Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 5, n°1; Juin/June 2009

www.jehps.net

Summary of the scientific work of Mr. Jean Ville (May 1955)[1]

Translation by Glenn Shafer

Contents

[1] *Translator's note:* The French original of this document, entitled "NOTICE sur les travaux scientifiques de M. Jean VILLE", was obtained by Pierre Crépel in 1984 from the University of Paris. I have occasionally shortened phrases, changed formatting, and silently corrected typographical errors. The original did not have a table of contents. Footnotes not labeled "Translator's note" are in the original. I am grateful for advice from Bernard Bru, Pierre Crépel, and Jean-Yves Jaffray.

1 Research on R. von MISES's theory of the "collective"

My first publications, extremely modest, were my contributions to the Mathematical Colloquium that met in Vienna under the direction of Mr. Karl MENGER, in which I participated thanks to an Arconati-Visconti scholarship.[2] Mr. MENGER's course dealt with geometry, especially the theory of dimension initiated by H. POINCARE. I published two small notes in the proceedings of the Colloquium. The first concerned the theory of so-called curves of quadratic length, in which one studies the limit of the sum of squares of the edges of an inscribed polygon [1]. The other concerned a proposition about the spaces derived from metric spaces by replacing the distance by its square root; these are also metric spaces, and if they contain two, three, or four points, they can be embedded in a Euclidean space of 1, 2, or 3 dimensions. I showed that the correspondence stopped at that number [2].

While attending the sessions of the Colloquium, I took part in discussions of R. von MISES's proposed new definition of probability, and the subject seemed worthy of study to me. To understand the state of the question at that time, one must pause a bit over the controversy concerning the definition of probability. It had long been well known that if one undertakes a sequence of independent trials that can each give rise to an event E or its contrary $\overline{E}$, and if one writes r_n for the number of occurrences of E in the first n trials, so that $f_n = r_n/n$ is the frequency of E in these initial trials, then one obtains, writing p for the probability of E, which is assumed to be constant,

$$\lim_{n\to\infty} \text{Prob}\{|f_n - p| < \epsilon\} = 1.$$

This constitutes the "weak" law of large numbers, conventionally expressed by saying that f_n tends in probability to p. This limit "in probability" is not a limit of the kind defined in analysis and obviously could not give any excuse for a definition of probability by means of a limit in the sense of analysis.

[2] *Translator's note:* In the original, Ville writes "Asconati" rather than "Arconati." He made this error on numerous occasions.

This situation changed completely when CANTELLI, using the results of E. BOREL's famous article on denumerable probabilities,[3] known as the "mémoire de Palerme," succeeded in showing that

$$\lim_{n\to\infty} \text{Prob}\{|f_n - p| < \epsilon \ \text{ and } |f_{n+1} - p| < \epsilon \ \text{ and } |f_{n+2} - p| < \epsilon \ \text{ and} \ldots \text{ad inf.}\} = 1,$$

making it appear that the probability of the event

$$\lim_{n\to\infty} f_n = p,$$

the limit now being in the sense of analysis, was equal to one.

Mr. von MISES then proposed the following two-part definition of probability:

1°) If we undertake a sequence of independent trials, the frequency of E tends to a limit.

2°) By definition, this limit is the probability of E.

The first part is an assertion having the same properties as a law of nature; the second is a definition in the proper sense of the term. Mr. von MISES's definition was completed by this condition:

3°) The limit p is not changed if one selects from the sequence of results of the trials an infinite partial sequence, using a "choice procedure" that decides whether to choose or omit a result based only on the values of preceding results.

1°) was called the axiom of the limit;

3°) was called the axiom of choice procedure.

The probability calculus thus became the study of the properties of certain sequences, those satisfying the axioms. R. von MISES called these sequences collectives. The axiom of choice procedure permitted the exclusion of regular sequences such as

$$E \ \overline{E} \ E \ \overline{E} \ E \ \overline{E} \ \ \ldots\ldots,$$

in which the frequency of E is obviously 0.5, but in which the value of the limit changes if one implements the very simple choice consisting of keeping only every second term.

R. von MISES's axioms were obviously contradictory, because he considered all choice procedures, which form a set with the cardinality of the continuum. The restriction that a choice should only take account of the preceding results was insufficient; there was no sequence satisfying his axioms.

So A. WALD, a participant in the Colloquium, proposed to limit the choice procedures to a countable number, which he could do by saying

> A collective in a deductive theory, which itself has only a finite number of axioms and consequently permits us to formulate only a countably infinite number of choice procedures, is a sequence such that. . . .

[3] *Translator's note:* Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo matematico di Palermo* 27:247–270, 1909.

The axioms became noncontradictory, for the following simple reason:

If we represent an infinite sequence by a point on the segment $(0, 1)$, the probability law that the sequence satisfies defines a countably additive measure on the segment. One can associate with every choice procedure the points representing the sequences for which the axiom of choice procedure is not satisfied. This set has measure zero with respect to the measure considered. Because we are considering only a countable set of choice procedures, we are excluding from the segment only a set of measure zero; the complementary set is not empty, and there is no contradiction.

The issue seemed to be settled. It appeared possible to consider sequences, defined starting with the deductive theory of "Principia Mathematica" for example, which would have the properties of a random sequence—for which it would be impossible, by the most ingenious manipulation, to modify the limiting frequency.

I then asked myself the following question. We are given this:

> Saying that a sequence of results enjoys a certain property with probability equal to one comes down to saying that the sequences not enjoying that property are represented by the points belonging to a set of measure zero.

It seems that the demonstration of the first assertion consists of covering the set of points representing sequences not enjoying the property with a set of measure zero. This being acknowledged:

> Do there exist sets of measure zero that one cannot cover with sets of measure zero defined by the choice procedures?

In other words, do there exist properties provable in the classical theory but not provable in the sense of R. von MISES (as completed by the work of A. WALD)?

I was able to show that this was the case, i.e., that the category of sets of measure zero defined as sets of points representing sequences that do not satisfy an axiom of choice procedure is not large enough to cover all the sets of measure zero. It is obvious that the sets of measure zero, countably infinite in number, associated with a countably infinite number of choice procedures, cannot cover all the sets of measure zero, but it is less obvious that there exists a set of measure zero independent of the choice procedures, such that no matter what infinitely countable set of choice procedures is adopted, one cannot cover it with sets of measure zero associated with these choice procedures.

I was guided in my search for a proof by LIOUVILLE's method of showing that there exist nonalgebraic numbers; I considered not sequences where the frequency was divergent but those where the frequency converged too rapidly. Once I found the proof, I was able assert a very simple result:

> To find a set of measure zero not coverable by the sets defined by R. von MISES, it suffices to consider the sequences in which f_n converges to p unilaterally. They can be collectives in the sense of

R. von MISES while possessing a peculiarity that the probability calculus excludes.

2 The notion of martingale

I proposed the preceding results to M. FRECHET as a thesis topic, and he was willing to accept it. I then proposed to replace R. von MISES's axioms with axioms that did not have the important gap I had brought to light. I noticed that the impossibility of changing the frequency could be interpreted as the impossibility of winning in a fair game using a martingale that excludes certain trials; I wondered if it would be possible to go farther by generalizing the martingale to permit continuous redistribution of the stakes. I had been struck by a result of LAPLACE, according to which a gambler playing heads and tails with a biased coin can gain an advantage on the first two flips with a judicious distribution of the stakes, using the information that the coin is biased but not knowing to which side.[4]

So I assumed that you start a game of heads and tails with a sum of money equal to unity, and before each flip you have the opportunity to redistribute the money you have at your disposition at that point depending on how the previous flips have come out. By the results I had obtained previously, going farther than R. von MISES's methods required being able to take advantage of fairly vague information, such as

> In the sequence of draws, the frequency will tend to 1/2 from one side, but we do not know which, nor from what point onward.

I realized that knowledge of any set of measure zero in which the point representing the sequence was located would permit the construction of a strategy that would win an amount of money exceeding any given limit. So I proposed substituting the notion of "martingale" for that of "choice procedure." Contradiction would be avoided if one talked about the (countable) set of assertable martingales, and one would achieve consistency with the set of assertable sets of measure zero, which was not the case with R. von MISES's theory.

I generalized the notion of martingale, which emerged as a very valuable tool for proof. Here is a simple example.

> In BERNOULLI trials with probabilities p and q, it is very easy to give a strategy guaranteeing one will have

$$\frac{r!(n-r)!}{(n+1)!}p^{-r}q^{r-n}$$

[4] *Translator's note:* In his *Essai philosophique sur les probabilités*, Laplace pointed out that a coin with a small unknown bias has a slightly elevated chance of producing two heads (or two tails) in two flips. For example, if the coin has probability 0.55 of coming up heads under hypothesis 1 and probability 0.45 under hypothesis 2, and both hypotheses have probability 1/2, then the probability of two heads is 0.2525 instead of 0.25. A gambler who is allowed to bet on heads at even odds can take advantage of this by betting \$1 on heads on the first flip and then betting \$2 or \$0 on the second, depending on whether the first came out heads or tails. The expected value of his net gain is $3 \times 0.2525 - 1 \times 0.7475 = \$0.01$.

at the end of the nth trial, r being the number of times the event of probability p happens.[5] This expression is therefore bounded.[6] STIRLING's formula permits us to deduce from this the strong law of large numbers and even the fact that $r/n-p$ tends to zero at least like $\sqrt{(\log n)/n}$.

I was also able to demonstrate, using the idea of a martingale, the following proposition that had been announced by KOLMOGOROV without proof:[7]

> A necessary and sufficient condition for the number r of successes in a sequence of BERNOULLI trials to satisfy
>
> $$|r - np| < \phi(n)\sqrt{2npq}$$
>
> from some point onward with probability 1, where $\phi(n)$ is the sequence of values taken for integral t by an increasing function $\phi(t)$, is that the integral
>
> $$\int \phi e^{-\phi^2} \frac{dt}{t}$$
>
> converges as t tends to ∞.

This idea of a martingale also turned out to be useful to other researchers. I was calling mathematicians' attention to an idea that existed already, because the word existed, but had not been considered important. The idea was taken up by J. L. DOOB, who speaks about it in these terms: "Although many authors had derived many martingale properties, in various forms, Ville was the first to study them systematically, and to show their wide range of applicability."[8]

The results of the studies mentioned above were published in various notes in the Comptes Rendus and collected in the thesis that I defended in 1939 [4]. This thesis, with some additional material, became a book in the collection "Monographies des Probabilités," edited by Mr. E. BOREL [5].

3 Contribution to the theory of strategic games

Before I defended my thesis, I was charged by Mr. E. BOREL with writing up the lectures he gave at the Sorbonne on the theme of games of chance. Mr. E. BOREL called to my attention the article by J. Von NEUMANN in the

[5] *Translator's note:* If we have seen r heads and $n-r$ tails, we bet the fraction $(r+1)/(n+2)$ of our current capital on heads on the next trial and the rest on tails. Our capital is then multiplied by $(r+1)/(n+2)p$ if heads happens, by $(n-r+1)/(n+2)q$ if tails happens. Ville's formula follows by induction, assuming that the initial capital is 1.

[6] *Translator's note:* The expression is the gambler's capital after the nth trial. By Ville's new principle, no strategy (martingale) will allow the gambler to get indefinitely rich (this is the generalization of von Mises's principle that no choice procedure can change the frequency of heads).

[7] *Translator's note:* See p. 103 of Ville's book.

[8] J. L. DOOB: Application of the theory of martingales, in Le Calcul des Probabilités et ses applications, Editions du C.N.R.S., 1949 (Colloques internationaux du C.N.R.S.).

Mathematischen Annalen, in which he proved the "Minimax Theorem," which plays a fundamental role in the theory of games involving both chance and the players' abilities. The principles of the theory and the definition of strategy had been formulated by Mr. E. BOREL, but there remained the question of whether the minimum gain for one of the players, the gain he can guarantee himself by maximizing the minimum that he might suppose the other player would allow him, was or was not equal to the maximum gain to which the other player might limit him by proceeding in the same way. In more concise terms, when A's gain G depends on the attitudes of both players, we want to know whether

$$\max_{(A)} \min_{(B)} G = \min_{(B)} \max_{(A)} G.$$

It seemed to me that J. Von NEUMANN's proof did not take account of the real nature of the problem, and I devised a new proof based on the theory of convex sets. The simplicity of this new aspect of the problem allowed me to generalize the theorem to the case where the strategies available to A and B form continuous sets. The relation between the problem of these games and the theory of convex sets, classical since MINKOWSKI, may seem obvious today, but at the time I wrote my note, it was not. I allow myself to cite on this point the opinion of J. Von NEUMANN, speaking about the evolution of the problem:[9]

> This connection may now seem very obvious to someone who first saw the theory after it had obtained its present form. . . . However, this was not at all the aspect of the matter in 1921–1938. The theorem, and its relation to the theory of convex sets were far from being obvious, witness the following facts:
>
> (a) In 1921, and thereafter, Borel surmised the theorem to be false or possibly false.
>
> (b) In 1928, I proved the theorem by observing its relation to the theory of fixed points and not yet to that one of convex sets.
>
> (c) In 1935, I generalized it (for the purposes of the theory of prices and production) by an even more explicit use of the fixed-point method.
>
> (d) It took ten years after my original proof, until J. Ville discovered, in 1938, the connection with convex sets.
>
> (e) Even now, this connection does not tell the entire, or the simplest, story about the theorem, as the work since 1945 of S. Kakutani, J. Nash, G. Brown, and myself shows.

I have subsequently worked on the theory of games only incidentally, in lectures given at the Institut Henri Poincaré, where I showed in a simple way

[9] *Translator's note:* Communication on the Borel notes, *Econometrica* 21(1):124–125, January 1953. Von Neumann was writing in response to a note by Fréchet that made a case for the importance of Borel's work on game theory. While reacting negatively to this (he had done his own work without knowing of Borel's), von Neumann went out of his way to praise Ville. Von Neumann wrote in English, and Ville quotes the English.

that the Minimax Theorem does not wrap up the question, even in a perfectly determined game like chess, because while it does show a perfect player how to play best against another perfect player, it does not give a perfect player the means to take maximum advantage of an imperfect player [6].

4 Research on the definition of entropy in information theory

Having had to learn information theory in connection with problems of electrical communication, I was displeased to see that the entropy

$$H = -\sum p_i \log p_i$$

of a set of probabilities $p_1, p_2, \ldots, p_n$ was introduced *a priori*.[10]

It was well known, obviously, that this function H had served statistical thermodynamics very well, and experience had also shown it to be useful in transmission problems, but I could not see the exact meaning of the expression. I tried to interpret it as the minimum number of binary trials needed to discover the state of a system that can take n states with the given probabilities, and I showed that if you take the logarithms to base 2, this average number does in fact fall between H and $H+1$. Because M. SCHUTZENBERGER had also undertaken the same investigation, we put the results together in a joint note in the Comptes Rendus [7].

In my search for another interpretation, I was inspired by the existence of machines that generate random binary digits, used in so-called Monte-Carlo methods of computation. The problem of generating a very long sequence that has the properties of a sequence of outcomes in the game of heads or tails is already a quite difficult technical problem, because approximate cycles are difficult to avoid. So if one needs a sequence of draws with probabilities p and q different from 1/2, it is too extravagant to imagine constructing a machine that provides this sequence directly; one should use an existing machine by coding its output appropriately. I showed that the average number of draws from the "standard" needed to obtain an element of a sequence of draws that should be provided by the probabilities

$$p_1, p_2, \ldots, p_n, \ldots$$

was precisely H.

This result brought my attention to the importance of the geometric mean for certain questions in probability. In fact, H is just the logarithm, with the sign reversed, of the geometric mean of the probability of the event that happens.[11] I was then able to show that the degree of verification of a statistical hypothesis is measured not by the probability of the event that has happened but by the deviation from its mean value (in the sense of geometric mean) of the probability

[10] See for example C. SHANNON: Mathematical Theory of Communication.

[11] *Translator's note:* $e^{-H} = e^{\sum p_i \log p_i} = \prod p_i^{p_i}$.

of the event that has happened. Thus it is not necessarily the hypothesis that gives the event the greatest probability that is best verified.

The results mentioned above were exposited in the lectures I gave at the Institut Henri Poincaré [6].

5 Theory of electrical networks

On the purely technical side, I worked on propagation along cables with irregular impedances [21], the phenomena of echoes and lags [22], and the combination of linear networks. Although its purpose was essentially didactic, I highlight the article on the synthesis of an impedance, where I put together results from CAUER, PILOTY, LEROY, and where I introduced a new method of presentation to bring the question within the grasp of technicians. This method allows the use of results on HURWITZ's theory of polynomials that are hardly classical. I showed that the inequality

$$|\arg Z(\lambda)| \leq |\arg \lambda|$$

is satisfied by a function $Z(\lambda)$ of a complex variable that is positive in the sense that $Z(\lambda^*) = [Z(\lambda)]^*$, and $\operatorname{Re}\lambda > 0$ implies $\operatorname{Re} Z(\lambda) > 0$. The proof uses the physical meaning of the condition. I also showed that if $h(\lambda)$ is a HURWITZ polynomial and if ρ_1 and ρ_2 are complex numbers of modulus 1, then the equations $h(\lambda)+\rho_1 h(-\lambda) = 0$ and $h(\lambda)+\rho_2 h(-\lambda) = 0$ have their roots on the purely imaginary axis, and these roots are intermingled. This allows one to show that every reactance (nondissipative impedance) is the quotient of the even and odd parts of a HURWITZ polynomial [8].

6 Bringing information theory into the formation of probability assumptions

My most recent work is in information theory, which I think should transform the probability calculus. In a lecture to the Congrès de Philosophie des Sciences, Paris (1949) [9], I explained the viewpoint that seems to me to emerge from the development of this young theory.

Consider the information that one learns that the true values of the probabilities of an event and its contrary are P and Q, whereas one had previously believed them to be p and q. I defined the "value" of this information by the expression, always positive or zero,

$$I = P \log \frac{P}{p} + Q \log \frac{Q}{q}.$$

This gives the information "the event happened" the value $\log(1/p)$, which becomes larger as p is made smaller. Because I is a measure of the advantage knowledge of the event's happening gives a gambler, these definitions provide a practical vision of reality.

They also permit the study of another question. Supposing p and q are given, one can evaluate the mean value of the information provided by the happening or failing of the event:

$$p \log \frac{1}{p} + q \log \frac{1}{q} \leq \log 2.$$

It is events that are equally probable for which that information has the greatest value. To put it a different way, it is trials that have one chance in two to succeed that are the most instructive. If $p \neq q$, there is a remainder

$$\log 2 - p \log \frac{1}{p} - q \log \frac{1}{q}$$

not found in the information about the event. I proposed to attribute this remainder to the information in the formation of the probability assumption. This allows the reestablishment of complete symmetry between two operations, prediction and making the assumption. We see in particular that the assumption $p = q = 0.5$, which corresponds to the most instructive experiment, is consequently the one that is never confirmed or disconfirmed as a probability assumption.

These considerations were developed in the lectures I gave at the Institut Henri Poincaré, in the chapter on probabilities and frequencies [6]. This is an effort to liberate the probability calculus as much as possible, from a foundational point view, from the notion of frequency. The goal is to base the probability calculus on direct interpretation of results. With these ideas, I succeeded in clarifying the following paradox: if tails have probability 0.501 and heads probability 0.499, and if tails come up a thousand times in a thousand flips, we are not satisfied, even though this is the most likely sequence of a thousand results. I have clarified the paradox without evoking either frequency or BERNOULLI's theorem.

7 Contribution to the definition of correlation coefficients

In addition to lecturing on classical topics at the Universities of Poitiers and Lyon, I have given some lectures where I have tried to introduce some personal results.

At the College of France in 1942, I gave a course on statistical correlation (PECCOT Foundation[12]). It was not published, but I have taken up and completed some of the results in lectures at the Institut de Statistique at the University of Paris [10].

In my course on matrix analysis, I generalized multiple and partial correlation from the case where it is relative to a single unknown variable, i.e., the case of the form

$$R_{1\cdot 23\ldots n} \qquad\qquad R_{1\cdot 2(34\ldots n)}.$$

[12] *Translator's note:* The Peccot Foundation awards a prize to a young mathematician annually. The recipient lectures on his work at the College of France. Ville gave his lecture in the first semester of the 1942–1943 academic year.

By appropriately interpreting matrices extracted from the variance-covariance matrix, I defined coefficients of the form

$$R_{\alpha\cdot\beta} \qquad\qquad R_{\alpha\cdot\beta(\gamma)},$$

where α, β, and γ are groups of indices for the variables.

8 Teaching matrix algebra

In the exposition of matrix theory [10], I introduced a new notation, in which Ψ represents a parallelepiped of contravariant vectors, and $\Phi = \Psi^{-1}$ represents the dual parallelepiped of covariant vectors, which allows us to express any matrix A in the form of a product

$$A = \Phi_1 \Psi_2$$

and any linear operator in the form

$$\mathcal{T} = \Psi_3 \Phi_4.$$

The theory of transforming coordinate systems becomes immediate, because $\mathcal{T}$'s matrix in the coordinate system Ψ is

$$A = \Phi \mathcal{T} \Psi,$$

while the operator with coordinates A in the coordinate system Ψ is

$$\mathcal{T} = \Psi A \Phi.$$

Apparently for the first time, I interpreted the operator associated with a variance-covariance matrix in this way, and I showed what group of transformations leaves that operator invariant. This is useful in regression theory, where the classical expositions are vague about the coordinate system of the deviations.

9 Teaching Boolean algebra

I gave a course on Boolean algebra that emphasized the solution of Boolean equations, a subject that seems to be neglected in classical expositions but is of great importance in the theory of logical machines. At the end of the course I added some ideas about lattices, and I showed how the classical theorem of logical intuitionism

> Triple negation is equivalent to simple negation (even though double negation is not equivalent to affirmation).

can be interpreted using a distributive lattice without complementation.

10 Contribution to the theory of price indices

As part of my teaching at the University of Lyon, I investigated the theory of price indices. I showed that one can compare two situations, one where goods are sold in quantities q_i at prices p_i, the other where the same goods are sold in quantities q'_i at prices p'_i, while distinguishing two elements in the variation of the value $\sum p_i q_i$, one corresponding to changes in the standard of living, the other to changes in the value of money. The proportions of the two elements vary between limits that can be calculated in some cases from the numbers p_i, q_i, p'_i, q'_i. I deduced conditions for the existence of a utility function from this theory, using considerations that involve the conditions for integrating a PFAFF form [12, 13].[13]

11 Theory of signals with bounded spectra; their extension

In connection with my technical work, I had to concern myself with signals with bounded spectra, i.e., functions $s(t)$ for which the FOURIER transform is zero outside a certain interval of frequencies. If the interval of frequencies is $(-f_s, +f_s)$, then the signal is known to be completely determined by its instantaneous values at time points in an arithmetic progression with common ratio $1/2f_0$ (Sampling Theorem[14]). I studied various properties of these signals, showing how they could all be derived from the fundamental proposition that the integrals of $s(t)$ and $s^2(t)$ are expressed exactly by the sums that one substitutes for them when calculating by the method of rectangles, provided the distance between the successive equidistant points is less than $1/2f_0$. I extended the theory to signals with bounded analytical spectrum, i.e., spectra of the form $\Psi(t)$ with $\Psi(t)$ a holomorphic function of t in the domain $\operatorname{Im} t > 0$.

I also considered the problem of analytically extending a signal with bounded spectrum. Being an analytic function, such a signal can theoretically be calculated for $t > 0$ if one knows the values for $t < 0$, but WEIERSTRASS's method of analytical extension is physically inadequate. I started from the following idea:

Because the signal is completely determined by its ordinates at intervals of $1/2f_0$ in the interval $(-\infty, \infty)$, one should be able, by tightening the grid in $(-\infty, 0)$, to define it sufficiently to be able to calculate the extension from those ordinates, defined for negative abscissas in an arithmetic progression. In fact, using the interval $1/6f_0$, I showed that the extension was possible by a procedure that uses only discrete sequences of abscissas. I also undertook to define a linear operator for the extension, which could be implemented in a prediction network, and I showed the structure of the operations needed, a sequence of filters and

[13] *Translator's note:* The contribution of this article has been analyzed by François Gardes and Pierre Garrouste in "Jean Ville's contribution to the integrability debate: The mystery of a lost theorem," *History of Political Economy*, 38(supplement):86–105, 2006.

[14] *Translator's note:* In English in the original.

modulations. In other words, I defined a linear functional

$$\mathcal{F}\begin{pmatrix} s(t) \\ t < t_0 \end{pmatrix} = s(t_0 + h) \qquad\qquad h > 0$$

by a sequence of operations that suppress frequencies or instantaneous values [14].

12 Research on the operator $\exp\left\{x + \frac{d}{dx}\right\}$ and related topics

It is well known that the operator d/dx, which differentiates, and the operator x, which multiplies by x, have a simple commutation law when applied to functions $f(x)$ with infinitely many derivatives. I considered the operator defined formally by

$$\exp\left\{x + \frac{d}{dx}\right\}$$

and developed it in a series. I showed that the operation results in the geometric mean of the results obtained by separately applying the operators

$$\exp\{x\} \cdot \exp\left\{\frac{d}{dx}\right\} \quad \text{and} \quad \exp\left\{\frac{d}{dx}\right\} \cdot \exp\{x\}.$$

I used this result to discuss SCHRÖDINGER's equation. In classical mechanics, we can associate with the joint distribution $F(q, \dot{q})$ of q and $\dot{q}$ (position and velocity) a characteristic function

$$\phi(u, v) = \int \exp\{uq + v\dot{q}\} F(q, \dot{q}) dq d\dot{q},$$

from which one can recover $F(q, \dot{q})$. In wave mechanics, we have not F but a wave function $\Psi(q)$ playing almost the same role, but with the difference that if position is associated with q, velocity is associated with the operator d/dq, so that to obtain the distribution of two variables (in the rectangular case) starting with a function of one variable $\Psi(q)$, we must use a trick. I proposed looking for a characteristic function and defining it as

$$\mathcal{M} \exp\{i(up + vq)\} = \int \Psi^* \exp\left\{uh\frac{d}{dq} + ivq\right\} \Psi dq.$$

Thus we use the operator mentioned earlier. We obtain in this way a convention that allows the calculation of the distribution of p and q, a real function of two variables, starting with Ψ, a complex function of one variable. The joint distribution of p and q thus obtained is consistent with HEISENBERG's uncertainty relation, which allows us to think that the operator we introduced formally corresponds to a physical reality [15].

This operator has also been useful to me in my work on instantaneous spectra.

13 Theory of instantaneous spectra

While studying the theoretical aspect of the demultiplication of frequencies, which consists of associating with a signal $s(t)$ whose spectrum occupies a certain band of frequencies a signal of the same duration whose spectrum occupies a narrower band, I had to consider the notion of instantaneous frequency. Inspired by GABOR's work in communication theory, I associated with $s(t)$ the signal in quadrature $\sigma(t)$ (HILBERT transform), which amounts to considering $s(t)$ as the real part of a function $\Psi(t)$ of a complex variable t, holomorphic in the upper half-plane. The instantaneous frequency is then the rotation velocity of the argument of $\Psi(t)$, so that the demultiplication consists of substituting the real part of $[\Psi(t)]^{1/n}$ for $s(t)$, where n is the demultiplication factor. Thanks to this theory of instantaneous frequency, I was also able to introduce the notion of the signal's instantaneous spectrum. For this I had to use the operator mentioned earlier. On this occasion I also gave the mathematical definition of the envelope of the signal, a notion particularly useful for theoretical calculations. The definitions I gave bring out clearly the nature of the two bands of modulation and the characteristics of the transmission on the unique lateral band [16].[15]

14 Signal theory

As a member of the U.R.S.I.,[16] I was asked to participate in the work of the C.C.I.R. (Comité Consultatif International des Transmissions Radioélectriques), in the section "Waves and Signals." My results on signaling theory were reproduced in the last chapter of my lectures at the Institut Henri Poincaré [6].

A channel's transmission capacity is known to be proportional to the width of the transmission band, and if the channel's equivalent depends on the frequency, it is tied to the curve of the amplitude as a function of the frequency, but this relation has never been spelled out rigorously. I asked the question in precise terms and showed that if the channel's capacity, measured as telegraphic capacity (number of signals discernible in a unit of time), was given by an integral involving a FOURIER transform, the question involved a property of the circulant determinants of a correlation matrix.

In the same line of studies, I served as a consultant at meetings of the C.C.I.F. (Comité consultatif international des téléphones), where I helped draft recommendations for the composition of quadripoles in transitory mode. I also had to study signals with finite duration $2T$ having the greatest fraction of their energy in a given frequency band. I gave an approximation for such a signal [17]. As for its rigorous determination, I was able to show that it satisfied the integral equation

$$s(t) = \lambda \int_{-T}^{+T} \frac{sin(t-u)}{t-u} s(u)du,$$

[15] *Translator's note:* For information on the influence of Ville's work in this area, see Patrick Flandrin's *Time-Frequency/Time-Scale Analysis*, Academic Press, 1999.

[16] *Translator's note:* Union Radio-Scientifique Internationale

and that if it was unique, it was self-conjugate.

15 Miscellaneous research

In the field of statistical estimation, I published in 1941 an article dealing with the following problem. Given a distribution $f(x;\theta)$ depending on an unknown parameter θ, estimation of θ comes down to choosing $f(x;\theta)$ from a family of permitted distributions. I studied the conditions under which the choice was invariant with respect to a change of variables affecting θ and x and showed that the invariant methods of estimation led to a metrization of the space of distributions [18].

In the field of the classical probability calculus, I studied, given the n variables $X_1, X_2, \dots, X_n$, the variable

$$Z = |\pm X_1 \pm X_2 \pm \cdots \pm X_n|$$

obtained by summing the variables, choosing the signs so that Z is minimized. Because the choice of signs depends on the values of the variables, and because Z is not an analytic function of the X_i, the problem is very difficult in its generality. In the case where the X_i are independent with a uniform distribution, I was able to show that Z is less dispersed than the smallest of the variables X_i [19].

I also published an article on the hydrolysis of polymers, which studied the distribution of the lengths of cellulose chains in solution [20].

Translator's note: The following list of publications was paginated separately from the preceding text. A more complete list of Ville's publications appears in this issue of the *Electronic Journal for History of Probability and Statistics*.

Work cited in the summary

(C & T stands for the journal Câbles et Transmissions)

1. Ueber Kurven mit quadratischen Lange. Ergebnisse einer Mathematischen Kolloquiums. Vienna 1935.

2. Ueber ein Satz von O. Blumenthal.[17] Ergebnisse einer Mathematischen Kolloquiums. Vienna 1935.

3. Sur la théorie générale des jeux où intervient l'habilité des joueurs. Note in fascicle II of volume IV of Mr. Emile Borel's Traité du Calcul des Probabilités. Paris (1938).

4. Etude critique de la notion de collectif (Thesis Paris 1939).

[17] *Translator's note:* The actual title was in French: "Sur une proposition de M. L. M. Blumenthal." The initials stand for Mr. Leonard M. Blumenthal.

5. Etude critique de la notion de collectif (Monographies des Probabilités, Paris 1939).

6. Sur quelques aspects récents de la théorie des Probabilités. Annales de l'I.H.P. Volume XIV, issue II, 1954.

7. In collaboration with Mr. Schützenberger, "Les problèmes de diagnostic séquentiel". C.R. Paris, 232, 206–207 (1951).[18]

8. Réseaux réactifs en échelle. Câbles et Transmissions – April 1948 – pp. 111–130.

9. La formation de la connaissance envisagée du point de vue probabiliste. Congrès Internationale de Philosophie des Sciences. Paris 1949.

10. Cours d'Analyse Matricielle (Publications de l'Institut de Statistique de l'Université de Paris).

11. Cours d'Algèbre de Boole (Publications de l'Institut de Statistique de l'Université de Paris).

12. Sur les conditions d'existence d'une ophélimité totale (Annales de l'Université de Lyon. 1946).

13. The existence condition of a total utility function (Translation of the preceding) The Review of Economic Studies Vol. XIX no. 49.

14. Signaux analytiques à spectre borné. First part, C & T January 1950. Second part, C & T January 1953.

15. Sur l'opérateur $\exp\{x + \frac{d}{dx}\}$. Note in the C.R. of the Academy of Sciences, séance of 5 November 1945.

16. Théorie et applications de la notion de signal analytique. C & T January 1948.

17. In collaboration with J. BOUZITAT. Sur un type de signaux pratiquement bornés en temps et en fréquence. (Bull. Tech. SOTELEC January 1955).

18. Sur la théorie invariante de l'estimation statistique. (Bull. Sc. Math. 1944).

19. Variables aléatoires équiréparties. C & T July 1949.

20. Essai d'une théorie de l'hydrolyse des polymères. Mémorial des Services chimiques de l'Etat 1944.

21. Etude statistique des irrégularités d'impédance des câbles coaxiaux. Bull. Soc. F. Elect. November 1944 and December 1944.

22. Limitation d'un traînage par limitation de l'écho. C & T July 1950.

[18] *Translator's note:* The initials "C.R." stand for the *Comptes Rendus* of the Academy of Sciences.

Other work

- Théorie des systèmes presque linéaires. R.G.E.[19] February 1945.
- In collaboration with P. Herreng: Sur la stabilité des réseaux linéaires. R.G.E. March 1945.
- Sur l'argument des fonctions positives. Bull. Tech. SOTELEC January 1947.
- In collaboration with R. Bélus and P. Herreng. Ecarts d'impédance admissibles sur les sections d'amplification de câble coaxial. C & T April 1947.
- Composition des distorsions en régime transitoire sur les différentes longueurs d'un câble coaxial (Onde électrique December 1948).
- In collaboration with P. Herreng. Etude des irrégularités d'impédance des câbles coaxiaux par observations oscillographiques des échos d'impulsions. C & T April 1948.
- Théorie du tétrapole. Bull. Tech. Sotelec July 1948.
- Sur une équation fonctionelle de la théorie de l'information (C & T January 1951).
- Interprétation des faibles valeurs de l'entropie d'un systéme. Bull. Tech. Sotelec January 1951.
- Décomposition d'un signal en composantes en cosinusoîde surélevé. C & T April 1951.
- Entropie d'un système de transmission. C & T July 1951.
- Sur les origines de la théorie des réseaux linéaires. R.G.E. January 1952.
- Calcul des quaternions et compositions de quadripôles en chaîne. C & T July 1952.
- In collaboration with R. Roch. Note sur les paramètres de transmission d'une section d'amplification. C & T April 1953.
- Expression analytique de l'exposant de transfert d'un circuit appelé à transmettre des impulsions. Onde électrique February 1954.
- Modulation conjuguée d'une démodulation linéaire. Onde électrique April 1954.
- Sur le prolongement des signaux à spectre borné (completing reference 14). Bull. Tech. Sotelec January 1955.

[19] *Translator's note:* Revue générale de l'Electricité.