on I and let us denote by F one primitive of f:

$$F(x)=K+\int_a^x f(\xi)\,d\xi,\ (a\in I,\ \ K\in\mathbb{C}).$$

Since F is a continuous function on I, there exists a distribution which is the derivative of F. We shall denote by F' $(=f)$ the derivative of F in the functional sense and by DF the derivative of F in the distributional sense. Subsequently, we shall identify DF with F', this identification being based on the following theorem:

3.2.1. THEOREM. *By assigning to each function $f\in\mathring{L}(I)$ the distribution $f^*=DF$ where F is a primitive of f, there is defined a one-to-one linear mapping of $\mathring{L}(I)$ into $\mathscr{D}(I)$ such that:*

- *(i) if f is continous on I, then $f=f^*$;*
- *(ii) if f if absolutely continuous on I, then Df^* corresponds to f'.*

PROOF. First of all, it must be observed that the distribution DF assigned to each function $f\in\mathring{L}$ does not depend on the choice of the primitive F of f. In fact, if G is another primitive of f, then $F-G$ is a constant function, so that $DF=DG$.

Now consider two functions $f,\ g\in\mathring{L}$; we have to prove that if, to f and g corresponds the same distribution, then $f=g$. Let F, G be primitives of f, g respectively and suppose $DF=DG$. Then $F-G\in\mathscr{P}_1$ is a constant on I, and therefore, F and G have the same derivatives in the functional sense (as a standard function), that is $f=g$.

For the remaining parts of the theorem, the proof is quite trivial. ♦

This theorem shows that we can identify every distribution DF, where F *is an absolutely continuous function,* with the locally summable function f, which is the derivative of F in the functional sense 3.1.1. We then write:

$$DF=F'=f.$$

Since every locally summable function f is a distribution, f will have derivatives of all orders (in distributions sense). *Conversely,* *every distribution may be expressed in the form $D^n f$ where $n \in /N_0$ and $f \in \mathring{L}$. For simplicity of notation, even if a locally summable function f is not a standard function, we shall denote by $D^n f$ the distribution $D^n \widetilde{f}$.*

For example, the δ distribution may be defined as the derivative of the (standardized) Heaviside function, and we may write in general:

$$\delta^{(n)} = D^{n+1}H, \text{ for } n = 0, 1, \dots .$$

3.3. Functions which are not distributions and pseudofunctions

Consider for example the function $\frac{1}{\hat{x}}$. Since this function is continuous on the set of all points $x \neq 0$, it is a distribution on this open set. But it is not a locally summable function *on /R,* and as we shall next see, it may not be interpreted as a distribution on /R. This function is the derivative *in the ordinary sense* (not defined at 0) of all functions f of the form:

$$f(x) = \begin{cases} log|x| + c_1, & \text{for } x > 0 \\ log|x| + c_2, & \text{for } x < 0 \end{cases}$$

or shortly:

$$f(x) = log|x| + aH(x) + b$$

with $a = c_1 - c_2$, $b = c_2$, where c_1 and c_2 are arbitrary complex numbers. Now, contrary to $\frac{1}{\hat{x}}$, *any function f of this form is locally summable on /R, hence it is a distribution on /R,* whose derivative is:

$$Df = D\,log|\hat{x}| + a\delta$$

where the symbol $D\,log|\hat{x}|$ denotes the derivative *of the locally summable function $log|\hat{x}|$ on /R, in distributions sense.*

Thus, there exist infinitely many *distinct* distributions on *IR* which are the derivatives of the functions f. But for any f, we have:

$$Df = \frac{1}{\hat{x}}, \text{ on the set of all } x \neq 0.$$

Therefore the function $\frac{1}{\hat{x}}$ is a distribution on this set, but may not be interpreted as a distribution on *IR*.

The distribution $D\,log|\hat{x}|$ on *IR* is called the **finite part** of $\frac{1}{\hat{x}}$ and denoted by $Pf\frac{1}{\hat{x}}$. But $Pf\frac{1}{\hat{x}}$ *is not* a *function, as* $\frac{1}{\hat{x}}$ *is not a distribution.*(4)

More generally the *finite part of* $\frac{1}{\hat{x}^n}$ is defined to be the distribution:

$$Pf\frac{1}{\hat{x}^n} = \frac{(-1)^{n-1}}{(n-1)!}D^n log|\hat{x}|, \quad n = 1, 2, \dots .$$

This belongs to an important class of distributions which are called **pseudofunctions** by L. Schwartz. We shall further see other examples of pseudofunctions.

3.4. Measures and functions of bounded variation

We have already discussed the concept of measure in chapter I. It is not dificult to see that an equivalent definition is the following:

A measure μ is defined on *IR* iff to every *bounded interval J* in *IR* is assigned a complex number, called the μ-**measure** of J and denoted by $\mu(J)$ or μJ, in such a way that:

---

(4) The expression "finite part" is connected with the concept of finite part of certain divergent integral which L. Schwartz used for defining this distribution. Note that there is no special reason to identify the function $1/x$ with $Pf1/x$ rather than with a distribution $Pf1/x + a\delta$, with $a \neq 0$.

M1. If J is expressed as the union of two disjoint intervals J_1 and J_2, then:

$$\mu(J)=\mu(J_1)+\mu(J_2).$$

M2. If J is the union of the intervals $J_1 \subset J_2 \subset \cdots$, then:

$$\mu(J)=\lim_{n\to\infty}\mu(J_n)$$

M3. For each *bounded* interval J, there exists a positive number $m(J)$ such that for every partition of J into a finite number of intervals $J_1, J_2, \ldots, J_n$, we have:

$$\sum_{k=1}^{n}|\mu(J_k)|\leq m(J).$$

Observe that the variable interval J which we are now considering may be a degenerate interval $[a, a]$.

We can define analogously the concept of measure on any open set $A \subset /R$ or even on a more extensive class. But then we must consider bounded intervals J, such that $\bar{J} \subset A$.

3.4.1. DEFINITION. If μ is a measure on the interval I on $/R$, a primitive of μ will be any function F defined on I by putting:

$$F(x)=\begin{cases} k+\mu[c, x], & \text{if } x\geq c \\ k-\mu]x, c[, & \text{if } x<c \end{cases}$$

where c is an arbitrary point of I and k an arbitrary complex number.

From this definition and M1 follows:

3.4.2. *$F(b)-F(a)=\mu]a, b]$, whenever $a<b$.*

On the other hand, from M1 and M2 we have:

3.4.2. $F(a)-F(a^-)=\mu[a, a]$, for all $a\in I$.

To see this we consider the case $c<a$ and it suffices to express $[c, a[$ as the union of a sequence of intervals $[c, x_n]$ such that

$c < x_1 < \cdots < x_n < \cdots < a$ and $x_n \to a$. Then by M2:

$\mu[c, a[= \lim\limits_{x_n \to a} \mu[c, x_n] = \lim\limits_{x_n \to a} F(x_n) - k = F(a^-) - k$. Formula 3.4.3. is analogously proved in the case $a \leq c$.

Finally, from 3.4.2. and 3.4.3. follows:

$$F(b) - F(a^-) = \mu[a, b]$$

3.4.4.

$$F(b^-) - F(a) = \mu]a, b[$$

for every pair of points *a, b* in *I* such that $a < b$. Consequently:

3.4.5. *If* μ *is a measure on I and if F is any primitive of* μ, *then* μ *is uniquely determined by F according to formulas* 3.4.2., 3.4.3. *and* 3.4.4.

Now we need a characterization of the functions which are the primitives of the measures on *I*. Let *F* be such a function. From M1 and M2 it can be easily deduced (as in 3.4.3.) that $F(a) = F(a^+)$ for any $a \in I$; i.e. *F* is *continuous on the right* at every point of *I*. In addition, M3 implies that to each compact interval $J = [a, b]$, there exists a number $m(J)$ such that, for every partition of *J* by means of points $a = x_0 < x_1 < \cdots < x_n = b$, we have:

$$\sum_{k=1}^{n} |F(x_k) - F(x_{k-1})| \leq m(J).$$

But this means that *F* is a function of locally bounded variation on *I*.

Conversely, it is easily seen that these two properties are sufficient to characterize primitives of measures. Thus:

3.4.6. *A necessary and sufficient condition for a function F on I to be a primitive of a measure* μ *on I, is that F be of locally bounded variation on I and continuous on the right at every point of I. Moreover two such functions* F_1 *and* F_2 *are primitives of the same measure* μ *if and only if* $F_1 - F_2$ *is constant on I.*

We shall denote by $\mathfrak{M}(I)$, or simply $\mathfrak{M}$, the set of all measures on I. The sum $\mu + \upsilon$ of two measures and the product $\alpha\mu$ of a complex number α by μ are defined by the formulas:

$$(\mu + \upsilon)(J) = \mu(J) + \upsilon(J),$$
$$(\alpha\mu)(J) = \alpha(\mu J).$$

for each bounded interval J such that $\bar{J} \subset I$.

Then $\mathfrak{M}(I)$ becomes a complex vector space.

3.5. Measures as distributions. Order of a distribution

Let I be any (non-degenerate) interval on $/R$. Observe that to each function $f \in \mathring{L}(I)$ and each bounded interval J such that $\bar{J} \subset I$, there corresponds the number $\int_J f$ and this correspondence $J \to \int_J f$ is a measure, μ_f, whose primitives are just the primitives of the function f. Thus, every function $f \in \mathring{L}(I)$ determines one measure $\mu_f \in \mathfrak{M}(I)$, and if $\mu_f = \mu_g$, then $f = g$, since f and g have the same primitives. Besides it obvious that $\mu_{(f+g)} = \mu_f + \mu_g$ and $\mu_{\alpha f} = \alpha\mu_f$ for any $\alpha \in \mathbb{C}$. Thus it is natural to identify each $f \in \mathring{L}(I)$ with μ_f so that $\mathring{L}(I)$ becomes a vector subspace of $\mathfrak{M}(I)$.

Now, remember that every function of locally bounded variation on I is Riemann integrable on each compact subinterval $J \subset I$; hence locally summable on I. Then taking 3.4.6. into account, it is easily shown that:

3.5.1. *By assigning to each measure μ on I the distribution DF, where F is any primitive of μ, there is defined a one-to-one linear mapping of $\mathfrak{M}(I)$ into $\mathscr{D}(I)$ such that if μ is a locally summable function f on I, then μ corresponds to $DF = f$.*

The proof is quite similar to the one of 3.2.1. It should however be observed that the primitive F of a measure is not in general a standard function; but since $\widetilde{F}$ is defined *only* at the continuity points of

F, with the same values, it is readily seen that the correspondence $F \to \widetilde{F}$, is one-to-one.

Recording 3.5.1., it is natural to identify every measure μ on I with the distribution DF, where F is a primitive of μ and to write $\mu = DF$. So $\mathfrak{M}(I)$ becomes a vector subspace of $\mathscr{D}(I)$ and more precisely of $C_2(I)$:

$$C \subset \mathring{L} \subset \mathfrak{M} \subset C_2 \subset \mathscr{D}.$$

For every integer $n \geq 0$, we shall denote by $\mathfrak{M}_n(I)$ the **set of all distributions** f such that $f = D^n\mu$ with $\mu \in \mathfrak{M}(I)$. It is obvious that

$$\mathscr{D} = \bigcup_{1}^{\infty} \mathfrak{M}_n .$$

3.5.2. DEFINITION. Given a distribution f on I the least n such that $f \in \mathfrak{M}_n$ is called the **order** of f.

For example δ, which is a distribution of rank 2 (see 2.3.) is of order 0 (i.e. is a measure). In general $\delta^{(n)}$ is of order n.

3.6. Product of a continuous function by a measure and the Stieltjes integral

Consider $f \in C(I)$ and $\mu \in \mathfrak{M}(I)$. Let J be any bounded interval such that $J \subset I$ and let P be any partition of J into a finite number of (mutually disjoint) intervals $J_1, \ldots, J_n$. Denote by $N(P)$ the greatest lenght of the intervals $J_1, \ldots, J_n$. Let x_k be an arbitrary point in J_k and put:

$$S_p(J) = \sum_{k=1}^{n} f(x_k)\, \mu(J_k) .$$

Then it is a classical result that $S_p(J)$ tends to a finite limit, $S(J)$, as $N(P) \to 0$; that is, to every $\delta > 0$, corresponds an $\varepsilon > 0$, such that:

$$N(P) < \varepsilon \text{ implies } |S(J) - S_p(J)| < \delta.$$

Moreover, it can be shown that the correspondence $J \to S(J)$ is a measure on I.

This measure is called the **product** of f by μ and denoted by $f\mu$. Thus, by definition:

$$(f\mu)(J) = S(J).$$

Previously (1.3.1.) we have adopted the convention that the μ-measure of an interval J *should be called the integral of* μ *on J and denoted by* $\int_J \mu$. Thus:

$$S(J) = (f\mu)(J) = \int_J f\mu \, .$$

Remember that $\int_J f\mu$ is usually called the Stieltjes integral of f with respect to μ. The notation $\int_J f d\mu$ is commonly used instead of $\int_J f\mu$, but that notation in the theory of distributions may induce in error. If F is a primitive of μ, it is quite natural to denote the integral of f with respect to μ by:

$$\int_J f(x)\, dF(x),$$

and since $\mu = F'$ (in the distribution sense) we could also write:

$$\int_J f(x)\, dF(x) = \int_J f(x) F'(x)\, dx = \int_J f\mu \, .$$

But then we should have:

$$\int_J f(x)\, d\mu(x) = \int_J f\mu',$$

and this is the integral of f with respect to the distribution μ' that we shall define later on.

As an example, let us calculate $f\delta$, where $f \in C(IR)$. If we consider a bounded interval J such that $0 \in J$ and a partition P of J, into intervals $J_1, \ldots, J_n$, then one and only one of these will contain 0, say J_j. Thus for every choice of $x_k \in J_k$:

$$S_p(J)=\sum_{k=0}^{n} f(x_k)\,\delta(J_k)=f(x_j)\,.$$

Therefore:

$$\lim_{N(P)\to 0} S_p(J)=\lim_{x_j\to 0} f(x_j)=f(0)\,;$$

that is $(f\delta)(J)=f(0)$, if $0\in J$. It is easily seen that $(f\delta)(J)=0$, if $0\notin J$. Hence:

3.6.1. $$f\delta=f(0)\delta.$$

More generally: $f(\hat{x})\,\delta(\hat{x}-a)=f(a)\delta(\hat{x}-a)$.

3.7. Derivatives of peace-wise smooth functions

We begin with the following proposition:

3.7.1. THEOREM. *Let f be a function on an interval $I=]a,b[$. Suppose that f is absolutely continuous on two intervals $]a,c[$ and $]c,b[$ and tends to finite limits as $x\to c^-$ and as $x\to c^+$. Then, denoting by f' the derivative of f in the ordinary sense (not necessarily defined at c) and putting $s=f(c^+)-f(c^-)$, we have:*

$$Df=[f']+s\delta(\hat{x}-c)\,.$$

PROOF: Since f is absolutely continuous on $]a,c[$ and $]c,b[$ and has finite limits $f(c^-)$ and $f(c^+)$ it is easily seen that $[f']$ is locally summable on I. Hence if we put:

$$g(x)=f(c^-)+\int_c^x [f'](\xi)\,d\xi\,,\qquad \forall x\in I$$

the function g will be absolutely continuous on I and:

$$g'=f',\quad g(x)=\begin{cases} f(x) & ,\text{ for } a<x<c \\ f(x)-s, & \text{for } c<x<b \end{cases}.$$

Hence $f(\hat{x})=g(\hat{x})+sH(\hat{x}-c)$ and $Df=[f']+s\delta(\hat{x}-c)$. ♦

Consider now a similar situation concerning a finite number of points $c_1, \dots, c_p$ in I and put $s_k = f(c_k^+) - f(c_k^-)$. Then:

$$Df = [f'] + \sum_{k=1}^{p} s_k \delta(\hat{x} - c_k).$$

Suppose more generally that:

i) *f has a derivative of order $n \geq 0$ in the ordinary sense except for a set of isolated points, $c_k (k = 0, \pm 1, \pm 2, \dots)$;*

ii) *$f^{(n-1)}$ is absolutely continuous on each subinterval of I not containing any point c_k;*

iii) *for every $k = 0, \pm 1, \dots$ and every $j = 0, \dots, n-1$ there exists finite limits $f^{(j)}(c_k^-)$ and $f^{(j)}(c_k^+)$.*

Then it is easily deduced from 3.7.1.:

3.7.2. $$D^n f = [f^{(n)}] + \sum_{-\infty}^{+\infty} \sum_{j=0}^{n-1} s_k^{(n-1)-j} \delta^{(j)}(\hat{x} - c_k),$$

where $s_k^j = f^{(j)}(c_k^+) - f^{(j)}(c_k^-)$. The last term in 3.7.2. (involving eventually a sum of infinitely many distributions) denotes the distribution whose restriction to each of the compact intervals $J \subset I$ is the sum of the distributions $\sum_{j=0}^{n-1} s_k^{(n-1)-j} \delta^{(j)}(\hat{x} - c_k)$ where $c_k \in J$ (in finite number).

For example, it is easily seen that:

$$D^2 |x| = 2\delta$$

$$D^3 |x^2 - 1| = 4(\delta_{(1)} - \delta_{(-1)} + \delta'_{(-1)} + \delta'_{(1)})$$

$$D^2 (x^{4/3} + |x|) = \frac{4}{9} x^{-2/3} + 2\delta.$$

Remarks about notation: In the preceding considerations when it has been necessary to distinguish the derivatives of a function f in the ordinary sense from its derivatives in the distributional sense, we have used the notation f' in the first case and Df in the second. But whenever no confusion is possible we shall consider $f^{(n)}$ and $D^n f$ as perfectly equivalent.