UNIVERSITE DE RENNES I
UER de Mathématiques et Informatique
Campus de Beaulieu
35042 - Rennes Cedex

PUBLICATION DES SEMINAIRES DE MATHEMATIQUES

SEMINAIRE DE PROBABILITES

RENNES 1984

QUELQUES MATERIAUX POUR L'HISTOIRE DE LA THEORIE DES MARTINGALES (1920-1940)

Exposé de Pierre CREPEL le 05.11.1984

.../...

I - INTRODUCTION

Cet exposé, un peu "érudit et technique" n'est qu'une étape pour servir à l'histoire des martingales : l'étape la plus facile pour un mathématicien, celle qui consiste à décrire les aspect explicites du développement de cette notion dans l'entre-deux-guerres. Le travail le plus intéressant reste à faire, on en dira ici seulement quelques mots en remarque.

L'article qui suit est écrit dans des termes familiers aux probabilistes d'aujourd'hui, et nous espérons que, si limité soit-il dans ses ambitions, il pourra leur être utile.

a) Le "corpus classique"

Quand on parle de "martingale" à un mathématicien de ces dernières décennies (éduqué dans l'esprit du livre de J.L. DOOB de 1953, "Stochastic processes"), cela évoque en gros les idées que voici :

Définition : on appelle "martingale" une famille (*) de variables aléatoires (v.a.) (S_t), indexée en général par le temps ($t \in \mathbb{R}$ ou $\mathbb{Z}$), adaptée à une famille croissante de tribus $\mathcal{F}_t$, et telle que

$$E(S_t / \mathcal{F}_s) = S_s \quad \text{p.s.} \quad \text{pour } s \leq t$$

L'exemple le plus classique est celui où $S_n = u_1 + \dots + u_n$, (u_n) désignant une suite de v.a. indépendantes et centrées.

Mais on pense aussi à d'autres exemples typiques : si S est une v.a. et si $\mathcal{F}_t$ est une famille croissante de tribus, $S_t = E(S / \mathcal{F}_t)$ est une martingale. Il y en a d'autres qui sont devenues classiques, celle qui donne la dérivée d'une mesure, ou bien encore en statistiques le rapport de vraisemblance, etc.

N.B. Toutes les définitions et tous les théorèmes énoncés ici ne sont exprimés ni sous les formes les plus rigoureuses, ni dans leur généralité maximale.

(*) Cette famille est d'ailleurs plutôt notée x(t) : l'exposé fera comprendre simplement cette différence de notation significative.

Le théorème d'arrêt

Soient σ et τ deux temps d'arrêt bornés ($\sigma \leq \tau$), alors $E(S_\tau / \mathcal{F}_\sigma) = S_\sigma$ p.s. (avec des définitions connues).

En d'autres termes, le caractère de martingale d'un processus n'est pas affecté par un temps d'arrêt borné, c'est-à-dire par une section aléatoire du temps tenant compte uniquement de la connaissance du passé. Dit encore autrement, pour un jeu équitable, il n'existe pas de stratégie (c'est-à-dire de façon de miser et de quitter le jeu en tenant compte uniquement des coups passés) qui permette de gagner, du moins lorsqu'on a une fortune finie.

Les théorèmes de convergence

Soit (S_n) une martingale à temps discret. Alors, sous diverses hypothèses exprimant qu'elle n'est pas trop grande, S_n converge p.s. C'est le cas, par exemple, si :

- $\sup_n E\, S_n^2 < \infty$, ou même si
- $\sup_n E\, S_n^- < \infty$ (qui inclut notamment le cas de toutes les martingales positives)

Si de plus, la famille (S_n) est équiintégrable, la limite p.s., S_∞, vérifie $E(S_\infty / \mathcal{F}_n) = S_n$ p.s., on dit que la martingale est "fermée".

Les martingales apparaissent comme la notion n° = 1 pour les convergences presque sûres.

Les inégalités

Les théorèmes de convergence reposent en fait sur des inégalités "maximales" dont l'intérêt est d'ailleurs capital en analyse pour l'étude des fonctions harmoniques. En voici un court échantillon. S_n désigne une martingale, λ un nombre positif quelconque.

.../...

- généralisation de l'inégalité de Bienaymé-Tchebycheff :

$$\mathbb{P}\,[\max_{1\leq k\leq n} |S_k| \geq \lambda\,] \leq \frac{1}{\lambda^2} ES_n^2$$

$$\mathbb{P}\,[\max_{1\leq k\leq n} S_k \geq \lambda\,] \leq \frac{1}{\lambda}\,[ES_n^- + ES_1]$$

- l'inégalité de Doob (*)

Soit $[a,b]$ un intervalle. Notons $N = N(a,b)$ le nombre (aléatoire) de passages en montant de la suite (S_n) par dessus $[a,b]$, c'est-à-dire le nombre d'indices $j = j(\omega)$ tels que

$S_{j-1} < a$ et $S_j > b$, alors :

$$E(N) \leq \frac{1}{b-a} E(|S_n| + |a|)$$

- et bien d'autres...

Martingales à temps continu

Les exemples-types de martingales à temps continu sont le mouvement brownien, le processus de Poisson "recentré". D'autre part, si Z_t est un processus de Markov et si h est une fonction harmonique, alors $h(Z_t)$ est une martingale. Lorsque t tend vers l'infini, certaines propriétés de convergence sont analogues à celles du cas discret.

D'autre part, sous des hypothèses de séparabilité, les trajectoires des martingales à temps continu ont des propriétés de régularité remarquables (existence presque partout de limites à gauche et à droite, ect.) sur lesquels nous n'insisterons pas.

Enfin, les martingales servent de base aux théories de l'intégrale stochastique et des équations différentielles stochastiques.

.../...

(*) En fait Doob a démontré de nombreuses inégalités de martingales et l'expression "inégalité de Doob" ne désigne pas la même chose chez tous les auteurs.

Remarques :

Nous arrêterons ici ce rappel semi-informel de "ce qui traîne spontanément dans la tête du probabiliste de base des années 60 à propos des martingales". Voir par exemple [1] ou [2].

Tout le monde n'a certes pas le même jugement sur l'importance relative des divers théorèmes, certains sont plus portés sur l'aspect convergence presque sûre, d'autres sur les inégalités, d'autres encore sur le "calcul stochastique". En tout cas, dans les années 60, le problème de l'extension du "théorème limite central", et d'autres résultats connexes, aux martingales rencontre fort peu d'écho (il y aura évolution dans les années 70, cf. [3]).

Qu'en était-il en 1930 ?

b) Les martingales dans le "folklore" des mathématiciens vers 1920-1930

La "préhistoire" n'est pas ici notre objectif. Nous nous contenterons de quelques observations tirées de l'exposé de B. Bru [4].

- L'idée qu'il est impossible de trouver un système de jeu permettant de gagner lors d'une succession de parties équitables est très ancienne : B. Bru l'a trouvée même dans l'Antiquité, chez Xénophon !

L'astuce qui consiste, dans un jeu de pile ou face équitable, à doubler sa mise après chaque perte et à s'arrêter à la première victoire est également connue depuis longtemps et, en tout cas, nommée "martingale" au moins depuis le 18e siècle (*) (voir par exemple Condorcet : "Eléments du calcul des probabilités", Royez 1805, p. 119).

Les traités sur les jeux de casinos abondent au 19e siècle et décrivent diverses méthodes analogues, laissant entrendre plus ou moins clairement qu'on ne peut être sûr de gagner ainsi que si l'on dispose d'une fortune infinie.

.../...

(*) Le mot "martingale" lui-même est plus ancien et employé en des sens divers ; son étymologie est controversée (voir les dictionnaires).

- D'un point de vue plus mathématique, tous "les grands noms" du calcul des probabilités ont étudié des problèmes tels que "la durée du jeu", "la ruine du joueur"..., où l'on peut trouver ce que nous appelerions aujourd'hui des techniques ou méthodes de martingale.

Parmi les plus explicites, citons :

- de Moivre : "De Mensura Sortis"
- Ampère : "Considérations sur la théorie mathématique des jeux" (1802) B.N.
- Ch. Babbage : "An examination of some questions connected with games of chance", Trans. Royal Soc. Edinburgh IX (1820), p. 153-177

Lacroix utilise le mot dans son Traité de Calcul des Probabilités.
On pourrait multiplier les citations, où sont formalisées, sur des jeux précis, les idées du caractère inévitable de la ruine du joueur ayant une fortune finie ou du caractère illusoire de la variation des mises. Mais, même si les principes sont énoncés philosophiquement de manière générale, il ne peut y avoir alors d'énoncés mathématiques généraux : rappelons que ne sont dégagées à ces époques ni la notion de variable aléatoire, ni celle de convergence presque sûre, ni celle d'espérance conditionnelle. (Pour mieux comprendre pourquoi, voir [6]).

Au début du 20e siècle, malgré les travaux de Bachelier(*), Markoff, Borel, Cantelli...., le concept de "martingale" en reste à peu près à ce point, dans la culture des probabilistes.

c) L'ambiance probabiliste des années 1920-1930

Ici encore, il ne s'agit que de points de repère pour la compréhension de la suite et non d'une analyse historique.
H. Cramér [5] définit la décennie 1920-1929 comme une décennie de préparation, et 1930-1939 comme celle des "grands changements".
Le "calcul des probabilités" n'est pas toujours considéré alors comme faisant partie intégrante des mathématiques ; ses charmes un peu flous sont diversement appréciés selon les pays, selon les convictions philosophiques, selon les circonstances de leur utilisation [6] : il y a peu de temps encore, la référence (philosophique et mathématique, contestée ou non) c'était Laplace.

.../...

(*) qui mériteraient une étude à part.

Le mot "probabiliste" ne semble guère utilisé, les mathématiciens qui s'adonnent aux probabilités ont en général une longue carrière d'analyste derrière eux.

- Les traditions nationales

En France, le calcul des probabilités, montré du doigt comme scandaleux au 19e siècle par Auguste Comte, Cauchy..., reste marginal, malgré les Traités de Bertrand, Poincaré, Borel ; jusque vers 1930 on ignore jusqu'au nom de Markov ; quant à la statistique mathématique, elle reste dans la vieille tradition.

En Angleterre, au contraire l'école statistique mathématique est florissante et originale depuis Galton, Pearson, Fisher..., mais la théorie des probabilités y intéresse fort peu de monde.

En Allemagne et dans les pays nordiques, l' "école continentale de statistique", occupée par l'actuariat, la démographie, l'astronomie n'a pas procédé au renouvellement britannique ; il n'empêche que ses apports aujourd'hui méprisés ou oubliés ne sont pas négligeables. Pour une large part les probabilités n'y ont de sens qu'au vu de problèmes pratiques : la théorie de Von Mises est déjà en germe dans cette école.

C'est en Russie, puis en U.R.S.S., que la recherche théorique en probabilités a une avance considérable, d'abord avec Tchebychev puis ses élèves Markov, Lyapounov, ensuite avec la jeune génération d'après la Révolution d'Octobre (Kolmogorov, Khintchine...).

On oublie souvent à tort les travaux des Italiens (sauf Cantelli) très en pointe en 1900 et 1920, ainsi que les tentatives "abstraites" en Pologne (Łomnicki, Steinhaus). Quant à la faiblesse des Etats-Unis elle tranche par rapport à ce que sera l'école américaine après la guerre.

.../...

Les grandes préoccupations (explicites) des "probabilistes", du moins conformément aux apparences, jusque vers 1935, voir par exemple [5].

Les théorèmes limites pour sommes de variables aléatoires indépendantes et quelques travaux pionniers pour se dégager petit à petit de cette condition d'indépendance.
Points de repère : loi forte des grands nombres (LFGN) et la loi du logarithme itéré (LLI) : Cantelli (1917), puis Kolmogoroff et Khintchine(1924-29) diverses versions du théorème limite central (TLC) et de sa réciproque : Lindeberg et P. Lévy (1922), Lévy et Feller (1935) ... ; les travaux de Markov à partir de 1906, puis ceux de S.Bernstein à partir de 1917-1918.

Les processus stochastiques : "de Markov", à temps discret puis continu, jugés fondamentaux à partir de la fin des années 20 : voir notamment les articles d'Hostinsky et surtout Kolmogorov (1931) ; le mouvement brownien : Wiener(1923) ; les processus à accroissements indépendants : P. Lévy et Khintchine (1934-1937) ; les processus stationnaires : Khintchine (1934).

Les fondements :discussions philosophiques et mathématiques sur les bases de la théorie des probabilistes sur son champ d' "applications", sur le lien entre probabilité et fréquence, sur l'axiomatique : S. Bernstein (1917), Von Mises (1919, →), Łomnicki et Steinhaus (1923), Kolmogorov (1933), etc.
A noter d'ailleurs que l'apport des "Grundbegriffe" de Kolmogorov n'est pas pour l'essentiel ce qu'en dit la légende : en fait les liens entre probabilités et théorie de la mesure sont familiers depuis longtemps aux probabilistes.
La nouveauté radicale vient plutôt de la place de l'indépendance dans cette axiomatique et dans la théorie des probabilités, de l'espérance conditionnelle, de la possibilité vraiment opératoire d'utiliser les probabilités dans les espaces de fonctions, donc d'une gestion rigoureuse des processus stochastiques, etc. (C'est d'ailleurs ce que Kolmogorov dit lui-même dans l'introduction de ce livre, qui marque certes une coupure dans l'histoire des probabilités).

.../...

Les liens entre probabilités et statistiques:

Avec l'estimation, les tests... le besoin est ressenti de mieux faire le lien entre les théorèmes limites du calcul des probabilités et les méthodes statistiques : notons par exemple dans ce cadre les théorèmes dits de Glivenko- Cantelli et de Kolmogorov-Smirnov.

Bien sûr, les considérations précédentes cumulent tous les défauts des notices historiques faites par les mathématiciens (attention portée exclusivement au développement interne des théories, jugement à l'aune de nos habitudes actuelles, anachronismes dans le langage etc.) ; mais rappelons qu'elles ne sont là que pour faciliter la lecture des probabilistes en fonction de leurs points de vue habituels.

d) "Martingale" : une notion qui "monte" de partout

Continuant sur le même ton, on pourrait dire que dans les années 20 et 30, petit à petit, des morceaux du "corpus classique" sont présents implicitement de plus en plus dans les travaux des analystes et des probabilistes. Notons, sans le détailler, un échantillon de ces "crypto-martingales" :

- la formalisation de l'impossibilité d'un système de jeu consistant à choisir, pile ou face, les coups où l'on mise et ceux où l'on ne mise pas (Von Mises, Wald...) : voir le chapitre IV.

- le sentiment d'une forte analogie, en même temps que de différences, entre les problèmes de convergence des séries de fonctions orthogonales, en particulier des séries de Fourier, et les théorèmes limites pour les séries de variables aléatoires indépendantes et centrées. Les mêmes mathématiciens démontrent des résultats voisins des deux types au moyen, souvent, de ce qu'on appellerait aujourd'hui des inégalités de martingales avant la lettre : ils ont clairement conscience de la nécessité de dégager des "notions intermédiaires" entre suite orthogonale et suite indépendante centrée : c'est très net chez Kolmogorov, P. Lévy, Kač, Steinhaus, Marcinkiewicz, Zygmund, Paley, Wiener, etc.

.../...

Mais ce n'est pas dit comme ça : on ne trouvera guère, même s'ils en sont tout à fait pénétrés, un énoncé du type suivant : "Soit $\Omega = [0,2\pi]$ muni de la probabilité uniforme, alors $\{\sin nx\}$ est une suite de v.a. orthogonales, mais non indépendantes".

- l'étude du développement en fractions continues d'un nombre pris "au hasard" dans $[0,1]$, à partir de motivations diverses, dont les petites perturbations des planètes [7] , [8]; voir le chapitre III.

- les problèmes de dérivation et le besoin d'une théorie rigoureuse de l'ingration dans les espaces de dimension infinie, d'extensions "propres" du théorème de Fubini qui reviennent (cf. chapitre III) à conditionner une v.a. par une suite croissante de tribus, dirions-nous aujourd'hui.

- les études sur le rapport de vraisemblance.

e) Présentation de la suite

Par souci de clarté, et de simplicité, nous centrerons la suite sur les contributions majeures de 4 auteurs intervenus directement dans l'élaboration de la théorie des martingales :

S. Bernstein, P. Lévy, J. Ville, J. L. Doob ; nous nous contenterons de décrire leurs travaux originaux et d'y ajouter quelques commentaires limités sur leurs circonstances.

A peu de choseprès, ces auteurs se connaissent, se citent : les choses sont claires ; il y a quand même quelques "querelles de priorité" (polies), comme disent les historiens des sciences. Par exemple certains travaux de S. Bernstein écrits en russe et non traduits, restent mal connus ou sous-estimés surtout à l'approche et au début de la guerre.

Pour harmoniser les notations, nous désignerons en général par $u_1, \ldots, u_n, \ldots$ une suite de v.a. vérifiant des conditions diverses d'indépendance ou de presque-indépendance et par $S_n = u_1 + \ldots + u_n$ la suite des sommes partielles.

.../...

Autant que possible nous énoncerons les théorèmes principaux à la fois dans le langage de l'auteur et dans un langage moderne.

Nous indiquerons, autant que faire se peut, la date de réception des articles par la revue de parution ainsi que l'auteur du résumé dans Zentralblatt (ZB), à partir de 1931, et dans les Maths. Reviews (MR) à partir de 1940.

II LES ARTICLES DE S. BERNSTEIN

a) Liste des travaux

En raison du caractère difficilement accessible de certains papiers originaux de S.Bernstein, nous donnons ici une traduction de leurs titres et revues de parution, établie à partir du tome IV des Oeuvres (en russe). Ceux qui nous intéressent le plus directement sont numérotés 4, 8, 22, 23, 24, 25, et 27, cités [B4], [B8] etc. dans la suite.

Ils sont d'ailleurs écrits en français, sauf [B25] qui contient cependant un résumé

[B22]	ZB 18	p. 032 (Khintchine)	
[B23]	ZB 22	p. 243 (de Finetti)	
[B24]	ZB 22	p. 061 (Hostinsky)	; MR 1 n° 340 (Doob)
[B25]	ZB 24	p. 263 (Hostinsky)	; MR 2 n° 107 (Doob)
[B27]			; MR 6 n° 88 (Feller)(*)

Articles historiques :

[B] Divers articles sur S.N. Berstein dans "Russian Math Surveys 24 (1969)

(*) A partir de la fin des années 30, Zentralblatt est tombé à peu près directement sous la coupe des nazis, certains de ses animateurs (comme Feller) qui ont d'ailleurs émigré ont fondé les Mathematical Reviews aux Etats-Unis, et l'intérêt des résumés d'ailleurs très irréguliers du Zentralblatt durant la guerre a beaucoup baissé.

.../...

OEUVRES DE S.N. BERNSTEIN
TOME 4 : PROBABILITES, STATISTIQUE MATHEMATIQUE
1911-1946

.../...

.../...

P. 276-285 — 17. "Sur les chaînes de Markov linéaires quasi-continues" Dkl. A.N. SSSR 2 n° 1 (1934), p. 1-9 et p. 361-365

P. 286-290 — 18. "Sur les diffusions avec absorption" Dkl. A.N. SSSR 1 (1934), p. 230-233

P. 291-315 — 19. "Principes de la théorie des équations différentielles stochastiques", Travaux de l'Inst. Phys.-Math. Steklov 5 (1934), p. 95-124 (Trudy ...)

P. 316-321 — 20. "Sur l'espérance mathématique des premiers ouvriers d'une unité pour un processus de production complexe" Ougol n° 117 (1935), p. 109-111

P. 322-330 — 21. "Détermination d'une limite inférieure de la dispersion des sommes de grandeurs liées en chaîne singulière", Math. Sbornik, 1 (43) : 1 (1936), p. 29-37

P. 331-333 — 22. "Sur quelques modifications de l'inégalité de Tchebychev" Dokl. A.N. SSSR 17 n° 6 (1937), p. 279-282 (parvenu le 14 Novembre 1937).

P. 334-357 — 23. "Equations différentielles stochastiques", Act. Sc. Ind. n° 738, p. 5-31 (1938)

P. 358-363 — 24. "Quelques remarques concernant le théorème limite de Lyapounoff" Dokl. A.N. SSSR 24 (1939), p. 3-7. (parvenu le 11 juin 1939).

P. 364-376 — 25. "Nouvelles applications des grandeurs aléatoires presqu'-indépendantes", Izv. A.N. SSSR, série math. 4 (1940), p. 137-150 (en russe, résumé en français) (parvenu le 5 février 1940).

.../...

p. 377-379 — 26. "Sur un problème du schéma des urnes à composition variable" Dokl. A.N. SSSR 28 n° 1 (1940), p. 5-7.

p. 380-385 — 27. "Sur les sommes de quantités dépendantes liées, de classes (A,N) et (B,N)" Dokl. A.N. SSSR 32 n° 5 (1941), p. 303-307 (parvenu le 31 mai 1941).

p. 386-393 — 28. "Sur les probabilités 'de confiance' de Fisher" Izv. A.N. SSSR, série math. 5 (1941), p. 85-93.

p. 394-395 — 29. "Sur une propriété caractéristique de la loi de Gauss" Travaux de l'Institut Polytechnique de Léningrad, n°3 (1941) p. 21-22 (Trudy...).

p. 396-408 — 30. "Retour au problème de l'exactitude de la formule de Laplace" Izv. A.N. SSSR, série math 7 (1943), p. 3-14. (en russe, résumé en français).

p. 409-433 — 31. "Sur les travaux de Tchebychev en théorie des probabilités" in "L'héritage scientifique de Tchebychev" t. 1 (1945), p. 43-68.

p. 434-447 — 32. "Sur le théorème limite de la théorie des probabilités" Izv. NII mat. i mekl. pri Tomski gas. univ. 3, 1 (1946), p. 174-189.

p. 448-454 — 33. "Un théorème réciproque du théorème de Laplace et sa généralisation" in"théorie des probabilités", 4e édition (1946), p. 458-464.

p. 455-483 — 34. "Classification des chaînes de Markov et de leurs matrices" in "Théorie des probabilités", 4e édition (1946), p. 203-213 pour l'introduction, et p. 465-484, pour les § 1-6.

.../...

P. 484-542	35. "<u>Equations stochastiques aux différences finies et équations différentielles stochastiques</u>" in "Théorie des probabilités", 4e édition (1946), p. 485-546
P. 543-575	<u>Commentaires</u> :
P. 543	sur le 1 (Prokhorov)
P. 543-544	sur le 5 (")
P. 544-546	sur le 7 (Sarmanov)
P. 547	sur le 8 (Sapogov)
P. 548	sur le 9 (")
P. 548-549	sur le 10 (")
P. 549-551	sur le 11 (Sarmanov)
P. 551-555	sur le 14 (")
P. 555-556	sur le 16 (Petrov)
P. 556	sur le 19 (Bernstein)
P. 556-557	sur le 20 (Petrov)
P. 557-558	sur le 21 (Linnik)
P. 558-560	sur le 22 (Zolotarev)
P. 560-565	sur le 23 (Blumenfeld-Bernstein)
P. 565-566	sur le 27 (Sarmanov)
P. 566-569	sur le 28 (Bolchev)
P. 569	sur le 29 (Linnick)
P. 570-574	sur le 30 (Prokhorov)
P. 574-575	sur le 34 (Sarymsakov)
P. 575-576	Table des matières
collé	errata

.../...

b) <u>Brèves remarques biographiques</u>

S. Bernstein(1880-1968) fait ses études, pour l'essentiel, à Paris, sous la direction de Picard et Hadamard, et y soutient sa thèse en 1904 sur le 9e problème de Hilbert (toutes les solutions de problèmes variationnels réguliers analytiques sont analytiques).

Retourné en Russie tsariste, victime de tracasseries diverses, il ne peut soutenir sa dissertation doctorale à Kharkov qu'en 1913, et n'a alors aucun contact avec l'école de Saint-Pétesbourg (Markoff, Liapounoff...), d'ailleurs il est même, au début, sceptique sur leurs travaux, pourtant son style et ses thèmes s'en rapprochent : "chaque papier de Bernstein, comme chaque papier des fondateurs de l'école de Saint-Pétersbourg, commençait par un problème spécifique difficile, demandant l'invention de nouvelles méthodes pour sa solution" [B].

c) <u>Description des articles</u>

c_1) <u>ceux des années 10-20</u>

En tout cas certains de ses travaux probabilistes <u>apparaissent</u> comme une continuation de l'oeuvre de Markoff et de Liapounoff, en particulier le célèbre article [B8], dont un résumé a paru en 1922 [B4] :

Je résume dans ces quelques pages, en omettant des démonstrations, mon étude (qui date de l'année 1917—1918) sur le théorème limite du calcul des probabilités ou sur les conditions de l'applicabilité de la loi de Gauss.(*) Le problème consiste dans la recherche des conditions pour que, S_n désignant une quantité dépendant de n dont l'espérance mathématique est nulle, la probabilité de l'inégalité

$$t_0\sqrt{2B_n} < S_n < t_1\sqrt{2B_n},$$

où B_n = Esp. Math. (S_n^2), ait pour limite, lorsque n croît indéfiniment,

$$\frac{1}{\sqrt{\pi}}\int_{t_0}^{t_1} e^{-t^2}\,dt.$$

Sans restreindre la généralité on peut poser $S_n = u_1^{(n)} + u_2^{(n)} + \ldots + u_n^{(n)}$. Dans la suite pour abréger l'écriture j'omettrai l'indice supérieur dans le terme $u_k^{(n)}$, en écrivant simplement u_k; mais pour que les énoncés des propositions qui suivent ne soulèvent pas de malentendus et soient compris dans toute leur généralité il faut bien se rappeler que nous n'admettons pas, en général, l'existence de la relation $S_n - S_{n-1} = u_n$.

(*) Notons incidemment ceci : ce résultat, appelé souvent alors "théorème de Laplace" ou "théorème de Liapounoff" est aujourd'hui connu sous le nom de "théorème limite central". Précisons que c'est le théorème (et non la limite) qui est central, ce que la traduction anglaise "central limit theorem" ne laisse pas voir. Cette expression (introduite par G. Polyá en 1920, dans "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung..." Math. Z. <u>8</u>, p. 171-181) signifie littéralement que ce théorème- limite (Grenzwertzatz) est central dans (au centre de) la théorie des probabilités. Elle n'est adoptée massivement que depuis la dernière guerre.

S. Bernstein explicite peu ses motivations, sauf à propos de la génétique [B8] § 16, p. 40-41. La consultation de ses archives serait sans doute éclairante.

Disons simplement ici que l'extension du théorème limite du calcul des probabilités à des quantités non indépendantes a été faite par Markoff, sous des hypothèses diverses, qui sont loin de se limiter au seul cas des chaînes dites de Markoff. Le travail de S. Bernstein en est une généralisation : il repose sur le lemme suivant, p. 238 de [B4].

> Lemme fondamental. *Si quel que soit l'ensemble de valeurs connues de $u_1, u_2, \ldots, u_{i-1}$, à l'exception d'un ensemble dont la probabilité est ε_i, les écarts que reçoivent les espérances mathématiques de u_i et u_i^2, respectivement, ne dépassent pas α_i et β_i, si de plus l'espérance mathématique de $|u_i|^3$ reste inférieure à c_i, le théorème limite est applicable à la somme S_n, pourvu que $\frac{\sum_1^n \alpha_i}{\sqrt{B_n}}$, $\frac{\sum_1^n \beta_i}{B_n}$, $\frac{\sum_1^n c_i}{\sqrt{B_n^3}}$, $\sum_1^n \varepsilon_i$ tendent vers 0 avec $\frac{1}{n}$.*

En termes plus faciles à comprendre pour les mathématiciens d'aujourd'hui (*), la première partie de l'énoncé signifie que :

$$\left.\begin{array}{ll} |E_{(i-1)}u_i - Eu_i| & \leq \alpha_i \\ |E_{(i-1)}u_i^2 - Eu_i^2| & \leq \beta_i \end{array}\right\} \text{ sur un ensemble de probabilité supérieure à } 1 - \varepsilon_i$$

Quant à la condition $\dfrac{\sum_1^n c_i}{B_n^{3/2}} \to 0$, où $c_i = \max E_{(i-1)} |u_i|^3$, nous l'appellerons pour des raisons évidentes : "condition de Liapounoff renforcée".

La première inégalité étant évidemment vérifiée si $E_{(i-1)} u_i = 0$, ce résultat contient tout simplement (en employant bien sûr un anachronisme) un théorème limite central pour martingales qui s'énonce ainsi (faisons $\varepsilon_i = 0$) :

(*) Nous noterons indifféremment $E_{(i-1)} u_i$ ou $E(u_i / u_1, \ldots, u_{i-1})$

Si S_n est une martingale ($S_n = u_1 + \ldots + u_n$) et si u_n vérifie la "condition de Liapounoff renforcée", une condition suffisante pour que $\frac{S_n}{\sqrt{B_n}}$ converge vers une loi gaussienne réduite est que les variances conditionnelles $E_{(i-1)}u_i^2$ soient presque non aléatoires au sens suivant :

$$\frac{1}{B_n} \sum_{i=1}^n \max | E_{(i-1)} u_i^2 - Eu_i^2 | \to 0$$

Donnons rapidement une idée de la démonstration de ce lemme en supposant $\varepsilon_i = 0$, $E_{(i-1)} u_i = 0$, pour simplifier : [B8] , p. 21-24

S. Bernstein, suivant Liapounoff, évalue la fonction caractéristique de $\frac{S_n}{\sqrt{B_n}}$

Notant y_k $(= y_{k,n}) = \dfrac{u_k}{\sqrt{B_n}}$, il s'agit de calculer $E(e^{i\xi \sum_1^n y_k})$, pour $m \leq n$

Pour cela, lorsque $|\xi| < N$, on écrit

$$\mathbb{E}(e^{i\xi y_k} | y_1, \ldots, y_{k-1}) = 1 + i\xi\, E(y_k | y_1, \ldots, y_{k-1}) - \frac{\xi^2}{2} E_{y_k^2}$$
$$+ \frac{\xi^2}{2} \{E_{y_k^2} - E_{(k-1)}\, y_k^2\} \quad + \text{termes d'ordre 3 environ}$$

Les hypothèses donnent immédiatement que $E(e^{i\xi y_k} | y_1, \ldots, y_{k-1})$ est peu différent de $1 - \frac{\xi^2}{2} E_{y_k^2}$

Par récurrence sur m, et grâce à un petit calcul sur les espérances conditionnelles on parvient au résultat.

<u>Remarques</u> : 1) S'il traite en fait un cas plus général que les martingales, S. Bernstein n'isole pas cette notion : son but est soumis à la recherche d'extensions variées des conditions de validité du théorème limite maniables pour la modélisation de certains phénomènes réels qu'il a en vue, et notamment s'appliquant aux chaînes de Markoff. Voici selon quel principe il utilise son lemme :

.../...

> Si les quantités u_i satisfont aux conditions du lemme, nous dirons, pour abréger, que les quantités u_i sont *presque indépendantes*. Ceci posé, si l'on a une somme quelconque $S_n = u_1 + u_2 + \ldots + u_n$, on pourra toujours affirmer l'applicabilité à cette somme du théorème limite, si l'on parvient à rassembler les termes de cette somme en $2l$ groupes: $y_1 + z_1 + y_2 + z_2 \ldots + y_l + z_l$ de telle sorte que $y_1, y_2, \ldots, y_l$ soient presque indépendantes, tandis qu'en même temps l'ordre de croissance de Esp. Math. $(z_1 + z_2 + \ldots + z_l)^2$ soit inférieur à celui de $B_n =$ Esp. Math. $(u_1 + u_2 + \ldots + u_n)^2$. En effet, il est facile de montrer alors que
>
> $$\frac{\text{E. M.}(y_1 + y_2 + \ldots + y_l)^2}{\text{E. M.}(u_1 + u_2 + \ldots + u_n)^2} \to 1$$
>
> et d'en conclure que le théorème limite étant applicable à la somme $y_1 + y_2 + \ldots + y_l$, doit s'appliquer également à la somme S_n.

2) S. Bernstein développe un exemple qui est, lui, une martingale- type puisqu'il s'agit d'un jeu équitable à enjeux variables : expliquons-le ici sur un cas simplifié du jeu de pile ou face symétrique.

Au premier coup le joueur mise 1 F (ou 1 rouble) sur pile. Soit x_1 son gain ($= \pm 1$)

Au deuxième coup, il mise $(1 + \varepsilon x_1)$F. Son gain est alors $x_2 = \pm(1 + \varepsilon x_1)$

Au troisième coup, il mise $(1 + \varepsilon x_2 + \frac{\varepsilon}{2} x_1)$ ...

Au $(i+1)^e$ coup, son gain x_{i+1} sera donc $(1+\varepsilon x_i + \frac{\varepsilon}{2} x_{i-1} + \ldots + \frac{\varepsilon}{2^{i-1}} x_1)$ avec probabilité 1/2, ou l'opposé de cette quantité avec même probabilité. Il résulte immédiatement de la définition que $E_{(i-1)} x_i = 0$. A noter que si $\varepsilon < \frac{1}{2}$, le gain est toujours compris entre 0 et 2 si on tire pile, de même pour la perte si on tire face : donc par rapport au jeu de pile ou face ordinaire, la modification apportée ici a "pour seul effet d'augmenter ou de diminuer dans une certaine proportion [l'enjeu] du joueur suivant qu'il ait été précédemment favorisé ou non par la fortune, l'influence de ces résultats antérieurs diminue en progression géométrique à mesure que la succession des jeux avance".

.../...

S. Bernstein montre alors que, dans ce cas intuitivement asymptotiquement indépendant, la convergence vers la loi de Gauss subsiste, par application de ses théorèmes généraux.

3) La fin de l'article, à partir de la p. 46, qui traite de la "corrélation normale", peut être comprise, en un certain sens, comme renfermant des théorèmes limites pour des martingales à valeurs dans $\mathbb{R}^d$, du moins avec notre façon actuelle de voir les choses.

4) Cet article, lu par tous les probabilistes de l'époque (y compris P. Lévy), a sur eux un profond impact. Il influence d'ailleurs, souvent à l'insu de leurs auteurs, de nombreux travaux actuels, cf. [3].

D'autre part, notons qu'il n'y a dans ce travail ni théorème de convergence presque sûre, ni étude sur les processus stochastiques au sens où on l'entend maintenant.

5) Comme nous l'avons indiqué dans l'introduction, l'idée d'utiliser des espérances conditionnelles nulles germe à cette époque. On lit ainsi dans une note de Kolmogoroff de 1929 [9], peu de temps donc après la parution de l'article de S. Bernstein (que Kolmogoroff connaît) :

"Considérons une suite de nombres réels $X_1, \ldots, X_n, \ldots$, où la valeur de X_n dépend du résultat des n épreuves consécutives $\mathcal{E}_1^{(n)}, \mathcal{E}_2^{(n)}, \ldots, \mathcal{E}_n^{(n)}$ (Nous ne supposons nullement que ces n épreuves soient mutuellement indépendantes) Il désigne par $E_k(\gamma)$ l'espérance mathématique d'une v.a. γ "dans l'hypothèse des résultats des k premières épreuves $\mathcal{E}_1^{(n)}, \mathcal{E}_2^{(n)}, \ldots, \mathcal{E}_k^{(n)}$ connus" et pose $Z_{nk} = E_k(X_n) - E_{k-1}(X_n)$. Il remarque alors (p. 472) que $E_{k-1}(Z_{nk}) = 0$ et que, par suite, les v.a. Z_{nk} sont orthogonales.

.../...

L'écart-type de Z_{nk}, à savoir $[E(Z_{nk}^2)]^{1/2}$ "est la moyenne quadratique de l'inflexion de l'espérance mathématique de X_n quand le résultat de l'événement $\mathcal{E}_k^{(n)}$ devient connu. [C'est donc] la mesure très naturelle de la dépendance de X_n du résultat de l'épreuve $\mathcal{E}_k^{(n)}$. (Cela suppose que les événements $\mathcal{E}_k^{(n)}$ se produisent dans l'ordre de leurs indices k)".
Ces considérations, d'ailleurs reprises au chapitre VI des "Grundbegriffe", débouchent sur une loi faible des grands nombres.
Il ne semble pas que Kolmogoroff pousse plus loin la réflexion à ce sujet.(*)

6) L'intervention de S. Bernstein au Congrès International de Zürich en 1932 (**), intitulée "Sur les liaisons entre les grandeurs aléatoires" est très importante pour la compréhension de l'oeuvre de Bernstein, mais il n'y a guère de traces de "martingales avant la lettre". Ce n'est à notre connaissance qu'en 1937 que S. Bernstein reviendra explicitement sur cette question ; visiblement il ignore alors (probablement en raison de l'isolement relatif de l'U.R.S.S. dans les années 30) les travaux récents de P. Lévy : c'est ce que nous allons maintenant décrire.

c_2) <u>ceux des années 30-40</u>

α) Dans [B22], S. Bernstein se place dès la première phrase sous l'hypothèse $E_{(i-1)}\, u_i = 0$.
En notant comme d'habitude $S_n = u_1 + \dots + u_n$ et $B_n = ES_n^2$, il s'agit d'étendre à ce cas deux raffinements de l'inégalité de Tchebycheff :

* "l'inégalité maximale de Kolmogoroff", démontrée par Kolmogoroff en 1928 dans le cas où les V.A.(u_i) sont indépendantes et centrées :
La probabilité que toutes les inégalités $|S_i| \leq t\sqrt{B_n}$ $(i=1,\dots,n)$ seront réalisées simultanément est supérieure à $1 - \frac{1}{t^2}$

.../...

(*) L'intervention de Kolmogoroff au Ier Congrès des mathématiciens hongrois en 1950 (comptes-rendus du Congrès - Budapest 1952, p. 367-376) semble montrer que celui-ci a pris connaissance très tard des travaux sur les martingales.

(**) Bernstein n'ayant pu se rendre à Zürich, cette intervention en français est lue par B. Hostinsky.

* "l'inégalité de Bernstein", démontrée par lui-même en 1924 [B5] dans le cas des v.a. u_i indépendantes et centrées, et exprimant ceci : si les u_i ont des moments de tous ordres ℓ tels que ces moments Eu_i^ℓ croissent modérément lorsque ℓ tend vers l'infini, alors la probabilité pour que toutes les inégalités $|S_i| \leq t\sqrt{B_n}$ $(i=1,\dots,n)$ soient réalisées simultanément est cette fois supérieure à $1 - e^{-t^2/4}$ si t n'est pas trop grand (ce qui est, lorsque ces conditions de moment sont réalisées, beaucoup plus précis que l'inégalité de Tchebycheff ou de Kolmogoroff).

Dans le cas qui nous intéresse, où $E_{(i-1)}\, u_i = 0$, sans supposer l'indépendance, le théorème de S. Bernstein s'énonce ainsi :

$$\text{si} \quad \frac{E_{(k-1)} u_k^2}{Eu_k^2} \leq R_k \quad \text{(non aléatoire)}$$

S'il existe un nombre H tel que

$$E_{(k-1)}\, u_k^\ell \leq \frac{1}{2}\; E_{(k-1)}\, u_k^2\, H^{1-2}\, \ell\,!,$$

la probabilité de réalisation simultanée de toutes les inégalités $|S_i| \leq 2t\sqrt{B'_n}$ (où $B'_n = \sum_1^n R_i\, Eu_i^2$) est supérieure à $1 - e^{-t^2}$ pourvu que $0 < t \leq \frac{\sqrt{B'_n}}{2H}$.

S. Bernstein remarque que la démonstration de l'inégalité maximale de Kolmogoroff, dans le cas où $E_{(i-1)}\, u_i = 0$, "est plus simple et entièrement analogue" à celle de l'autre inégalité. Il s'agit en fait d'une simple adaptation de la démonstration de Kolmogoroff qui revient, dirions-nous aujourd'hui (mais on était bien loin de dire cela) à introduire, comme dans le cas indépendant, un temps d'arrêt.

.../...

L'argument supplémentaire à utiliser pour démontrer l'inégalité de Bernstein consiste à appliquer la même méthode à la suite de v.a. $e^{\alpha S_i}$ (il s'agit en fait d'un argument ancien qu'on trouve déjà chez de Moivre, et qui est utilisé de manière beaucoup plus systématique par Esscher en 1932 pour les assurances, par H. Cramér en 1938 pour la loi des grands écarts, puis par Wald dans l'analyse séquentielle, etc.)

β) L'article [B24] s'occupe de la convergence des moments dans le théorème limite central, mais uniquement dans le cas des variables indépendantes ; toutefois, avec un peu d'imagination, on peut penser, au vu de la remarque de la p. 7, que Bernstein estime qu'elle s'étendrait au cas presque indépendant (donc à celui des martingales) (?). Mais c'est probablement par une confusion bibliographique avec la note précédente que Hall et Heyde citent cet article comme un travail sur les martingales avant la lettre.
Notons au passage que la "mini-querelle de priorité", contenue dans cette remarque de la p. 7, ne cite que Feller (et non P. Lévy qui a montré la même chose que Feller au même moment), ce qui semble confirmer que S. Bernstein n'a pas lu les travaux de P. Lévy d'après 1934.

γ) Venons-en maintenant à l'article [B25] écrit en russe, avec résumé en français.
L'auteur y élargit la notion de presque indépendance. Pour le cas particulier qui nous intéresse ici (celui où $E_{(i-1)}u_i = 0$), il obtient que, sous la condition de Liapounoff simple (et non plus renforcée), il suffit, pour avoir la convergence vers la loi gaussienne réduite de $\frac{S_n}{\sqrt{B_n}}$, que

.../...

$$\frac{1}{B_n} \sum_1^n \left| E_{(i-1)}\, u_i^2 - E\, u_i^2 \right| \to 0 \text{ (il a enlevé le "max")}$$

S. Bernstein applique ce résultat à un problème du schéma des urnes à compostion variable.
Plus loin, supposant maintenant réalisée la condition de Liapounoff renforcée, il aboutit à une condition nécessaire et suffisante de convergence de $\frac{S_n}{\sqrt{B_n}}$ vers la loi gaussienne réduite, à savoir : que "la probabilité de l'inégalité

$$\left[\frac{\sum_1^n E_{i-1}\, u_i^2}{B_n} - 1\right] < \varepsilon$$

ait pour limite 1 quand $n \to \infty$, quel que soit $\varepsilon > 0$ donné"
Il montre aussi dans ce cas une CNS de convergence en loi de $\frac{S_n}{\sqrt{B_n}}$ vers une loi non précisée à priori, à savoir que "la probabilité de l'inégalité

$$\frac{\sum_1^n E_{i-1}\, u_i^2}{B_n} < \sigma \qquad (\sigma \geq 0)$$

tende vers une limite déterminée $h(\sigma)$".

Il précise alors la fonction de répartition $F(t)$ de la loi limite de $\frac{S_n}{\sqrt{B_n}}$:

$$F'(t) = \int_0^\infty \frac{e^{-t^2/2\sigma}}{\sqrt{2\pi\sigma}}\, dh(\sigma) \qquad t \geq 0,\ t \leq 0$$

$$F(+0) - F(-0) = h(+0)$$

N.B. : on comparera ces conditions à celles obtenues plus tôt par P. Lévy (mais que S. Bernstein ne connaît visiblement pas).

.../...

δ) Enfin, dans [B27] , rappelant que dès le début de ses travaux il a supposé que ses v.a. u_k pouvaient dépendre de n (comme nous l'avons vu plus haut), en posant $u_{k,n} = \frac{u_k}{\sqrt{\sum_1^n u_k^2}}$, il démontre toujours dans le cas $E_{(i-1)}\, u_i = 0$, que sous la condition de Liapounoff simple, on a :

$$P\left(\frac{S_n}{\sqrt{\sum_1^n u_i^2}} < t\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2}\, dt$$

Il note l'intérêt de ce théorème et de ses corollaires en statistiques.
Il s'intéresse aussi au cas où $E(u_k \mid u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n) = 0$

ε) Le tableau ne serait pas complet si l'on omettait de signaler que dans son papier du colloque de Genève de 1937 [B23], S. Bernstein utilise l'inégalité maximale de Kolmogoroff relative aux martingales, pour étendre ses résultats sur les équations différentielles stochastiques. Mais il n'y a évidemment aucune théorisation sur les liens entre martingales (S. Bernstein n'emploie d'ailleurs jamais le mot) et les équations différentielles stochastiques .

Pour conclure provisoirement, disons que S. Bernstein a tout à fait explicitement, et dès 1917-1918, un théorème limite central pour les martingales, mais que jusqu'à la fin des années 30, il ne dégage pas nettement ce concept.

.../...

III - LES TRAVAUX DE P. LEVY

a) Liste des publications

Livre et articles relatifs aux variables enchaînées :

[L1] : "Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue", Bull Soc. Math. France 57, 1929 p. 178-194 (envoyé en fait en janvier 1930)

[L2] : " L'addition des v.a. enchaînées et la loi de Gauss", Bull Soc. Math. France 62 (1934), Communications et Conférences, p. 42-43, 23 mai.

[L3] : "Propriétés asymptotiques des sommes de v.a. enchaînées" CRAS 199 (1934) Ier octobre, p. 627-629

[L4] : "CNS pour l'application asymptotique de la loi de Gauss à la somme d'un grand nombre de v.a. indépendantes ; extension au cas de variables enchaînées". C.R. des séances de la SMF 48-49, 28.11.1934.

[L5] : "Propriétés asymptotiques des sommes de v.a. enchaînées", Bull. Sci. Math. (2) 59 (1935), p. 84-96 et 109-128 (ZB 11-262 : Hostinsky).

[L6] : "Propriétés asymptotiques des sommes de v.a. indépendantes ou enchaînées", J. Math. Pures Appl. (9) 14, p. 347-402 (ZB 13.028 : Feller)

[L7] : "La LFGN pour les v.a. enchaînées" CRAS 201 (1935), p. 493-495 Errata 800 (26 août) (ZB 12.111 : Feller)

.../...

[L8] : "La LFGN pour les v.a. enchaînés" J Math. Pures Appl. (9) 15 (1936), p. 11-24 (ZB 13.273 : Khintchine)

[L9] : "Sur le développement en fraction continue d'un nombre choisi au hasard", Composition Math. 3 (1936), p. 286-303 (ZB 14.268 : Kolmogoroff)

[L10] : "Théorie de l'addition des variables aléatoires", Paris, Gauthier-Villars 1937 : 1ère édition ; 1954 : 2ème édition en particulier § 41, chapitres VIII et IX : (parution en janvier 1937) (rédaction au printemps et à l'été 1936) (ZB 16.170 : Khintchine)

Ouvrages et articles à caractère historique :

[L11] : "Notice sur les travaux scientifiques" juin 1935 + 3 suppléments Archives de la bibliothèque de Mathématiques, Université de Paris VI-VII. Jussieu.

[L12] : "Quelques aspects de la pensée d'un mathématicien", Albert-Blanchard (1970)

[L13] : Notice nécrologique : "Paul Lévy 1886-1971" par M. Loève, Ann. Proba 1 (1973), p. 1-18.

.../...

b) Sur la genèse de ses idées

P. Lévy écrit dans [L8], p. 11-12

L'idée qui est à la base de ces travaux, que j'ai indiquée pour la première fois en 1929 à propos d'une application à l'étude des fractions continues (¹), est que la plupart des théorèmes relatifs aux suites de variables aléatoires indépendantes les unes des autres peuvent s'étendre à une suite de variables enchaînées

$$u_1, \ u_2, \ \dots, \ u_n, \ \dots,$$

si l'on prend soin d'introduire, pour chacune de ces variables u_n, non sa loi de probabilité *a priori*, mais sa loi *a posteriori*, celle dont elle dépend lorsqu'on connait $u_1, u_2, \dots, u_{n-1}$, et qui, en pratique, caractérise les conditions de l'expérience qui doit déterminer u_n. Il est bien connu que, sans cette précaution, l'extension des théorèmes asymptotiques les plus simples est impossible; elle devient au contraire facile grâce à l'introduction de ces lois *a posteriori*.

L'application la plus simple de cette remarque conduit à penser que l'on a, sous des conditions assez peu restrictives, une bonne évaluation de la somme

$$S_n = u_1 + u_2 + \dots + u_n,$$

en remplaçant chaque terme u_ν, non par $\mathcal{E}\{u_\nu\}$, mais par $\mathcal{E}_{\nu-1}\{u_\nu\}$(¹). Sans doute objectera-t-on que la valeur approchée ainsi obtenue est une variable aléatoire, et n'a pas la valeur pratique d'une évaluation *a priori*. Mais en calcul des probabilités, du moins dans une théorie générale, on ne peut que préciser la relation probable entre la loi de probabilité et le résultat de l'expérience, entre la cause et l'effet; les énoncés obtenus pourront seulement conduire à des conclusions plus précises dans les applications particulières où l'on précisera la manière dont les conditions de chaque expérience dépendent des résultats des précédentes. L'application déjà mentionnée à l'étude des fractions continues suffit à montrer l'intérêt de cette méthode.

Il est d'ailleurs intéressant de citer exactement le passage de son article sur les fractions continues : [L1] où il parle "du théorème suivant du calcul des probabilités :

Dans une série illimitée d'expériences donnant à un événement A des probabilités

$$\alpha_1, \ \alpha_2, \ \dots, \ \alpha_n, \ \dots$$

sa fréquence au cours des n premières expériences diffère de la probabilité moyenne

$$\alpha_n' = \frac{\alpha_1 + \alpha_2 + \dots + \alpha_n}{n},$$

d'une quantité presque sûrement infiniment petite pour n infini, c'est-à-dire qu'elle tend vers zéro, sauf dans des cas dont la probabilité totale est inférieure à n'importe quel nombre positif donné.

> Il faut remarquer que cet énoncé ne suppose pas l'existence l'une limite pour z_n : il importe d'ailleurs peu que les probabilités considérées soient ou non indépendantes; si elles sont successives, chaque probabilité z_n étant appréciée au moment de l'expérience, compte tenu du résultat des expériences antérieures, le théorème s'applique sans difficulté

Toujours dans son article de 1936 sur la loi des grands nombres, P. Lévy ajoute en note p. 13 :

> Si je m'étais contenté d'un énoncé sans démonstration, c'était en partie pour ne pas interrompre par une trop longue digression un travail consacré à l'étude des fractions continues, en partie parce que, n'étant pas sûr d'avoir lu tous les travaux publiés sur la loi forte des grands nombres, je pensais qu'un résultat si simple pouvait être connu; je suis arrivé depuis à la conclusion qu'il était nouveau, et je ne crois pas que sa démonstration ait été publiée jusqu'ici.

Quelques commentaires sont nécessaires

1° - On doit d'abord revaloriser l'importance des fractions continues dans la motivation des probabilistes et des analystes de la fin du 19^e siècle et de la première moitié du 20^e : Stieltjes, Markov, Borel, Lévy, Khintchine...

2° - Le mémoire de 1929 de P.Lévy, ainsi que beaucoup d'autres de cette période, puisent leur inspiration de départ dans le célèbre article de Borel de 1909 [7] et dans la discussion qui s'est instaurée entre celui-ci et Félix Bernstein en 1911-1912 [8] , [10] :

Si ω est un nombre choisi "au hasard" entre 0 et 1, la suite $a_n(\omega)$ des "quotients incomplets" de son développement en fraction continue

$$\omega = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}}$$

.../...

est une suite de variables aléatoires non indépendantes (contrairement à la suite donnant le développement décimal), mais qu'on peut traiter à certains points de vue comme une suite de v.a. indépendantes.

Nous n'entrons pas ici dans les détails renvoyant à [11] , ou au chapitre IX de [L10].

3° - Dans la présentation de P. Lévy, le temps est orienté : losqu'il parle d'une suite (u_n) de variables enchaînées, cela signifie que celles-ci sont déterminées par des expériences successives et que la valeur de u_n peut dépendre du résultat des expériences précédentes $u_1, \dots, u_{n-1}$.

P. Lévy dira dans la 2e édition de son livre, à propos de la notation $x(t,\omega)$ pour les processus : "C'est une notation souvent commode, bien qu'elle donne comme un tout né en un instant ce qui, pour moi, est essentiellement un perpétuel devenir" (p.360), explicitant ainsi sa différence de sensibilité avec le point de vue de Doob.

4° - Je crois P. Lévy sincère lorsqu'il dit pourquoi il s'est contenté dans son article de 1929 d'un "énocé sans démonstration". Toutefois, il y a quelque chose qu'il n'avait pas saisi à l'époque : c'est l'idée claire que cette situation est à dégager et à encadrer parce qu'elle peut servir dans des problèmes très variés et même révolutionner la théorie des probabilités. D'ailleurs, la condition ($\mathcal{C}$) n'y est pas écrite noir sur blanc, même si elle est en germe. P. Lévy écrit aussi dans [L8] , p. 24 qu'à cette époque il ne connaissait pas l'inégalité maximale de Kolmogoroff. Il serait intéressant d'étudier de plus près comment ces idées ont mûri de 1929 à 1934.

.../...

c) Description des mémoires

Venons-en maintenant aux trois articles principaux, dont le contenu essentiel est élaboré aux cours de l'année 1934.

P. Lévy, qui lisait relativement peu les travaux des autres, est cependant loin de les ignorer tous. Dès la première phrase de [L5] , il qualifie le mémoire de S. Bernstein de 1927 (qu'il ne citait pas en 1929) de "progrès important" dans l'étude des sommes de variables enchaînées, et situe ses propres travaux comme "une nouvelle contribution à cette étude".

Décrivons rapidement les théorèmes essentiels qui s'y trouvent

- Autour de la "loi 0-1"

> 38. — J'ai démontré en 1934 le théorème suivant : si la réalisation d'un événement A dépend d'une suite infinie d'épreuves, il est presque sûr qu'une des deux circonstances suivantes sera réalisée : ou bien sa probabilité x_n évaluée après n épreuves tend vers un pour n infini, et il est réalisé, ou bien elle tend vers zéro et il n'est pas réalisé. Autrement dit, on peut, avec une certitude croissante, prévoir le résultat des épreuves.

L'énoncé dans l'article original (lemme 1 de [L5] p. 88) est exactement le suivant :

"P(E) et P_n(E) désignent respectivement la probabilité d'un événement E avant la détermination des x_ν , et après la détermination de $x_1, x_2, \ldots, x_n$ et en fonction des valeurs de ces variables supposées connues. Cet événement E dépend de la suite indéfinie des x_ν" :

<u>Lemme</u> : "Si un événement E a une probabilité α, les suites réalisent cet événement, sauf dans des cas de probabilité nulle, réalisent aussi la condition $\lim_{n\to\infty} P_n(E) = 1$".

N.B. Pour peu que α existe, la limite est donc indépendante de α.

.../...

On reconnaît là un cas particulier du théorème de convergence des martingales (voir chapitre I) selon lequel si z est une v.a., et si $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ est une suite croissante de tribus $\mathbb{E}(z|\mathcal{F}_n) \to \mathbb{E}(z|\mathcal{F}_\infty)$ p.s.

Ici $z = 1_E$. Ceci fait dire à M. Loève [L13] qu'il s'agit tout simplement du "premier théorème de convergence des martingales". Il ajoute : "c'est peut-être l'un des plus beaux résultats de la théorie des probabilités".

> Ce théorème comporte un cas particulier important. Si x_n est indépendant de n, donc égal à la probabilité à priori $x = x_0$ de l'événement E, x est égal à zéro ou un (autrement $x_n = \alpha$ ne pourrait pas tendre vers une des ces limites). C'est le *théorème de l'alternative zéro ou un de Kolmogorov.* Il est antérieur à mon travail de 1934, mais je ne le connaissais pas en rédigeant ce travail, qui parut en 1935.

En fait ce genre d'idées est progressivement de plus en plus présent chez divers auteurs, y compris chez P. Lévy depuis quelques années dans une note rajoutée lors de la correction des épreuves, P. Lévy, qui vient d'avoir connaissance de l'article de B. Jessen de 1934 [12], le signale : "Si la probabilité de E n'est pas modifiée par la connaissance d'un nombre fini de variables x_n, c'est-à-dire si $\mathbb{P}(E/x_1, \ldots, x_n) = \mathbb{P}(E)$, il résulte de notre lemme que $\mathbb{P}(E) = 0$ ou 1".

<u>N.B.</u> : L'article de B. Jessen a été imprimé le 6 juillet 1934 (la date de réception n'est pas indiquée).

La démonstration de P. Lévy est courte (une page), mais non évidente. A noter qu'il énonce le lemme pour le cas de v.a. indépendantes et annonce à la page suivante que cette condition est en fait inutile. Le lien avec la propriété de martingale n'est pas indiqué explicitement.

.../...

- (lemme 2 de [L5]) : c'est l'extension du lemme dit de Borel-Cantelli "au" cas dépendant :

Rappelons l'énoncé classique dans le cas indépendant :

Soit A_n une suite d'événements indépendants, notons $u_n = 1_{A_n}$

et $\alpha_n = P(A_n)$, alors les séries $\Sigma\, u_n$ et $\Sigma\, \alpha_n$ sont p.s. de même nature (Dit autrement $P(\limsup A_n) = 0$ ou 1 selon que $\Sigma\, P(A_n)$ converge ou diverge).

Le lemme 2 dit ceci :

On ne suppose plus que les u_n sont indépendants, et ici α_n désigne la probabilité que $u_n = 1$ "évaluée[non pas a priori, mais] au moment de l'expérience qui détermine u_n"

(i.e. $\alpha_n = \mathbb{P}\,(A_n/u_1,\dots,u_{n-1})$),

alors la conclusions reste valable : $\Sigma\, u_n$ et $\Sigma\, \alpha_n$ sont p.s. de même nature.

- "L'hypothèse essentielle" ($\mathcal{C}$) : $\mathbb{E}_{n-1}(u_n) = 0$ est introduite p. 93 de [L5] pour des v.a. réelles u_n quelconques.

($\mathcal{C}$) : comme "centré" ? ou comme "condition" ?

Il démontre d'abord l'extension, immédiate à ce cas, de l'inégalité maximale de Kolmogoroff :

$$\mathbb{P}\,\{\max_{k\leq n} |S_k| \geq C\sqrt{ES_n^2}\,\} \leq 1/_{C^2} \qquad \text{pour tout } C > 0,$$

et s'attaque à la convergence de la série Σu_n.

Répétons ce que nous avons dit plus haut : P. Lévy, commme S. Bernstein, a en vue l'extension des théorèmes limites du calcul des probabilités (LGN , TLC, LLI) au cas des termes non indépendants. Ce n'est pas un hasard,

.../...

si l'hypothèse essentielle n'est pas exprimée sous la forme $\mathbb{E}_{n-1}(S_n) = S_{n-1}$ et si l'on parle de la convergence de la série Σu_n et non de la convergence de la suite S_n .

Théorème I de [L5]

Si la suite (u_n) vérifie la condition ($\mathcal{C}$) et est uniformément bornée par un nombre U , alors Σu_n et $\Sigma\, \mathbb{E}_{n-1}\, u_n^2$ sont de même nature p.s.

C'est une généralisation naturelle du résultat correspondant de Kolmogoroff dans le cas des v a indépendantes. L'hypothèse que les u_n sont bornées est évidemment trop stricte, c'est pourquoi P. Lévy indique différents cas où on peut l'adoucir.

-La suite des résultats de [L5] et l'essentiel de ceux de [L6] concernent ce que nous appellerions le théorème limite central pour les martingales. Expliquons le problème de manière naïve. Pour conserver le résultat classique, avant toute réflexion approfondie, on souhaiterait avoir le théorème suivant : si (u_n) vérifie la condition ($\mathcal{C}$) et si, chaque terme étant individuellement négligeable, la suite vérifie la condition de Lindeberg, alors $S_n/\sqrt{ES_n^2}$ converge en loi vers une v.a. gaussienne.

Malheureusement les choses sont plus compliquées, ce que montre P. Lévy par une série d'exemples.

D'abord faut-il utiliser les variances Eu_n^2, ES_n^2 ou les variances conditionnées $E_{n-1}\, u_n^2$, $\sigma_n^2 = \sum_{i=1} E_{i-1}\, u_i^2$, pour obtenir des généralisations adaptées ?

(Dans le cas des v.a. indépendantes: $Eu_n^2 = E_{n-1}\, u_n^2$)

.../...

Il montre :

Sous l'hypothèse ($\mathcal{C}$), et si de plus

($\mathcal{C}_1$) : $E_{n-1}\, u_n^2 = E\, u_n^2$

($\mathcal{C}'$) : $|u_k| \leq \varepsilon \sqrt{ES_n^2}$ $\quad \forall\, k = 1,\dots,n$

alors $\dfrac{S_n}{\sqrt{ES_n^2}}$ converge en loi vers une v.a. gaussienne réduite

c'est le théorème 67.1 de son livre, qu'il qualifie de "cas de généralité intermédiaire" entre le cas indépendant et le meilleur qu'on puisse atteindre sous l'hypothèse ($\mathcal{C}$).

La démonstration est une modification de celle de Lindeberg.

Il ajoute ceci : "Cette étude est un cas particulier important de l'étude des sections [voir plus loin, mais P. Lévy l'a effectuée avant dans son texte] ; c'est même l'importance de ce cas qui a jusqu'ici empêché d'apercevoir les services que pouvait rendre l'étude préalable des sections à t constant" p. 126.

De quoi s'agit-il ? Revenons-en à l'idée de base :

S_n, sous l'hypothèse ($\mathcal{C}$) qui est en quelque sorte une hypothèse de centrage, doit se comporter en loi comme la racine de la somme partielle des variances conditionnées $E_{n-1}\, u_n^2$, c'est-à-dire S_n^2 " $\simeq$ " $\sigma_n^2 = \sum_{i=1}^{n} E_{i-1}\, u_n^2$; par suite il est intéressant de couper les trajectoires $S_n(\omega)$ à σ_n constant.

En d'autres termes, soit t donné, appelons $N_t(\omega)$ le premier entier n tel que σ_n^2 atteigne t, que peut-on dire du comportement de $\dfrac{S_{N_t}}{\sqrt{t}}$ lorsque $t \to \infty$?

(N_t est défini p.s. si la série $\Sigma\, E_{i-1}\, u_n^2$ est p.s. divergente).

.../...

La réponse de P. Lévy est que cette forme (selon des "sections à t constant") est la mieux adaptée au problème et que, sous des conditions raisonnables, la limite est effectivement une v.a. gaussienne. Il montre aussi qu'on peut aller plus loin et étudier le cas des sections à t aléatoire : §9 et 10 de [L6].

Dans [L6], § 11 et 22, P. Lévy revient au cas où n est non aléatoire, et aux § 21-22, il s'attaque à la réciproque du théorème limite central, dans l'espri de ce qu'il a fait (et de ce que fait Feller indépendamment de lui) pour le cas indépendant : en d'autres termes, toujours sous l'hypothèse ($\mathcal{C}$), il s'agit d'obtenir des conditions nécessaires et suffisantes pour que $\dfrac{\mathcal{S}(t)}{\sqrt{t}}$

ou $\dfrac{S_n}{\sqrt{ES_n^2}}$ converge vers une loi gaussienne.

Il parvient à des résultats qu'on pourrait comparer à ceux antérieurs et à ceux postérieurs de S. Bernstein. Nous laisserons de côté ici cette question assez technique.

- Enfin l'article [L8] est consacré, toujours en vue de généraliser les théorèmes limites classiques du calcul des probabilités à la loi forte des grands nombres et à la loi du logarithme itéré, sous l'hypothèse ($\mathcal{C}$).

Il montre au § 7 que la LFGN peut se démontrer directement à l'aide de l'inégalité maximale de Kolmogoroff étendue, mais ne s'occupe pas de l'obtenir sous la seule condition d'existence du moment d'ordre 1.

Il démontre la LLI suivante :

Sous l'hypothèse ($\mathcal{C}$) et si de plus :

($\mathcal{C}''$) $\quad$ la série $\sum_{1}^{\infty} E_{i-1}\, u_i^2$ diverge p.s. ($\sigma_n^2 \to \infty$)

.../...

$(\mathscr{C}')$ la valeur $\mathscr{S}(t)$ de la chaîne (S_n), arrêtée au moment où $\sigma_n^2 = t$, est majorée par $\phi(t)\sqrt{t}$ où ϕ est une fonction monotone telle que $\int_1^\infty \frac{\phi(t)}{t}\, dt < \infty$,

alors $P\left(\frac{\mathscr{S}(t)}{\sqrt{t}} > c\sqrt{2 \log\log t} \text{ une infinité de fois}\right) = \begin{cases} 0 & \text{si } c > 1 \\ 1 & \text{si } c < 1 \end{cases}$

La démonstration s'inspire du cas indépendant. Certaines extensions possibles sont suggérées à titre d'hypothèse dans le livre § 72.

d) Brèves remarques

1° - On trouvera dans [L11] et [L12] diverses prévisions sur les circonstances de rédaction de ces articles et du livre qui en reprend l'essentiel.

2° Parmi les idées qui montreront leur fécondité ultérieurement, notons celle de temps aléatoire que P. Lévy utilise ici systématiquement :
il n'est pas fortuit que cette notion apparaisse précisément à cette époque.

3° - Comme continuation des idées de P. Lévy, on peut mentionner la thèse de M. Loève (qui est partiellement son élève), soutenue le 21 juin 1941 [13], publiée seulement de façon détaillée à la fin de la guerre.
L'étude y est assez systématique, tant à propos de la loi des grands nombres que du théorème limite central (voir en particulier le "théorème fondamental" de la p. 290 dont M. Loève dit, p. 286, qu'il "renferme comme cas particulier toutes les propositions obtenues, depuis 1900, et relatives aux conditions suffisantes de tendance centrale") (*).

.../...

(*) M. Loève n'a pas connaissance des résultats de 1940 de S. Bernstein. Il ne cite non plus ni Ville, ni Doob.

Toutefois, quelle que soit l'intérêt de cette thèse, on ne peut pas dire qu'elle apporte de nouveauté relativement aux martingales : elle critique même, en quelque sorte, cette notion considérée comme "assez restrictive" et propose de lui substituer des "hypothèses asymptotiques" (p.249-250...). Ce n'est pas ce genre d'appréciation qui prévaut aujourd'hui, même lorsqu'on se limite aux convergences en loi, cf [3], p. 11.

IV - LA THESE DE J. VILLE

a) Travaux cités

Articles originaux

[V1] : "Sur les suites indifférentes", CRAS 202 (1936), p. 1393-1394
(ZB 13.408 de Finetti)

[V2] : "Sur la notion de collectif", CRAS 203 (1936), p. 26-27
(ZB 14.168 Feller)

[V3] : "Etude critique de la notion de collectif", Thèse Paris, mars 1939
(jury : Borel, Fréchet, Garnier) (ZB 21.145 de Finetti)

[V4] : "Etude critique de la notion de collectif", Monographie des probabilités, Gauthier-Villars (1939) (ZB 21.146 Kamke)
(N.B. : Ce livre reprend exactement la thèse en y supprimant l'introduction (d'une page) et en y ajoutant un chapitre préliminaire de 17 pages : on passe donc de la pagination de la thèse à celle du livre en ajoutant 16).

Articles à caractère historique

[V5] : "Notice sur les travaux scientifiques de M. Jean Ville", mai 1955, Archives Fréchet, Laboratoire de probabilités, Université de Paris VI

[V6] : "Théorie des jeux, dualité, développement", Economie Appliquée 36 (1983), p. 595-610.

.../...

b) Indications biographiques

J. Ville, né en 1910, est étudiant au début des années 30. Voici comment il relate ce qu' "on" pensait des probabilités à cette époque en France :

> au temps où j'étais étudiant, le calcul des probabilités lui-même était considéré comme un passe-temps honorable pour mathématiciens chevronnés, qui s'étaient distingués déjà par d'autres travaux de mathématiques pures, pour ne pas dire véritables mathématiques, telles que l'analyse.
>
> D'autres passes-temps tenaient compagnie au Calcul des Probabilités. On y comptait l'algèbre de Boole, la numération binaire, l'économique rationnelle de François Divisia, la logique mathématique, alors appelée logistique. Cette attitude de dédain se manifestait par la quasi-inexistence des enseignements, et la méfiance envers les jeunes gens attirés par ces disciplines, soupçonnés d'être séduits par le divertissement plutôt que par la recherche sérieuse.
>
> Emile Borel venait de la théorie des fonctions et de la mesure, Maurice Fréchet de la topologie des espaces abstraits, George Darmois de la relativité; ils avaient suffisamment excellé dans ces occupations sérieuses pour avoir gagné le droit de s'amuser.

S. Bernstein (né en 1880) et P. Lévy (né en 1886) font également partie des "mathématiciens chevronnés" quand ils abordent le calcul des probabilités ; de plus leurs résultats sur les martingales "avant la lettre" ne sont pas leurs premiers travaux probabilistes.

La situation de J. Ville est toute différente, il a donc quelque mérite à choisir un domaine alors peu considéré. D'ailleurs, comme il écrit dans son Introduction : "M. Fréchet, qui m'avait d'abord proposé un autre sujet, a bien voulu, sur ma demande, accepter l'étude de la théorie de M. de Mises comme sujet de Thèse".

A titre de document, nous reproduisons ici le début de la notice sur les travaux scientifiques faite par J. Ville lui-même en 1955 [V5].

.../...

1 - RECHERCHES sur la THEORIE du "COLLECTIF" de R.V. MISES.

Les premières publications que j'ai faites, extrêmement modestes, furent ma contribution aux travaux d'un Colloque Mathématique qui tenait ses séances à Vienne, sous la direction de M. Karl MENGER, dont j'étais allé suivre les cours grâce à l'octroi, qui m'avait été fait, d'une bourse Asconati-Visconti. Les cours de M. MENGER traitaient de géométrie, tout particulièrement de la théorie de la dimension, au sens initié par H. POINCARE; je publiai, dans les comptes rendus du Colloque deux petites notes, concernant l'une, la théorie des courbes dites à longueur quadratique, dans laquelle on étudie la limite de la somme des carrés des côtés d'une ligne polygonale inscrite, l'autre une proposition concernant les espaces déduits d'un espace métrique en y remplaçant la distance par la racine carrée de la distance; de tels espaces sont également métriques, et peuvent, s'ils contiennent deux, trois, quatre points être plongés dans un espace euclidien à 1, 2, 3 dimensions. Je montrai que cette correspondance s'arrêtait à ce nombre de points.

En assistant aux séances du Colloque, je pris part à des discussions sur la nouvelle définition que R. Von MISES proposait pour les probabilités, et le sujet me parut digne d'étude. Pour rendre compte de l'état de la question à cette époque, il est nécessaire d'insister un peu sur les controverses relatives à la définition de la probabilité. Il était bien connu, depuis longtemps, que si on procède à une suite d'épreuves indépendantes, pouvant donner lieu chacune à une alternative, à savoir l'apparition d'un évènement E ou de son contraire $\overline{E}$, et si on appelait r_n la répétition de E dans les n premières épreuves, et par conséquent $f_n = r_n : n$ la fréquence de E dans ces premières épreuves, on avait, en appelant p la probabilité de E supposée constante :

$$\lim_{n=\infty} \operatorname{Prob}\{|f_n - p| < \varepsilon\} = 1 .$$

Ceci constituait la loi "faible" des grands nombres, que l'on exprimait conventionnellement en disant que f_n tendait en probabilité vers p. Cette limite "en probabilité" n'était évidemment pas la limite telle qu'elle est définie en analyse, et ne pouvait donc pas donner prétexte à une définition de la probabilité <u>par</u> une limite au sens de l'analyse. Cette situation se trouva entièrement modifiée lorsque CANTELLI, utilisant les résultats du célèbre mémoire de E. BOREL sur les probabilités dénombrables, connu sous le nom de "Mémoire de Palerme", réussit à démontrer que :

$$\lim_{n=\infty} \operatorname{Prob}\{|f_n - p| < \varepsilon \text{ et } |f_{n+1} - p| < \varepsilon \text{ et } |f_{n+2} - p| < \varepsilon \text{ et... ad inf.}\} = 1 ,$$

d'où il apparaissait que la probabilité de l'évènement

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

la limite étant cette fois la limite au sens de l'analyse, était égale à 1.

M. Von MISES proposa alors une définition de la probabilité, qui était la suivante, et qui se décomposait en deux :

1°) Si on procède à une suite d'épreuves indépendantes, la fréquence de E tend vers une limite.

2°) Cette limite est par définition la probabilité de E .

La première partie était une affirmation ayant mêmes caractéristiques qu'une loi de la nature; la seconde une définition au sens propre du terme. La définition de M. Von MISES était complétée par la condition :

3°) La limite p n'est pas modifiée si on extrait de la suite des résultats de tirage une suite partielle, infinie, par un "procédé de choix" ne tenant compte, pour décider du choix d'un résultat, que de la valeur des résultats <u>précédents</u>.

1°) était appelé l'axiome de la limite ;
3°) l'axiome de procédé de choix.

Le calcul des probabilités devenait ainsi une étude des propriétés de certaines suites, satisfaisant à des axiomes, suites que R.Von MISES appelait des collectifs. L'axiome du procédé de choix permettait d'exclure des suites régulières telles que

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

dans laquelle la fréquence de E est manifestement 0.5, mais dans laquelle, si on pratique le choix très simple qui consiste à ne conserver qu'un terme sur deux, la valeur de la limite est modifiée.

Les axiomes de R. Von MISES étaient manifestement contradictoires, parce que si l'on considère tous les procédés de choix, qui forment un ensemble ayant la puissance du continu, la restriction qu'un choix ne doit tenir compte que des résultats antérieurs est insuffisante, et il n'existe aucune suite satisfaisant aux axiomes.

Un participant au Colloque, A. WALD, proposa alors de se limiter à une infinité dénombrable de procédés de choix, ce qui pouvait s'obtenir en disant :

On appelle collectif dans une théorie déductive n'ayant elle-même qu'un nombre fini d'axiomes, et ne permettant par conséquent d'énoncer qu'une infinité dénombrable de procédés de choix, une suite telle que.....

Les axiomes devenaient non contradictoires, pour la raison simple suivante :

Si on représente une suite infinie par un point du segment (0,1), la loi de probabilité à laquelle satisfait la suite définit sur ce segment une mesure complètement additive. A tout procédé de choix, on peut associer les points représentant les suites pour lesquelles l'axiome de procédé de choix n'est pas satisfait; cet ensemble est, dans la mesure considérée, de mesure nulle. Comme on ne considère qu'un ensemble dénombrable de procédés de choix, on n'exclut du segment qu'un ensemble de mesure nulle, l'ensemble complémentaire est donc non vide, il n'y a pas contradiction.

La cause semblait donc entendue, et il apparaissait donc possible de considérer des suites, définies par exemple à partir de la théorie déductive des "Principia Mathematica", qui présenteraient les caractères d'une suite au hasard, et pour lesquelles il serait impossible, par manipulation si ingénieuse que ce soit, de modifier la fréquence limite.

La question que je me posais fut alors la suivante :

Etant donné que :

Dire qu'une suite de résultats jouit, avec probabilité égale à 1, d'une certaine propriété, revient à dire que les suites qui ne jouissent pas de cette propriété sont représentées par des points appartenant à un ensemble de mesure nulle;

Il apparaît que la démonstration de la première affirmation consiste à couvrir l'ensemble des points représentatifs des suites ne jouissant pas de la propriété par un ensemble de mesure nulle;

Ceci étant reconnu,

existe-t-il des ensembles de mesure nulle que l'on ne puisse pas couvrir avec les ensembles de mesure nulle définis à partir des axiomes de son choix.

Autrement dit, existe-t-il des propriétés démontrables dans la théorie classique, et indémontrables au sens de R. Von MISES (complété par les travaux de A. WALD) ?

Je pus montrer qu'il en était ainsi, c'est-à-dire que la catégori des ensembles de mesure nulle définis comme ensemble des points représentatifs de suites ne satisfaisant pas à un axiome de procédé de choix n'était pas assez étendue pour couvrir tous les ensembles de mesure nulle. Il est évident que les ensembles de mesure nulle, en infinité dénombrable, associés à une infinité dénombrable de procédés de choix, ne peuvent couvrir tous les ensembles de mesure nulle; mais il est moins évident qu'il existe un ensemble de mesure nulle, indépenda des procédés de choix, tel que, quels que soient les procédés de choix adoptés, en infinité dénombrable, on ne puisse le couvrir avec le ensembles de mesure nulle associés à ces procédés de choix.

Je fus guidé dans la recherche de la démonstration par la méthode de LIOUVILLE pour démontrer qu'il existe des nombres non algébriques; je m'attachai non pas aux suites où la fréquence était divergente, mais aux suites où la fréquence convergeait trop rapidement. Une fois la démonstration trouvée, je pus énoncer un résultat très simple :

Pour trouver un ensemble de mesure nulle non couvrable par les ensembles définis par R. Von MISES, il suffit de considérer les suites pour lesquelles f_n tend vers p unilatéralement. Elles peuvent être des collectifs au sens de R. Von MISES, et présentent une particularit que le Calcul des probabilités exclut.

La NOTION de MARTINGALE.

Je proposai les résultats précédents à M. FRECHET pour constituer un sujet de thèse. Il voulut bien l'accepter. Je me proposai ensuite d remplacer les axiomes de R. Von MISES par des axiomes ne présentant pa la lacune importante que j'avais mise en évidence. Je remarquai que l'impossibilité de modifier la fréquence par un choix pouvait s'interpréter comme une impossibilité de gagner à un jeu équitable par une ma tingale excluant certains coups; je me demandai si en généralisant la martingale de manière à permettre une répartition continue des mises, il ne serait pas possible d'aller plus loin. J'avais été frappé par u résultat de LAPLACE, d'après lequel si on jouait à pile ou face avec u pièce non symétrique, on pouvait, sur les deux premiers coups, s'assur l'avantage par une répartition judicieuse des mises, à partir de l'information que la pièce était dissymétrique, sans savoir de quel côté.

J'admis donc que dans une partie de pile ou face on avait le loisi avant chaque coup, de répartir la somme que l'on avait à sa dispositio et provenant d'une somme initiale égale à l'unité, selon les résultats des coups précédents. D'après les résultats que j'avais obtenus, il fallait, pour pouvoir aller plus loin que les méthodes de R. Von MISES, pouvoir tirer parti d'une information assez vague, comme la suivante :

Dans la suite des tirages, la fréquence tendra vers $\frac{1}{2}$ unilatéralement, mais nous ne savons pas de quel côté, ni à partir de quel rang.

Je constatai que la connaissance d'un ensemble de mesure nulle quelconque dans lequel devait se trouver le point représentatif de la suite permettait de bâtir une règle de jeu permettant de gagner au-del de n'importe quelle limite.

Je proposai donc de substituer la notion de "martingale" à celle de "procédé de choix", ce qui permettait d'éviter la contradiction si on parlait de l'ensemble (dénombrable) des martingales énonçables, et assurait la concordance avec les ensembles de mesure nulle énonçables, ce qui n'était pas le cas pour la théorie de R. Von MISES.

Je généralisai la notion de martingale, qui apparut un instrument précieux de démonstration. Je donne un exemple simple :

Dans un tirage de BERNOULLI, avec probabilités p et q, il est très simple de donner une règle de jeu assurant au n^e coup le gain

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

r étant la répétition de l'évènement de probabilité p. Cette expression est donc bornée. La formule de STIRLING permet d'en déduire la loi forte des grands nombres, et même que $\frac{r}{n} - p$ tend vers zéro comme $\sqrt{\frac{\log n}{n}}$, au moins.

Je pus également démontrer, en utilisant la notion de martingale, une proposition énoncée, sans démonstration, par KOLMOGOROFF, à savoir

La condition nécessaire et suffisante pour que dans une suite de tirages de BERNOULLI la répétition r satisfasse avec probabilité 1, à partir d'un certain rang, à l'inégalité :

$$|r - np| < \varphi(n)\sqrt{2npq}$$

où $\varphi(n)$ est la suite des valeurs prises pour t entier par une fonction $\varphi(t)$ croissante, est que l'intégrale

$$\int \varphi\, e^{-\varphi^2} \frac{dt}{t}$$

soit convergente pour $t = +\infty$.

Cette notion de martingale a pu servir utilement à d'autres chercheurs; nous avons attiré l'attention des mathématiciens sur une notion existante, puisque le mot existait, mais qui n'avait pas été considérée comme une notion importante. Elle a été reprise par J.L. DOOB, qui en parle en ces termes : "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 " (1).

---

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

c) Précisions et description de la Thèse

1° La situation du problème est très bien expliquéedans diverses publications, nous renvoyons donc :

- à M. Fréchet [14], qui expose une partie de la Thèse de J. Ville alors non publiée, et montre bien les interconnexions dans les débats d'alors entre les problèmes mathématiques et les problèmes philosophiques,
- à C. Dellacherie [15] , pour les questions mathématiques que soulèvent les suites aléatoires, ainsi qu'à C.P. Schnorr [20],
- à J. Bonitzer [6], pour une vue tout à fait pénétrante sur les débats philosophiques relatifs au hasard.

2° Le séminaire de K. Menger

K. Menger est le fils du célèbre économiste "marginaliste" autrichien. Dès les années 20 il s'intéresse au rôle des probabilités et de l'incertain en économie, en relation avec le "jeu de Saint-Pétesbourg". Dans le "Mathematische Kolloquium" qu'il organise de 1928 à 1937 à Vienne (avant que la montée du nazisme n'oblige lui-même, et bien d'autres, à émigrer), on se réunit toutes les semaines pour discuter de manière libre et ouverte (*) de sujets variés : il y a Gödel, Nöbeling, Alt, Beer, Tausky, Čech, Knaster, Tarski, A. Wald... En liaison avec les philosophes du Cercle de Vienne, on y étudie la théorie de Von Mises : K. Popper y fait un exposé, mais il n'y a pas de "probabiliste de métier" . A. Wald, frappé par ces questions en 1934 y consacre alors une partie de son travail. cf[16].

.../...

(*) Cela tranche avec l'enseignement français extraordinairement scolaire et hiérarchisé, où l'étudiant, dressé par les classes de spéciales, n'ose rien faire (dit J. Ville).

3° La définition de la martingale telle qu'elle est donnée par J. Ville

DÉFINITION 1. — *Soit $X_1, X_2, \ldots, X_n, \ldots$ une suite de variables aléatoires, telle que les probabilités*

$$\Pr.\{X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n\} \quad (n = 1, 2, 3, \ldots)$$

soient bien définies et que les X_i ne puissent prendre que des valeurs finies.

Soit une suite de fonctions $s_0, s_1(x_1), s_2(x_1, x_2), \ldots$ non négatives telles que

$$(14) \begin{cases} s_0 = 1, \\ \mathfrak{M}_{x_1, x_2, \ldots, x_{n-1}}\{s_n(x_1, x_2, \ldots, x_{n-1}, X_n)\} = s_{n-1}(x_1, x_2, \ldots, x_{n-1}), \end{cases}$$

où $\mathfrak{M}_X\{Y\}$ représente d'une manière générale la valeur moyenne conditionnelle de la variable Y quand on connaît la position du point aléatoire X, au sens indiqué par M. P. Lévy.

Dans ces conditions, nous dirons que la suite $\{s_n\}$ définit une martingale ou un jeu équitable.

[V3] p. 83 et p. 73 dans un cas particulier.

N.B. : La définition classique d'une martingale positive, aujourd'hui, est la suivante : on note $\mathcal{F}_n$ la suite de tribus engendrées par les v.a. $X_1, \ldots, X_n$ et Z_n une suite de v.a. ≥ 0 $\mathcal{F}_n$ - mesurables. Alors Z_n est une martingale si $E(Z_n | \mathcal{F}_{n-1}) = Z_{n-1}$ p.s.

On constate la différence avec la définition de J. Ville :

"Chez moi, sont données les X_n, sont à trouver $s_1(x_1)$, $s_2(x_1, x_2)$, ...etc. Maintenant on suppose données les $\mathcal{F}_n$ et les Z_n. On insistait sur le fait que Z_n ne pouvait devenir infinie. Mon but était, les X_n étant données, et étant donné un ensemble tel que la probabilité soit nulle que la suite X_n lui appartienne, de former les s_n tels que sur cet ensemble les s_n tendent vers l'infini. J'insite là-dessus parce qu'il m'a fallu très longtemps pour faire comprendre cette démarche" (lettre de J. Ville à l'auteur,1985)

.../...

On notera aussi que J. Ville étudie directement la suite $s_n(X_1,\ldots,X_n)$ sans la considérer (comme S. Bernstein et P. Lévy) comme résultat des sommes de v.a. presque indépendantes. Ce point de vue lui permet de généraliser la notion de martingale au cas du temps continu.

1. Jeu continu équitable. — Généralisons les définitions de la page 83. Au lieu de la suite $\{s_i\}$ de fonctions $s_n(X_1, X_2, \ldots, X_n)$, nous aurons une suite $\{S_\tau\}$ continue ($\tau \geq 0$) de fonctionnelles $S_\tau(X)$, c'est-à-dire de fonctions du point dans E_ω telles que

(a) $S_0(X) = 1$;

(b) $S_\tau(X) \geq 0$;

(c) $S_\tau(X)$ *ne dépend que des valeurs prises par $X(t)$ pour $t \leq \tau$;*

(d) *La valeur moyenne de $S_{\tau_1+\tau_2}(X)$ ($\tau_2 > 0$), quand on sait que $X(t) \equiv X_0(t)$ pour $t \leq \tau_1$ est égale à $S_{\tau_1}(X_0)$.*

Ces conditions ont besoin d'être précisées dans le cas continu, ajoute-t-il (le mot "martingale" n'est plus utilisé ici)

4° Les résultats obtenus par J. Ville

Il démontre d'abord une inégalité maximale (sans hypothèse de moment d'ordre 2) appelée "théorème de la ruine des joueurs"

Théorème 1 Soit $\{s_n\}$ une suite définissant une martingale et λ un nombre > 1. Nous avons l'inégalité

$$\mathbb{P}\{\sup_{(n)} s_n(X_1,\ldots,X_n) \geq \lambda\} \leq \frac{1}{\lambda}$$

N.B : comme la martingale est positive et d'espérance 1, c'est l'inégalité maintenant classique :

$$\mathbb{P}(\max_{1\leq k\leq n} S_k \geq \lambda) \leq \frac{1}{\lambda} E|S_n|)$$

[V3], p. 84 et sur un cas particulier, p. 36

Il en déduit que $\sup_n s_n(X_1,\ldots,X_n)$ est p.s. borné, mais n'explicite pas de théorème de convergence.

.../...

Ensuite il démontre seulement une moitié du théorème de Kolmogoroff cité dans la notice : si l'intégrale $\int \phi e^{-\phi^2} \frac{dt}{t}$ est convergente, alors $|r-np| < \phi(n)\sqrt{2npq}$ p.s. à partir d'un certain rang.

La démonstration, outre la notion de martingale, fait intervenir ce que nous appellerions aujourd'hui une représentation des fonctions harmoniques pour des marches aléatoires sur un semi-groupe.

Dans le cas continu, l'extension de la "formule de la ruine des joueurs" ne peut se faire sans précautions, et il y a lieu de préciser dans ce cas ce qu'on entend par probabilité conditionnelle, par borne supérieure, par continuité. Doob dira à propos de ce passage de J. Ville : "His discussion of the meaning of a continuous process, and of generalized upper bounds is somewhat obscure". En tout cas, il est certain qu'à cette époque, la théorie des processus stochastiques à temps continu est pour la plupart des gens assez floue. J. Ville sent qu'il faut "se placer uniquement sur le terrain de M. Kolmogoroff" et que les conclusions peuvent être différentes "en suivant le point de vue de M. Paul Lévy" [V3], p. 100. Mais ces considérations ne sont qu'embryonnaires et c'est seulement Doob qui clarifiera les choses.

Soit $X(t)$ un processus de Markov à temps continu, on peut lui associer un processus "$S_\tau(X)$ qui ne dépend que de $X(\tau)$. En d'autres termes, il existe une fonction $s(t,x)$ telle que $S_\tau(X) = s[\tau, X(\tau)]$".

Alors les conditions pour que $S_\tau(X)$ soit une martingale positive se réduisent à : $s(0,x) = 1$, $s(t,x) \geq 0$, et

$$s(t_1, x_o) = \int s(t_2, x)\, P_{t_1,t_2}(x_o, dx), \text{ en notant}$$

$P_{t_1,t_2}(x_o,.)$ la probabilité de transition

(N.B. il utilise des fonctions de répartition, en réalité).

.../...

En d'autres termes s est une fonction harmonique ≥ 0 sur l'espace-temps. (Ville n'insiste pas en fait sur cette relation entre fonctions harmoniques et martingales qui est probablement inconsciente chez d'autres auteurs, mais, semble-t-il, non écrite avant lui).

Il étend alors à ce cas, <u>en un certain sens</u>, la formule de la ruine des joueurs, et <u>sous une hypothèse supplémentaire</u>.

Il démontre ensuite un analogue continu du théorème de Kolmogoroff, vu plus haut sur l'ordre de grandeur p.s., au mouvement brownien : si $\int \phi\, e^{-\phi^2} \frac{dt}{t}$ converge, $\phi(t)$ est "une fonction limite supérieure stochastique" de $\frac{S_t}{\sqrt{2t}}$, en un sens qu'il précise. (Il retrouve ainsi de manière plus simple un résultat de Petrowsky) [17].

Enfin il donne un théorème voisin pour la suite des $Y_n = \max_{1 \leq k \leq n} X_k$, lorsque X_k est une suite de v. a. indépendantes et de même loi.

- L'apport de J. Ville à la théorie des martingales ne se réduit donc pas à avoir donné un nom à une notion mathématique connue. Sa Thèse marque une étape, un renversement entre le moment où une condition telle que $E_{n-1}(X_n) = 0$ ne pouvait être envisagée que comme adaptée à l'extension de théorèmes limites classiques, et le moment où un processus stochastique discret ou continu vérifiant $E_t(S_{t'}) = S_t$ $(t < t')$ peut être étudié systématiquement et reconnu comme une nouvelle manière de considérer un pan entier, et en formation, de la théorie des probabilités. Mais J. Ville ne donne que quelques exemples, il n'applique pas cette idée de façon systématique, comme le fera J.L. Doob.

.../...

V - L'ARTICLE DE DOOB DE 1940

a) Travaux cités

[D1] : "Note on probability", Ann. Math 37 (1936), p. 363-367 (ZB 13.408 (1936) B. Jessen) (parvenu le 16 septembre 1935)

[D2] : "Regularity properties of certain families of chance variables" T.A.M.S. 47 (1940), p. 455-486 (reçu le 11 décembre 1939) (MR 1.343 : Feller, ZB 23.6 - 12 décembre 1940, B. Hostinsky)

[D3] : "Topics in the theory of Markoff chains", T.A.M.S. 52 (1942), p. 37-64 (reçu le 24 mai 1941) (MR 4.17 : Kač)

[D4] : "Stochastic processes", Wiley (1953), en particulier le chapitre VII, mais aussi le II, le III etc., voir l'appendice historique.

[D5] : "Classical potential theory and its probabilistic counterpart", Springer (1984). Notice historique, p. 793-817.

b) Remarques diverses

Nous ne nous intéressons ici qu'aux travaux de Doob jusqu'à 1940 et non ceux d'après la guerre.

- Les premiers articles de Doob, et notamment sa Thèse, portent sur les fonctions analytiques et leurs valeurs au bord, sujet dont on comprendra clairement, bien après la guerre, le lien avec les martingales.
- Il vient ensuite à la statistique puis aux probabilités, vastes domaines qui, du point de vue mathématique, étaient fort peu développés aux Etats-Unis. Dès ses premiers articles de statistique (en 1934)(*) , il adopte à fond le point de vue de Kolmogoroff dans les Grundbegriffe et utilise encore davantage les espaces abstraits :la puissance de ce cadre ne lui échappe pas et ses travaux de probabilités ultérieurs en sont une défense et illustration.

.../...

(*) Un article contient une démonstration erronée (mais l'erreur est rattrapable) de la consistence de l'estimé du maximum de vraisemblance ; peut-être cette question a-t-elle fait partie de ses motivations conscientes ou non pour étudier la théorie des martingales ? voir [D4] appendice.

- Ses articles de probabilités (à partir de 1936-1937) visent à unifier la théorie naissante des processus, en particulier à temps continu, à l'aborder en toute rigueur. Comme le dira P. Lévy [L10] 2^{e} édition, p. 360 :
"Doob part d'une théorie abstraite, et reste volontiers aussi longtemps que possible dans l'abstraction, de manière à donner à ses théorèmes la forme la plus générale possible. Je préfère au contraire arriver aussi vite que possible à des définitions concrètes..." ;
et plus loin : "Comme toujours, Doob suit rigoureusement, j'allais dire impitoyablement, l'ordre logique" p. 378.
Le bourbakisme n'est pas forcément dans le pays où on l'attendrait. D'ailleurs dans ses articles, jusqu'en 1940 en tout cas, Doob ne donne quasiment jamais la moindre motivation, ni le moindre indice sur l'importance ou l'utilité des notions qu'il introduit (par exemple sur celle de martingale...)
Pour mieux comprendre la différence de point de vue entre P. Lévy et J.L. Doob, sur la théorie des processus, il n'est pas inutile de citer longuement encore une fois la note II ajoutée à la fin de la 2^{e} édition de [L10]

Pour les analystes modernes (notamment Kolmogoroff et Doob), un problème de probabilité est défini par la donnée d'une fonctionnelle additive P des sous-ensembles B de l'ensemble $\mathcal{E}$; elle est non négative, égale à un pour l'ensemble $\mathcal{E}$ lui-même, et bien définie dans un corps borélien K. La théorie des processus stochastiques doit rentrer dans ce cadre, qui implique qu'on se donne

.../...

au coup toute la fonction $X(t)$; considérée comme élément d'un espace fonctionnel $\mathfrak{G}$.

Au point de vue des principes, cet aspect de la probabilité est fondamental, et je songe d'autant moins à le contester que je l'avais pris pour point de départ dans un travail reproduit en 1925 à la fin de mon calcul des probabilités, à cela près que je n'avais pas précisé que le champ de définition de la fonctionnelle P doit être un corps borélien (1). Mais ce point de vue ne montre pas ce qu'il y a d'essentiellement nouveau dans la théorie des processus stochastiques, qui, comme nous l'avons déjà dit, est l'idée de l'intervention continue du hasard. Dans certains cas simples, comme celui du processus de Poisson, on peut conserver à cette idée sa forme initiale en déterminant successivement les instants

des sauts. Il en est de même pour les fonctions additives dans le cas où les sauts constituent le seul élément aléatoire. Si le nombre probable des sauts est fini dans tout intervalle fini, il n'y a aucune difficulté; s'il n'en est pas ainsi, $X(t)$ est du moins la somme d'une série presque sûrement convergente dont chaque terme peut s'étudier sans aucun retour en arrière.

Dans le cas du mouvement brownien, on imagine plus difficilement ce qu'est l'intervention continue du hasard, et il faut bien considérer d'abord une suite de valeurs croissantes du temps, et revenir en arrière pour effectuer des interpolations successives

p.369-371

Remarquons d'autre part que la notion de probabilité conditionnelle ne joue pas du tout le même rôle dans les deux points de vue. Dans la méthode constructive, les lois dont dépendent les X_n sont données successivement; il s'agit bien entendu pour chacune d'elles d'une loi conditionnelle, fonction des X_ν antérieurement déterminés. Mais elle sera donnée, et c'est par le théorème des probabilités composées qu'on en déduit les lois à n variables $X_1, X_2, \ldots, X_n$, et, à la limite, la loi qui définit le processus. Si au contraire on part du point de vue abstrait, on suppose cette dernière loi donnée en bloc, et, pour en déduire une définition constructive, il faut déterminer les lois dont dépendent les choix successifs des X_n. Sauf la première, ce sont des lois conditionnelles, et la définition d'une telle loi devient particulièrement importante.

.../...

-Il y a donc un renversement d'objectif et de direction d'étude :

. Pour P. Lévy, il s'agit encore de généraliser les théorèmes limites classiques (autant en loi que p.s.) à des v.a. "enchaînées", au moyen d'idées cousines de celles de Markov ou de S. Bernstein ; il ne cherche guère (dans ce cas), malgré l'utilisation de temps aléatoires, à étudier globalement les processus et ne fait d'ailleurs pas de martingales à temps continu.

. Pour J. Ville, les martingales restent d'abord un outil ; même dans le chapitre V, où il asseoit la notion et envisage de l'étudier pour elle-même, l'objectif est de montrer des théorèmes limites, à temps continu ou discret, voisins des théorèmes classiques, éventuellement pour d'autres processus que les martingales.

. Pour J.L. Doob, il s'agit d'une étude systématique à partir de la définition a priori, pour obtenir tant des théorèmes de convergence que des régularités de trajectoires (*) (les inégalités ne semblent pas encore un but en soi en 1940). Un grand principe est dégagé, qui va traverser la théorie des probabilistes.

c) Sur les publications de J.L. DOOB avant 1940

- Avant de décrire le contenu de l'article de 1940, mentionnons une note de 1936 : il y démontre "de manière élégante" (Feller), que "l'énoncé selon lequel un système gagnant est impossible dans un jeu de hasard correspond à un théorème mathématique dans Ω_∞" (Ω_∞ est l'espace des suites infinies) [D1] .

.../...

(*) le mot n'est pas employé.

Plus précisément, notons (x_n) une suite de v.a. indépendantes et de même loi, alors :

> THEOREM. *Let $\tau_1(\omega)$, $\tau_2(\omega)$, $\cdots$ be a sequence of measurable functions of ω:$(x_1, x_2, \cdots)$ on Ω_∞ with the following properties:*
> (i) *$\tau_n(\omega)$ only takes on the values 0 and 1;*
> (ii) $$\limsup_{n\to\infty} \tau_n(\omega) = 1$$
> *almost everywhere on Ω_∞;*
> (iii) *$\tau_1(\omega) \equiv 0$ or $\tau_1(\omega) \equiv 1$; if $n > 1$, $\tau_n(\omega)$ depends only on $x_1, \cdots, x_{n-1}$. Let $\theta(n, \omega)$ be the n^{th} value of j for which $\tau_j(\omega) = 1$ and let $x'_n(\omega) = x_\theta(\omega)$ for all $n = 1, 2, \cdots$ and every ω at which (ii) is true. The transformation T taking ω:$(x_1, x_2, \cdots)$ into ω':$(x'_1(\omega), x'_2(\omega), \cdots)$ is then measure-preserving in the sense that if Λ' is a measurable set in Ω_∞, and if Λ is the set of all points of Ω_∞ taken by T into points of Λ', Λ is measurable and has the same measure as Λ'.*[6]

P. R. Halmos appelle "system" une suite de fonctions mesurables telle que (τ_n) et "system transformation" une transformation telle que T. Dans ce langage, dont l'inspiration "ergodique" est explicite, le théorème devient : "On an independant, stationary, stochastic process every system transformation is measure preserving" [18] , p. 465.

En langage probabiliste, en parlant de "temps d'arrêt" (mais on ne faisait pas cela à l'époque), on dirait ceci :

Si (n_j) est une suite strictement croissante de temps d'arrêt finis > 1, tels que $\{n_j=k\}\in \mathcal{F}_{k-1}$ définie par $x'_j = x_{n_j}$ a les mêmes prorpiétés que x_j : indépendance, loi commune égale à celle de x_1. (cf. [D4], VII 2.3).

Il est clair que cet énoncé n'est autre que le principe d'irrégularité de Von Mises (que Doob ne cite pas), sous sa forme originelle, sans y inclure la modification de J. Ville (à savoir la possibilité non seulement de choisir les instants où l'on joue, mais aussi celle de varier le montant des mises).(*)

.../...

(*) jeu de mots certainement classique à l'époque.

Démontrer rigoureusement cette impossibilité d'un système de jeu dans un cadre assez large est une idée dans l'air, nous l'avons noté dans le chapitre sur J. Ville, mais il y a bien d'autres références (voir la bibliographie de [18] ; Doob cite d'ailleurs Copeland et sa théorie des nombres admissibles.

Il est nécessaire de noter que l'article de P. R. Halmos déjà cité envisage cela de manière très explicite et l'étend considérablement. Son théorème 8 dit clairement que le caractère d' "accroissement de martingale" n'est pas affecté par une suite de temps d'arrêt vérifiant la propriété précédente :(p.469

> THEOREM 8. *Let Ω be a real-valued stochastic process in which the functions $x_n(\omega)$ are uniformly bounded. Let T be a system transformation on Ω taking $\omega = \{x_1, x_2, \cdots\}$ into $\omega' = \{x'_1, x'_2, \cdots\}$. If $E(x_1, \cdots, x_{n-1}; x_n)$ vanishes for all n and almost all $x_1, \cdots, x_{n-1}$, then $E(x'_1, \cdots, x'_{n-1}; x'_n)$ vanishes for all n and almost all $x'_1, \cdots, x'_{n-1}$.*

Il généralise ces propriétés à diverses hypothèses d'indépendance asymptotique et de martingale approchée, fort utiles en théorie ergodique, mais qu'il ne compare pas aux hypothèses de S. Bernstein.

L'article de Doob n'est cité ni par P. Lévy, qui ne l'a sans doute pas lu, ni par Ville(*) qui cite pourtant un autre article de Doob de 1937.

La généralisation par Halmos de la note de Doob semble ignorer les travaux antérieurs ou quasi-simultanés de Lévy et de Ville, et réciproquement. Il est plus étonnant que l'article de 1940 de Doob ne souffle mot ni de sa propre note, ni du travail d'Halmos qui est son élève à Urbana (Illinois), sauf erreur de ma part. Par contre, il est mentionné et amélioré dans le livre de 1953, [D4] VII 2.3.

.../...

(*) "L'article de Doob de 1936 n'a pas retenu mon attention. L'impossibilité de gagner par choix des coups était par trop voisine de l'axiome de choix de von Mises. Quant à l'article de Halmos, je ne l'ai pas lu. 1939 était une année très critique. J'avais déjà été mobilisé plusieurs mois fin 1938". (Lettre de J. Ville à l'auteur).

Certes l'article que nous allons examiner en est "logiquement indépendant", il est toutefois probable qu'une partie de son intuition et de sa motivation y est liée.

d) L'article-clé

L'article de 1940 [D2] commence brutalement par la définition suivante(*) :

> Let t vary in any simply ordered set, and let $\{x_t\}$ be a family of chance variables. We shall say that the chance variables have the property $\mathcal{E}$ if whenever $t_1 < \cdots < t_{n+1}$,
> $$E[x_{t_1}, \cdots, x_{t_n}; x_{t_{n+1}}] = x_{t_n}\dagger$$
> with probability 1. The subscript n here is any positive integer. It will always be supposed that the given chance variables have expectations, so that the conditional expectations involved in the definition of the property $\mathcal{E}$ will always exist.

On notera que, bien que Ville soit cité à plusieurs reprises, le mot "martingale" n'est ni adopté, ni même prononcé de tout le texte.

Doob montre d'abord quelques conséquences assez simples de la définition, par exemple le fait que $E|x_t|$ est non-décroissant en t.

Le § 1 est consacré aux inégalités et aux théorèmes de convergence dans le cas du temps discret.

Il suit d'une certaine façon, en les affinant, les idées de Lévy et de Ville,e les applique aussi bien aux martingales dans le sens usuel qu'aux martingales "à rebrousse-poil", c'est-à-dire qu'il considère à la fois le cas où n tend vers $+\infty$ et celui où n tend vers $-\infty$, lorsque la suite (x_n) vérifie la propriété $\mathcal{E}$. Mais ici, l'inégalité de martingales (appelée par Ville "formule de la ruine des joueurs") et le théorème de convergence presque sûre sont démontrés dans ces deux cas sans hypothèse de positivité, et sans hypothèse de moment d'ordre supérieure à 1. C'est un pas essentiel dans la théorie.

.../...

(*) Le conditionnement est fait ici par rapport à des variables aléatoires, et non pas, plus "abstraitement", par rapport à des tribus.

Doob indique que, si la famille de v.a. est en outre équiintégrable, la martingale est fermée (l'équiintégrabilité étant à mettre dans les hypothèses, pour n tendant vers $+\infty$, et étant automatique pour n tendant vers $-\infty$). Il donne également une sorte de réciproque. Enfin il donne une inégalité plus précise dans le cas d'une martingale fermée $x_1, x_2, \ldots, x$ pour laquelle $E\,|x\log|x|| < \infty$, en s'inspirant d'un article de N. Wiener.

Le § 2 "One-parameter families of chance variables", où la condition de martingale n'est pas supposée, vise en 12 pages à préciser rigoureusement ce qu'on peut entendre par des "termes aussi descriptifs" que "continuité" ou " bornitude" des trajectoires : il approfondit ainsi et simplifie des travaux antérieurs en se plaçant d'emblée, dans le cadre de Kolmogorov et de Khintchine.

Le § 3 est consacré aux martingales à temps continu. L'objectif déclaré est d'étendre à ce cas les propriétés de régularité "bien connues" établies par P. Lévy pour les processus à accroissements indépendants (PAI).

"Le résultat principal de cette section est qu'il existe un espace Ω d'un processus stochastique avec la mesure de probabilité donnée dont les éléments sont continus sauf peut-être sur un ensemble dénombrable [D] de valeurs de t e[t] continus partout à droite, sauf peut-être aux points de D_1 [$\subset$ D] (la limite $\lim_{t'\downarrow t} x(t')$ existe même sur D_1)" . p. 478 (D et D_1 sont discutés dans le texte).

Il note que ces résultats sont moins forts que dans le cas des PAI, "où les seules discontinuités sont des sauts", et ajoute : "Une différence de ce genre pouvait être suspectée car la propriété $\mathcal{E}$ est essentiellement dyssimétrique en t". Il y a là encore une étape beaucoup plus délicate à franchir que ce court résumé ne le laisserait supposer.

.../...

Le cadre est donc prêt pour attaquer une variété de problèmes nouveaux en théorie des probabilités, mais on remarquera qu'il n'y a dans cet article ni "théorème d'arrêt" (*), ni allusion à l'analyse stochastique, ni sur - (ni sous) martingale. Tout cela viendra 10 ans après. D'ailleurs, apparemment, il y a peu de travaux sur les martingales dans les années 40 (que ce soit par J.L. Doob ou par d'autres). Nous étudierons cette question dans un article ultérieur. (**)

Signalons seulement ici l'utilisation par Doob de la condition de martingale, toujours appelée condition &, pour l'obtention de théorèmes de régularité relatifs aux processus de Markoff à temps continu $x(t)$, à valeurs dans un espace d'états dénombrable : [D3], à partir de la page 57. :

"We shall use a somewhat indirect method in examining in more detail the transitions of the system, that is, the discontinuities of $x(t)$. This method has the advantage of exhibiting analytically the relation between the regularity of the matrix function $p_{i,j}(t)$ and the discontinuities of $x(t)$" p. 56.

L'idée de départ consiste à remarquer que si T est fixé, alors, pour $0 < t < T$, $P(X_T = j \,/\, x(t))$ est une martingale ; il améliore ainsi des résultats de Doeblin et de Feller.

.../...

(*) Le résultat de Halmos est différent du théorème d'arrêt, bien entendu

(**) voir déjà l'appendice historique de J.L. Doob : [D4]

VI - REMARQUES ET QUESTIONS

Le descriptif que nous venons de faire laisse de nombreuses questions dans l'ombre, à propos de l'émergence du concept de martingale, par exemple sur les liens avec l'analyse, mais aussi sur le rapport avec le mouvement des idées et celui de la vie sociale :

Contentons-nous ici de quelques remarques tout juste embryonnaires.

M. Loève estime [19] que la notion de martingale naît pour des raisons internes à la théorie des probabilités. C'est vite dit. Certes on ne peut nier ces raisons internes qui sont évidentes. Mais la théorie des fractions continues, l'intégration dans les espaces de dimension infinie, la théorie ergodique, voire l'analyse harmonique qui servent de points de départ, d'intuition, d'appels d'air à de nombreuses contributions ne sont pas purement internes à la théorie des probabilités. Et la discussion sur les fondements encore moins ! Non seulement elle a des aspects philosophiques évidents, mais surtout c'est toute la question du champ de pertinence et d'application du calcul des probabilités qui est en jeu : cela est on ne peut plus clair chez von Mises ou chez Wald.

En tout cas, il nous semble peu judicieux d'isoler artificiellement l' "interne" de l' "externe" dans l'élaboration d'un concept, comme si cette distinction avait un sens absolu. Donnons-en une illustration à propos des martingales : l'idée de temps aléatoire qui traverse toute cette histoire (par exemple notion de "système de jeu") puis celle de temps d'arrêt, avec les "sections" de P. Lévy, avec les changements de temps dans les processus ou avec l'analyse séquentielle de Wald) n'est-elle pas étroitement liée à la

.../...

conception du temps en mathématiques et ailleurs qu'on peut avoir dans les années 1930-1940 ? Peut-on vraiment comprendre l'histoire des martingales sans se pencher sur une telle question ?

Nous aborderons ces questions et quelques autres dans un prochain travail. Nous espérons entre temps recevoir les suggestions les plus diverses des lecteurs.

.../...

BIBLIOGRAPHIE COMPLEMENTAIRE

[1] Article "Martingale" dans l' "Encyclopaedia Universalis".

[2] Article "Martingale" dans l' "Encyclopedic Dictionary of Mathematics" n° 263 (édition de 1977)

[3] P. HALL - C. HEYDE : "Martingale limit theory and its applications", Academic Press (1980)

[4] B. BRU : Exposé sur l'histoire des martingales. Séminaire d'histoire des mathématiques de l'Institut Henri-Poincaré - Notes non publiées.

[5] H. CRAMER : "Half a century with probability theory : Some personal recollections" Ann. Proba 4 (1973), p. 509-546.

[6] J. BONITZER : "Philophie du hasard", Editions Sociales (1984).

[7] E. BOREL : "Les probabilités dénombrables et leurs applications arithmétiques", Rend. Circ. Mat. Palermo 28 (1909), p. 247-271.

[8] F. BERNSTEIN : "Über eine Anwendung der Mengen lehre auf ein aus der Theorie der säkularen Störungen herrührendes Problem", Mat. Ann. 71 (1911), p. 417-439.

[9] A.N. KOLMOGOROFF : "Sur la loi des grands nombres", Rend. Accad. Lincei 9 (1929), p. 470-474.

[10] E. BOREL : "Sur une propriété de probabilités relatives aux fractions continues", Math. Ann. 72 (1912), p. 578.

[11] A. YA. KHINTCHINE : "Continued fractions", The University of Chicago Press (1964).

.../...

[12] B. JESSEN : "The theory of integration in a space of an infinite number of dimensions", Acta Math 63 (1934), p. 249-323.

[13] M. LOEVE : "Etude asymptotique des sommes de variables aléatoires liées", Journ. de Math. 24 (1945), p. 249-318, et notes aux CRAS citées.

[14] M. FRECHET : "Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités", Colloque consacré à la théorie des probabilités (Genève), Actualités Scientifiques et Industrielles, Hermann, n° 735 (1938), p. 23-55

[15] C. DELLACHERIE : "Nombres au hasard", La Gazette des Mathématiciens, octobre 1978.

[16] K. MENGER : "Ergebnisse eines mathematischen Kolloquiums", B.G. Teubner (8 cahiers entre 1931 et 1938) ; "Selected Papers in Logic and Foundations, Didactics, Economics", Reidel (1979) ; et "The formative years of A Wald and his work in geometry", Ann. Math. Stat. 23 (1952), p. 14-20.

[17] I.G. PETROWSKY : "Zur ersten Randwertaufgabe der Wärmecleitungsgleichung" Compositio Math. 1 (1935), p. 383-419.

[18] P. R. HALMOS : "Invariants of certain stochastic transformations : the mathematical theory of gambling systems", Duk Math. J. 5 (1939), p. 461-478.

[19] M. LOEVE : partie relative aux probabilités dans "Abrégé d'histoire des mathématiques" (Dieudonné et Cie), Hermann (1978).

[20] C.P. SCHNORR : "Zufälligkeit und Wahrscheinlichkeit", Lecture Notes in Math. n°218 , Springer (1970).

N.B. Les références de (et sur) S. Bernstein, P. Lévy, J. Ville et J.L. Doob sont incluses dans le corps du texte.