Paul Lévy et les fondements du calcul des probabilités

Michel Willem
Professeur à l'Université catholique de Louvain

I. Introduction

Nous lisons dans les *Éléments d'histoire des mathématiques* de Nicolas Bourbaki que la théorie des probabilités «autrefois prétexte à devinettes et à paradoxes» est «devenue une branche de la théorie de l'Intégration depuis son axiomatisation par Kolmogoroff» (1). Beaucoup d'auteurs partagent ce point de vue. Par exemple Adams et Guillemin affirment que «probability theory become a respectable mathematical discipline only in the early 1930» (2).

Or personne n'a jamais écrit que le calcul différentiel et intégral est devenu respectable depuis que Cauchy et Weierstrass lui ont apporté un fondement rigoureux.

Comme Paul Lévy a contribué de manière essentielle au calcul des probabilités avant l'axiomatisation de Kolmogoroff, il nous a semblé intéressant d'étudier son œuvre entre 1920 et 1925.

L'exposé comprend quatre parties : une description sommaire du calcul des probabilités avant 1920, la biographie de Paul Lévy, ses travaux entre 1920 et 1925 et la question du fondement du calcul des probabilités.

(1) N. BOURBAKI, *Éléments d'histoire des mathématiques*, p. 258.
(2) M. ADAMS - V. GUILLEMIN, *Measure theory and probalility*, p. VII.

II. Histoire du calcul des probabilités

Le calcul des probabilités est né de l'étude des jeux de hasard (pile ou face, dés, cartes, roulette). Les premiers traités sont dus à Jérôme Cardan (1501-1576) et à Galilée (1564-1642). Toutefois, Pascal (1623-1662) et Fermat (1601-1665) sont reconnus comme les véritables fondateurs du calcul des probabilités. (Observons que le *Traité de la roulette* de Pascal est consacré à une courbe nommée cycloïde et non au jeu de la roulette).

Dans le cas des jeux de hasard, il s'agit de probabilités finies, car le nombre de cas possibles est fini, et, pour des raisons de symétrie, les différents cas sont également probables.

Nous ne décrirons que trois résultats classiques : 1) la loi des grands nombres de Jacques Bernoulli ; 2) le théorème de la limite centrale et 3) la loi forte des grands nombres de Borel.

Nous considérons n variables aléatoires $X_1, \ldots, X_n$ indépendantes, mais soumises à la même probabilité. En nous limitant au jeu de pile ou face, nous associons pile à -1 et face à +1. Nous pouvons alors énoncer les résultats suivants :

Loi des grands nombres

Pour chaque $\epsilon > 0$,

$$\lim_{n\to\infty} p[\,|\frac{1}{n}\,(X_1 + \cdots + X_n)| > \epsilon\,] = 0.$$

La loi des grands nombres signifie qu'une longue suite de lancers où la fréquence de pile s'éloignerait sensiblement de 1/2 est très peu probable, ce qui ne signifie pas impossible. En effet, une suite constituée uniquement de piles est aussi probable que toute autre suite donnée à l'avance ! Par contre, le nombre de suites telles que la fréquence s'éloigne sensiblement de 1/2 devient très petit par rapport au nombre total de suites.

Théorème de la limite centrale

Pour tout nombre réel c,

$$\lim_{n\to\infty} p\,[\,\frac{1}{\sqrt{n}}\,(X_1 + \cdots + X_n) < c\,] = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} \exp\,(\frac{-x^2}{2})\,dx.$$

Le théorème de la limite centrale précise la loi des grands nombres. Il est dû à de Moivre et à Laplace.

Loi forte des grands nombres

Pour presque toute suite (X_n),

$$\lim_{n\to\infty} \frac{1}{n} (X_1 + \cdots + X_n) = 0.$$

L'énoncé du théorème précédent fait intervenir la mesure de Lebesgue. Nous associons pile à 0 et face à 1. À chaque suite (X_n) de variables aléatoires correspond alors le développement binaire

$$0.\ X_1 X_2 \cdots X_n \cdots$$

d'un point de l'intervalle [0,1]. Au sens de la mesure de Lebesgue, l'ensemble des nombres qui ne vérifient pas

$$\lim_{n\to\infty} \frac{1}{n} (X_1 + \cdots + X_n) = \frac{1}{2}$$

est négligeable.

En introduisant la théorie de la mesure, Borel révolutionne le calcul des probabilités.

III. Biographie de Paul Lévy

Paul Lévy est né à Paris le 15 septembre 1886. Son grand-père était professeur de mathématique. Son père, Lucien Lévy, fut pendant 23 ans examinateur de l'École polytechnique. L'une de ses filles, Madame Laurent Schwartz, est une mathématicienne connue.

Paul Lévy a obtenu au concours général un premier prix de thème grec et un prix de mathématique. Reçu premier à l'École normale et second à l'École polytechnique, il choisit cette dernière. Sorti dans le corps des mines, il s'oriente vers la recherche mathématique. Il soutient, en 1911, une thèse sur l'analyse fonctionnelle devant Jacques Hadamard, Émile Picard et Henri Poincaré. Hadamard apprécie beaucoup ses travaux. En 1919, Paul Lévy est chargé du cours Peccot. En 1920, il est nommé professeur d'analyse à l'École polytechnique. Il y enseignera jusqu'à l'âge de 73 ans en 1959. Les lois

raciales, promulguées par le régime de Vichy, l'ont privé de la chaire entre 1940 et 1944. Laurent Schwartz, beau-fils de Paul Lévy, sera professeur d'analyse à l'École polytechnique de 1959 à 1980.

C'est par hasard que Paul Lévy s'est intéressé au calcul des probabilités. En 1919, Carvallo, directeur des études à l'École polytechnique, lui demande de faire trois conférences sur la loi de Gauss et la théorie des erreurs. En consultant les traités classiques de Bertrand, de Poincaré et de Borel, Lévy ne trouve aucune démonstration correcte. Il introduit alors les notions de fonction de répartition et de fonction caractéristique qui lui permettent de démontrer le théorème de la limite centrale. Le texte des conférences est perdu, mais les résultats sont repris dans trois notes aux *Comptes rendus de l'Académie des Sciences de Paris* publiées en 1922 et dans un traité de calcul des probabilités publié en 1925.

Paul Lévy s'intéresse ensuite aux processus aléatoires et au mouvement brownien. Tous ces travaux lui valent une renommée mondiale.

En 1967, à 81 ans, il publie un magistral article sur les mathématiques et l'Académie des sciences de Paris entre 1866 et 1966. D'après Fréchet, «par sa hauteur de vue, il se compare avec avantage à l'histoire des mathématiques de Bourbaki» (3).

En 1970, paraît *Quelques aspects de la pensée d'un mathématicien*. Dans ce livre extraordinaire, Paul Lévy raconte sa vie mathématique et expose ses opinions philosophiques. Il s'éteint le 15 décembre 1971.

Les 10 livres et les 268 articles de Paul Lévy ont été écrits dans des conditions très favorables. Son cours à l'École polytechnique n'avait lieu qu'une année sur deux et comprenait quarante conférences. Bien que très populaire, il n'avait pratiquement pas de contact avec les élèves et très peu de réunions. On devine que des technocrates vigilants ont mis fin à cette situation scandaleuse.

IV. Travaux de Paul Lévy entre 1920 et 1925

Les trois notes au *Comptes rendus de l'Académie des Sciences de Paris* de 1922 sont développées dans le *Calcul des probabilités* publié en 1925. Nous suivrons cet ouvrage.

(3) M. FRECHET, *Lettres à Paul Lévy*, p. 55.

Par définition, «une loi de probabilité à une variable peut être représentée par une répartition sur l'axe des x de masses positives et de sommes égales à l'unité... Analytiquement, nous définirons une loi de probabilité, ou la répartition de masse correspondante, par la fonction F(x) qui représente la somme des masses situées dans l'intervalle] - ∞, x [».

La *fonction de répartition F* de X est donc définie sur $\mathbb{R}$ par

$$F(x) := p\ (X < x).$$

Elle jouit de trois propriétés :

(F1) F est croissante,

(F2) F est continue à gauche,

(F3)
$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

Pour toute fonction continue $u : \mathbb{R} \to \mathbb{R}$ nulle en dehors de [-a,a], nous pouvons calculer l'espérance

$$E(u) := \int_{\mathbb{R}} u(x)\, dF(x)$$

en passant à la limite sur les sommes de Riemann-Stieltjes :

$$S_n := \sum_{k \in \mathbb{Z}} u\left(\frac{k}{2^n}\right) \left(F\left(\frac{k+1}{2^n}\right) - F\left(\frac{k}{2^n}\right) \right).$$

Selon un résultat classique de Lebesgue, F est la somme de 3 fonctions croissantes :

$$F = F_d + F_c + F_s.$$

La fonction F_d est constante en dehors des points de discontinuité de F et correspond aux masses discrètes. La fonction F_c est absolument continue et sa dérivée F'_c correspond à la densité. La fonction singulière F_s ne varie que sur un ensemble négligeable.

Alors que les masses discrètes et les densités étaient bien connues en probabilité, les mesures singulières sont tout à fait nouvelles. Paul Lévy parvient ainsi à la notion générale de probabilité sur la droite réelle.

La *fonction caractéristique* φ de X est définie sur $\mathbb{R}$ par

$$\varphi(y) \ := E(e^{ixy}) = \int_{\mathbb{R}} e^{ixy}\, dF(x).$$

Le lecteur reconnaît, au signe près, la transformée de Fourier. Le théorème de Lévy affirme que, si pour tout $y \in \mathbb{R}$,

$$\varphi_n(y) := \int_{\mathbb{R}} e^{ixy}\, dF_n(x) \to \varphi(y) := \int_{\mathbb{R}} e^{ixy}\, dF(x)$$

alors, pour toute fonction bornée et continue u,

$$\int_{\mathbb{R}} u(x)\, dF_n(x) \to \int_{\mathbb{R}} u(x)\, dF(x).$$

La convergence des fonctions caractéristiques assure donc la convergence étroite des mesures. (En fait, le théorème de Levy est encore plus fort).

Soient X et Y des variables aléatoires indépendantes, la fonction de répartition et la fonction caractéristique de $X + Y$ sont données par

$$F_{X+Y} = F_X \star F_{Y,\ \Phi X+Y} = \varphi_X \varphi_Y\ .$$

En effet, la transformation de Fourier échange la convolution et la multiplication.

Les outils précédents permettent à Lévy de prouver le théorème de la limite centrale. Considérons une suite de variables aléatoires indépendantes et identiquement réparties (X_n) telle que, après normalisation,

$$E(X_1) = 0, E(\ (X_1 - E(X_1)\ )^2\ ) = 1.$$

Alors, pour tout nombre réel c,

$$\lim_{n\to\infty} p[\frac{1}{\sqrt{n}}\ (X_1 + \cdots + X_n) < c] \ = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}}\ \exp(\frac{-x^2}{2})dx.$$

En fait, Lévy retrouve en les généralisant des résultats de l'école russe. En 1934, il obtient les conditions nécessaires et suffisantes pour la convergence, après normalisation, vers la loi de Laplace-Gauss.

L'approche indiquée conduit naturellement à l'axiomatique de Kolmogoroff et aux distributions de Schwartz.

Considérons un ensemble de partie B d'un ensemble Ω. Supposons que B contienne Ω et ϕ et soit stable par intersection, passage au complémentaire et union dénombrable.

Une *mesure de probabilité* sur Ω associe à chaque élément B de B un nombre réel $\mu(B)$ tel que

(M1) $\mu(\phi) = 0$, $\mu(\Omega) = 1$.

(M2) Pour toute suite (B_n) de parties disjointes de B,

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n).$$

L'axiome (M2) n'est autre que le principe des probabilités totales. Une remarque du *Calcul des probabilités* de Paul Lévy préfigure l'axiomatique de Kolmogoroff :

> «Dans le cas général, en empruntant le langage de la théorie des ensembles abstraits, on peut dire que le principe des probabilités totales exprime que la probabilité totale est une *fonctionnelle additive* de l'ensemble dont chaque élément est un groupe de cas possibles ; elle doit être toujours positive ou nulle, et avoir la valeur unité pour le groupe de tous les cas possibles. Se donner une loi de probabilité, c'est se donner les éléments suffisants pour la définition de la fonctionnelle ; s'ils sont surabondants, ils devront être compatibles» (4).

Si $\Omega = \mathbb{R}$, la relation entre la mesure de probabilité et la fonction de répartition est donnée par la formule $F(x)= \mu\,(\,]-\infty, x[\,)$.

Fréchet découvre la mesure abstraite des ensembles en 1915 dans un article prophétique qui ne figure pas dans la bibliographie de l'histoire des mathématiques de Bourbaki.

Or, depuis 1921, Fréchet s'intéresse aux probabilités. Pourquoi n'a-t-il pas axiomatisé cette théorie avant Kolmogoroff ?

(4) P. LÉVY, *Calcul des probabilités*, p. 11.

Considérons maintenant l'espace $K(\mathbb{R})$ des fonctions continues sur $\mathbb{R}$ et nulles en-dehors d'un intervalle [-a,a] (qui dépend de la fonction). *Une mesure de Radon positive* associe à chaque fonction u de $K(\mathbb{R})$ un nombre $<\mu, u>$ tel que

(R1) $<\mu, \alpha u + \beta u> = \alpha <\mu, u> + \beta <\mu, u>$.

(R2) $u \geq 0 \Rightarrow <\mu, u> \geq 0$.

Il est clair que l'intégrale de Riemann-Stieltjes associée à une fonction de répartition définit une mesure de Radon positive. D'après un théorème célèbre de Riesz, toute mesure de Radon positive telle que $<\mu, 1> = 1$ est l'intégrale de Riemann-Stieltjes d'une fonction de répartition. Ainsi apparaît la dualité qui conduit naturellement aux distributions définies par Schwartz en 1945. En effet, une distribution est une forme linéaire continue sur un certain espace fonctionnel. Sans quitter le cadre de l'intégrale de Riemann-Stieltjes, nous observons que la mesure de Dirac δ associée à la fonction de répartition

$$F(x) \begin{array}{ll} := 0, & x < 0, \\ := 1, & x \geq 0, \end{array}$$

a pour fonction caractéristique

$$\varphi(y) = \int_{\mathbb{R}} e^{ixy}\, dF(x) = 1.$$

Nous retrouvons ainsi la formule classique $\hat{\delta} = 1$. La théorie des distributions permet de dériver les mesures de Radon et de prouver, par exemple, que

$$\hat{\delta}^{(k)} = (iy)^k.$$

V. Fondements du calcul des probabilités

Nous distinguerons les axiomes et leur interprétation.

Paul Lévy est résolument platonicien. Pour lui, l'axiome du choix est évident et l'hypothèse du continu «vraie pour Dieu, indémontrable pour

l'homme» (5). Il renforce l'exigence classique d'évidence des axiomes : «Les travaux où l'on se contente d'énoncer des axiomes, qui souvent ne sont pas plus évidents que le théorème que l'on veut démontrer, et d'en déduire la démonstration de ce théorème, sont sans intérêt» (6).

Tout autre est, par exemple, la doctrine de Buffon :

> «Nous avons fait les suppositions, nous les avons combinées de toutes les façons, ce corps de combinaison est la science mathématique ; il n'y a rien dans cette science que nous n'y avons mis, et les vérités qu'on en tire ne peuvent être que des expressions différentes sous lesquelles se présentent les suppositions que nous avons employées [...] Et comme les définitions sont les seuls principes sur lesquels tout est établi, et qu'elles sont arbitraires et relatives, toutes les conséquences qu'on en peut tirer sont également arbitraires et relatives... Ce qu'on appelle vérités mathématiques se réduit donc à des identités d'idées et n'a aucune réalité ; nous supposons, nous raisonnons sur nos suppositions, nous en tirons des conséquences, nous concluons, la conclusion ou dernière conséquence est une proposition vraie relativement à notre supposition, mais cette vérité n'est pas plus réelle que la supposition elle-même» (7).

Cet extrait du *Premier discours* de *L'Histoire naturelle* date de 1749. L'histoire semble donner raison à Buffon. L'axiomatique de Kolmogoroff n'est pas l'explication *a priori* de la notion de probabilité qui s'est longtemps passé d'un tel fondement. Son mérite est de permettre au calcul des probabilités d'aller de l'avant en mettant entre parenthèse les problèmes philosophiques. L'axiomatique se justifie donc *a posteriori* par son succès. Elle montre aussi clairement que le calcul des probabilités appartient aux mathématiques et non aux sciences de la nature.

En ce qui concerne l'interprétation des probabilités, Paul Lévy est résolument subjectiviste : «Quand je dis que deux événements sont également probables, cela veut dire que je n'ai aucune raison de m'attendre à l'un

(5) P. LÉVY, *Quelques aspects de la pensée d'un mathématicien*, p. 221.
(6) P. LÉVY, *Calcul des probabilités*, p. 12.
(7) BUFFON, *Premier discours*, p. 247.

plutôt qu'à l'autre. C'est là une notion essentiellement subjective, et la science reposant sur cette matière est une science de raisonnement» (8). Rappelons que, pour Laplace, le hasard n'est que «l'expression de l'ignorance où nous sommes des véritables causes» (9). Paul Lévy, qui est strictement déterministe, se situe dans la tradition de Laplace. Selon l'interprétation objective, qui remonte à Cournot, la probabilité représente une réalité objective indépendante de l'état de nos connaissances.

Ici encore citons le *Premier discours* de Buffon :

> «Une suite de faits semblables ou, si l'on veut, une répétition fréquente et une succession non interrompue des mêmes événements, fait l'essence de la vérité physique : ce qu'on appelle vérité physique n'est donc qu'une probabilité, mais une probabilité si grande qu'elle équivaut à une certitude [...]. L'évidence mathématique et la certitude physique sont donc les deux seuls points sans lesquels nous devons considérer la vérité; dès qu'elle s'éloignera de l'une ou de l'autre, ce n'est plus que vraisemblance et probabilité. Examinons donc ce que nous pouvons savoir de science évidente ou certaine, après quoi nous verrons ce que nous ne pouvons connaître que par conjecture, et enfin ce que nous devons ignorer» (10).

(8) P. LÉVY, *Le raisonnement et l'expérience dans les fondements des mathématiques*, p. 330.

(9) LAPLACE, *Essai philosophique sur les probabilités*, p. 2.

(10) BUFFON, *Premier discours*, p. 248.

Bibliographie

ADAMS (M.) - GUILLEMIN (V.), *Measure theory and probability*. – Monterey : Wadsworth and Brooks, 1986.

BOURBAKI (N.), *Éléments d'histoire des mathématiques*. – Paris : Hermann, 1960.

BUFFON (G.L.), *Premier discours*, dans *Buffon, 1788-1988*. – Paris : Imprimerie Nationale, 1988. – pp. 219-253.

DIEUDONNÉ (J.), *Notice nécrologique sur Paul Lévy*, in *Comptes rendus de l'Académie des Sciences de Paris : Vie Académique*, t. CCLXXIV, 1972, pp. 137-144.

FRECHET (M.), *Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait*, in *Bulletin de la Société Mathématique de France*, t. XLIII, 1915, pp. 248-265.

—, *Lettres à Paul Lévy*, in *Cahiers du Séminaire d'Histoire des Mathématiques*, t. I, 1980, pp. 54-57.

LAPLACE (P.S.), *Essai philosophique sur les probabilités*. – Paris : Courcier, 1814.

LÉVY (P.), *Calcul des probabilités*. – Paris : Gauthier-Villars, 1925.

—, *Le raisonnement et l'expérience dans les fondements des mathématiques*, in *Scientia*, t. XLVII, 1930, pp. 325-334.

—, *Les mathématiques*, dans *Académie des sciences : Troisième centenaire, 1666-1966*. – Paris : Gauthiers-Villars, 1967. – pp. 143-212.

—, *Quelques aspects de la pensée d'un mathématicien*. – Paris : A. Blanchard, 1970.

LOEVE (M.), *Paul Lévy, 1886-1971*, in *Annals of Probability*, t. I, 1973, pp. 1-18.

SCHWARTZ (L.), *Quelques réflexions et souvenirs sur Paul Lévy*, in *Astérisque*, t. CLVII-CLVIII, 1988, pp. 13-28.