LE CALCUL DES PROBABILITÉS, LES GRANDES LIGNES DE SON DÉVELOPPEMENT ET SES PRINCIPAUX CHAMPS D'APPLICATION *

par R. FORTET
Professeur à la Sorbonne

A. — AXIOMATIQUE ET GÉNÉRALITÉS

1° La notion de probabilité est intuitive ; pourtant, elle n'est devenue une notion scientifique, plus précisément mathématique, que tardivement ; cette élaboration s'est opérée principalement par deux voies à certains égards bien différentes : les jeux de hasard et les dénombrements statistiques. Dans les jeux de hasard, les événements intéressant le joueur sont combinaisons d'un nombre fini d'événements fondamentaux ; les dispositions sont prises pour que ces événements fondamentaux soient équivalents du point de vue de la question de savoir lequel d'entre eux se produira, ce qu'on peut traduire en disant qu'ils ont la même probabilité : c'est ce qu'on a appelé le « principe de symétrie » ; la probabilité Pr (A) d'un événement A quelconque est alors repérée par le rapport n/N du nombre n de ceux des événements fondamentaux qui réalisent A au nombre total N des événements fondamentaux ; et rien n'empêche de poser conventionnellement $\Pr(A) = \frac{n}{N}$. La théorie des « événements équivalents » de de Finetti peut être considérée comme une forme moderne et généralisée du « principe de symétrie ».

Les premiers dénombrements statistiques portant sur des populations assez larges et prolongés assez longtemps concernaient surtout des caractères sociaux ou économiques ; un tel dénombrement fait apparaître, par exemple, que, sur les N individus recensés, n ont le caractère C étudié (d'être célibataire, chômeur, etc...) ; il est arrivé que, pour une population donnée (mais qui évolue nécessairement au cours du temps), la fréquence empirique $\frac{n}{N}$ garde une stabilité frappante ; on peut traduire cette observation concrète en disant qu'il y a une probabilité Pr (C) qu'un individu de la population en question ait le caractère C ; la fréquence empirique $\frac{n}{N}$ diffère peu de Pr (C), si N est assez grand, et, intuitivement, d'autant moins que N est plus grand : $\frac{n}{N}$ fournit donc une estimation plus ou moins précise de la valeur inconnue

---

* Conférence prononcée le 20 novembre 1958, à l'Institut Henri-Poincaré, et organisée par la Société Mathématique de France, en accord avec l'A.P.M., à l'intention spéciale des professeurs.

de Pr (C) : ce qui conduit naturellement à attribuer aux probabilités Pr (C) les propriétés mathématiques des rapports $\frac{n}{N}$.

Les probabilités auxquelles on accède par l'un ou l'autre des deux procédés ci-dessus sont-elles bien de même nature ? En tout cas, elles ont les mêmes propriétés formelles ; ce sont ces propriétés qui, exprimées sous une forme mathématique rigoureuse et imposées par axiomes aux probabilités, constituent l'axiomatique du calcul des probabilités ; je vais indiquer cette axiomatique, sous la forme aujourd'hui classique dite de « Kolmogorov ».

2° *Axiomatique de Kolmogorov :* Un problème de calcul des probabilités concerne un système $\sum$ quelconque, mais déterminé, qui a plusieurs configurations distinctes possibles, parmi lesquelles il en adopte une déterminée sous l'influence du hasard. Soit $\mathcal{U}$ l'ensemble de toutes les configurations possibles, u un élément quelconque de $\mathcal{U}$; dans la terminologie Fréchet, $\mathcal{U}$ est la catégorie d'épreuves, u une épreuve ; dans la terminologie de Kolmogorov, u est un « événement élémentaire », $\mathcal{U}$ l'ensemble des « événements élémentaires ». Un événement A (concernant $\sum$) est toujours définissable sous la forme suivante : A est réalisé si l'épreuve u réalisée possède un certain caractère ou propriété C ; si alors on appelle A le sous-ensemble de $\mathcal{U}$ constituée par les u qui ont le caractère C, on peut identifier l'événement A et le sous-ensemble A : *les événements sont les sous-ensembles de* $\mathcal{U}$. Les éléments généraux de la théorie des ensembles sont donc applicables aux événements : notion de réunion, intersection, limite, etc... ; des sous-ensembles disjoints de $\mathcal{U}$ sont des événements incompatibles, etc...

La probabilité p(A) d'un événement A $\subset \mathcal{U}$ est une fonction d'ensemble A ; l'un des axiomes comporte que c'est une fonction complètement additive ; comme le domaine de définition d'une telle fonction est nécessairement une σ-algèbre (corps de Borel), nous pouvons dire finalement que :

Axiome 1 : Tout problème de calcul des probabilités porte :

a) sur un ensemble arbitraire $\mathcal{U}$ (catégorie d'épreuves) d'éléments u de nature quelconque (épreuves) ; tout sous-ensemble A de $\mathcal{U}$ est un événement ;

b) sur une σ-algèbre (corps de Borel) $\mathcal{B}$ de sous-ensembles A $\subset \mathcal{U}$, contenant $\mathcal{U}$ lui-même, mais pouvant être quelconque par ailleurs ;

c) sur une mesure finie, ou probabilité, p définie sur $\mathcal{B}$, c'est-à-dire sur une fonction d'ensemble p(A) réelle non négative, complètement additive (σ-additive), définie sur $\mathcal{B}$, telle que $p(\mathcal{U}) = 1$, mais pouvant être quelconque par ailleurs ; p est la loi de probabilité de la catégorie $\mathcal{U}$.

Les jeux de hasard et les relevés statistiques dont j'ai parlé au § 1 invitent évidemment à imposer à p d'être additive ; aucune évidence

expérimentale n'oblige à imposer à p la restriction supplémentaire d'être *complètement* additive, mais sans elle trop de questions devraient rester sans réponse.

On peut imposer à p la condition supplémentaire d'être une mesure *complète*, mais ce n'est pas à proprement parler une restriction, puisque si p n'est pas complète, il est toujours possible de la compléter, par un élargissement convenable du σ-corps $\mathcal{B}$.

On notera qu'en général il existe des sous-ensembles A de $\mathcal{U}$ qui ne font pas partie de $\mathcal{B}$, donc il existe des événements A qui n'ont pas de probabilité : non-probabilisés.

Probabilités conditionnelles : Un point très important est que la probabilité d'un événement n'est pas une propriété intrinsèque de cet événement ; elle dépend de lui, mais aussi de la catégorie d'épreuves $\mathcal{U}$ à laquelle on réfère l'événement.

Soit $\mathcal{U}'$ une partie quelconque de $\mathcal{U}$: c'est aussi une catégorie d'épreuves, plus restreinte que $\mathcal{U}$; comme sous-ensemble de $\mathcal{U}$, $\mathcal{U}'$ représente un événement B, et $\mathcal{U}'$ est la catégorie des épreuves qui réalisent B.

Supposons un événement ε consistant en ce que l'épreuve réalisée u possède un certain caractère C ; par rapport à la catégorie $\mathcal{U}$, ε est représenté par le sous-ensemble $A \subset \mathcal{U}$ constitué par les $u \in \mathcal{U}$ qui possèdent le caractère C ; par rapport à la catégorie $\mathcal{U}'$, ε est représenté par le sous-ensemble A' des u de $\mathcal{U}'$ qui ont le caractère C ; évidemment :

$$A' = A \cap B ;$$

supposons ε probabilisé dans $\mathcal{U}$ [$\varepsilon \in \mathcal{B}$; $\Pr(\varepsilon)$ ou $\Pr(A) = p(A)$] ; pour que ε soit aussi probabilisé dans $\mathcal{U}'$, il faut d'abord que $\mathcal{U}'$ soit munie d'une loi de probabilité ; donc :

— qu'il existe un σ-corps $\mathcal{B}'$ de sous-ensembles de $\mathcal{U}'$, contenant $\mathcal{U}'$; avec, définie sur $\mathcal{B}'$, une mesure de probabilité $p'[p'(\mathcal{U}') = 1]$;

— ensuite, il faut que $A' = A \cap B \subset \mathcal{B}'$; alors la probabilité de ε par rapport à $\mathcal{U}'$ est $p'(A \cap B)$; on l'appelle la probabilité conditionnelle de ε (ou de A) par rapport à B [quand B est réalisé] et on la note souvent : $\Pr(A/B)$.

Les jeux de hasard et les dénombrements statistiques suggèrent que $\Pr(A/B)$ est normalement différente de la probabilité $\Pr(A)$ de ε dans $\mathcal{U}$; et plus précisément que :

(2,1) $\Pr(B)\Pr(A/B) = \Pr(A \cap B)$ [$p(B)\,p'(A \cap B) = p(A \cap B)$] ;

on est donc amené à imposer la propriété (2,1) par axiome ; mais on s'aperçoit rapidement que cet axiome est insuffisant. Le problème typique que l'on rencontre est le suivant : en considérant le triplet ($\mathcal{U}$, $\mathcal{B}$, p) comme donné, soit donnée également une famille $\mathcal{F}$ de sous-ensembles $e \in \mathcal{B}$, deux à deux disjoints et dont la réunion est $\mathcal{U}$; A désignant un événement quelconque de $\mathcal{B}$, la probabilité conditionnelle $q(e ; A) = \Pr(A/e)$, si elle existe pour tout $e \in \mathcal{F}$ et tout $A \in \mathcal{B}$, doit

être, pour tout e fixe, une fonction complètement additive de A $\mathcal{B}$, avec $q(e\ ;\mathcal{U}) = 1$. Complété par (2,1), cela ne suffit pas à déterminer q ; il est naturel de substituer à (2,1) la propriété plus stricte suivante ; soit $\mathcal{B}_{(\mathcal{G})}$ le plus petit σ-corps contenant tous les e ; pour tout événement E $\mathcal{B}_{(\mathcal{G})}$, on doit avoir :

$$(2,2) \qquad \int_E q(e\ ;\mathrm{A})\,d\mu = p(e \cap \mathrm{A}),$$

où μ est la mesure induite par p sur $\mathcal{B}_{(\mathcal{G})}$: nous adopterons (2,2) comme second axiome (axiome des probabilités composées).

Alors q(e ; A) est déterminée par le théorème de Radon-Nicodym ; mais elle n'est pas ainsi déterminée de façon unique ; et, en général, parmi les solutions ainsi offertes par le théorème de Radon-Nicodym, il n'y en a aucune qui, comme fonction de A, soit complètement additive.

On s'est alors demandé quelles conditions, au moins suffisantes, il convient d'imposer à ($\mathcal{U}$, $\mathcal{B}$, p), — et à la famille $\mathcal{G}$ —, pour qu'il y ait au moins une solution complètement additive ; les résultats dans ce sens les plus intéressants semblent ceux obtenus par Irgina, qui a utilisé le concept de « mesure parfaite » introduit par Gnedenko et Kolmogorov. Il serait utile d'avoir des conditions assurant l'existence, mais aussi *l'unicité* d'une solution complètement additive ; pour le moment, on n'a de telles conditions que dans des cas assez particuliers.

J'ai, ici, considéré le problème de la probabilité conditionnelle en partant de p (ou probabilité *a priori*) comme donnée, la probabilité conditionnelle étant à déduire de p : c'est ainsi que le problème se présente naturellement en statistique. Mais, dans d'autres domaines (processus stochastiques), ce qui est naturellement donné, c'est un jeu de probabilités conditionnelles, à partir duquel il faut construire p ; dans les cas simples, le théorème de Fubini règle la question ; dans des cas compliqués, on se trouve en présence d'un problème non encore résolu.

3° *Commentaires sur l'axiomatique de Kolmogorov :* On a critiqué (et d'ailleurs Kolmogorov lui-même) l'axiomatique précédente, en particulier sur les points suivants :

1) elle prend comme base les notions d'épreuve et de catégorie d'épreuves : or ces notions sont très abstraites ;

2) un événement de probabilité 1 n'est pas forcément certain, mais seulement presque-certain ; cette distinction entre événements certains et événements presque-certains n'est pas immédiatement intuitive, ou suggérée par l'expérience ;

3) l'additivité complète, adoptée pour les raisons dites plus haut, garde cependant un certain caractère d'arbitraire.

On peut éviter ces objections, en prenant comme base les événements probabilisés eux-mêmes ; ils forment une algèbre de Boole, et finalement tout problème de probabilité concerne alors une algèbre de Boole métrique, qu'on peut toujours supposer complète ; divers auteurs ont étudié cette autre forme d'axiomatique, montrant qu'elle échappe effectivement aux objections 1), 2), 3) ci-dessus, tout en étant en fait

strictement équivalente à l'axiomatique de Kolmogorov (les objections 1), 2), 3) sont donc de forme plutôt qu'essentielles) ; jusqu'à présent, on utilise peu cette axiomatique booléenne.

4° *Indépendance :* Revenant au triplet $(\mathcal{U}, \mathcal{B}, p)$, deux événements A et B sont indépendants si :

$$\begin{cases} \Pr(A \cap B) = \Pr(A) \times \Pr(B) \\ \Pr(\overset{\vee}{A} \cap B) = \Pr(\overset{\vee}{A}) \times \Pr(B) \\ \Pr(A \cap \overset{\vee}{B}) = \Pr(A) \times \Pr(\overset{\vee}{B}) \\ \Pr(\overset{\vee}{A} \cap \overset{\vee}{B}) = \Pr(\overset{\vee}{A}) \times \Pr(\overset{\vee}{B}) \end{cases}$$

en désignant par $\overset{\vee}{E}$ le contraire (complémentaire) d'un événement E ; ces quatre conditions se réduisent à une seule si ni Pr (A), ni Pr (B) ne valent 0 ou 1 ; on définit de façon analogue l'indépendance « *mutuelle* » d'un nombre fini quelconque d'événements ; noter que trois événements A, B, C peuvent être indépendants deux à deux sans l'être mutuellement.

5° *Notion générale d'élément aléatoire :* Soit $\mathcal{X}$ un espace quelconque et $x = x(u)$ une application quelconque de $\mathcal{U}$ dans $\mathcal{X}$: c'est un élément aléatoire, à valeurs dans $\mathcal{X}$, qu'on désignera souvent par une majuscule X. En général, on se limite aux éléments aléatoires X, ou applications $x = x(u)$, tels que les événements intéressants concernant X soient probabilisés. Un événement concernant X est toujours du type : $X \in e$, où e est un certain sous-ensemble de $\mathcal{X}$. L'ensemble des $\Pr(X \in e)$ pour tous les e tels que l'événement $X \in e$ soit probabilisé constitue *la loi de probabilité de* X.

B. — Variables aléatoires

6° *Définition et description d'une variable aléatoire :* Naturellement, la première catégorie d'éléments aléatoires qu'on a considérés sont ceux pour lesquels $\mathcal{X}$ est l'ensemble R des nombres réels ; il s'agit alors des nombres (réels) ou variables aléatoires ; pour une variable aléatoire X, les événements naturellement intéressants sont ceux du type $X < x$, x nombre réel donné quelconque ; on supposera donc que ces événements sont probabilisés [$x(u)$ est une fonction mesurable-p de $u \in \mathcal{U}$], et on posera :

$$\Pr(X < x) = F(x) :$$

$F(x)$ est la *fonction de répartition* de X ; c'est une fonction monotone non-décroissante, continue à gauche. On écarte en général le cas où il y aurait une probabilité positive que X soit infini ; on a alors :

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0 ;$$

$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$$

$F(x)$ détermine $\Pr(X \in e)$, du moins pour tout sous-ensemble $e \subset R$ d'intérêt pratique, en particulier si e est un ensemble de Borel.

Lorsque $F(x)$ est dérivable, sa dérivée $f(x)$ est *la densité de probabilité* de X.

Espérance mathématique et moments : L'espérance mathématique E(X) de X est définie par :

$$E(X) = \int_{\mathcal{U}} x(u)p(du) = \int_{-\infty}^{+\infty} x\, dF(x),$$

où, par définition, l'intégrale existe si, et seulement si, elle est absolument convergente.

Le moment algébrique d'ordre K (K entier $\geqq 0$), rapporté à l'abscisse a, $a^m K$ est défini par :

$$a^m K = E[(X-a^K)] = \int_{-\infty}^{+\infty} (x-a)^K\, dF(x);$$

le moment absolu d'ordre K (K réel $\geqq 0$) rapporté à l'abscisse a, $a^{m*}K$, est défini par :

$$a^{m*}K = E(|X-a|^K) = \int_{-\infty}^{+\infty} |x-a|^K\, dF(x).$$

Quels que soient $K > 0$ et $h > 0$, on a :

$$(6,1) \qquad Pr(|X-a| \geqslant h) \leqslant \frac{a^{m*}K}{h^K}$$

(inégalités de Bienaymé-Tschebischev).

Extensions :

a) *Variables aléatoires complexes :* Si $\mathcal{X}$, au lieu d'être l'espace R des nombres réels, est l'espace C des nombres complexes, X est une variable aléatoire complexe.

b) *Variables aléatoires réelles à plusieurs dimensions :* Si $\mathcal{X}$ est un espace E_n affine proprement euclidien, X est une variable aléatoire réelle à n dimensions ; si on rapporte E_n à un repère cartésien, si $X_1, X_2, ..., X_n$ sont les coordonnées de X par rapport à ce repère, il apparaît qu'une variable à n dimensions n'est autre qu'un système de n variables aléatoires (en général non indépendantes). Une variable complexe peut être considérée comme une variable réelle à deux dimensions.

Si $X = (X_1, X_2, ..., X_n)$ est une variable réelle à n dimensions, $F(x_1, x_2, ..., x_n) = Pr(X_1 < x_1, X_2 < x_2, ..., X_n < x_n)$ est sa fonction de répartition ; c'est une fonction de n variables.

Fonction caractéristique : Revenons au cas d'une variable réelle X à une dimension. Sa fonction caractéristique $\varphi(u)$ est la fonction de u réel définie par :

$$\varphi(u) = E[e^{iux}] = \int_{-\infty}^{+\infty} e^{iux}\, dF(x);$$

autrement dit, $\varphi(u)$ est la transformée de Fourier de $F(x)$. $\varphi(0) = 1$, $\varphi(u)$ est uniformément continue sur $(-\infty, +\infty)$ et est définie positive ; toute fonction $\varphi(u)$ continue et définie positive avec $\varphi(0) = 1$ est la transformée de Fourier d'une fonction de répartition et d'une seule ; la

correspondance biunivoque entre $F(x)$ et $\varphi(u)$ est bicontinue (avec des topologies convenables).

Avec des adaptations convenables, les notions d'espérance mathématique, de moments, de fonction caractéristique s'étendent aux variables complexes, ou aux variables à plusieurs dimensions.

Si X et Y sont deux variables réelles (à 1 seule dimension chacune), on a toujours :

(6,1) $$E(X + Y) = E(X) + E(Y).$$

Si on pose :

$$\sigma_X^2 = E(|X - E(X)|^2)$$
$$\sigma_Y^2 = E(|Y - E(Y)|^2)$$
$$\sigma_X \sigma_Y r = E\{[X - E(X)][Y - E(Y)]\},$$

on a : $0 \leq |r| \leq 1$ d'après l'inégalité de Schwarz ; $|r| = 1$ si, et seulement si, il y a entre X et Y une relation linéaire à coefficients non aléatoires : r s'appelle le *coefficient de corrélation* de X et Y.

X et Y sont indépendants si, quels que soient les ensembles e_1, e_2 de nombres réels, les deux événements $X \in e_1$, $Y \in e_2$ sont indépendants. Si X et Y sont indépendants, on a :

(6,2) $$E(XY) = E(X) \times E(Y),$$

et par suite le coefficient de corrélation de X et de Y est nul.

7° *Addition des variables aléatoires indépendantes :* Si X et Y sont indépendantes, et si $S = X + Y$ est leur somme, on a :

(7,1) $$E(S) = E(X) + E(Y)$$

d'après (6,2) ; si $E(X) = E(Y) = E(S) = 0$, on voit avec (6,2) que :

(7,2) $$E(S^2) = E(X^2) + E(Y^2)\ ;$$

d'une façon générale, il n'est pas difficile de calculer un moment algébrique quelconque $E(S^K)$ de S, en fonction des moments, d'ordre $\leq K$, de X et de Y ; les formules obtenues sont de plus en plus compliquées quand K augmente.

Si $F(x)$, $G(y)$, $H(s)$ sont les fonctions de répartition de X, Y, S respectivement, on a :

(7,3) $$H(s) = \int_{-\infty}^{+\infty} G(s-x)\, dF(x) = \int_{-\infty}^{+\infty} F(s-y)\, dG(y),$$

autrement dit, H est le produit de convolution de F et G ; par suite, si $\varphi_1(u)$, $\varphi_2(u)$, $\varphi_3(u)$ sont les caractéristiques de X, Y, S, on a :

(7,4) $$\varphi_3(u) = \varphi_1(u) \times \varphi_2(u).$$

8° *Convergence en probabilité, presque-sûre, en moyenne :* Les convergences en mesure, presque-partout, en moyenne d'ordre α, usuelles dans la théorie de la mesure et des fonctions mesurables, se transforment immédiatement en calcul des probabilités, où elles prennent les noms, respectivement, de convergence en probabilité, presque-sûre, en moyenne d'ordre α ; par exemple, une variable aléatoire X_K $(K = 1, 2, 3)$ tend, lorsque $K \to +\infty$, vers une variable aléatoire limite X en moyenne d'ordre 2 (en moyenne quadratique) si $E[(X_K - X)^2] \to 0$; alors, d'après (6,1), X_K tend vers X en probabilité. Si X_K tend vers X en pro-

babilité, cela ne signifie pas que la différence $X_K - X$ est petite, mais seulement qu'il y a, pour K assez grand, une probabilité arbitrairement voisine de 1 pour que $X_K - X$ soit arbitrairement petite.

9° *Lois asymptomatiques de l'addition des variables aléatoires indépendantes :* Le chapitre le plus anciennement développé du calcul des probabilités consiste à étudier la somme S_n d'un grand nombre n de variables aléatoires indépendantes $X_1, X_2, ..., X_n$:

$$S_n = X_1 + X_2 + ... + X_n ;$$

on dégage ainsi des lois asymptotiques (pour $n \to +\infty$) qui appartiennent à deux types principaux :

1° *Lois des grands nombres :* Les lois des grands nombres proprement dites sont des théorèmes établissant, sous des conditions convenables, que lorsque $n \to +\infty$ la moyenne arithmétique S_n/n tend (par exemple en probabilité) vers une limite L.

Exemple : Soient une succession de parties indépendantes de pile ou face, numérotées 1, 2, ..., n ; à chaque partie, les probabilités de pile et face ont les mêmes valeurs respectivement p et $q = 1 - p$ (on suppose $0 < p < 1$). Si X_j vaut 1 ou 0 suivant que la partie N° j donne pile ou face, S_n/n est la fréquence relative de pile sur les n parties ; on prouve facilement que S_n/n tend en moyenne quadratique (donc en probabilité), et aussi presque-sûrement vers p ; car

$$(9,1) \qquad E(|S_n/n - p|^2) \sim \frac{1}{n} ;$$

ce qui rejoint l'idée intuitive que la fréquence empirique, sur un grand nombre n d'expériences, de l'événement pile doit être en un certain sens une valeur approchée de sa probabilité p.

2° *Lois de probabilités limites :* Reprenons l'exemple de pile ou face ; pour n grand, S_n est probablement grand ; sa fonction de répartition $F_n(x)$ tend donc vers 0 pour tout x fini ; mais cette constatation n'apporte rien de nouveau ; mais normons S_n convenablement, après l'avoir rapporté à une origine également convenable ; en fait, considérons, au lieu de S_n,

$$U_n = (S_n - np)/\sqrt{pqn},$$

qui pour tout n a une espérance nulle et un moment d'ordre 2 constamment égal à 1. On démontre que la fonction de répartition $G_n(u)$ de U_n tend pour tout u vers :

$$\Phi(u) = \frac{1}{\sqrt{2}} \int_{-\infty}^{u} e^{-\frac{x^2}{2}} dx,$$

fonction de répartition connue sous le nom de loi de Laplace (ou de Gauss).

De telles lois-limites peuvent être mises en évidence dans des cas beaucoup plus généraux que celui de pile ou face ; l'usage des caractéristiques et de (7,4) est fondamental pour cette étude, aujourd'hui à peu près terminée ; elle a mis en évidence les notions de lois stables, quasi-stables, indéfiniment divisibles (la loi de Laplace est stable, et par suite indéfiniment divisible) ; les efforts les plus récents ont visé à préciser la rapidité des convergences [de $G_n(u)$ vers $\Phi(u)$, par exemple], ou à

savoir si la convergence des fonctions de répartition s'accompagne de la convergence des densités de probabilités (en supposant que celles-ci existent).

C. — Les grands chapitres du calcul des probabilités

10° *Les processus stochastiques :* En dehors de l'addition des variables aléatoires indépendantes, le calcul des probabilités, dans son état actuel, est principalement consacré à l'étude des processus stochastiques, ou évolutions aléatoires au cours du temps d'un système quelconque. On peut concevoir une infinie variété de processus stochastiques ; pour des raisons théoriques et, plus souvent, pratiques, on a étudié principalement :

a) les processus à accroissements indépendants, dont la théorie se fonde sur l'addition des variables aléatoires indépendantes ;

b) les processus de Markov, dont l'étude peut être envisagée comme une application de la théorie des semi-groupes d'opérations linéaires dans un espace de Banach ; les processus de Wiener-Lévy, qui sont à la fois de Markov et à accroissements indépendants, présentent d'intéressantes connexions avec la théorie du potentiel ;

c) les processus stationnaires du second ordre, dont l'analyse harmonique est un chapitre de l'analyse harmonique générale.

Je ne donne pas d'autres indications sur ces questions qui seront reprises dans des conférences suivantes.

Je note seulement que les processus stochastiques sont des éléments aléatoires d'un certain type, évidemment beaucoup plus complexes que les variables aléatoires, et qu'une façon indirecte de les étudier est de développer la théorie des éléments aléatoires les plus généraux possibles [c'est-à-dire celle des applications $x(u)$ dans des espaces $\mathcal{X}$ aussi généraux que possible] ; cette théorie, encore à son début, comporte déjà des résultats prometteurs.

D. — Les applications du calcul des probabilités

11° *Les applications du calcul des probabilités :* Les premières applications, historiquement, du calcul des probabilités concernèrent la démographie, les assurances, divers phénomènes économiques et sociaux. Ce domaine reste important avec une branche nouvelle : la psychométrie ; toutefois, l'introduction des processus stochastiques en économétrie, bien qu'apparemment souhaitable, reste très timide.

Mais le calcul des probabilités a trouvé, plus récemment, un immense champ d'application en physique ou dans des techniques dérivées de la physique, dans tous les cas où intervient un ensemble considérable de particules : diffusion et mouvement brownien, mécanique statistique (classique ou quantique), phénomènes d'émission (bruit de fond dans les tubes électroniques), semi-conducteurs... ; je citerai aussi, d'un type différent et également très importantes, la théorie de la transmission de l'information et celle de la détection. En retour, ces applications ont

notablement influé sur le cours du développement moderne du calcul des probabilités.

Je ne puis donner aucun détail sur ces applications ; d'ailleurs, leur existence au moins est bien connue ; je préfère m'arrêter sur une autre application, très curieuse, qui peut éventuellement être utilisée dans n'importe quel domaine, et qui est généralement ignorée en dehors des spécialistes : les méthodes de Monte-Carlo.

12° *Les méthodes de Monte-Carlo :* Je veux d'ailleurs simplement faire comprendre le principe de ces méthodes, sans entrer dans la technique de leur utilisation.

Pour obtenir la valeur numérique θ_0, supposée inconnue *a priori*, d'un paramètre θ dont dépend la fonction de répartition d'une variable aléatoire X, il n'y a pas d'autre moyen que le procédé empirique : on procédera à n déterminations, par exemple indépendantes, de X et on déduira de leurs résultats une estimation de θ_0 qui sera adoptée comme valeur approchée de θ_0, généralement en application d'une loi des grands nombres [comme, d'après (9,1), S_n/n peut être adopté comme estimation de p] ; sous le nom de « théorie de l'estimation des paramètres », les statisticiens ont de longue date développé des méthodes perfectionnées (notion de résumé exhaustif, analyse séquentielle, etc...) pour réduire l'erreur e à craindre, et le temps nécessaire (c'est-à-dire n) ; le calcul des probabilités montre [cf. (9,1)] que e est ordinairement de l'ordre de $n^{-\frac{1}{2}}$:

$$(12,1) \qquad e \sim n^{-\frac{1}{2}} .$$

Soit θ_0 un nombre ; supposons que l'on possède de θ_0, non sa valeur, mais une définition mathématique qui le détermine entièrement : par exemple, θ_0 est la plus petite racine d'une équation algébrique donnée à racines toutes positives ; de la définition mathématique de θ_0, on peut déduire en général une méthode de calcul au moins approchée de sa valeur : de telles méthodes reposent souvent sur l'itération d'une opération plus ou moins complexe, de sorte que l'erreur e a en général un ordre de grandeur du type :

$$(12,2) \qquad e \sim a^n \qquad (a < 1),$$

où n est le nombre de fois qu'on a répété l'opération.

Mais soit une famille de variables aléatoires $\{X_\theta\}$ dont la fonction de répartition $F(\theta ; x)$ dépend d'un paramètre θ ; et supposons qu'on sache procéder à des déterminations empiriques de X_{θ_0} ; comme la valeur de θ_0 n'est pas connue, il faudra que la définition mathématique de θ_0 suffise à permettre de telles déterminations empiriques ; s'il en est ainsi, on voit qu'on pourra évaluer θ_0 par une méthode d'estimation de paramètre : on dira alors qu'on évalue θ_0 par une *méthode de Monte-Carlo.*

Exemple 1 : Soit $\theta_0 = \pi$; π possède des propriétés mathématiques qui le définissent absolument et dont on a depuis longtemps déduit sa valeur numérique avec une précision considérable ; mais imaginons qu'on n'ait pas encore effectué ce calcul ; on pourrait obtenir la valeur de π par une méthode de Monte-Carlo, par exemple de la façon suivante :

— On connaît le « jeu de l'aiguille » de Buffon ; soit X le nombre des intersections de l'aiguille et des parallèles tracées sur le plan ; si la longueur de l'aiguille est la moitié de l'écart des parallèles (de sorte que X ne peut valoir que 0 ou 1), on a :

$$\Pr(X = 1) = \frac{1}{\pi} :$$

c'est pile ou face, avec $p = \frac{1}{\pi}$; la fréquence S_n/n correspondante (cf. § 9) sera donc une estimation de π.

Exemple 2 : u variant de 0 à 1, soit $f(u)$ une fonction continue donnée, avec : $0 \leqq f(u) \leqq 1$; et soit :

$$\theta_0 = \int_0^1 f(u)du.$$

Soient U et V deux variables aléatoires indépendantes de loi uniforme (densité constante) sur (0,1), et soit X la variable aléatoire qui vaut 1 si $V \leqq f(U)$ et 0 si $V > f(U)$; on a :

$$E(X) = \theta_0.$$

On peut procéder à des déterminations empiriques de U et de V, donc de X ; pour n déterminations indépendantes $x_1, x_2, ..., x_n$ de X, $\frac{1}{n} \sum_j x_j$ tend en moyenne quadratique vers $E(X) = \theta_0$ (loi des grands nombres) ; on peut donc estimer θ_0 par $\frac{1}{n} \sum_j x_j$, et l'erreur à craindre est de l'ordre de :

$$\left\{ E(|\frac{1}{n}\sum_j x_j - \theta_0|^2) \right\}^{\frac{1}{2}} = \sqrt{\frac{\theta_0(1-\theta_0)}{n}}.$$

On a recouru (von Neumann, Ulam) aux méthodes de Monte-Carlo en 1942 pour des problèmes de diffusion de neutrons, difficiles à étudier par les procédés mathématiques ordinaires et se rattachant à la fabrication de la bombe atomique ; dans ce domaine, ce recours est demeuré usuel.

Naturellement, pour pouvoir évaluer un nombre θ_0 dont on possède une définition mathématique par une méthode de Monte-Carlo, il faut trouver un schéma stochastique auquel rattacher θ_0 et réalisable sur la base de la définition mathématique possédée de θ_0 ; cela peut être difficile ; mais, souvent aussi, un tel schéma est en évidence parce que θ_0 est précisément apparu à propos d'un phénomène aléatoire (cas de la diffusion des neutrons, du fonctionnement des centraux téléphoniques, etc...) ; il suffit alors de reproduire ce phénomène, naturellement, ou artificiellement avec une machine électronique.

L'usage de méthodes statistiques pour évaluer des quantités relevant normalement d'un calcul strictement mathématique semble cependant paradoxal ; la comparaison de (12,1) avec (12,2) met en évidence que, pour le même nombre n d'opérations, l'erreur est ordinairement beaucoup plus forte pour une méthode de Monte-Carlo que pour une méthode mathématique ; à vrai dire, les opérations à faire ne sont pas

les mêmes dans les deux cas, et peuvent être beaucoup plus simples et rapides pour la méthode de Monte-Carlo.

Il reste qu'avec une méthode statistique, n devra généralement être pris très grand, ce qui pose le problème pratique de fabriquer du hasard, en quelque sorte, en grande quantité ; plus correctement à opérer (assez rapidement), un grand nombre n de tirages au sort. On peut toujours se ramener à des tirages au sort (mutuellement indépendants) d'un nombre ou variable aléatoire compris entre 0 et 1, selon une loi uniforme sur (0,1) [densité de probabilité de la variable aléatoire constante et égale à 1, sur (0,1)]. On a eu l'idée d'observer empiriquement de tels nombres produits par un phénomène naturel aléatoire (bruit de fond dans un tube électronique), et on a dressé et publié une table de $1,2.10^6$ tels nombres. Cette table est difficile à utiliser si on emploie par ailleurs une machine à calculer électronique, et on préfère généralement faire fabriquer les nombres aléatoires par la machine elle-même, selon un procédé arithmétique qui, naturellement, fournit des nombres qui sont en réalité des nombres déterminés, qu'on ne peut considérer comme aléatoires que par approximation.

N.D.L.R. — Nous croyons utile de signaler l'ouvrage récent de M. FORTET : « Some Aspects of Analysis and Probability », John Wiley and Sons, New-York (en collaboration avec I. KAPLANSKY, E. HEWITT et MARSHALL HALL), ainsi que les suivants, en préparation :

Collection universitaire de Mathématiques (Dunod) :

VIII. R. FORTET : « Matrices ».

X. R. FORTET : « Processus stochastiques ».