QUELQUES GRANDES THEORIES MATHEMATIQUES INTERVENANT DANS LE CALCUL DES PROBABILITES

par E. Mourier

Maître de Conférences à la Faculté des Sciences de Poitiers

Le calcul des probabilités a souvent ses méthodes qui lui sont propres et ceci, en particulier, est dû au fait que les probabilités sont des nombres compris entre zéro et un ; toutefois, il fait également intervenir de grandes théories mathématiques, et des plus modernes. Tout d'abord, il résulte de l'axiomatique même du calcul des probabilités, développée dans la conférence précédente, que toute loi de probabilité *est* une mesure finie ; les liens entre le calcul des probabilités et la théorie de la mesure sont alors si clairement en évidence qu'il est inutile d'insister sur ce point. Soulignons seulement le fait que c'est en tant que fonction d'ensembles complètement additive, définie sur une σ-algèbre (ou corps de Borel), de sous-ensembles d'un ensemble E, qu'une mesure s'identifie à une loi de probabilité sur E ; que l'existence et la définition d'une loi de probabilité sur E sont indépendantes de la définition d'une topologie sur E.

Un des premiers, et des plus importants problèmes qui se soient posés en calcul des probabilités, est celui de l'étude de la fréquence de la réalisation d'un événement A au cours de n expériences indépendantes, le résultat de chaque expérience étant soit la réalisation de A, de probabilité p, soit sa non-réalisation, de probabilité $q = 1 - p$ (cas de Bernoulli). Si on associe à chaque expérience la variable aléatoire X prenant les valeurs 1 ou 0 selon que A est réalisé ou non, la fréquence de A au cours des n expériences indépendantes est :

$$F = \frac{1}{n}[X_1 + X_2 ... + X_n].$$

Or :

$$EX = p$$
$$EX^2 = p$$

d'où :

$$E(F) = \sum_i \frac{EX_i}{n} = p.$$

$$\sigma^2(F) = E[F - E(F)]^2 = \frac{1}{n^2}\sum_i E(X_i - EX_i)^2 = \frac{p(1-p)}{n} = \frac{pq}{n}.$$

Il résulte donc de l'inégalité de Bienaymé :

$$\Pr[|X - EX| \geqslant \tau] \leqslant \frac{\sigma^2}{\tau^2}.$$

que $\Pr[|F - p| \geqslant \tau] \leqslant \dfrac{n\tau^2}{pq}$ tend vers zéro lorsque $n \to +\infty$, c'est-à-dire que F converge en probabilité vers p.

Ce résultat constitue la première et la plus simple « *loi des grands nombres* ». Le même raisonnement s'étend immédiatement à une variable aléatoire X quelconque ayant des moments du 1^er^ et du 2^e^ ordre.

$EX = m$, $E(X - EX)^2 = \sigma^2$, si $X_1, \ldots, X_n$ sont n variables aléatoires indépendantes de même loi que X,

$Y_n = \dfrac{X_1 + \ldots + X_n}{n}$ converge en probabilité vers m, en effet

$EY_n = m$ et $\sigma^2 Y_n = E[Y_n - EY_n]^2 = E[Y_n - m]^2 = \dfrac{\sigma^2}{n}$

et par conséquent :

$$\Pr\left[|Y_n - m| \geqslant \tau^2\right] \leqslant \frac{\sigma^2}{n\tau^2}$$

donc Y_n converge en probabilité vers m.

Remarquons d'ailleurs que, pour démontrer cette convergence en probabilité, nous utilisons la propriété plus stricte :

$\sigma^2 Y_n = E[Y_n - m]^2 = \dfrac{\sigma^2}{n}$ tend vers zéro lorsque $n \to +\infty$,

qui, par définition même, exprime que la variable aléatoire Y_n *converge en moyenne quadratique vers* m ; ce résultat est la « *loi des grands nombres en moyenne d'ordre 2* ».

En se rappelant qu'une variable aléatoire X est une application mesurable, $x(u)$, de l'espace probabilisé des épreuves ($\mathcal{U}$, S μ) dans l'espace R des nombres réels, la convergence en moyenne quadratique d'une suite de variables aléatoires est la convergence forte dans l'espace $L^2(\mathcal{U}, \mu)$ des fonctions réelles $x(u)$, $u \in \mathcal{U}$, de carré sommable par rapport à la mesure μ, la norme d'un élément $x \in L^2(\mathcal{U}, \mu)$ étant définie par :

$$\|x\| = \left[\int_{\mathcal{U}} |x(u)|^2 \, d\mu\right]^{1/2}$$

L'énoncé et la démonstration des lois des grands nombres rappelées ci-dessus semblent propres au calcul des probabilités. Un des problèmes essentiels du calcul des probabilités classique, c'est-à-dire concernant les variables aléatoires numériques, a été de préciser de telles lois des grands nombres et de remplacer les conditions énoncées ci-dessus : variables aléatoires indépendantes, de même loi, ayant des moments du 1^er^ et du 2^e^ ordre, par des conditions plus faibles.

On dit que des éléments aléatoires X_n forment une suite strictement stationnaire si, quel que soit l'entier positif s, quels que soient les entiers $n_1, n_2, \ldots, n_s$ et quel que soit l'entier h, la loi de probabilité de l'élément complexe $(X_{n_1+h}, X_{n_2+h}, \ldots, X_{n_s+h})$ ne dépend pas de h. Kolmogorov a montré que si des variables aléatoires X_i forment une suite strictement stationnaire et si EX_i existe (et alors EX_i ne dépend pas de i, du fait de la stationnarité), la moyenne arithmétique $Y_n = \dfrac{X_1 + \ldots + X_n}{n}$ tend, lorsque $n \to +\infty$, presque-sûrement vers une variable aléatoire limite Y telle que EY existe et que $EY = EX_i$.

Dans le cas particulier où les X_i sont des variables aléatoires indépendantes et de même loi, Y se réduit à l'élément certain EX_i.

Ce théorème constitue la « *loi forte des grands nombres* ».

On s'aperçoit alors que résultats et méthodes font intervenir *la théorie ergodique*, dont l'origine ne se trouve pas en calcul des probabilités, mais dans l'étude des systèmes dynamiques.

La première *hypothèse* ergodique est relative au comportement des trajectoires, dans l'espace des phases, d'un système dynamique ; elle fut énoncée par Maxwell en 1850 : pour les systèmes ayant un très grand nombre de degrés de liberté, par exemple gaz de la théorie cinétique, la moyenne statistique est égale à la moyenne temporelle. Les premiers théorèmes ergodiques furent démontrés par G.-D. Birkhoff, I. Carleman, D.-O. Koopman, J. von Neumann, etc..., au sujet du comportement de moyennes de transformations. Depuis, cette théorie ergodique a connu un développement considérable, qu'il est impossible de résumer ici ; j'indiquerai seulement deux théorèmes essentiels, à savoir :

Théorème ergodique de Birkhoff : Soit Ω un ensemble d'éléments ω, pourvu d'une mesure bornée μ, on supposera $\mu(\Omega) = 1$. Soit φ une transformation (application biunivoque de Ω sur lui-même) conservant la mesure μ. Soit $\mathcal{G}$ l'espace de Banach des fonctions numériques réelles, $f(\omega)$, sommables μ sur Ω, avec la norme :

$$\|f\| = \int_\Omega |f(\omega)|\mu(d\omega).$$

Soit T l'opération linéaire dans $\mathcal{G}$ définie par :

$$Tf = f[\varphi(\omega)],$$

T est une opération bornée, $\|T\| = 1$.

Théorème : Pour presque tout ω (au sens de μ),

$$\lim_{n \to +\infty} \frac{1}{n} (T^1 + T^2 + \ldots + T^n) \quad f(\omega) = \lambda(\omega)$$

existe, et $\lambda \in \mathcal{G}$.

Remarquons que les entiers $1, 2, \ldots, n, \ldots$ peuvent s'interpréter comme des temps (temps discret), et alors la moyenne arithmétique est la moyenne temporelle.

Théorème ergodique de Yosida et Kakutani : Soit T un opérateur linéaire borné qui transforme un espace de Banach $\mathcal{X}$ en lui-même, et tel que T^n soit uniformément borné, c'est-à-dire qu'il existe un nombre C, indépendant de n, tel que $\|T^n\| \leqslant C$, $n = 1, 2, \ldots$ Si pour tout $x \in \mathcal{X}$, la suite

$$\{x_n\} \text{ où } x_n = \frac{1}{n}\,[T + T^2 + \ldots + T^n]x, n = 1, 2, \ldots,$$

contient une sous-suite qui converge *faiblement* vers un point $\bar{x} \in \mathcal{X}$, la suite $\{x_n\}$ converge *fortement* vers $\bar{x}$.

Il faut noter que le théorème de Yosida et Kakutani est de nature beaucoup plus général que celui de Birkhoff, puisque les opérateurs linéaires T sont seulement astreints à ce que T^n est uniformément borné, alors que les opérateurs de Birkhoff sont de type très particulier et de norme 1 ; d'autre part, le théorème de Yosida et Kakutani ne fait pas intervenir de mesure.

Kolmogorov eut le premier le mérite de montrer qu'il y a équivalence entre la loi forte des grands nombres pour des variables aléatoires et le théorème de Birkhoff.

Les recherches modernes en calcul des probabilités ont principalement pour objet de formuler des « lois des grands nombres », c'est-à-dire d'étudier le comportement asymptotique de

$$Y_n = \frac{1}{n}\,(X_1 + \ldots + X_n),$$

non plus seulement pour des variables aléatoires, mais pour des éléments aléatoires X de nature quelconque et sous des hypothèses aussi générales que possible. La nature même du problème impose alors de considérer des éléments aléatoires à valeurs dans un espace vectoriel topologique ; d'autre part, une des généralisations les plus importantes de la notion de variable aléatoire est celle de fonction aléatoire qui intervient dans de nombreux problèmes tant pratiques que théoriques : problèmes posés par la théorie des télécommunications, de l'information, etc... Or, un moyen commode d'étudier une fonction aléatoire est de la considérer comme élément aléatoire à valeurs dans un espace approprié, dans de nombreux cas un espace L_p de Lebesgue. Par conséquent, la théorie des espaces vectoriels topologiques, et en particulier des espaces de Banach, intervient naturellement dans toute étude générale de lois des grands nombres. D'autre part, l'hypothèse de l'existence de EX_i intervenant dans la loi forte des grands nombres, pour des variables aléatoires, nécessite la définition de l'espérance mathématique d'un élément aléatoire à valeurs dans un espace vectoriel topologique $\mathcal{X}$, et ainsi intervient la théorie de l'intégration d'une fonction définie sur un ensemble $\mathcal{U}$ quelconque, à valeurs dans un tel espace $\mathcal{X}$.

Lorsqu'on considère des éléments aléatoires à valeurs dans un espace de Banach $\mathcal{X}$, il convient de se limiter aux applications $x(u)$ de $\mathcal{U}$ dans $\mathcal{X}$ telles que $< x^*, x(u) >$ soit mesurable pour tout $x^* \in \mathcal{X}^*$ ($\mathcal{X}^*$ dual fort de $\mathcal{X}$) ; on dira alors que $X = x(u)$ est un L-élément aléatoire. Alors, par définition, l'espérance mathématique de X, EX, est l'élément de $\mathcal{X}$, s'il existe, tel que $< x^*, EX > = E < x^*, X >$ pour tout $x^* \in \mathcal{X}^*$, c'est-à-dire que EX est l'intégrale de Pettis de $x(u)$.

Alors, sans grandes difficultés, le théorème ergodique de Birkhoff permet de démontrer la *« loi forte des grands nombres dans un espace de Banach » :*

Si $\mathcal{X}$ est un espace de Banach, séparable, si $\{X_i\}$ est une suite strictement stationnaire de L-éléments aléatoires à valeurs dans $\mathcal{X}$ et si $E(\|X_i\|) < +\infty$, presque-sûrement, quand $n \to +\infty$, la moyenne temporelle $Y_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ tend fortement vers une limite Y qui est un L-élément aléatoire à valeurs dans $\mathcal{X}$ et tel que $EY = EX_i$.

Si les X_i sont des éléments aléatoires indépendants de même loi, Y se réduit encore à l'élément certain EX_i.

Le théorème ergodique de Yosida et Kakutani ne faisant pas intervenir de mesure ne peut évidemment pas permettre d'obtenir une loi *forte* des grands nombres, c'est-à-dire une convergence presque-sûre, mais il permet d'établir une « *loi des grands nombres en moyenne d'ordre* α *dans un espace de Banach* ».

Si $\mathcal{X}$ est un espace de Banach séparable, si $\{X_i\}$ est une suite strictement stationnaire de L-éléments aléatoires à valeurs dans $\mathcal{X}$ et tels que $E(\|X\|^\alpha) < +\infty$, avec $1 \leqslant \alpha < +\infty$; il existe un L-élément aléatoire Y à valeurs dans $\mathcal{X}$, tel que $E(\|Y\|^\alpha) < +\infty$, $EY = EX_i$, et que

$$\lim_{n \to +\infty} E(\|\frac{1}{n}(X_1 + \dots + X_n) - Y\|^\alpha) = 0.$$

Toutefois, ce théorème est loin de découler directement de celui de Yosida et Kakutani ; sa démonstration nécessite en particulier l'étude d'un espace $\overset{a}{\mathcal{X}}$ associé à l'espace de Banach, séparable, $\mathcal{X}$, et des fonctionnelles linéaires définies sur $\overset{a}{\mathcal{X}}$.

$\overset{a}{\mathcal{X}}$ est défini de la façon suivante : tout L-élément aléatoire X à valeurs dans $\mathcal{X}$ et tel que $E(\|X\|^\alpha) < +\infty$ est considéré comme un élément $\overset{a}{X}$ d'un espace $\overset{a}{\mathcal{X}}$ normé en posant :

$$\|\overset{\alpha}{X}\| = [E(\|x\|^\alpha)]^{1/\alpha}$$

ou encore $\overset{a}{\mathcal{X}}$ est l'espace constitué par les fonctions $X = x(u)$ telles que

$$\int_{\mathcal{U}} \|x(u)\|^\alpha \, d\mu < +\infty \text{ avec } \|\overset{a}{X}\| = [\int_{\mathcal{U}} \|x(u)\|^\alpha \, d\mu]^{1/\alpha}$$

ainsi défini, $\mathcal{X}$ est un espace de Banach.