L. SCHWARTZ

Historical roots and basic notions in the theory of distributions

Artémiadis (N.), éd., *Proceedings of the General Mathematics Seminar of the University of Patras*, University of Patras, 1982.

Extrait des *Œuvres de Laurent Schwartz*
publiées par la Société mathématique de France, 2011.

Historical Roots and Basic Notions in the Theory of Distributions

by L. Schwartz

October 1982

Historical Roots and Basic Notions in the Theory of Distributions

by L. Schwartz

Οκτώβριος 1982

NOTE

A tape recorder was used to record Prof. Schwartz' lecture. Then some members of the Math. Department had to work hard to come up with this final text, approved by the speaker.

The same lecture was given in several other foreign Universities but it is published for the first time, now, in the Proceedings of the General Mathematics Seminar of the University of Patras.

N.Artémiadis

It is a great pleasure for me to give this lecture here. It is a difficult lecture, though, because you ask me to speak about the history of Distributions, in which I am much involved, so that I cannot avoid speaking about myself, which is always a difficult problem. So I decided to do exactly that: to show you in what way I could discover, or I discovered, the Distributions, seeing some things quite soon and some things very late; in what way my strength and my weakness can be compared with other people's.

Because, when one examines this discovery, as well as many others, one can see that there were many people before me who had something, even may be a large part of the subject, but could not go far enough. And then, at some time, something, imposed by the circumstances, comes out, which was not recognized before.

I shall try to give other examples, because I want to show you in a way, the strength and the weakness of human spirit for discovery.

Let me give a very famous example: Einstein discovered Relativity, but Lorentz had already completely written, what is called the "Lorentz Group", namely the group of transformations of space-time which leave invariant Maxwell's Equation of Electromagnetism. So he wrote that! but it was more or less ignored, because it was not assumed that the electromagnetism and the velocity of light had to be invariant. Why should they be invariant? When one has the light as something propagated in a matter like ether, then it does not have to be invariant with respect to the various observers, moving with respect to each other.

It was only a negative physical experiment that Michelson made, trying to measure the velocity of light with respect to various places moving mutually, which indicated that there was no observable difference. After Michelson's experiment had been repeated, it was admitted that the velocity of light seemed to be constant. Again nobody considered that as a fundamental fact. And then Einstein(1) came. It was already known by experience that the velocity of light was invariant and that there was a group leaving invariant

(1) I don't want to compare me with Einstein, but just to show how, usually, a discovery arrives through various steps and various people.

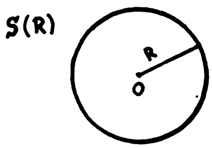
Maxwell's Equation and velocity of light, i.e. the Lorentz group. But he had to impose the new conception of space-time, and in the same way a lot of new formulae and concepts, on Mechanics (force, energy, mass) and Physics; nobody would believe it at the beginning. And there was a very strong contravening current. There are still people who want to fight Relativity, but they are really crazy. At that time it was the opposite, I mean it was Einstein who wasn't believed immediately. Yet afterwards, he was completely recognized and the system became universal. But it was considered to be so difficult, that it was said in many circles that only two or three people could understand relativity. Nowadays Relativity is taught in all Universities in the world; and even Highschool pupils know something about Relativity.

Usually, when a new concept is established, one can find that this concept had already existed but was not recognized as something useful and fundamental.

This is also true for the Distributions. It is difficult to imagine how many people, who lived much earlier, already knew something about them. Even I had known a lot of things which I could not bring together. And when I did, I found very oblique ways to get the results and not immediately the definition which exists now. So one has to overcome external and internal difficulties in order to establish new theories. This I wanted to show.

Let us return to the past. You find some traces of future distributions in Riemman, in Gauss, in Dirichlet, in the Theory of Harmonic Functions.

For instance it has been known that a harmonic function f ($\Delta f = 0$) in $\mathbb{R}^n$ (in order to be harmonic it has to be twice differentiable) is ipso facto indefinitely differentiable. Then people observed that in order to define a harmonic function it was not even necessary to have derivatives of the second order and a harmonic function could be defined as a continuous function such that for every sphere S(R) the value f(0) at the center was equal to the mean value of f on the sphere S(R). This could be taken as a definition:



If a continuous function has the property that the mean value on every sphere is equal to the value at the center, then it is ipso facto twice differentiable and is harmonic. So, one has a definition of harmonic functions without derivatives apriori.

The same happened with holomorphic functions. If you have a holomorphic function f(z) of one complex variable in a complex plane, Cauchy proved that it is indefinitely differentiable and if C is a closed rectifiable curve then:

$$\int_C f(z)dz = 0$$

There is also a converse theorem by Morera saying that if a continuous function has integral zero over every closed rectifiable curve, then it is holomorphic, so it has derivatives. Therefore one can also define a holomorphic function without using derivatives. Of course this does not lead immediately to the distribution but it is implying that sometimes it is not necessary to assume the derivatives in the usual sense to get a property which finally comes back to derivatives.

At the end of the 19th and the beginning of the 20th century, people considered very much integral equations, convolution integral equations, Volterra convolution, Mercer convolution of the type:

$$h(t) = \int_0^t f(t-\tau)g(\tau)d\tau$$

Convolution played an important role at that time in the studies of electricity, in differential equations, in partial differential equations and, in particular,workers in electricity were led to consider phenomena which started at the time zero (one switches on the current, for instance) and then they observe the reactions. So there are functions which are zero for negative t and just different from zero for non-negative t. This produces an integral from o to t.

It was known that if one had a complex electrical system which contained resistances, self-inductions, capacities, it had an impedance Z(t) which was a function of t, so that if you had put electromotive force e(t) you obtained a current with intensity i(t). Then one has the relationship

$$e(t) = \int_0^t Z(t-\tau)i(\tau)d\tau$$

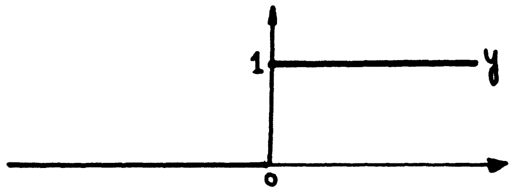
i.e., the convolution of the intensity and the impedance of the electromotive force. This was known at the end of the 19th and the beginning of the 20th century.

Then the British engineer and mathematician Heaviside introduced for that a *symbolic calculus* (1893): instead of writing e(t), Z(t), i(t) he wrote i(p), Z(p), e(p) and he wrote that in the following algebraic rule: $e(p) = Z(p) \cdot i(p)$.

This was justified later on using the notion of Laplace transform, which was not systematically used at that time[1]. It was purely symbolic. He introduced a δ impulsion (it was not called δ, it was called a "unity impulsion") much before Dirac. This δ impulsion represented for him an imaginary, very strange, electromotive force acting for a very short time, so that the integral was equal to one. It was exactly the future notion of Dirac function.

(1) Carson 1926, van der Pol 1932.

Then he introduced the so called unit scale y which was a current equal to zero for negative t and equal to one for positive t.



He observed that the derivative in some sense of this unit scale should be equal to δ:$y' = \delta$, (in what sense was not very precise) but he observed that δ must be zero except at the origin, and that:

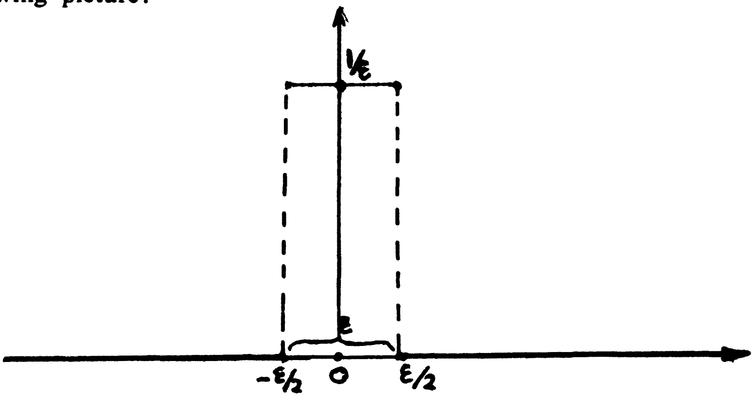
$$\int_{-\varepsilon}^{+\varepsilon} \delta(x)dx = y(\varepsilon) - y(-\varepsilon) = 1$$

So he said everything about Dirac's δ-function at that time: δ is the derivative of the function y above, it is zero everywhere exept at the origin and its integral equals one. But this cannot be possible because δ is almost everywhere equal to zero, so that its integral is zero, according to Lebesgue theory. So δ didn't really exist, but he wrote that. He called it, 1; he called δ', p; and y he called $\frac{1}{p}$. That $\delta' * y = \delta$ (here is the modern writing with distributions) Heaviside wrote as $p \cdot \frac{1}{p} = 1$. He made a complete symbolic calculus with these functions of p. He constructed a symbolic algebra in which he used polynomial multiplications, decomposition of rational fractions into simple elements and it became a fantastic construction with asymptotic developments and so on, but no mathematician had accepted that! It was completely rejected by the community of mathematicians who said he was just crazy. It was a little strange because it led him mostly to perfectly true results; he also found theorems which were absolutely exact (for instance the extension of the fundamental solution E of a differential equation on $\mathbb{R}$ with constant coefficients, $P(D)E = \delta$, written as $P(p)E(p) = 1$ or $E(p) = \frac{1}{P(p)}$, computed by decomposition of the rational fraction $\frac{1}{P(p)}$ into simple elements), and which nobody could reasonably" explain. One could prove these theorems, but he found them using this symbolic computation and nobody would admit that it ought to have something true in it, if it led to true results.

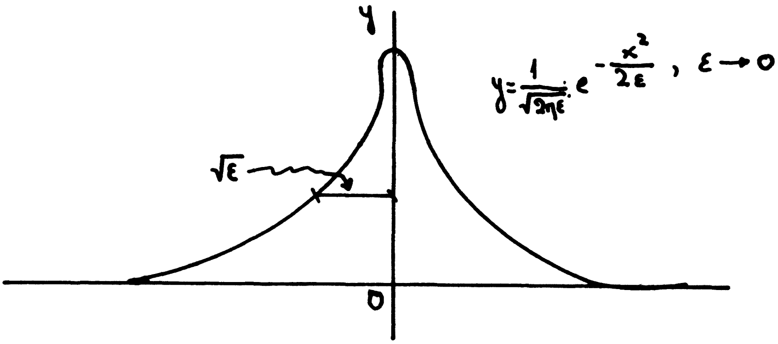
He was refused, considered insane; and in the same way he was so rejected that at the end of his life he lost partly his equilibrium. It may be that to be a mathematician is as dangerous as to be a man of politics! You may be rejected if you are not in the normal current!

This was a first experience and it is now part of the Theory of Distributions, but it was then rejected.

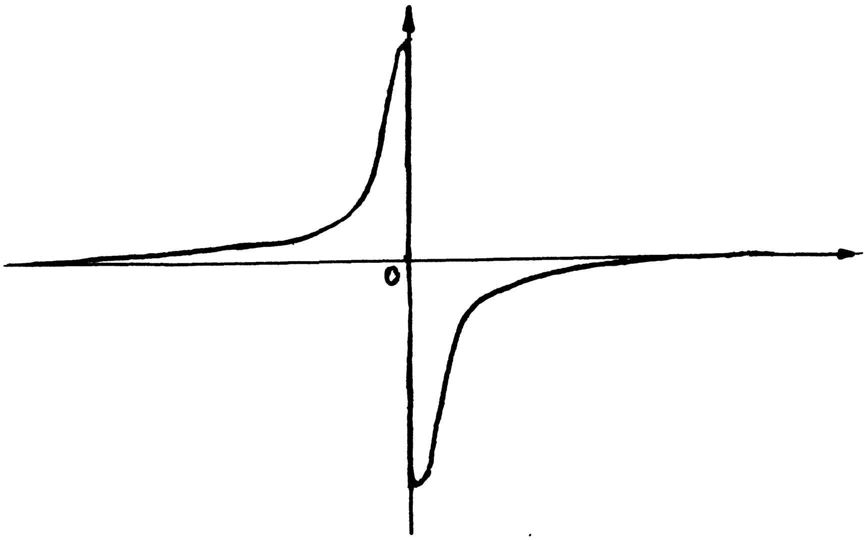
Later, Dirac introduced in 1927 the δ-function which is called the "Dirac function", with the same definition: δ is a function which is zero everywhere except at the origin and whose integral has to be one. Then he said another thing about δ: We have an approximation of δ as shown in the following picture:



This is a discontinuous approximation of δ with the value 1/ε on an ε-interval. He also gave another, this time continuous, approximation which was a Gauss-lobe, very concentrated, called a Gauss-lobe with a small parameter. In both cases the integrals are equal to one (see Fig. 4).



He then introduced the derivative δ′, something which was completely unacceptable for mathematicians, but his explanation was that: if the function in Fig. 4 is an approximation of δ, you just differentiate it and you obtain a function as follows:



which should be an approximation of δ'. He then differentiated δ-function many times. But not only did he introduce the δ-function, he also made changes of variables: for instance he considered $\delta(t^2-x^2-y^2-z^2)$ which introduces a distance in terms of Relativity. This was a change of variables in Distributions before they actually existed! Then he manipulated all these in such a way that in 1940 in all books of Theoretical Physics there were fantastic computations with distributions prior to the existence of distributions in Mathematics! These computations were sometimes wrong, since they were based on intuition which often is misleading. Still, there were a lot of true formulae which are now well proved. So Dirac was not only the inventor of Dirac function, but he and many physicists also did a lot of computations so that in every book of Physics there were many δ-functions and distributions with no justification at all from the mathematical point of view (which came at least 15 years later). The mathematicians felt only contempt for that and they rejected the whole method. Yet Dirac did not become insane, but the Physicists formed a world outside of Mathematics.

I followed a course, when I was a student, in the year 1935, in which I heard about δ-function and its derivatives, I thought about it with some friends, so it remained in my mind that this existed in Physics and had absolutely no sense in Mathematics. I discussed it with the mathematicians of that time, my professors, my classmates and we came to the conclusion that it was nothing, we had nothing to do with it. It was absolutely impossible to find any kind of justification, especially for δ' which was the first true obstacle. Because δ could be a measure, but then a measure has not derivative. But I kept that open in my mind.

One of the first completely coherent things on the subject has been introduced by Sobolev just before the war (1936). Sobolev introduced some kinds of generalized functions which were very near to my definition of distributions. Sobolev instead of considering the space C^{∞}_{comp} (of infinitely differentiable functions with compact support) (that I called $\mathfrak{D}^{\infty} = \mathfrak{D}$), considered C^{m}_{comp} which he called $\mathfrak{D}^{m}$, and introduced continuous, linear functionals of $\mathfrak{D}^{m}$ and manipulated some derivatives, so that he had a correct definition, possible to use. Something completely lacking in his theory was the support, the carrier of a distribution; and the carrier of a distribution depends on the Theorem of Unity Partition which was discovered in France during the War by Dieudonné. Sobolev had no convolution, no Fourier transform, no strong topological properties, but he had many formulae and, in particular, he was able to see that a partial differential equation with a boundary value could be solved in this way by putting the boundary value in the second hand-side, a decisive fact. It was quite a good approximation of distributions. However in the world of Mathematicians it remained a curiosity in the sense that it was just a publication among others, for a particular aim and not presented as a general tool for a lot of things. Most of the mathematicians ignored it (including myself). Of course this appeared just before the War and Sobolev himself did not continue much afterwards in this direction. He just had it published and then went to other subjects. However he introduced very interesting objects which still carry his name, the spaces H^m, $m \geqslant 0$ integer, called the Sobolev spaces. With his generalized derivative he defines H^m, for integer m, as the space of L^2 functions whose derivatives, in the weak sense of generalized functions, belong to L^2; they need not be continuous, they are differentiable in the sense of distributions and not in the usual sense. That was a really considerable work, and the Sobolev spaces remained because they are an essential tool for the study of partial differential equations. The beginning was a little forgotten; it is difficult to know exactly why, but this is the case (as often happens) and it is comparable with the Lorentz group. Lorentz found the group, but he forgot it and he continued on other directions. In order that a new *essential* idea becomes *universally accepted*, it's not sufficient that somebody introduces it casually, it's necessary that one or several people introduce it at a *massive* scale, and be strongly persuasive.

Bochner made another very good approach. The case of Bochner is extremely interesting because in some way he was at the least possible distance from distributions, although he did not go further. So he published at the end of his book about Fourier Integrals (1932) a chapter in which he introduces "formal functions" on the real line $\mathbb{R}$. A formal function was defined as follows: Consider a square integrable function f, multiply it by a polynomial Q and apply P(D), a partial differential operator with constant coefficients. So it is the derivative of a slowly increasing function:

$$P(D)[Q \cdot f]$$

These beings, introduced by Bochner, are exactly the same with the temperate distributions which I introduced later on, on $\mathbb{R}^N$, N arbitrary. But nobody perceived it, I ignored then myself. So it remained also a curiosity and it was so difficult to be explained, that he put that as an isolated chapter at the end of the book (24 pages), as a curiosity. He did make some computations with that; he did know that Fourier operation, transforms convolution into multiplication and multiplication into convolution. So he knew everything in some way, without topology. Yet, even he himself did not recognize it as a new fundamental tool.

For instance if you want to define δ in this way (which he did not do) you have to define it as:

$$\delta = D^2\left(\frac{1}{2}|x|\right)$$

(the second derivative of the function one-half of modulus of x).

But the second derivative of $\frac{1}{2}|x|$ is not usable as δ! You cannot handle δ in this way. So he did not define the δ of the physicists. And so there was no relationship with the experiments and with the physicists; no relationship with Heaviside; it was just a chapter flying over the remainder as a kind of curiosity. It is a quite exceptional scope on what is the introduction of a theory, because in some way everything was contained in that chapter and nobody recognized it; even Bochner himself didn't estimate it at its correct value. He himself considered it as a curiosity, published at the end of his book; and he never spoke about it any more. You see how things advanced exactly as in the case of Lorentz. Lorentz found the Lorentz group but there was no conclusion about that.

A second time Bochner considered a partial differential equation with constant coefficients: $P(D)f = 0$ in $\mathbb{R}^N$, and he had to know what was a "generalized" solution of the partial differential equation. This could be done by using his Fourier Transform, but he did not do it this way! He said: f is a generalized solution of $P(D)f = 0$, if there exist infinitely or several times differentiable functions f_n which converge uniformly to f on every compact subset of $\mathbb{R}^N$, and such that: $P(D)f_n = 0$, in the usual sense. This could be done only for an operator with constant coefficients, by regularization: the f_n are of the form $f*\varphi_n$, φ_n smoolth functions with compact support, φ_n was differentiable enough, so that $P(D)\ (f*\varphi_n)$ was meaningful, not indefinitely differentiable. So there is a second article in which he discovers in a way the derivative in the sense of distributions, but he establishes no relationship between both articles!

The main defect is that one could define f verifying $P(D)f = 0$, but one could not compute the values of the derivatives. For instance in the hyperbolic equation of a vibrating string:

$$\frac{1}{\upsilon^2}\frac{\partial^2 f}{\partial t^2} - \frac{\partial^2 f}{\partial x^2} = 0$$

it was known a long time ago that the general solution was:

$$f = u(x + \upsilon t) + w(x - \upsilon t)$$

In order to be a solution, u and w must be twice differentiable. What about the case of a function which is a sum of this kind, where u and w are just continuous functions without derivatives? That was one of my obsessions since already 1934: in what sense is it possible to say that a $u(x + \upsilon t)$ and a $w(x - \upsilon t)$, where u and w are continuous functions, can form a solution of the wave equation, although they have no derivatives in the usual sense? So one can not actually compute either $\frac{\partial^2 f}{\partial t^2}$ or $\frac{\partial^2 f}{\partial x^2}$, but in some sense their combination is zero. I ignored also this definition of Bochner. I found it again myself in 1944. I shall describe that below. Anyway it didn't solve my problem: Namely f can be approximated by $f_n = u_n + w_n$,

$$f_n = u_n(x + \upsilon t) + w_n(x - \upsilon t)$$

Then one could say that f was a generalized solution, but one could compute neither $\frac{\partial^2 f}{\partial t^2}$ nor $\frac{\partial^2 f}{\partial x^2}$. Just the combination was zero by a limiting procedure, and one could not compute intermediate terms. Jean Leray had introduced early 1934, the weak derivatives, and the weak solutions of partial differential equations, by just integrating by parts (as it is now in distributions), and I attended his lectures in 1935. He defined the weak derivative *when it was a function,* and also a weak *solution* of $\frac{1}{\upsilon^2}\frac{\partial^2 f}{\partial t^2} - \frac{\partial^2 f}{\partial x^2} = 0$, *without each term having any meaning*! Always the same difficutly! I had now this objective which remained always in my mind for the future: how to define a generalized solution of a partial differential equation so that one can say that f is a generalized solution, but also that $\frac{\partial^2 f}{\partial t^2}$ has some meaning and $\frac{\partial^2 f}{\partial x^2}$ too, in such a way that if you combine them you find zero?

Now, Hadamard introduced finite parts of divergent integrals. In his very remarkable book on Hyperbolic Partial Differential Equations he studied fundamental solutions of the equations. The very definition of a fundamental solution had never been written in its full generality. The following definition comes with distributions: if P(D) is a partial differential operator with constant coefficients, E is called a fundamental, or an elementary solution, if $P(D)E = \delta$ in the sense of distributions. So that now it is very easy, but because δ did not exist at that time, the elementary solution was a solution of P(D)E = 0 in some open set, and with some kind of singularity at the origin or on the wave core. With these singularities in the way, it was difficult to define the fundamental solution. A lot of computations have been done by Hadamard himself and by other scientists on fundamental solutions of P.D.E'.s by Fourier transform, or some other method, without a fundamental solution being anywhere correctly defined. They have now a meaning with the theory of distributions. Hadamard also defined the finite parts of divergent integrals which played an enormous role in the theory of hyperbolic partial differential equations (1932). So something was ready for me that I knew, the finite parts of divergents integrals of Hadamard, and fundamental solutions. I knew them well and I knew that something had to be done also in this domain. Now, I remember that Georges de Rham came to Clermont-Ferrand during the war (1942) and gave a lecture about this notion of "currents". He had a notion of currents, which was quite fascinating, according to which a current of degree p, on a manifold of dimension N could be either a differential form of degree p, or a submanifold of dimension N-p with boundary, or a sum of a differential form w and a manifold with boundary: $\Gamma = w + \upsilon$. And then there was a notion of coboundary $d\Gamma$ of this current, which was the coboundary of the differential form plus or minus the boundary of the manifold: $d\Gamma = dw \pm \beta\upsilon$, $dF = dw + (-1)^{p+1}\beta\upsilon$.

But this was just formal and he wanted to have something more than that, and we had a discussion. He then told me: "I should like to find a kind of generalization of Lebesgue integration". In Lebesgue integration on $\mathbb{R}^N$ you have absolutely continuous measures with respect to dx and singular measures as a unit mass. But you have a lot of intermediates, as measures on surfaces, and you have a space of measures, which is a complete vector space with very known properties. I should like some more generalized currents in which one may have differential forms corresponding to the absolutely continuous functions; I should also like some manifolds corresponding to δ (which was not called δ, but ε) or measures on surfaces; so that one may have also a lot of intermediates and this could be a true complete vector space with nice properties. But it seems to be so considerably more difficult than Lebesgue theory, that probably we are very far from that. "It's not for our generation".

I was also impregnated with this. This was in 1942 and I said: "I think it is probably impossible, since we have no good tools to do it". Two and a half years later, I found the distributions, allthough at that time I considered it as an impossible task! I had threrefore a lot of things by that time, not all related to each other. I worried about P.D.Es. with generalized solutions; I worried about δ, δ' of the physicists and whether they can be related to de Rham's currents. I had also functional analysis and duality. It was known after André Weil's book of integration on topological groups (1940) that instead of considering abstract measures on a sigma field it was possible to define a Radon measure μ on a locally compact topological space X as a continuous linear functional on the space C(X) of continuous functions vanishing outside compact subsets of X, μ was a functional, so that for every continuous function with compact supprt φ, $\mu(\varphi)$ was a number; μ was a linear form on C(X), continuous on each $C_K(X)$ (space of continuous functions on X with supprt CK, K compact subset of X) equipped with the sup-norm.

Frederic Riesz′ theorem is that every measure gives birth to such a functional, but A. Weil reverted the process and he defined in his book a Radon measure as being a continuous linear functional on the space of continuous functions with compact support. He had the $\mu(\varphi)$ and he had the carrier or support of μ, introduced by Henri Cartan in potential theory. Henri Cartan called "noyau fermè des masses" what is now called the support of μ. I learned that in his book in 1940. I had these functionals and I knew they were very important!

So it appeared to me that the dual of a nice functional space might be a very important space too. That was a fundamental step! I also put it in my mind and tried to study other cases. I thus made for myself (just for myself), in the year 1943, during the worst of the war, a theory of duality not for Banach spaces but for Fréchet spaces or for other topological spaces. I did not know the work of Mackey who had done many things in the United States but they were not known in France at that time.

I had a topological vector space E, for instance $E = C([0,1])$. I had the dual E′ which in this case is $M([0,1])$, the space of measures: $E' = M$. It could be very interesting to consider E′, for non-Banach spaces E, but the theory of duality for non-Banach spaces, at that time, at least in France, practically did not exist.

I did that for myself. I took E, a topological vector space, considered its dual E′ and tried to define the usuial things.

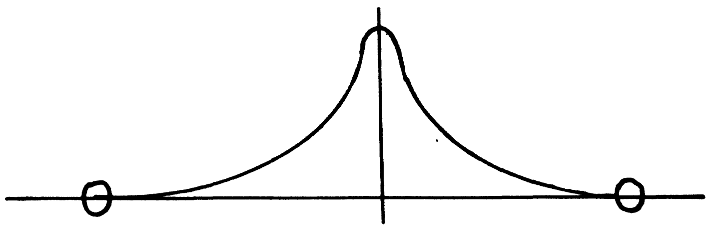
Hahn-Banach theorem, Banach Steinhaus theorem were known. I studied the bounded subsets, the weak and strong topology on the dual E′ (the strong one being the topology of uniform convergence on the bounded subsets of E), the bidual, the reflexivity. I remember I took the particular case $E = \mathcal{D}([0,1]) = C^\infty([0,1])$, and considered its dual E′, which is, in the distributions of today, the space $\mathcal{D}'([0,1])$ of distributions on [0,1] (more difficult than the space of distributions on $\mathbb{R}^N$ $\mathcal{D}'(\mathbb{R}^N)$, because of the singularities at the boundary $\{0,1\}$; but $\mathcal{D}(\mathbb{R}^N) = C^\infty_{comp}(\mathbb{R}^N)$ is much more complicated than $\mathcal{D}([0,1]) = C^\infty([0,1])$, that was my reason!). I didn't think at all of the possibility to introduce differentiation on $\mathcal{D}'([0,1])$ by transposing the differentiation on $\mathcal{D}([0,1])$! I just considered the topological properties (for instance $\mathfrak{D}([0,1])$ is reflexive). "Well, I said, I have a dual, I have reflexivity, it's nice. But it's probably completely uselless". This was in 1943, one and a half year before the distributions! I forgot it, I didn't think of it anymore. But the war prevented me from doing mathematics!

This proves that when people find various new concepts (as did not only Heaviside, who was rejected, but Sobolev and Bochner, and myself) they may often take certain steps without *realizing* they are finding something very useful. Some further work is necessary in order to see that it may be used, and that it is important.

Now comes an article by Deny and Choquet (October 1944), in which they prove something about polyharmonic functions (they appeared as functions which are solutions of an iterated Laplacian). But the theory is easy with the iterated Laplacian because polyharmonic functions, exactly as harmonic functions, can be defined without derivatives and they are immediately infinitely differentiable functions. Then I made some generalization, a small article of three pages which has been partly the start of the whole thing, called "Sur certaines familles non fondamentales de fonctions continues" (November 1944). In this, instead of finding polyharmonic functions, I found *generalized* solutions of a P.D.E. with constant coefficients. And the generalized solutions which I used were in the exact sense of Bochner, as the limit of usual solutions, by regularization (I invented them myself, ignoring Bochner's article).

Immediately this trick upset me: I had generalized solutions of a P.D.E. but I could just say that a function f is a generalized solution of the wave equation $\frac{\partial^2 f}{\partial t^2} - \Delta f = 0$ without $\frac{\partial^2 f}{\partial t^2}$ and Δf having any particular sense, once again!

But in order to define that, I used convolution with a C^∞ function φ with compact support: $f*\varphi$, and took the limit, if f is a function, of $f_n = f*\varphi_n$. The φ_n's were $\geqslant 0$, C^∞ functions as in Fig. 6, converging to δ in the sense that their support converged to zero and their integral was equal to one. And I said that f is a generalized solution of $P(D)f = 0$, if the $P(f*\varphi_n)$ are 0.



Then I showed my work to Cartan, but he told me: "C^∞ functions with compact support are a bit monstruous beings because they are not analytic. If you take a C^∞ function with compact support it is not analytic at the points marked in Fig. 6 (with all the terms being zero, and all the derivatives equal to zero) and therefore it has no Taylor expansion. These functions will be horrible, maybe you should better avoid to consider them".

I said, "yes of course, I know, I must be careful but they give the good trick here".

So actually he was a little sceptical about the use of them, but I was not sceptical at all because they were giving here the shortest proof. In fact, I found only a few years ago an article by N. Wiener in 1926, in which he says that although the C^∞_{comp} functions are monstruous beings, they are very useful indeed; and he proves a theorem with this, which can be considered as a theorem of Distributions.

Then, in November 1944, suddenly I got a good complete theory for everything, a theory which gave derivatives to all beings and made unification between a lot of things. They were the Distributions.

I could then justify δ and δ', the generalized solutions of the P.D.Es. with separate meaning for the terms. Now when I have:

$$\left(\frac{\partial^2}{\partial t^2} - \Delta\right)f = 0$$

each term $\left(\frac{\partial^2 f}{\partial t^2}\text{ and } \Delta f\right)$ separately can be a distribution and the sun of the two distributions gives zero.

I had exactly what I wanted and that for me was the beginning of the revelation. During a long night of thinking I found most of the theorems. Just in one night! Cartan was now immediately enthusiastic, also the other members of Bourbaki, whose support helped me very much.

However my definition *was not* the definition which is adopted now, it was a different one, *much more complicated*!

It was not yet $T(\varphi)$, it was something else. I did not call them distributions, but operators, because of their origin in Choquet and Deny's problem. I considered the space $\mathcal{D}(\mathbb{R}^N)$ of C^∞_{compl} functions and $\mathcal{E}(R^N)$ the space of C^∞ functions with arbitrary support; I put the good topologies on the $\mathcal{D}'_K$'s, and on $\mathcal{E}$; then an operator T was not a distribution, but a linear map from $\mathcal{D}$ into E, continuous on each $\mathcal{D}_K$: $\mathcal{D} \xrightarrow{T} \mathcal{E}$, commuting with convolution with $\mathcal{D}$.

I called this map: $\varphi \longrightarrow T \,.\, \varphi$. This was a convolution operator: to every test function $\varphi \in \mathcal{D}$ it assigned a function $T \cdot \varphi \in \mathcal{E}$, with the commutation $T \cdot (\varphi * y) = (T \cdot \varphi) * y$ · So I called them operators. There operators, if T is a function, gave exactly the usual convolution:

$$(f \cdot \varphi)(x) = \int f(x-\xi)\varphi(\xi)d\xi;$$

a locally integrable function was an operator. I put on the space $\mathcal{D}$ of operators the topology of uniform convergence on bounded subsets. One could differentiate the operators infinitely many times. The derivative of an operator, DT, is defined by: $DT \cdot \varphi = T \cdot D\varphi$ (DT convoluted by φ equals T convoluted by $D\varphi$) *(no minus sign)*. So every operator, in particular every continuous function, is infinitely differentiable in the space of operators. A function which has no usual derivative has a derivative operator.

I immediately saw δ: the convolution with δ is just the identity operator: $\varphi \xrightarrow{\delta} \varphi$. δ' is just the derivative: $\varphi : \xrightarrow{\delta'} \varphi'$.

So I had all that concerned the Physicists. I had the convolution of two distributions: $(S*T) \cdot \varphi = T \cdot (S \cdot \varphi)$, if S has a compact support: if S has a compact support, φ is indefinitely differentiable with compact support, $S \cdot f$ is also infinitely differentiable with compact support, and $S*T$ is an operator. So that, apart from the restriction of compact support, you could convolute arbitrary operators.

I remember that, some day later on, after a long work in the night, I had a dream. I was so enthusiastic about the possibility of having convolution of two arbitrary operators (but of course the result of convolution is an operator), that I dreamed I was explaining to somebody that you could convolute whatever you wanted, for instance you could say Mozart* Beethoven, but of course it would not be a musician, but an operator; and you could convolute Nancy and Strasbourg, but of course it would not be a city, but an operator. It may seem extremely strange that, having the $\mu(\varphi)$ of Weil to define a Radon measure, having studied the duality E-E′ in 1943, I defined an operator $\varphi \longrightarrow T \cdot \varphi$ from $\mathcal{D}$ into $\mathcal{E}$, instead of a linear form $\varphi \longrightarrow T(\varphi)$, a distribution in the today meaning. It comes of the origin of the invertion Choquet-Deny's problem. I didn't realize that a duality formula with $T(\varphi)$ a number, would have given also a differentiation, with $T'(\varphi) = -T(\varphi')$, $(P(D)T)(\varphi) = T({}^tP(D)\varphi)$. This came only several months later! My $T \cdot \varphi$ of November 1944 is the $T*\varphi$ of today. This proves the important negative role of the inhibitions; to discover is sometimes to break inhibitions!

A strong obstacle was to define a product αT, the multiplicative product of a distribution by a C^∞ function α. It is very difficult to define a product because multiplication does not commute with convolution! But I had also a theorem at this time: that every distribution T, locally speaking, is a derivative of a continuous function f (may be without usual derivative) i.e. $T = D^P f$, so that all you have to do is to define $\alpha D^P T$ formally and then you can use Leibnitz formula to get it. But it was necessary to have this finiteness theorem, that every operator, locally, is a derivative of a continuous function (without usual derivative). Of course one had to prove that αT is independent of the representation $T = D^P f$; this was done by a limiting procedure, it was true for a C^∞ function, and the C^∞ functions were dense in $\mathcal{D}'$. I had the topology, and all my theorems of 1943 about the duality and about the usual topological vector spaces came back. However I was unable to find Fourier Transform. Here again the definition as convolution operators made the things awkward.

I couldn't find it, because I tried to define the $T \cdot \varphi$ as a function but not $T(\varphi)$, a complex number. Moreover, it was necessary to introduce this space $J(\mathbb{R}^N)$ of C^∞ functions, rapidly decreasing at infinity as well as their derivatives and its dual $J'(\mathbb{R}^N)$, the space of temperate distributions. It was unusual, at this time to introduce *many* functional spaces. But it came five months later. I was in Grenoble 1944-45. Cartan was in Paris and we had a correspondence and we met often. During several months I was stubborn with this $T \cdot \varphi$; manipulated it better and better but I wasn't able to get the result. Eventually one day I said, well why not take $T(\varphi)$? It was in February or March 1945 (about four or five months after a beginning); I wrote immediately a letter to Cartan, saying I had the trick! One must consider distributions as linear forms, and not as operators. Why the name distributions? Because $\mu(\varphi)$, when μ is a measure, is a distribution of charges in the universe; electric charges for instance; with distributions you have the dipoles, you have the magnets, double layers, you have distributions which are exactly physical distributions of masses, magnetic or electric distributions with positive and negative charges; and if you take the space, introduced by Deny later, of all distributions which are of finite energy and therefore may intervene in physics, they are distributions, not measures. As soon as de Rham knew my article on distributions he found immediately the general notion of currents, which he was looking for, for many years! It was still difficult at the time of the operators, but with the distributions it worked perfectly. Although I had that also in my mind for years, I was essentially preoccupied at that time with $\mathbb{R}^N$, multiplication, convolution, Fourier transform (with the functional spaces S, S', O_M, O_C'); I knew more or less how to manage to get currents, but really didn't think of them seriously; but he considered immediately these currents, their coboundary, their Laplacian on a Riemannian manifold, their cohomology, etc... He published his book "Formes, currents, formes has moniques", in 1955. I found also distributional sections of vector bundles, but published them only much later.

After two articles in the "Annales de l'Institute Fourier" (1945, 1947), I slowly wrote the book; it was achieved in 1950-51. Then, of course, everything was arranged. I think I had the first elements when I was a student, in 1934 (Dirac introduced his function in 1926-27), I had the definite form in 1945 and I published the book in 1950-51. 15 years!

I also want to mention that in doing research, one can waste a lot of time on easy things and find immediately very difficult ones. For instance, there has been a theorem about which I was thinking one whole week, eight hours every day, and it became more and more difficult to prove it (I believed the theorem was true). Each day, at the end of the day, I believed I had the proof, but I was too tired and went to sleep and in the morning I found the proof was false. I tried to prove that a distribution carried by a compact subset K can be expressed as a finite sum of derivatives of measures carried by K. It's false, K has to be "Whitney regular". This lasted one week, one entire week of suffering. Research is not only enjoyment; it is enjoyment when you find the results, but you may suffer for a long time and the enjoyment is the result of this suffering!

I became more and more upset about that theorem and in the morning of the last day I found a counterexample of four or five lines. The theorem was false! This I wrote in my book using some kind of irony against myself. It goes aw follows: "One could believe that the following theorem is true (and I wrote the theorem); it is not true though, as proved by the following very trivial counterexample". And I gave the very trivial counterexample and nobody knew I had suffered one week to find it.

So that was a little of this history. I gave a course in Paris in 1945 about Distributions (cours Peccot) which was attended by about 25 people; half of them were Physicists and in particular workers in electricity.

I also gave a lecture about that in the Symposium of Harmonic Analysis in Nancy in 1947. There were many foreign mathematicians present. Among them Harald Bohr was very enthusiastic, and he invited me to Copenhagen where I gave several lectures on distributions, to a large audience; that was my first travel abroad in my professional life, the beginning of a long series! At the same occasion, I went to Lurd, where I made acquaintance with Marcal Riesz and Gording. I lectured also in Oxford and London in 1947, and was invited to the Canadian Mathematical Congress in 1949 at Vancouver, also to speak about distributions. At the International Congress of Mathematicians, Cambridge, Mass, U.S.A. 1950, I received the Fields Metal for my work on distributions (before the publication of the book!), and I gave a talk on the theorem of kernels, which was published only later on. So the resonance of the distributions grew reasonably!

However, when the book appeared in 1950, it was not yet quite accepted.

I had to fight some battle against two quite opposite categories of critics: some people said it was so simple that it could not really be useful, and some other said it was such a complicated definition of a generalized function that it could not be handled and could not be used.

So some people found it too simple to give useful results while some others found it too complicated to be usable. Sometimes I thought myself one or the other of these contradictory things! So a battle had to be fought which I remember very well. Also Gording told me he kept it in his memory, after I gave my lecture in Lund, in 1947. Hörmander was about sixteen years old then, I guess. He only entered the University a little later so he learnt distributions in the first years of his studies, and found rapidly applications.

I remember the battle I fought to make the distributions universally accepted, something which is part of the researcher's world. Some of my younger students, as Lions, or Malgrange, may say that it was immediately accepted, but this is not quite true. They came to the distributions as students, in Nancy in 1945-50, they accepted them immediately; other young people started r[illegible]arch in the years 1953-54-55, when the battle was over. The work of th[illegible]unger generation has been an essential part of the success.

In the years after 1950 came many articles or books; Gelfand-Shylov's books on generalized functions are very complete and self-contained. Now every mathematician introduces easily new vector spaces; and later, distributions were generalized by Sato's hyperfunctions.

Remark. Distributions can be considered as a generalization of functions in the sense in which real numbers are generalization of rational numbers. How do we generalize the rationals; by the cuts defined by Dedekind on the straight line. After all, at some time, it became necessary to define correctly the irrational numbers! and Dedekind introduced cuts on the real line, a rational number defined a cut, so that the reals are generalization of the rationals.

And there is an article by Peano, in 1912, in which he considers Heaviside's computations and he says: "I am sure that something is to be found now; there must be a notion of generalized functions which are to functions what the reals are to the rationals". Peano writes this in 1912, so very much earlier to 1944!

Reading that text again, I think I exaggerate the convergence of a lot of vario[illegible]stions to the same solution: distributions. It has been so, and this expla[illegible] the strength of my own enthusiasm for distributions, because it solve[illegible]ether many problems I had before. But it is sure that till the final colle[illegible]solution came, these problems were not bound together in my mind, they were not thought as having the same solution: partial differential equations (wave equation) Dirac's δ of the physicists, de Rham's currents, du[illegible] dual of C^∞ [0,1]) were considered by me as *independent* things. T[illegible]fication by the distributions causes my enthusiasm!