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

Doob at Lyon

On his lecture, *Application of the Theory of Martingales*, at the Lyon Colloquium, June 28 – July 3, 1948

Bernard LOCKER[1]

Translated from the French by Ronald Sverdlove[2]

The colloquium

From June 28 to July 3, 1948, an international colloquium on probability theory and its applications was held at Lyon by the CNRS with financial support from the Rockefeller Foundation.[3] In 1946, in the framework of its Post War Program for France, the Foundation had decided to give two grants to the CNRS, the first for $250,000 for the purchase of materials, the second for $200,000 for organizing "Special Conferences", about ten per year for three years, that would include foreign scholars as participants. According to Warren Weaver,[4] these conferences were to be small and informal, "the attendance of mature contributors restricted to say 15". They were to cover all fields of scientific research, and one third of them were to take place in the provinces. Zallen [1989] tells the history of these grants, with interesting details on the roles of Weaver and Louis Rapkine, as well as Frédéric Joliot,

[1]UFR de Mathématiques et Informatique, Université Paris Descartes, Laboratoire MAP5 UMR CNRS 8145, 45 rue des Saints Pères 75006 Paris, Bernard.Locker@math-info.univ-paris5.fr.

[2]School of Management, New Jersey Institute of Technology, 3000 Central Avenue Building, University Heights, Newark, NJ 07102-1982, USA.

[3]The CNRS (Centre National de la Recherche Scientifique) is a research agency of the French government.

[4]Zallen [1989], page 9.

Georges Teissier and Pierre Auger.[5] Dosso [1998] particularly emphasizes Rapkine's role. In the end, thirty-six colloquia were held between 1947 and 1952.[6] For mathematics, the colloquia began in 1947 with harmonic analysis at Nancy (June 15–22) and algebraic topology at Paris (June 26–July 2). Probability theory and its applications was among the subjects chosen for 1948.

Maurice Fréchet (1878–1973, academician in 1956) was in charge of organizing the colloquium on probability at Lyon and editing the proceedings, which appeared in 1949. In his preface [Fréchet 1949a], he began by justifying the choice of Lyon:

> Lyon was chosen for the site of the colloquium dedicated to this subject because the University of Lyon has constantly sustained the initiatives of Professor Eyraud in this area. It liberally welcomes articles on the 'Theory of Probability' in the mathematical section of the 'Works of the University of Lyon'; it has created and continues to sustain a very useful Institute of Financial Sciences and Insurance (ISFA); finally it has created a certificate of higher economic studies, thus taking the lead of the movement that, despite resistance, moves political economy toward the use of mathematics.

Was this simply politeness toward Henri Eyraud (1892–1994), who had defended his thesis in 1926 and had been professor at the University of Lyon since 1930? No. According to Armatte [2006], ISFA owed its success to Eyraud's course, and to the support of Fréchet and Emile Borel since its creation in 1930.

There was also there a nod to Georges Darmois (1888–1960), who had been director of studies at ISUP[7] since 1945 and academician since 1955. Darmois was ten years younger than Fréchet but was already an old campaigner for statistics and its instruction, on which Borel and Fréchet's activism is described in Meusnier [2006]. We add that Fréchet and Darmois had participated in the meeting of the Commission on Econometrics on December 16, 1946, which created two seminars in econometrics, one in the provinces. Lyon was chosen over three other provincial cities, Rennes, Strasbourg, and Lille [Bungener and Joël 1989]. The choice of Lyon fit perfectly into the constant action of Fréchet, sustained by his vision of the relations among probability, statistics, social sciences, research and education since his course at Strasbourg in 1919 [Armatte 2001, Locker 2001, Barbut, Locker, and Mazliak 2004, Havlova, Mazliak, and Sisma 2005, Siegmund-Schultze 2005].

[5] In an appendix, we provide biographical information about these individuals and some others mentioned in the course of this article.

[6] Archives of the CNRS: International Colloquia Supported by the Rockefeller Foundation or the CNRS 1946–1967 [ART 141–173]. The colloquia are also listed in Zallen [1989].

[7] Institute of Statistics of the University of Paris.

Including Fréchet, Darmois, and Eyraud, sixteen participants produced papers for the proceedings.[8] As Fréchet explained,

> In addition to the sixteen scholars who were going to present communications…, more than thirty people (from all parts of France and the world, most of whom were specialists in the questions treated) asked to listen to the lectures and participated very helpfully in their discussion.[9]

Fréchet thus respected perfectly Weaver's wishes to keep the colloquia small; this was not always the case in the other colloquia, particularly in physics.

Here we list the titles, translated into English when necessary, of the sixteen contributors. Five, including the American Joseph Leo Doob, were foreigners:

1. G. Ottaviani (Italy): *The Uniform Law of Large Numbers in the Classic Theory of Probability.*

2. J. L. Doob (USA): *Application of the Theory of Martingales.*

3. D. van Dantzig (Holland): *On the Method of Generating Functions.*

4. H. Wold (Sweden): *On Stationary Point Processes.*

5. J. Wishart (UK): *Test of Homogeneity of Regression Coefficients.*

Ten were French:

1. G. Darmois: *On Certain Forms of Relations of Probabilities.*

2. M. Fréchet: *The Typical Values of Order Zero or Infinity of a Random Number and Their Generalization.*

3. P. Lévy: *Double Markov Processes.*

4. A. Blanc Lapierre: *Considerations on the analysis of random functions.*

5. J. Kampé de Fériet: *Stationary Random Functions and Transformation Groups in an Abstract Space.*

6. E. Halphen: *On the Problem of Estimation.*

7. P. Delaporte: *On the Use of Systematics of Mathematical Statistics in Factorial Analysis.*

8. R. Fortet: *Probability of Loss of a Telephone Call.*

[8]Fréchet prepared a communication [Fréchet 1949b] to present in case of the absence of one of the lecturers.

[9]C. R. Rao, who did not present a lecture, participated in all the meetings and intervened several times, once following the exposition of Doob.

9. J. Ville: *Random Functions and Transmission of Information.*
10. G. Malécot: *Stochastic Processes and Genetics.*
11. H. Eyraud: *Pure Economy. Credit and Speculation.*

For the history of martingales, one remembers that Paul Lévy, Jean Ville, and Joseph Leo Doob were among the contributors. Martingales, of which Lévy and Ville were the pioneers in prewar France, thus returned from America after the war in the form of a theory elaborated by Doob and moved into the framework for probability and stochastic processes, inspired by Kolmogorov's *Grundbegriffe* of 1933, that Doob had persistently developed and applied since 1935. Under the title "*Application of the Theory of Martingales*", Doob would show how this "Theory of Martingales" could be applied on one hand to the (strong) law of large numbers and on the other hand to statistical estimation.

Yet the history of martingales must also take into account that Lévy did not recognize has own prewar techniques in Doob's martingales at Lyon, and that Ville did not meet Doob there. Although Fréchet's preface leaves the impression that all the contributors to the proceedings had been present at Lyon, Ville made it clear in a letter to Pierre Crépel in 1985 [Crépel 2009, p. 15] that he was not present at Lyon and never met Doob. So Lyon symbolizes a missed opportunity. Of the three pioneers of martingales supposedly there, two were left at the side of the road, while the third made of martingales one of the masterpieces of his work.

Paul Lévy

If Lévy did not isolate the concept of a martingale, it was his work on sums of dependent variables[10] that links him to the history of martingales before Ville and Doob. On the place of Lévy in the history of martingales, we refer to two articles in this issue of JEHPS: Mazliak's very complete article [2009] and the marvelously argued exposé by Bru and Eid [2009], concerning the correspondence between Lévy and Jessen on the rarely discussed subject of the pair "Lévy's Lemma" and "Jessen's Theorem".[11] Here we recall only that from 1935 to 1936, Lévy extended the strong law of large numbers, the three-series theorem of Kolmogorov and certain results on convergence to

[10] In the French terminology of the period, used by Lévy, one spoke not of dependent variables but of variables in chains: *variables enchaînées*. This term was used whenever the hypothesis of independence was weakened or removed, both when relations between each new variable X_n and one or more preceding variables were given and when only transition probabilities were given, as for Markov chains. But here we translate *variables enchaînées* as "dependent variables".

[11] In modern terms, for an integrable random variable X and a filtration F_n, the martingale $\mathrm{E}(X|F_n)$ converges to X almost surely in L_1. Lévy proved a version of this result where X is the indicator of a set; this is often called "Levy's Lemma" or "Lévy's zero-one law".

the normal distribution beyond the case of independent variables to dependent variables "*satisfying condition C*" and, when needed, for part of Kolmogorov's three series theorem for example, certain second-order hypotheses. Condition C, in the notation used in Lévy [1936a, 1937] says that

$$\mathcal{M}_{\nu-1}\{X_\nu\} = 0, \tag{1}$$

where $\mathcal{M}_{\nu-1}$ is "*the probable value of* X_ν*, given the values of* $X_0, \dots, X_{\nu-1}$", that is, the conditional expectation[12] relative to the variables with indices $1, 2, \dots, \nu-1$. In 1935, 1936, and 1937, Lévy made use of the sum of dependent variables X_ν subject to his condition C

$$S_n = \sum_1^n X_\nu\,. \tag{2}$$

The sum S_n is clearly a martingale (with respect to the filtration associated with X_ν, which is also the filtration associated with S_ν). In Lévy's notation, explained above:

$$\mathcal{M}_{n-1}\{S_n\} = S_{n-1}\,. \tag{3}$$

It is interesting to add that in Lévy [1937], the conditional probabilities permitting the definition of conditional expectations of the type in formulas (1) and (3) are understood without reference to Radon's theorem [Radon 1913] or Nikodym's theorem [Nikodym 1930], even though the latter was published in French (in the most general form that is now classical[13]), nor even any allusion to the presentation "à la Radon-Nikodym" of Kolmogorov [1933].[14] In Lévy [1937], one is far from such a presentation. Of course, we know that Lévy read little and preferred to rediscover for himself results he heard discussed,[15] as the need arose and using his own methods, but it is appropriate to add that with Lévy, when the probabilist took precedence over the analyst, the results of measure theory were arranged in a strange catalogue, dedicated only to his needs in probability, translated once and for all into his language of probability.

Although Lévy had taken care Lévy [1936b][16] to spell out his "*notion of conditional probability*", affirming that this was the viewpoint he would "*always take in his research on dependent variables*",[17] he limited himself to

[12]Following an old practice, Lévy in 1937 still sometimes used *probable value* in place of *expectation*. These two terms appear together on several pages of Lévy [1937].

[13]Pages 166 and 168 of Nikodym [1930], which repeats the communication presented by Otton Nikodym in September 1929 to the *First Congress of Mathematicians of the Slavic Countries*.

[14]On Radon [1913] and Nikodym [1930], see Bourbaki [1969] and Michel [1992].

[15]On this subject, see Barbut, Locker, and Mazliak [2004] and especially Lévy [1970], in which Lévy describes his psychology and his method of working.

[16]Lévy [1936a] also appears mysteriously hidden in the last volume (volume VI) of Lévy's complete works.

[17]As usual with Lévy, the argument is dressed up with an appeal to an authority, preferably one of the great Russian probabilists [Locker 2001]; in this case Bernstein is invoked.

showing, for an event A of probability α, that if its probability becomes $\lambda(x)$ when one knows that $X = x$, then $\alpha = \mathcal{E}(\lambda(x))$[18] equally well whether A, X is determined by a "*unique trial*" or by two "*successive trials*",[19] a result later generalized to the case of n variables. The requirements of measurability for conditional probability are specified and the whole is presented in terms of Lebesgue's integral and measure. In Lévy [1937], it is the Lebesgue-Stieltjes integral that comes to the forefront. Given an event B, perhaps of the form $\{Y < y\}$, and a random variable X with distribution function F, the "*conditional probability of B given the hypothesis* $X = x$" must then be a function $g(x)$ admitting a Lebesgue-Stieltjes integral with respect to F satisfying

$$\int_{-\infty}^{x-0} g(x)dF(x) = \Pr\{B \text{ and } X < x\}. \tag{4}$$

The existence and the properties of g are obtained, up to a set of F-measure zero, by a method of differentiation of the function obtained from $\Pr\{B$ and $X < x\}$ by the change of variables $\xi = F(x)$,[20] probably inspired (because Lévy preferred distribution functions) by applications that he had made to probability theory of the theorem on the decomposition of functions of bounded variation and of the Lebesgue-Stieltjes integral, adapting Lebesgue's theorem (on the almost everywhere differentiability of nondecreasing functions).[21]

So when the joint distribution of Y and X is known, formula (4) is valid.[22] Starting from the conditional probability $\Pr\{Y < y|X < x\}$, he designates

[18]In Lévy [1936b], $\mathcal{E}$ denotes "*the probable value*", i.e., the expectation (see note 12).

[19]That is to say depending on whether one gives a priori the probability $\mathrm{P}[A \text{ and } X < x]$ or the probability for x followed by the probability $\lambda(x) = \mathrm{P}[A|X = x]$, a dichotomy very dear to Lévy and which influences all his considerations in Lévy [1937] having to do with conditional probabilities and/or dependent variables. This dichotomy is to be put in parallel with his vision of stochastic processes in which chance intervenes at each instant t after having constructed the points of the trajectory before t. (See Barbut, Locker, and Mazliak [2004], pages 53 and 92.)

[20]No. 23 of Lévy [1937]:

> *Set* $Pr\{B \text{ and } X < x\} = F_1(x)$. *The probability of B under the hypothesis that* $a < X < b$ *is* $(F_1(b) - F_1(a))/(F(b) - F(a))$, *and since it is always between 0 and 1,* $F_1(x)$ *can be considered a function of the variable* $\xi = F(x)$, *with a derivative that is well defined except on a set of measure zero and between 0 and 1; this derivative is then a function* $g(x)$ *of* x, *well defined except on a set of probability zero (that is to say, to which the distribution of* X *assigns probability zero.*

[21]This theorem, amply discussed in No. 12 of Lévy [1937], concerns the additive decomposition of the distribution function F into three parts, the sum F_1 of jumps + the absolutely continuous part F_2 + the difference $(F - F_1 - F_2)$.

[22]Lévy then discusses the case of the determination by "*two successive trials*" (see note 19), giving the measurability conditions required to reconstitute from a function $g(x)$ a "*well determined probability*" for a pair Y, X of random variables such that (4) is valid, thus making g a conditional probability.

by $\mathcal{M}_x$ "*a probable value calculated when* X *is known*" and adds "*it is therefore a function of* X". The argument as a whole shows the measurability of $\mathcal{M}_x(Y)$. The conditional probabilities $\mathrm{P}[X_\nu|X_0,\ldots,X_{\nu-1}]$ and the corresponding conditional expectations $\mathcal{M}_{\nu-1}[X_\nu|X_0,\ldots,X_{\nu-1}]$ are obtained by iteration from the preceding ones (Lévy [1937], no. 64) but also by recourse to "denumerable probabilities" and to the representation of all the random variables considered as measurable functions of a single variable. All the properties of conditional probabilities and expectations that Lévy would use in Lévy [1937] flow from his "*notion of conditional probability*".[23] He long remained faithful to it,[24] never bothering in the rest of his work (even for the difficult questions he attacked, for example in the theory of Markov processes in continuous time) to check his bearings, as if it were obvious that the properties he had derived for his "notion of conditional probability" were universally and eternally valid.

Jean Ville

It is in his thesis, *Etude critique de la notion de collectif* [Ville 1939][25] that Ville introduced the term "*martingale*". In the debates on the foundations of the theory of probability, Ville's martingales, opposed by the "frequentist" theory of collectives of von Mises [1931], which was presented at the Institute Henri Poincaré in 1931, were to play a significant role.[26] We point out the

[23]Later, Lévy would remain equally ignorant of the counterexamples given by Jean Dieudonné [1948], although Fréchet was made aware of them by a letter of November 1951 from Robert Fortet, recounting that Dieudonné had showed that conditional probabilities do not always exist. See Locker [2001].

[24]Lévy returned to conditional probabilities in a letter of January 9, 1962, number 90 in Barbut, Locker, and Mazliak [2004], to affirm the agreement of his "notion" of 1936–37 with Doob and his differences with de Finetti.

[25]In English: Critical study of the concept of a collective. He had earlier mentioned martingales in a note reporting some of his results, in the *Comptes rendus* in 1936 (volume 203, pp. 26–27).

[26]It is interesting to note that Lévy returned twice to the question of von Mises' collectives and Ville's thesis in his correspondence with Fréchet, partially contradicting himself at an interval of 25 years. On October 22, 1939, he writes:

> *I arrived myself at the idea that, without sharing the ideas of Mises, it was not necessary to absolutely proscribe models and collectives: they constitute a useful language for study from an axiomatic point of view that is different from the one I have used until now.*

On April 28, 1964:

> *I never understood well the first definition of collective given by Ville. Loève and Khintchine told me and wrote to me that they did not understand it. . . . As to the theory of collectives, I have always found it absurd, in spite of the merits that I recognize in von Mises, and I did not hide this from Wald when he presented it at Grenoble. I am grateful to Ville for having helped me to combat it.*

Barbut, Locker, and Mazliak [2004].

very useful commentaries of Nualart [1987] and Von Plato [1994][27] on Ville's thesis. Bru, Bru, and Chung [1999] contains an analysis of Ville's thesis (pp. 203–207) and a note on Ville and the Lyon Colloquium (pp. 232–233). On all that concerns Ville and Ville [1939], the reader should refer first to Shafer [2009] in this issue of JEHPS. Here we recall, with Bru, Bru, and Chung [1999] that Ville

> *begins by defining the general notion of a (positive) martingale adapted to an arbitrary sequence* (X_n) *of random variables using the now classical martingale property (e.g. [Neveu 1972]), the conditional expectation being 'defined' 'in the sense indicated by Mr. Paul Lévy'.*

One notes also that these three authors, in their well-argued paper find in Ville's text a "*type of reasoning 'by stopping' . . . used by all sound authors*" when he proves a martingale inequality (the inequality of the gambler's ruin) in discrete time, which we will see again below, because it did not escape Doob, who generalized it in Doob [1940a].

We add here several details on the framework in which Ville defined his conditional expectations. First of all, Ville [1939], in section 1 of his chapter V (which concerns martingales and the gambler's ruin in discrete time) relies exclusively on the definition of Lévy [1937 pages 96–99], to the point of reproducing identically the "proof" of the existence of the conditional probability of Lévy [1937] that we described above. Later, in section 2 of Chapter V (*Continuous Game*) (pages 111–129), Ville discusses the generalization of the inequality of the gambler's ruin for (positive) martingales in continuous time, and for this he works in the space that he calls E_0 of all real functions of a real variable:

> *a distribution function will then be a completely additive function* ≥ 0 *defined on certain subsets of* E_0 *[Kolmogoroff [1]]. Letting* $P(L)$ *be this function, we naturally assume that* $P(E_0) = 1$.

More precisely (page 112):

> *The condition (d)*[28] *brings in the notion of conditional mean that we attach to that of conditional probability as it was defined by Mr. Doob [Doob [1] p. 123]; this notion is the generalization to a function space of the notion of conditional probability (due to Mr. P. Lévy) that we already used on page 87. . .*

[27]In his chapter 6, "*Von Mises' Frequentist Probabilities*".

[28]Ville defined a positive martingale in continuous time by introducing a family X_t ($t \geq 0$) of points of E_0 and a family S_τ (τ real and positive) of "*positive functionals*" defined on E_0 satisfying four conditions, (a), (b), (c), and (d). Conditions (a) and (b) are related to the initial values and positivity. Condition (c) "S_τ *depends only on the* X_t *for* $t \leq \tau$" and condition (d) "*the mean value of* $S_{\tau+\tau'}(X)$ *when one knows that* $X(t) = X_0(t)$ *for* $t \leq \tau$ *is equal to* $S_\tau(X_0)$". In today's less exotic terms, one recognizes in conditions (c) and (d) respectively the notion of adapted process and the condition on conditional expectation defining a martingale in continuous time.

He adds that the existence of the conditional probability that he defines using Fréchet's integral "*results from a proof of Mr. Nikodym (p. 168–179)*". So his references are now Fréchet [1915], Kolmogorov [1933], and finally Doob [1938]. This presentation of martingales in continuous time did not escape Doob in his critique [Doob 1939] of Ville [1939].

Joseph Doob

It is at the top of the first page of Doob [1940a] that Doob announces that he will study "*certain families of chance variables*"[29] x_t having the property he denotes $\mathcal{E}$ – i.e., verifying for all $t_1 < \cdots < t_{n+1}$, and with probability 1, the relation (written here in his own notation[30])

$$\mathrm{E}[x_{t1}, \ldots, x_{tn}; x_{tn+1}] = x_{tn}. \tag{5}$$

This equation is followed immediately by a footnote:

> *We shall use the notation* $\mathrm{E}(y)$ *for the expectation of the chance variable* y *and* $\mathrm{E}([y_1, \ldots, y_n; y]$ *for the conditional expectation of* y *for given* $y_1, \ldots, y_n$*, a function of* $y_1, \ldots, y_n$*. If the* y_i *are not finite in number, the notation will be modified accordingly. We shall assume the definitions of Kolmogoroff for these conditional expectations.*[31]

Still on this first page, Doob makes precise the sense he wants to give to his random variables and to the probability P that he will use throughout:

> *In the following, we shall always suppose that the* x_t *are measurable functions defined on a space* Ω*, on certain sets of which a measure function is defined. That this can always be done, and how this is to be done, was shown by Kolmogoroff. The space* Ω*, following Kolmogoroff, will be taken to be the space of real-valued functions of* t*. ... The qualification "with probability 1" will be used interchangeably with "almost everywhere on* Ω*."*

The conditioning of relation (4) and property $\mathcal{E}$ are immediately generalized on the next page to an infinite set of indices T (which can be N, $-N$, a section of Z or of R) by considering the "*Borel field*"[32] of "x_t*-sets*" generated by all

[29] In 1940, Doob still used the term *chance variable* that he later replaced by *random variable*.

[30] The expectation of X given A is denoted $\mathrm{E}[A; X]$ in Doob [1940a] and $\mathrm{E}\{A \backslash X\}$ in Doob [1949]. This may confuse today's readers, because it is now standard to put X first, as in $\mathrm{E}[X|A]$.

[31] As in Doob [1937] and Doob [1938], this is a reference to *Grundbegriffe der Wahrscheinlichskeitrechnung* of 1933 for conditional expectations constructed using Nikodym's theorem.

[32] "*Borel field*", translation of the French "*corps de Borel*", which is the established expression designating the Borelian ancestor of "σ-algebras". In the first edition of Lévy [1937], Lévy manipulated and constructed "*Borel fields*" in "abstract sets" (terminology borrowed from Fréchet). See Lévy [1937] in the edition of 1954 on page 17 for the distinction that he made later between "*Borel fields*" and "*closed Borel fields*".

finite sets of the variables considered. The "chance variables with property $\mathcal{E}$" of Doob [1940a] will be called "martingales" in Doob [1949].

In all the preceding cases, Doob's martingale, in modern terms, is adapted to the family of tribes $(F_t, t \in T)$ generated by the random variables with indices smaller than t, and this family is increasing in t. When the set of indices is a subset of Z bounded above, an obvious translation of index leads to the set $-N$ of integers ≤ 0, and the results obtained by Doob (when the index tends to $-\infty$) are concerned rather with what are now widely called "backward martingales".[33]

In this article, Doob [1940a], which consists of three sections, Doob will therefore, evidently and permanently, rely explicitly on measure theory, which he had not ceased to do since his first article on probability, Doob [1934], in the manner of Doob [1935], Doob [1936b], Doob [1937], Doob [1938], and also Doob and Ambrose [1940], which slightly preceded Doob [1940a],[34] all of which together form the framework that will permit him, after having established in the first section the theorems that are classical today on discrete time martingales (relative to uniform integrability, convergence and closure), to put in place in the last two sections, the tools and the first results for martingales in continuous time. For our focus on the state of affairs before the Lyon colloquium, we will remain with the first section, in which Ville and Lévy are cited or commented on four times.

We note first, as mentioned above, that Doob [1940a] in a footnote on page 458 cites Lévy [1937] and Ville [1939] when he is proving the inequalities

$$\int_{\Lambda \cdot N} x dP \geq k\mathrm{P}\{\Lambda \cdot N\} \quad \text{and} \quad \int_{M \cdot N} x dP \leq k\mathrm{P}\{M \cdot N\} \tag{6}$$

established for a family "$\ldots, x_{-1}, x_0, \ldots, x$" satisfying $\mathcal{E}$, of which the set of indices, if it is infinite, is a subset of Z either bounded above or bounded below, N is a P-measurable set (where P is the probability measure fixed at the beginning of the article on the space of all functions of a real variable), and the products $\Lambda \cdot N$ and $M \cdot N$ designate the intersections $\Lambda \cap N$ and $M \cap N$[35] of the sets $\Lambda = \left\{\sup_{m \leq j \leq n} x_j \geq k\right\}$ and $M = \{\inf_{m \leq j \leq n} x_j \geq k\}$.

> *These inequalities are implicit in the work of Ville [9, pp. 100–101]*[36] *who discussed sequences of non-negative chance variables*

[33]Recall that a backward martingale indexed by the positive integers is basically a sequence $(Y_n; n \geq 0)$ of random variables adapted to a decreasing sequence $(G_n; n \geq 0)$ of tribes satisfying $\mathrm{E}[Y_n|G_{n+1}] = Y_{n+1}$. With the replacement of each n by $-n$, supposing that $X_{-n} = Y_n$ and $F_{-n} = G_n$, this becomes $\mathrm{E}[X_{-n}|F_{-n-1}] = X_{-n-1}$ and the sequence of tribes $(F_{-n}; -n \leq 0)$ is increasing $(F_{-n-1} \subset F_{-n})$, giving a martingale that is "ordinary but indexed by the negative integers".

[34]Doob and Ambrose [1940] was received by the *Annals of Mathematics* on September 25, 1939. Doob [1940] was received by the *Transactions of the American Mathematical Society* on December 11, 1939.

[35]Classical notations in set theory at the time.

[36]i.e., Ville [1939]

with the property $\mathcal{E}$. The method of proof we use was used by Lévy [8, pp. 129][37] in a related discussion.

The inequalities (6) are important since one can deduce from them, with Doob and after some gymnastics, a theorem numbered 1.2 (page 458) for sequences indexed by the set $-N$ of negative integers satisfying $\mathcal{E}$ (which is in fact a "backward martingale" theorem according to a remark already made):

> *Let $\ldots, x_{-1}, x_0$ be a sequence of chance variables with the property $\mathcal{E}$. Then $\lim_{n\to\infty} x_n = x$ exists with probability 1, and the chance variables $x, \ldots, x_{-1}, x_0$ have the property $\mathcal{E}$. The chance variables are uniformly integrable, and $\mathrm{E}[x_0] \geq \mathrm{E}[x_{-1}] \geq \cdots \geq \mathrm{E}[x]$; $\mathrm{E}[x_n] \to \mathrm{E}[x]$.*

Ville had originally used a form analogous to the first of the inequalities (6) in the case of a martingale indexed by the integers and with $k = 1$ (inequality of the gambler's ruin!). He had then extended it to the continuous case as early as 1938 in a note in the CRAS. See Bru, Bru, and Chung [1999]. Lévy [1937] also uses this first of the inequalities (6) for his martingale (3), aiming especially to extend Kolmogorov's inequality[38] from the case of independent variables to his martingale differences, which satisfy his condition C of formula (1).

We note next Theorem 1.3 on page 460 (convergence a.e. of integrable martingales and closure in the equi-integrable case that today is called "regular"):

> *Let $x_1, x_2, \ldots$ be a sequence of chance variables with the property $\mathcal{E}$. Then $\mathrm{E}|x_1| \leq E|x_2| \leq \cdots$. If $\lim_{n\to\infty} \mathrm{E}|x_n| = l < \infty$, then $\lim_{n\to\infty} x_n = x$ exists, with probability 1, and $\mathrm{E}|x| \leq l$. If the x_j are uniformly integrable, $\lim_{n\to\infty} x_n$ exists, with probability 1, and the chance variables $x_1, x_2, \ldots, x$ have the property $\mathcal{E}$.*

Doob makes this comment:

> *Ville has studied sequences of non-negative chance variables with the property $\mathcal{E}$. Since, by the corollary to Theorem 0.2,[39] Ville's hypotheses imply that*
>
> $$Ex_1 = Ex_2 = \cdots = E|x_1| = E|x_2| = \cdots ,$$
>
> *the hypotheses of the first part of Theorem 1.3 are satisfied, in Ville's case. Ville proved that in his case $\sup_{j\geq 1} |x_j| < \infty$, with*

[37] i.e., Lévy [1937]. The page reference is to the first edition of Lévy [1937]. The pagination of the second edition (1954) differs here from that of the first edition.

[38] In the notation of Lévy, under condition C for the X_ν with $T_n = \max_{\nu\leq n} |S_\nu|$, $c > 0$ and $b_n^2 = \mathcal{M}\left\{S_n^2\right\}$, one has $\mathrm{Prob}\left\{T_n > cb\right\} < 1/c^2$.

[39] This corollary of "theorem 0.2" is at the beginning of the article. In current terms: if X_t is a martingale, $\mathrm{E}(X_t)$ is constant and $\mathrm{E}(|X_t|)$ is an increasing function of t.

probability 1 (implied by our conclusion that $\lim_{n\to\infty} x_n$ *exists with probability 1, and that the limit is integrable) and applied this fact to the study of certain games of chance.*

Here Doob goes too far. In fact the almost sure convergence of "Ville's martingales" does not appear in Ville [1939]. Moreover, according to Bru, Bru, and Chung [1999], neither Ville nor Borel nor Lévy seem to have noticed this almost sure convergence (a result therefore due entirely to Doob). We also note that in this first part (page 462), Doob cites Lévy two more times for results on almost sure convergence from Lévy [1937], where Lévy gives his version of the 0-1 theorem and the "martingale version" of Kolmogorov's three series theorem.

At the colloquium

The hypotheses advanced by Bru, Bru, and Chung [1999] to explain Ville's retreat from the martingale scene and his likely absence from the Lyon colloquium are clear. Ville may have lost interest in martingales after reading Doob [1940a] which seemed to close the question. After his return from captivity in June 1941, he considered the application of martingales to the geometry of vector Brownian motion but soon found that Lévy had left him behind on this topic.[40] In 1946, after the faculty at Lyon chose Gustave Malécot over him for a chair, Ville had left academics to work in industry. Remaining on very cool terms with the academics at Lyon, he concentrated on the transmission of information in telecommunications.[41] To the hypotheses advanced by these three authors, one can add another. In November 1939, Doob had written a review [Doob 1939] of Ville [1939],[42] of which the whole second half was very negative:[43]

> *It is unfortunate that this book, which contains much material which clarifies the subject, should contain so much careless writing. This ranges from uniformly incorrect page references to mathematical errors. Thus (p. 46) it is claimed (and used in*

[40] Fréchet wrote to Lévy about this. See in Barbut, Locker, and Mazliak [2004] the letter from Lévy to Fréchet dated June 1, 1943.

[41] Note that Ville's presentation at Lyon had the title *Random Functions and the Transmission of Information*, and that Ville [1948] treated analytic signals. Finally, we add to the list of Ville's works after his withdrawal from martingales and before the colloquium an article [Ville 1946] treating the existence conditions for a total utility and an index of prices; this article is discussed by Gardes and Garrouste [2004].

[42] In December 1939 there appeared another review in English of Ville [1939] by Henry Thomas Herbert Piaggio (English, 1884-1967, assistant, then professor of mathematics at the University of Nottingham from 1908 to 1950). The word "martingale" never appears in Piaggio [1939].

[43] In his celebrated conversation with J. Laurie Snell [1967], Doob speaks of this review and of the direction in which reading Ville pointed him, but he makes no allusion to this negative second part.

a proof) that every denumerable set is a G_δ.[44] *The author's main theorem on systems is not as strong as earlier results with which he is apparently unfamiliar. (Cf. Z. W. Birnbaum, J. Schreier, Studia Mathematica, vol. 4 (1933), pp. 85–89; J. L. Doob, Annals of Mathematics, (2), vol. 37 (1936), pp. 363–367.) His discussion of random functions is inadequate and obscure, for example, his demonstration that his main theorem on martingales does not go over to the continuous process uses as an example a measure on function space not in accordance with the usual definition of probability measures on this space.*[45] *A specialist who can overlook such slips will find many stimulating ideas in this book. Other readers can profit by the comparative analysis of the different criteria for collectives, and by the discussion of martingales.*

Ville, after that, would have no more wanted to meet Doob than the academics of Lyon.

Lévy and Doob certainly did meet at Lyon, but between them it was not at all a question of martingales. The proceedings of the colloquium show that Doob, following Lévy's presentation on double Markov processes [Lévy 1949, where linear time is replaced by curves] agreed with Van Dantzig[46] that "*Mr. Lévy has not made the definition of double Markov processes sufficiently precise*", which provoked a response from Lévy on his analytic hypotheses and the conditional independence of two parts of the plane relative to a curve. As for Lévy, he did not ask Doob any questions about his presentation,[47] in

[44] This error by Ville does occur on the page indicated and, moreover, at the beginning of an important proof that generalizes a result of the American A. H. Copeland (well known for his work on the question of collectives). This is Theorem 1 of Chapter 2 of Ville [1939]:

> *If L is the system of $G_\delta s$, there exists no collective belonging to L that does not have the following property: there exist among the sets of L a countable infinity of sets, each reduced to a point, such that the sum of their probabilities is equal to 1.*

[45] On martingales in continuous time, Doob [1940a, page 476] simply points out in a footnote on the subject of Ville, "*His discussion of the meaning of a continuous process and the generalized upper bounds is somewhat obscure*" before himself extending the inequalities (6) to continuous time as Ville had done, but for positive martingales. The "*main theorem on martingales*" of Ville's to which Doob alluded in 1939 is without any doubt his inequality of the gambler's ruin and the extension of it that he made to continuous time, after having found a counterexample in order to better show hypotheses needed to make this extension valid. In this counterexample, Ville assumes a probability on a set of curves which is no longer the probability P (that he had presented a "à la Kolmogorov") with which he had furnished the space E_0 of all functions of real variables.

[46] The contribution of Lévy was on June 28. He had to face several other questions requiring an elaborate response. He was only able to write responses to Kampé de Fériet's and Fréchet's questions of June 29 and 30.

[47] Only Rao asked Doob a question, relating to a problem of nonparametric statistics in the framework of martingales.

which, we may note, Lévy is not cited.[48]

Making this all the more extraordinary, part 2 of Doob's contribution touched on the application of martingale techniques to the strong law of large numbers, but only in the case of independent variables, and it is clear that neither Lévy nor Fréchet (who knew Lévy [1937] perfectly well[49]) made the connection between Lévy [1937] and Doob [1949]. Moreover, it is clear that Lévy did not read Doob [1940a],[50] where he is cited several times.

According to a letter he wrote to Fréchet on April 28, 1964, Lévy did not make the connection between his work and martingales until 1950:

> *It was in 1950 at Berkeley that I learned from Loève[51] that the processes called martingales were those that I had considered starting in 1935; according to your letter his second definition, p. 99, coincides with mine. (Or at least becomes the same when constants are added.)*

In the introduction to the second edition (1964) of Lévy [1948], Lévy says, referring the reader to Doob [1953], that he has "*renounced introducing the important notion of separable processes and speaking of martingales.*"

The Lyon colloquium ended on Saturday, July 3, 1948. Several participants got together for lunch in Paris. Lévy, who had finished writing *Stochastic Processes and Brownian Motion* [Lévy 1948], the appearance of which was planned for the new school year, returned to Paris and left on vacation. Doob quickly left France; we find him again at a Congress in Madison, Wisconsin, on September 7, 1948 [Doob 1948]. There remained four years until the publication of his *Stochastic Processes* [Doob 1953], with the 100 pages of its Chapter VII devoted to martingales. So martingales would need more round trips to America before a new generation of French probabilists took them up in their turn.

---

[48]Ville is cited from the fifth line on, and Doob's bibliography contains only three references, in the order: Ville [1939], Doob [1940a], and von Mises [1931].

[49]Fréchet had had in his hands the proofs of Lévy [1937], read by Wolfgang Doeblin and laid out for printing by another of his protégés, the Hungarian refugee Ervin Feldheim. On this topic, see the letter of December 21, 1936, in Barbut, Locker, and Mazliak [2004]. On Feldheim, see Bru [1992].

[50]It is well known that Lévy read little. This was all the more the case for the "Jew Lévy" during the period of the armistice and then the "Pétain regime". Freed from hiding, Lévy devoted himself to the publication and extension of the theorems he had proven during this period (stochastic integrals "à la Lévy" and Brownian motion), then to Markov processes (after a return to the law of "the Lévy surface"). He discovered belatedly that he had anticipated Kakutani on "Kakutani's Theorem", and he did not discover the work of Ito [1944] until 1954 [Locker [2001, Barbut, Locker, and Mazliak 2004].

[51]Michel Loève (1907-1979) was not at the 1948 Lyon colloquium, much to the regret of Fréchet, who said so in Fréchet [1949a]. Several years later, right after the appearance of Doob's *Stochastic Processes*, Loève would visit Paris and would take Paul André Meyer to the USA, where the connection would be made that signaled a new development of martingales (and also probabilistic potential theory) in France.

Doob's lecture

At Lyon, Doob's "*families of chance variables with property* $\mathcal{E}$" of [Doob 1940a] became "*martingales*".[52] Doob's paper consisted of three parts, the first a review of Doob's definition of martingales and his four results in Doob [1940a] (including those we mentioned above), which would be used in the two following parts, *Application to the strong law of large numbers* and *Application to inverse probabilities.* The "law of large numbers" and "inverse probability" are two themes that run through the history of probability theory and statistics,[53] the law of large numbers at the heart of controversies about foundations, and the problem of inverse probabilities at the center of polemics on the interpretation of statistical inference.[54] Moreover, the law (strong)[55] of large numbers, with its connections to foundations, was presented at Lyon by Ottaviani, who advocated "*Cantelli's classical theory*" against von Mises' collectives.

Doob had already tried, in a single 1934 article that cast probability and statistics together in the mold of analysis [Doob 1934], to make the law of large numbers a consequence of Birkhoff's ergodic theorem[56] and to propose "*for the first time a complete proof of the validity of maximum likelihood of R. A. Fisher.*"[57] In 1936 he had to return to Fisher, giving conditions for the consistency of the maximum likelihood estimator [Doob 1936a].[58]

The proof at Lyon of the strong law of large numbers for identically distributed, independent, and integrable variables (the iid strong law) is a model of the type of "spectacular application of martingale theory" that Doob took pleasure in repeatedly presenting, in particular in 1949–50 at Feller's seminar at Cornell [Chung 1998].

To the sequence u_n of variables (iid), Doob associates the "*martingale*" indexed by the negative integers $-n$

$$X_{-n} = \mathrm{E}\{\ldots, y_{-n-1}, y_{-n} \backslash y_{-1}\}, \tag{7}$$

[52]Relative to a filtration.

[53]It is not possible to give a complete bibliography on these themes, so much have they been studied by historians of probability and statistics. The references in the following notes are only the author's "heartthrobs".

[54]See Chapter 3 of Stigler [1986], the introduction to Robert [2006], and Fienberg [2006].

[55]On the history of the strong law, which had already been made clear by Borel in 1909, see Seneta [1992]. The first appearance of the term "strong law" was in French ("loi forte"), in a note in CRAS by Khinchin for January 30, 1928; see Locker [2001].

[56]In 1971, Doob returned to the relations between ergodic theory and martingales, affirming that "*It is true that in a reasonable sense there are two qualitative convergence theorems in measure theory, the ergodic theorem and the martingale convergence theorem.*" [Doob 1971].

[57]On Fisher and maximum likelihood, see especially Bartlett [1965], Aldrich [1999], and Stigler [2007]. Edwards [1997] discusses the sense in which Fisher uses "inverse probability".

[58]Presented in December 1934, received April 4, 1935.

where y_{-m} designates the partial sum of the u_n for $n \geq m$.[59] After a change from $-n$ to n, as we have already noted, this is a backward martingale (relative to the decreasing tribes G_n generated by the partial sums of order m with $m \geq n$). From the iid hypothesis, one has

$$X_{-n} = \frac{1}{n} \sum_{1}^{n} u_k. \tag{8}$$

Finally, when $-n$ tends to $-\infty$, one has almost sure convergence to $\mathrm{E}(u_1)$ from Theorem 1.2 of Doob [1940a], cited above and recalled at Lyon in property iv on page 24.

For the Lyon application of martingales to "inverse probability" one has a result of (almost sure) consistency[60] of the Bayesian procedure[61] for parametric estimation. The parameter θ is subject to the a priori density law $f(\theta)$. For each θ the variable Y admits a law $F(\theta, y)$ with density $f(\theta, y)$ – see "*Hypothesis A*" in Doob's text, and the mapping that associates each θ with its law is injective ("*Hypothesis B*"). The frequency $v_n(y)$ of observations smaller than y in a sample of size n tends almost surely to $F(\theta, y)$ (for each value y) when n tends to infinity (by a direct application of the iid law of large numbers), which allows us to view θ as a function $\hat{\theta}$ defined (up to a negligible set) on the sequences $(y_1, y_2, \ldots, y_j, \ldots)$.

Doob remarks that the measurability hypothesis made in his "*Hypothesis A*" and "*Hypothesis B*" implies his "*Preliminary Hypothesis C*", which says that $\hat{\theta}$ is a measurable function of the sequences (y_j), a "*random variable on* $y_1, y_2, \ldots, y_j, \ldots$ *sample space*"[62] He finds this "*somewhat surprising*".

The almost sure convergence of $\mathrm{E}[\hat{\theta}|y_1, y_2, \ldots, y_n]$[63] to $\hat{\theta}$ and the almost sure convergence of the conditional variance to 0 are obtained by application of the martingale results in Doob [1940a], recalled in the introduction to the presentation at Lyon. The remainder is a discussion of the hypotheses and their interpretation.

After Doob's presentation, only Calyampudi Radhakrishna Rao rose to ask questions, on the possibility of applying the method without an a priori distribution for θ, as in the nonparametric case. One can read Doob's responses in two parts, (i) and (ii), in the discussion that closes the document we have presented.

---

[59] An expression of the form $\mathrm{E}\{A|X\}$ now designates the expectation of X given A. See note 30.

[60] In Doob [1936a] the consistency of a sample statistic is defined by the convergence in probability to the "true value" of the parameter. He also draws attention to the interest of almost sure convergence.

[61] On Bayesian inference, see Fienberg [2006].

[62] One will have noticed that the "chance variables" of Doob [1940a] have now become "random variables". See note 29.

[63] Here we use the now standard notation for the conditional expectation of z given y, $\mathrm{E}[z|y]$ rather than Doob's notation, in which the order of the variables is the opposite: $\mathrm{E}[y \backslash z]$.

For a complementary analysis and for the connection with relatively recent results and the extensions of Doob's consistency theorem, one can refer to Ghosal [1999], and for the nonparametric case to Lijoi, Prünster, and Walker [2004].

Appendix (Brief Biographies)

Pierre Auger (1897–1993). Physicist (atomic physics, nuclear physics, cosmic radiation), professor at the Faculty of Sciences of Paris, and at the École Normale Supérieure, was named Director of Higher Education at the liberation. He occupied this post until 1948. Pierre Auger was the creator of the third cycles and of the consultative committee of the universities. He became a member of the Academy of Sciences in 1977. One can consult at http://picardp1.mouchez.cnrs.fr the entrancing "*Interview with Pierre Auger on April 23, 1986*" conducted by J. F. Picard and E. Pradoura.

Frédéric Joliot (1900–1958). Took charge of the CNRS on August 20, 1944 (Parisian insurrection), ousting the Petainist geologist Charles Jacob.

Paul Lévy (1886–1971). From the time of his thesis of 1911 until "returning to probability" in 1919, Lévy was an analyst, working mainly on functional analysis and on the calculus of variations in infinite dimensions. See Barbut, Locker, and Mazliak [2004] for the whole of Lévy's work. For the period that interests us, which goes until 1937, let us say that Lévy knew Fréchet's integral in "abstract spaces" [Fréchet 1915], that he had become familiar with Daniell's [Daniell 1918] with some delay (he says so in Lévy [1937]), and that he had been able to come very close to Wiener's measure (with his means on the L^2 sphere). He is cited, along with Gateaux and Daniell, for his measures in infinite dimensions in Norbert Wiener's fundamental article on Brownian motion of 1923 [Wiener 1923, *Differential Space*]. He also helped Wiener in 1924 in editing a French article translating "Differential Space" into the language of Borel's denumerable probabilities [Wiener 1924]. In 1924, Lévy created his own approach to measure with his "theory of partitions", that he then extended to "abstract spaces" and which is presented in Chapter II of Lévy [1937] as a kind of foundation for the possibility of giving a general procedure for effectively constructing all probability laws on sets having the power of the continuum. Very satisfied with his theory, he never used it in practice. Lévy had to protect himself, each time it was necessary from the

> *prejudices of certain analysts, and not the least of them, with respect to the theory of probability, or at least with respect to probabilists, supposed not to have the sense of rigor.*

Thus, in the author's preface to the first edition of Lévy [1937], he responded in advance to "*these analysts, and not the least*", by recalling his contributions since 1919 to "*translating well known theorems of analysis into the language of probability*" and marveling

> *...that one could think that an argument, to be rigorous, needed to be translated from one language to another... ...this sounds to me like saying that my French text had to be translated into German in order for my arguments to appear rigorous.*"

He added that the translation was "*within reach of a beginner*".

Richard von Mises (1883–1953). Von Mises presented his theory in his lectures given at the IHP in November 1931, published in 1932 [von Mises [1932]. In order to show the superiority of his system (second lecture, page 157), he criticized the concepts of random variables that he attributed to "*Mr. Fréchet and several others*" who "*do not give an exact definition of this new notion [and] pretend that it is known a priori*" before bringing up "*the ideas established by M. Borel ... on a sort of mathematical probability whose object does not belong to the physical world*". We note also that in May 1935, on the opposite side of the philosophical and logico-mathematical chessboard on the foundations of probability, it was Bruno de Finetti's turn to present, in the course of five lectures at the IHP, retouched in De Finetti [1937], in which he developed his point of view, that he said himself was (op. cit., page 3) "*considered the most extreme solution on the subjectivist side*", then rejecting (op. cit., page 23) "*as illusory*" the idea that "*the impossibility of making precise the relations between probability and frequencies is analogous to the practical impossibility that one encounters in the experimental sciences of exactly connecting the abstract notions of the theory with the empirical realities.*" In footnotes, he attributed this idea to "*modern*" treatises: Castelnuovo [1925], Fréchet-Halbwachs [1924], Lévy [1925], and von Mises [1928]. Lévy's "moderate subjectivist probability" position, which accommodates a realist interpretation of frequencies [Lévy 1925, Lévy 1937], thus found itself caught between two fires, which explains his self-contradictions on von Mises [1931] and Ville [1937], perhaps as much as the poor opinion he had of Ville according to Bru, Bru, and Chung [1999], Locker [2001], and Barbut, Locker, and Mazliak [2004].

Louis Rapkine (1904–1948). Biologist (biochemistry and cellular biology). Originally from Belarus, he emigrated to Paris in 1911 with his family. From 1934 on, he worked to help scientists fleeing Nazism (Welcome Committee for Foreign Scholars). In 1940, he went to New York and, in collaboration with the Rockefeller Foundation, organized the transfer of scientists to the USA (including Hadamard, Perrin, André Weil...). In 1941, he directed the scientific office of Free France in New York with the approval of General De Gaulle. In August 1944, in London, he prepared for the return of scientific exiles to France. It was in September 1945 that he returned to the Rockefeller Foundation in New York to obtain assistance for the reconstruction of French science. He died of cancer in Paris in December 1948.

Georges Teissier (1900–1972). Biologist (marine biology, then evolutionary genetics), having belonged, like Joliot, to the French Communist Party (PCF) during the war, but in the area of armed internal resistance in the Franc Tireurs Partisans, became director of the evolutionary genetics laboratory in 1945, and then succeeded Joliot as director of the CNRS when Joliot took over the atomic energy commission, on February 3, 1946, after the resignation of De Gaulle on January 20, 1946. The PCF ministers were ousted on May 4, 1947, but Teissier remained director of CNRS until 1950, at which time he was relieved of his duties because of his political positions (unofficially because he organized petitions and demonstrations in his role as honorary president of the French University Union, close to the PCF) by the minister Yvon Delbos. (Already twice Minister of National Education before the war, Delbos would again hold this position under the three governments of Marie, Queuille, and Bidault from 1948 to 1950; he would be a candidate for the presidency of the republic in 1953.) Teissier's firing was the object of a decree of the Council of State, "the Teissier decree", made by the office of jurisprudence in the matter of the government being able to fire managers without giving reasons.

Warren Weaver (1894–1978). American mathematician, very influential administrator of research at the Rockefeller Foundation. In 1944 he worked with Claude Shannon on information theory.

References

[1] Aldrich, J. (1997): R. A. Fisher and the making of maximum likelihood 1912–1922, *Statistical Science* 12:162–176.

[2] Armatte, M. (2001): Maurice Fréchet statisticien, enquêteur et agitateur public, *Revue d'Histoire des Mathématiques* 7(1):7–65.

[3] Armatte, M. (2006): Enseignement de la statistique économique (1885–1925). Présentation de quelques documents, *JEHPS* 2(2).

[4] Barbut, M., Locker, B. and Mazliak, L. (2004): *Paul Lévy, Maurice Fréchet. 50 ans de correspondance mathématique*, Hermann.

[5] Bartlett, M. S. (1965): R. A. Fisher and the last fifty years of statistical methodology, *Journal of the American Statistical Association* 60(310):395–409.

[6] Bourbaki, N. (1969): *Eléments d'histoire des mathématiques*, Hermann.

[7] Bretagnolle, J. (1987): Notes de lecture sur l'œuvre de Paul Lévy, *Annales de l'Institut Henri Poincaré, section B* 23(S2):239–243.

[8] Bru, B. (1992): La vie et l'œuvre de W. Doeblin (1915–1940) d'après les archives parisiennes, *Mathématiques et sciences humaines* 119:5–51. English version: Doeblin's life and work from his correspondence, pp. 1–64 of *Doeblin and Modern Probability*, edited by Harry Cohn, American Mathematical Society, Providence, 1993.

[9] Bru, B., Bru, M. F., and Chung, K. L. (1999): Borel et la martingale de Saint-Petersbourg, *Revue d'histoire des mathématiques* 5(2):181–247. English version: Borel and the St. Peterburg martingale, *JEHPHS* 5(1).

[10] Bru, B. and Eid, S. (2009): Jessen's theorem and Lévy's lemma, *JEHPS* 5(1).

[11] Bungener, M. and Joël, M. E. (1989): L'essor de l'économétrie au CNRS, *Cahiers pour l'histoire du CNRS* 4:45–78.

[12] Castelnuovo, G. (1928): *Calcolo delle probabilità*, Zanichelli.

[13] Charle, C. (1989): Le personnel dirigeant du CNRS (1937–1966), *Cahiers pour l'histoire du CNRS* 4:1–36.

[14] Chaumont, L. Mazliak, L. and Yor, M. (2006): A. N. Kolmogorov : Some aspects of the probabilistic works, *in* Kolmogorov's heritage in mathematics, Springer, 2006.

[15] Crépel, P. (2009): Jean Ville's recollections, in 1984 and 1985, concerning his work on martingales, *JEHPS* 5(1).

[16] Chung, K. L. (1998): Probability and Doob, *The American Mathematical Monthly* 105(1):28–35.

[17] Daniell, Percy J. (1918): A general form of integral, *Ann. Math.* 19:279–294.

[18] De Finetti, B. (1937): La prevision, ses lois logiques, ses sources subjectives, *Annales de l'IHP* 7(1):1–68.

[19] Dieudonné, J. (1948): Sur le théorème de Radon-Nicodym III, *Annales de l'Université de Grenoble* 23:25–53.

[20] Doob, J. L. (1934): Probability and statistics, *Transactions of the American Mathematical Society* 36(4):759–775.

[21] Doob, J. L. (1936a): Statistical estimation, *Transactions of the American Mathematical Society* 39(3):410–421.

[22] Doob, J. L. (1936b): Note on probability, *Annals of Mathematics* 37(2):363–367.

[23] Doob, J. L. (1937): Stochastic Processes depending on a continuous parameter, *Transactions of the American Mathematical Society* 42(1):107–140.

[24] Doob, J. L. (1938): Stochastic processes with an integral parameter, *Transactions of the American Mathematical Society* 44(1):87–150.

[25] Doob, J. L. (1939): Review of *Etude critique de la notion de collectif*, by Jean Ville, *Bulletin of the American Mathematical Society* 45(11):824.

[26] Doob, J. L. (1940a): Regularity properties of certain families of chance variables, *Transactions of the American Mathematical Society* 47(3):455–486.

[27] Doob, J. L. (1940b): The law of large numbers for continuous stochastic processes, *Duke Math. J.* 6:290–306.

[28] Doob, J. L. (1942): What is a stochastic process? *The American Mathematical Monthly* 49(10):648–653.

[29] Doob, J. L. (1947): Probability in function space, *Bulletin of the American Mathematical Society* 53:15–30.

[30] Doob, J. L. (1948): The transition from stochastic differences to stochastic differential equations, in Report of the Madison Meeting, *Econometrica* 17(1):63–78.

[31] Doob, J. L. (1949): Application of the theory of martingales, *Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 juin–3 juillet 1948)*, Paris CNRS, 23–27.

[32] Doob, J. L. (1953): *Stochastic Processes*, Wiley, New York.

[33] Doob, J. L. (1971): What is a martingale? *The American Mathematical Monthly* 78(5):451–453.

[34] Doob, J. L. and Ambrose, W. (1940): On two formulations of the theory of stochastic processes depending upon a continuous parameter, *Annals of Mathematics* 41(4):737–745.

[35] Dosso, D. (1998): *Louis Rapkine et la mobilisation de la France Libre*, Thèse de l'Université Paris Diderot (Paris VII).

[36] Dieudonné, J. (1948): Sur le théorème de Radon-Nicodym III, *Annales de l'Université de Grenoble* 23:25–53.

[37] Dugué, D. (1937): Application des propriétés de la limite au sens du calcul des probabilités a l'étude de diverses questions d'estimation, *J. l'École Polytechnique* 3(4):305–373.

[38] Edwards, A. W. F. (1997): What did Fisher mean by inverse probability? *Statistical Science* 12(3):177–184.

[39] Eid, S. (2008): *Martingales et géométrie borelienne des probabilités*, Thèse de l'Université Paris Diderot (Paris VII).

[40] Fienberg, S. E. (2006): When did Bayesian inference become Bayesian? *Bayesian Analysis* 1(1):1–40.

[41] Fréchet, M. (1915): Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait, *Bulletin de la Société Mathématique de France* XLIII, 248–265.

[42] Fréchet, M. (1949a): Avant propos, *Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 juin–3 juillet 1948)*, Paris CNRS, 5–10.

[43] Fréchet, M. (1949b): Les valeurs typiques d'ordre nul ou infini d'un nombre aléatoire et leur généralisation, *Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 juin–3 juillet 1948)*, Paris CNRS, 47–51.

[44] Fréchet, M. and Halbwachs, M. (1924): *Le calcul des probabilités à la portée de tous*, Dunod.

[45] Gardes, F., and Garrouste, P. (2004): Ville's contribution to the integrability debate: the mystery of a lost theorem, *History of Political Economy* 38(5):87–106.

[46] Ghosal, S. (1999): A review of consistency ad convergence rate of posterior distribution, *Proceedings of Varanashi Symposium*, Banaras Hindu University.

[47] Hald, A. (2004): *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713 to 1935*, Department of applied mathematics and statistics, University of Copenhagen.

[48] Havlova, V. Mazliak, L. and Sisma, P. (2005): Le début des relations franco-tchécoslovaques vu à travers la correspondance Hostinsky-Fréchet, *JEHPS* 1(1).

[49] Ito, K. (1944): Stochastic integrals, *Proceedings of Imperial Academy Tokyo* 20:519–524.

[50] Jessen, B. (1935): The theory of integration in a space of an infinite number of dimensions, *Acta mathematicae* 63:249–323.

[51] Kolmogorov, N. (1933): Grundbegriffe der Warscheinlichkeitrechnung, *Ergebenisse der Mathematik* II(3).

[52] Lévy, P. (1935a): Propriétés asymptotiques des sommes de variables aléatoires enchaînées, *Bull Sci. Math.* 84–96 and 109–128.

[53] Lévy, P. (1935b): La loi forte des grands nombres pour les variables enchaînées, *CRAS* 201:493–495.

[54] Lévy, P. (1935c): Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées, *Journal de Mathématiques pures et appliquées* 14:347–402.

[55] Lévy, P. (1936a): La loi forte des grands nombres pour les variables enchaînées, *Journal de Mathématiques pures et appliquées* 15:11–24.

[56] Lévy, P. (1936b): Sur la notion de probabilité conditionnelle, *Bulletin des Sciences Mathématiques* 60:66–71.

[57] Lévy, P. (1937): *Théorie de l'addition des variables aléatoires*, Gauthier-Villars.

[58] Lévy, P. (1948): *Processus stochastiques et mouvement brownien*, Gauthier-Villars.

[59] Lévy, P. (1949): Processus doubles de Markov, *Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 juin–3 juillet 1948)*, Paris CNRS, 53–59.

[60] Lévy, P. (1970): *Quelques aspects de la pensée d'un mathématicien*, Albert Blanchard.

[61] Lijoi, A. Prünster, I. and Walker, S. G. (2004): Extending Doob's consistency theorem to nonparametric densities, *Bernoulli* 10(4):651–663.

[62] Locker, B. (2001): *Paul Lévy: "La période de guerre". Intégrales stochastiques et mouvement brownien*, Thèses, Université René Descartes Paris V.

[63] Mazliak, L. (2003): Andrei Nikolaevitch Kolmogorov (1903-1987), Un aperçu de l'homme et de l'œuvre probabiliste, *Les cahiers du CAMS*, Série Histoire du Calcul des Probabilités et de la Statistique 53.

[64] Mazliak, L. (2009): How Paul Lévy saw Jean Ville and martingales, *JEHPS* 5(1).

[65] Meusnier, N. (2006): Sur l'histoire de l'enseignement des probabilités et des statistiques, *JEHPS* 2(2).

[66] Michel, A. (1992): *Constitution de la théorie moderne de l'intégration*, Vrin.

[67] von Mises, R. (1928): *Wahrscheinlichkeit, Statistik und Wahrheit*, Springer.

[68] von Mises, R. (1931): *Wahrscheinlichkeitrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Deuticke.

[69] von Mises, R. (1932): Théorie des probabilités, fondements et applications, *Annales de l'IHP* 3(2):137–190.

[70] Neveu, J. (1972): *Martingales à temps discret*, Masson.

[71] Neveu, J. (1973): Note sur les travaux de Paul Lévy, *La Jaune et la Rouge*.

[72] Nikodym, O. (1930): Sur une généralisation des intégrales de M. Radon, *Fundamenta Mathematicae* 15:131–179.

[73] Nualart, D. (1989): Les martingales i les seves applicacions des d'una perspectiva historica, *Butlleti de la Societat Catalana de Mathemàtiques* 4:33–46.

[74] Ottaviani, G. (1949): La loi uniforme des grands nombres dans la théorie classique des probabilités. *Actes du Colloque International sur le Calcul des probabilités et ses applications Lyon (28 juin–3 juillet 1948)*, CNRS, 11–17.

[75] Piaggio, HTH. (1939): Review of *Etude critique de la notion de collectif*, by Jean Ville, *The Mathematical Gazette* 23(257):490–491.

[76] Radon, J. (1913): Theorie und Anwendungen der absolut additiven Mengenfunctionen, *Sitz. Akad. Wiss. Wien.* 122:1295–1438.

[77] Robert, C. P. (2006): *Le choix bayésien: Principes et pratique*, Springer.

[78] Seneta, E. (1992): On the history of the strong law of large numbers and Boole's inequality, *Historia Mathematica*, 19:36–48.

[79] Shafer, G. (2009): The education of Jean André Ville. *JEHPS* 5(1).

[80] Siegmund-Schultze, R. (2005): Maurice Fréchet à Strasbourg: les mathématiques entre nationalisme et internationalisme, in *La science sous influence: L'université de Strasbourg, enjeu des conflits franco-allemands, 1872–1945*, edited by E. Crawford and J. Ollf-Nathan, Strasbourg, La Nuée Bleue, 185–196.

[81] Snell, J. Laurie (1997): A conversation with Joe Doob, *Statistical Science* 12(4):301–311.

[82] Stigler, S. M. (1986): *The History of Statistics, The Measurement of Uncertainty before 1900*, The Belknap Press of Harvard University, Cambridge, Massachussets.

[83] Stigler, S. M. (2007): The epic story of maximum likelihood, *Statistical Science* 22(4):598–620.

[84] Ville, J. (1939): *Etude critique de la notion de collectif*, Gauthier-Villars.

[85] Ville, J. (1942): Sur un problème de géométrie suggéré par l'étude du mouvement brownien, *CRAS Paris* 215:51–52.

[86] Ville, J. (1946): Sur les conditions d'existence d'une ophélimité totale et d'un indice du niveau des prix, *Annales de l'Université de Lyon* 9:32–39.

[87] Ville, J. (1948): Théorie et application de la notion de signal analytique, *Câbles et Transmissions* 2.

[88] Ville, J. (1949): Fonctions aléatoires et transmission de l'information, *Actes du Colloque International sur le Calcul des probabilités et ses applications Lyon (28 juin–3 juillet 1948)*, CNRS, 115–119.

[89] Von Plato, J. (1994): *Creating Modern Probability*, Cambridge University Press.

[90] Wiener, N. (1923): Differential space, *J. Math-Phys Massachussets Institute* 2:131–174.

[91] Wiener, N. (1924): Un problème de probabilités dénombrables, *Bulletin de la Société mathématique de France* 11:569-578.

[92] Yor, M. (2007): Joseph Leo Doob (27 février 1910–7 juin 2004), *SMF Gazette*.

[93] Zallen, D. (1989): The Rockefeller Foundation and French research, *Cahiers pour l'histoire du CNRS* 5.

APPLICATION OF THE THEORY OF MARTINGALES

by J. L. DOOB.

(Urbana, Ill., U. S. A.)

I. Introduction and definitions.

Let $\ldots, x_{-1}, x_0, x_1, \ldots$ be a sequence of random variables with $E\{|x_n|\} < \infty$ [1] for all n. The sequence is called a martingale if for every n

$$(1.1) \qquad E\{\ldots, x_{n-2}, x_{n-1} \backslash x_n\} = x_{n-1},$$

with probability 1. The sequence x_n may be infinite in only one direction.

Although other authors had derived many martingale properties, in various forms, Ville [2] was the first to study them systematically, and to show their wide range of applicability. In the following sections new applications of martingale theory, using more recent results, will be made, to the strong law of large numbers and to statistical theory. The following theorems are stated for later reference [3]. In all cases it is supposed that the variables involved satisfy the martingale property (1.1).

(i) $E(x_n)$ does not depend on n, but

$$\ldots \leq E\{|x_n|\} \leq E\{|x_{n+1}|\} \leq \ldots$$

(ii) $\lim_{n \to -\infty} x_n = x_{-\infty}$ exists with probability 1; $E\{x_{-\infty}\}$ exists and

$E\{x_{-\infty}\} = E\{x_n\}$. The sequence $\{x_n\}$, $n \leq 0$, is uniformly integrable.

(iii) If $\lim_{n\to\infty} E\{|x_n|\} = l < \infty$, it follows that $\lim_{n\to\infty} x_n = x_\infty$ exists with probability 1; $E\{x_\infty\}$ exists and $E\{|x_\infty|\} \leq l$. In particular if the x_n' s are all ≥ 0, $l = E\{x_n\} < \infty$ and the stated conclusions hold.

(iv) If $\ldots, y_0, y_1, \ldots$ y are any random variables, with $E\{|y|\} < \infty$, and if x_n is defined by

$$x_n = E\{\ldots, y_{n-1}, y_n \backslash y\}$$

then the sequence $\{x_n\}$ is a martingale. Moreover since $E\{|x_n|\} \leq E\{|y|\}$, (iii) is applicable. In this case

$$x_\infty = E\{\ldots, y_0, y_1, \ldots \backslash y\}.$$

2. Application to the strong law of large numbers.

Let $u_1, u_2, \ldots$ be mutually independent random variables with a common distribution function, and suppose that $E\{u_n\}$ exists. Then

$$(2.1) \qquad \lim_{n\to\infty} \frac{u_1 + \cdots + u_n}{n} = E\{u_1\}$$

with probability 1. This particular case of the strong law of large numbers for stationary processes can be derived from martingale theory as follows.

Define $\ldots, y_{-2}, y_{-1}$ by

$$y_{-n} = u_1 + \ldots + u_n$$

and x_n by

$$x_{-n} = E\{\ldots, y_{-n-1}, y_{-n} \backslash y_{-1}\}$$
$$= E\{y_{-n} \backslash y_{-1}\} = \frac{u_1 + \cdots + u_n}{n}$$

Then by (iv) the sequence $\{x_n\}$ is a martingale, and hence $\lim_{n\to-\infty} x_n$ exists with probability 1, giving the existence of the limit in (2.1). The limit is easily identified as $E\{u_1\}$.

3. Application to inverse probabilities.

Suppose that for each value of parameter θ, which varies in some Borel linear set $f(\theta, y)$ is a probability

---

(1) In the following $E\{x\}$ will denote the expectation of x and $E\{y \backslash x\}$ will denote the (conditional) expectation of x for given y. In most cases below y will be replaced by infinitely many conditioning variables.

(2) Etude critique de la notion de collectif, Paris 1939.

(3) Doob, transactions of the American Mathematical Society 47 (1940) 455-486.

density in y, corresponding to the distribution function F (θ, y),

$$(3.1)\qquad F(\theta, y)=\int_{-\infty}^{y} f(\theta, z)\, dz.$$

Hypothesis (A) : *F (θ, y) is a Baire function of (θ, y)*. For this to be true it is necessary and sufficient that F (θ, y) be a Baire function of θ for each value of y. This is the only regularity hypothesis to be imposed on F (θ, y); it implies that $f(\theta, y)$ can also be assumed to be a Baire function of (θ, y).

Suppose that θ is chosen in accordance with some probability distribution having density $f(\theta)$ and that n sample values of y, $y_1, \ldots, y_n$ are then drawn independently using the $f(\theta, y)$ distribution. We shall apply martingale theory to the classical problem of estimating the value of θ chosen, from a knowlecdge of the sample values $y_1, \ldots, y_n$.

Hypothesis (B). *To different values of θ correspond different distributions*, that is if $\theta_1 \neq \theta_2$, then F (θ_1, y) and F (θ_2, y) are not identically equal.

Hypotheses (A) and (B) are the only hypotheses to be imposed on F (θ, y). The estimation problem stated above can only be solved if for large n the value of θ drawn is very nearly a function of $y_1, \ldots, y_n$. This relationship can be formalized as follows. For every n the quantities $\theta, y_1, \ldots, y_g$ are random variables with joint density $f(\theta) \prod_{j=1}^{n} f(\theta, y_j)$. If $n=\infty$ we have a sequence of random variables $\theta, y_1, y_2, \ldots$, and we now show that θ is a function of $y_1, y_2, \ldots$. To show how to calculate F (θ, y) (and therefore θ, by Hypothesis (B) this we show in terms of the sequence $y_1, y_2, \ldots$. This is simply done using the form of the strong law of large numbers proved in Section 2. In accordance with this law, if $\nu_n(y)$ is the number of the first n y_j s which are $\leq y$

$$(3.2)\qquad F(\theta, y)=\lim_{n\to\infty} \frac{\nu_n(y)}{n}$$

with probability 1, for each y. Hence (3.2) holds with probability 1 simultaneously for all rational values of y. Excluding the exceptional y_j sequences, the *nth* term on the right, for a given y_j sequence, is a distribution function in y which converges for all rational y to F (θ, y). Thus (3.2y) as now interpreted states the distribution function of the first n y_j's converges to F (θ, y) with probability 1. It thus follows from Hypothesis (B) that $\hat\theta$ is a function of y_j sequences (neglecting zero probabilities as usual), and in the following we shall write $\hat\theta$ or $\theta(y_1, y_2, \ldots)$ for the variable considered in this connection.

Preliminary Hypothesis (C), (implied by (A) and (B), see below).

$\hat\theta$ is a measurable function of y_j sequences, that is a random variable on $y_1, y_2, \ldots$ sample space. The significance of this measurability hypothesis will be discussed below. It is essentially a restriction on the regularity of F (θ, y) in θ.

Assuming Hypothesis (A) the conditional distribution of θ for given $y_1, \ldots, y_n$ has density (roughly the probility that $\theta=\alpha$)

$$(3.3)\qquad \frac{f(\alpha)\prod_{1}^{n} f(\alpha, y_j)}{\int f(\varphi)\prod_{1}^{n} f(\varphi, y_j)\, d\varphi}$$

where the integral is taken over the whole range of the parameter. Now when $n\to\infty$, $y_1, \ldots, y_n$ becomes $y_1, y_2, \ldots$; any sequence $y_1, y_2, \ldots$ (excluding a class of sequences of total probability 0 determines a value of $\hat\theta$ and the problem is to see how the distribution determined by (3.3) becomes concentrated at $\hat\theta$ when $n\to\infty$. This is the full formalization of the estimation problem described in intuitive language at the beginning of this section. Hypothesis (B) is obviously a necessary condition if $\hat\theta$ is to be estimated from $y_1, \ldots, y_n$. In fact what is estimated is not θ but the distribution function F $(\hat\theta, y)$ and without Hypothesis (B) even if F$(\hat\theta, y)$ were known precisely $\hat\theta$ would not be uniquely determined. It is clear that some sort of regularity condition on F $(\hat\theta, y)$ in θ is also necessary since we can at best hope to approximate the true distribution function F $(\hat\theta, y)$ more and more closely as $n\to\infty$ and must make enough hypotheses to insure that approximating F $(\hat\theta, y)$ means approximating $\hat\theta$. It is somewhat surprising that Hypotheses (A) and (B) are sufficient.

There are several statements it would be desirable to prove. For example it would be desirable to prove that the conditional distribution has as $n\to\infty$ the «true value» $\hat\theta=\theta(y_1, y_2, \ldots)$ as its expectation that is

$$(3.4)\qquad \lim_{n\to\infty} E\{y_1, \ldots, y_n \backslash \hat\theta\}=\lim_{n\to\infty} \frac{\int \alpha f(\alpha)\prod_{1}^{n} f(\alpha, y_j)\, d\alpha}{\int f(\varphi)\prod_{1}^{n} f(\varphi, y_j)\, d\varphi}=\hat\theta$$

with probability 1. Secondly it would be desirable to prove that the distribution becomes concentrated around the true value, as $n\to\infty$, in the sense say that variance goes to 0, that is

$$(3.5)\qquad \lim_{n\to\infty} E\{y_1, \ldots, y_n \backslash \hat\theta^2\}-E^2\{y_1 \ldots, y_n \backslash \hat\theta\}=0$$

with probability 1. These two statements imply that the conditional probability of differing from the true value $\hat\theta$ by at least ε goes to 0, as $n\to\infty$ for every $\varepsilon>0$. This fact can itself be formulated as follows : let $z_{\lambda,\mu}=1$ when $\lambda \leq \hat\theta \leq \mu$ and $z_{\lambda,\mu}=0$ otherwise. Then

$$(3.6)\qquad \lim_{n\to\infty} \Pr\{y_1, \ldots, y_n \backslash \lambda \leq \hat\theta \leq \mu\}=\lim_{n\to\infty} \frac{\int_{\lambda}^{\mu} f(\alpha)\prod_{1}^{n} f(\alpha, y_j)\, d\alpha}{\int f(\varphi)\prod_{1}^{n} f(\varphi, y_j)\, d\varphi}=z_{\lambda\mu}$$

that is

$$(3.6')\qquad \lim_{n\to\infty} E\{y_1, \ldots, y_n \backslash z_{\lambda,\mu}\}=z_{\lambda,\mu}$$

with probability 1. Now equations (3.4), (3.5), and (3.6) are all trivial applications of the martingale theorems stated in Section 1. Equation (3.4) requires the existence of

$$E\{\hat\theta\}=\int_{-\infty}^{\infty} \theta f(\theta)\, d\theta;$$

equation (3.5) requires that of

$$E\{\hat\theta^2\}=\int_{-\infty}^{\infty} \theta^2 f(\theta)\, d\theta;$$

but equation (3.6) requires nothing beyond Hypotheses (A) and (B) (and (C) which will be shown below to be implied by them). We can even go further than (3.6) which states that there is convergence of the distributions, by examining the densities. In fact the conditional densities also converge in the sense that excluding a set of $\hat\theta$ values of probability, 0,

$$(3.7)\qquad \lim_{n\to\infty} \Pr\{y_1, \ldots, y_n \backslash \hat\theta=\alpha\}=\lim_{n\to\infty} \frac{f(\alpha)\prod_{1}^{n} f(\alpha, y_j)}{\int f(\varphi)\prod_{1}^{n} f(\varphi, y_j)\, d\varphi}=0$$

with probability 1. The first term in (3.7) is only suggestive, and explains why the sequence of random variables involved is a martingale, a fact which is easily checked directly (although it may be necessary to exclude a set of $\hat\theta$ values of probability 0). The fact that there is convergence then follows from (iii), Section 1, since the random variables are non-negative. The fact that the limit is 0 follows easily from the integrated version (3.6). A refinement of the argument can be used to prove that there is a limit (≥ 0) in (3.7) for all α.

We must still discuss the significance of Hypothesis (C), a condition which must be translated into a condition on the function F (θ, y). Without Hypothesis (C) the limits in (3.4), (3.5) and (3.6) would still exist, but would be different. For example that in (3.4) would become E $\{y_1, y_2, \ldots \backslash \hat\theta\}$.

It is quite possible mathematically to have a pathological situation in which $\hat\theta$, although it is a function of the y_j's, is not measurable on $y_1, y_2, \ldots$ space, and hence in which E $\{y_1, y_2, \ldots \backslash \hat\theta\}$ is not $\hat\theta$. If this occurred here the limit in (3.5) would be

$$E\{y_1, y_2, \ldots \backslash \hat\theta^2\}-E^2\{y_1, \ldots, y_n \backslash \hat\theta\}$$

which would be positive, with positive probability. In order to analyze Hypothesis (C) further we shall consider distribution functions as points of a space Δ defining the distance between two distribution functions as usual as the maximum distance between their graphs (filled in with vertical lines at the jumps) measured along lines with slope -1.

The space Δ is then separable, since the distribution functions which are rational valued and constant except for finitely many jumps, which occur only at rational values, are everywhere dense in Δ. The family of distribution functions F (θ, y) is a curve T in Δ, and the regularity of F (θ, y) in θ can be described in terms of the regularity properties of this curve. For each θ there is a single point of this curve, and (Hypothesis (B)) distinct values of θ correspond to distinct points of the curve. The transformation $\theta \longleftrightarrow$ *point of* Γ can be described by the function F (θ) which for each θ is the point of Γ corresponding to θ, so that F (θ) is simply the function F (θ, y) from a different point of view. We now prove that (C) is implied by (A) and (B); the latter hypotheses are assumed true throughout the following discussion.

a. F (θ) is a Baire function of θ. In fact since Δ is separable, with the step functions described above dense in ΔF (θ) is a Baire function if for each point G of Δ which is a step function increasing only at a finite number of points, and for each positive ρ, the values of θ for which the Δ distance from F (θ) to G *is* $\leq \rho$ form a Borel set. This fact is an immediate consequence of Hypothesis (A).

(*b*) *F $(\hat\theta)$ is a measurable function of the y_j's.* In factd if $\nu_n(y)/n$ is defined as in (3.2), it is a distribution function, and therefore is a point of Δ depending on the y's; in other words $\nu_n(y)/n$ defines a function $G_n(y_1, \ldots, y_n)$ taking on values in Δ. In terms of Δ distance

$$(3.8)\qquad \lim_{n\to\infty} G_n(y_1, \ldots, y_n)=F(\hat\theta)$$

with probability 1; this is merely (3.2) in a different form. Hence it is sufficient to prove that $G_n(y_1, \ldots, y_n)$ is measurable, and for this, since Δ is separable, it is sufficient to prove that the $y_1, \ldots, y_n$'s for which the distance from G_n to any given point G of Δ is $\leq \rho$ is measurable, for every ρ. It is even sufficient to prove this when G is a step function with a finite number of jumps. In this case the statement is obvious.

(c) *To a Borel θ set corresponds a Borel set on* Γ. This follows from the fact that a Baire function [F (θ) in the present case] which has a single valued inverse takes Borel sets into Borel sets.

(d) *(A) and (B) together imply* (θ). In fact the inequality $\hat{\theta} < k$ defines a Borel set on Γ, by (c), which in turn defines a measurable $y_1 y_2, \ldots$ set by (b). Hence θ is measurable as a function of the y_j's, as was to be proved.

Thus the only hypotheses necessary to ensure the truth of the inverse probability theorems discussed in this section are (A) and (B) (except for the existence of moments of the θ distribution needed in (3.4) and (3.5) as already noted. The condition (B) is necessary and (A) is not far from being necessary. Note that these theorems are «probability 1 theorems». The estimate of the value of θ drawn may not be good for a θ set of probability 0. This distinguishes the present discussion from that of von Mises who solved the problem in the Bernoulli case for individual θ values, with necessarily stronger hypothese [4]. Finally we note that although the problem was stated in terms of density functions, this formulation only served to simplify the notation; the restriction is entirely unnecessary.

It is interesting to consider the special case when θ varies through a finite or denumerable set, assigning positive probability to each θ value. In this case Hypothesis (A) is automatically satisfied, and we find that [assuming only Hypothesis (B)] θ can always be estimated accurately no matter how F (θ, y) varies with θ. As a striking special case, suppose that θ varies through all positive integers, and that F $(1, y) = \lim_{n\to\infty}$ F(n, y) for all y. In this case one might expect trouble in estimating $\hat{\theta}$ when $\theta = 1$, but the discussion shows that this is not true. This is of course to be taken as a theoretical statement. Pratically the situation would give considerable trouble.

4. Discussion,

In answer to questions asked by Dr. Rao, the following additional points are noted.

(i) The estimation problem was stated in § 3 as a problem in inverse probabilities. The results can be interpreted however in a way which makes unnecessary the hypothesis of an *a priori* distribution. We apply this only to one case, suggested by (3.4). Let $f(\theta, y)$ be as in § 3, and for some $\theta = \theta_0$ let $y_1, y_2, \ldots$ be mutually independent, with the distribution determined by $f(\theta_0, y)$. The problem is to determine the unknown θ_0 from the sample $y_1, y_2, \ldots$ Here θ_0 is *not* a random variable. We assume hypotheses (A) and (B) on $f(\theta, y)$. Let $f(\theta)$ be a density function in θ and suppose that $\int |\theta| f(\theta)\, d\theta < \infty$. We do not use $f(\theta)$ to make θ a random variable, but merely use it as a weighting function to define [cf. (3.6)]

$$\theta_n(y_1, \cdots, y_n) = \frac{\int \alpha f(\alpha) \prod_1^n f(\alpha, y_j)\, d\alpha}{\int f(\varphi) \prod_1^n f(\varphi, y_j)\, d\varphi}. \tag{4.1}$$

This function of $y_1, \ldots, y_n$ is to be taken as an estimate of θ_0, the «true value». Now, according to § 3 if $f(\theta)$ were used to define θ probabilities θ_n and θ would be random variables and $\lim_{n\to\infty} \theta_n = \theta$ with probability 1 in $\theta, y_1, y_1, \ldots$ sample space. Hence for fixed θ (neglecting possibly a θ set of θ probability 0) $\lim_{n\to\infty} \theta_n = \theta$ with probability 1 in $y_1, y_2, \ldots$ sample space. Dropping the probability interpretation of $f(\theta)$ we can now finally say that if any weight function $f(\theta)$ is chosen (4.1) defines an estimate of θ_0, the true value, with the following consistency property. Defining $y_1, y_2, \ldots$ probabilities by $f(\theta_0, y)$, $\lim_{n\to\infty} \theta_n = \theta_0$ with probability 1 for each value of θ_0 except possibly for a set of values over which the integral of f vanishes. In other words $\{\theta_n\}$ is a consistent statistic except possibly for a θ_0 set over which the integral of $f(\theta)$ vanishes. This exceptional set has Lebesgue measure 0 if $f(\theta)$ is always positive, but may depend on the choice of f.

(ii) The estimation problem was stated in § 3 in parametric form; θ varied in a Borel linear set. The parametric form is unnecessary. It can be supposed that all possible distributions are considered, that is θ is now supposed to vary over all points of Δ, the space of distribution functions. An *a priori* distribution of θ is then a probability distribution over the Borel sets of Δ. Hypotheses (A) and (B) are now satisfied automatically, and the proof that (C) is also true θ is almost immediate, since the only delicate point was the proof of (c) which is now true by definition; the fact that a Borel θ set corresponds to a Borel Δ set is now true because θ space is Δ. The arguments around this point in § 3 were really aimed at proving that $f(\theta)$ defined a probability measure of Borel Δ sets, which we are now making a hypothesis in the non-parametric form. The results of § 3 all go over, with the obvious modifications due to the fact

[4] Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Leipzig 1931, p. 188-192.

that θ is now not numerically valued. In conclusion we remark that the arguments of § 3 are valid whenever θ varies on a Borel set of a complete metric space. The difference between parametric and nonparametric estimation is rather subtle at this level. If θ space is a line, or an n-dimensional Euclidean space, the problem would be called parametric. If θ space is all of Δ the problem would presumably be called nonparametric, at least if the θ probability measure did not confine θ with probability 1 to the homeomorphic image of a line or of n space, in which case we should be again in the parametric problem just described.