

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MAURICE FRÉCHET

**Les espaces abstraits et leur utilité en statistique théorique
et même en statistique appliquée**

Journal de la société statistique de Paris, tome 88 (1947), p. 410-421

http://www.numdam.org/item?id=JSFS_1947__88__410_0

© Société de statistique de Paris, 1947, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/*

VI

LES ESPACES ABSTRAITS ET LEUR UTILITÉ EN STATISTIQUE THÉORIQUE ET MÊME EN STATISTIQUE APPLIQUÉE⁽¹⁾

INTRODUCTION. — L'objet principal des recherches en calcul des probabilités et en statistique mathématique a été jusqu'ici le nombre aléatoire. Relégué dans les chapitres des probabilités géométriques apparaissait aussi le point aléatoire sur la droite, dans le plan ou dans l'espace à trois dimensions. A une époque plus récente, on a dû considérer des éléments aléatoires plus généraux, déterminés chacun par un nombre fini fixe, s , de paramètres numériques aléatoires. On a trouvé ensuite plus simple et plus suggestif de considérer ces systèmes comme des points aléatoires choisis au hasard dans un (ainsi nommé) espace à s dimensions. Enfin, dans ces dernières années, s'est introduite la théorie des séries aléatoires et des fonctions aléatoires. N'y a-t-il pas des traits communs à ces études successives de catégories d'éléments aléatoires de natures diverses? Et même, n'y aurait-il pas une théorie commune non seulement à ces éléments aléatoires particuliers, mais encore à tous les éléments aléatoires, quelle que soit leur nature; c'est-à-dire une théorie applicable aux éléments aléatoires abstraits? Nous nous proposons dans cette communication d'apporter à ces questions une réponse affirmative et d'en souligner d'autre part le côté pratique.

Les éléments abstraits. — Nous savons tous ce qu'on entend en disant de deux ensembles d'éléments que le nombre d'éléments du premier est plus grand que le nombre d'éléments du second. Que ces deux ensembles soient deux tas de pommes ou deux troupes de soldats, l'assertion se vérifiera de la même façon. Il-en sera de même pour la somme des nombres des éléments de ces ensembles; ce sera le nombre des éléments de la réunion de ces ensembles. Autrement dit, on sait, et cela depuis des siècles, faire dans l'arithmétique élémentaire des raisonnements sur des ensembles d'éléments qui sont indépendants de la nature de ces éléments. Autrement dit, on sait raisonner sur des ensembles abstraits, c'est-à-dire sur des ensembles d'éléments abstraits. Et ici, élément abstrait n'est pas pris dans un sens mystérieux et métaphysique. Pour nous, un élément abstrait est un élément dont on ne connaît pas la nature, ou plus généralement dont la nature, inconnue ou *même connue*, n'intervient pas dans nos raisonnements.

Nous venons de montrer dans l'arithmétique élémentaire une branche très ancienne des mathématiques qui raisonne souvent sur des éléments concrets pour se mettre à la portée des enfants, mais dont tous les raisonnements sont applicables à des éléments abstraits. Il n'y aura donc pas à s'effrayer si d'autres chapitres des mathématiques s'adressent à des éléments abstraits, ni à déclarer d'avance que cela est perdre son temps et encore moins à croire qu'il est impossible de raisonner sur des éléments dont on ne sait exactement ce qu'ils sont.

(1) Communication présentée à la séance du 21 mai 1947.

Les éléments aléatoires abstraits. — Mais dira-t-on, que nous importent vos éléments abstraits? Ne nous intéressent que les éléments que vous avez déjà cité et qui sont les seuls que rencontre le statisticien. Encore celui-ci ne voit-il déjà aucune utilité à considérer les séries aléatoires, les fonctions aléatoires dont vous nous avez parlé.

Mais ici, il faut distinguer; je ne considère pas seulement comme statisticien celui qui travaille dans un organisme officiel de statistique, mais encore, en un sens plus large, celui qui fait usage et usage utile de la statistique théorique et pratique dans au moins l'une de leurs applications. Parmi celles-ci figure, par exemple, l'aérodynamique qui dans l'étude de la turbulence, fait déjà grand usage des fonctions aléatoires.

Il faut surtout dire que si le statisticien, pris au sens large où je me suis placé, est encore, et d'ailleurs très naturellement, disposé à croire qu'il n'a pas à se soucier de ces nouveaux éléments aléatoires, c'est avant tout, parce qu'il ne disposait jusqu'ici, non seulement d'aucun moyen mais même d'aucun langage pour les étudier. Sans nous adresser uniquement aux statisticiens spécialisés dans telle ou telle science de la nature, ne croyez-vous pas que la statistique administrative s'intéressera un jour aux contours des villes, des communes, qu'elle voudra comparer leurs formes variées, les lois de leur développement, vérifier par exemple, d'une manière scientifique, la fameuse loi de leur poussée vers l'Ouest. Elle aura à étudier ainsi des éléments aléatoires qui ne sont ni des nombres ni des fonctions mais des *courbes aléatoires*. L'anthropologue qui veut définir une race, cherchera à définir une forme typique des crânes d'individus de cette race ainsi que leur dispersion, etc... Il aura affaire à des *surfaces aléatoires*. Dans les sciences sociales, on a affaire à des éléments aléatoires qui sont à première vue rebelles à toute représentation mathématique : l'intelligence, la beauté, etc... On arrive cependant à les soumettre aux évaluations statistiques en faisant intervenir non plus des grandeurs mesurables, mais des repères numériques. Par exemple, les professeurs notent, par un chiffre, l'ensemble des qualités mentales déployées par un élève dans une composition ou une interrogation. Mais qui ne confessera qu'il est difficile de comparer par des chiffres les valeurs intellectuelles de Cromwell et de Napoléon, de Pasteur et de Poincaré. Déjà dans le domaine infiniment plus simple des courbes, nous ne pouvons dire qu'une courbe est plus... qu'une autre; nous pouvons dire qu'elle est plus longue, qu'elle est plus régulière, etc... mais nous ne pouvons les placer dans une suite ordonnée suivant l'ensemble de leurs propriétés.

Nous devons donc prévoir un développement ultérieur des sciences humaines où certains éléments aléatoires ne seront plus considérés seulement au point de vue de telle ou telle qualité repérable par un chiffre, ni même au point de vue d'un ensemble de ces qualités représenté par un nombre fini de chiffres mais dans leur totalité non représentable par des moyens mathématiques classiques. On aura affaire à des éléments aléatoires qui ne seront ni nombres, ni points, ni courbes, ni surfaces, et auxquels on s'efforcera pourtant d'adopter entre autres les notions statistiques fondamentales d'éléments typiques et de dispersion.

Mais dans tous ces cas, nous manquions du langage approprié ou plus exacte-

ment, nous avions l'intuition de ce que doivent signifier, par exemple, la dispersion, la position typique d'un élément aléatoire sans être à même d'en préciser les sens. C'est cette lacune que nous nous proposons de combler ici sous l'hypothèse fondamentale de l'existence d'une distance ou plutôt de la possibilité d'un choix, conforme à notre intuition, d'une définition de la distance des éléments aléatoires considérés.

Cependant, avant de passer à cette extrême généralité et de montrer que la considération des éléments abstraits est possible et utile en dehors de l'arithmétique élémentaire et, singulièrement, en statistique, arrêtons-nous pour préparer cette extension et ne pas sauter tous les échelons à la fois, à un stade intermédiaire.

Avant d'étendre les définitions et les propriétés des nombres de l'arithmétique ou de l'algèbre et des points de la géométrie, à des éléments de nature indéterminée, commençons simplement à les étendre à des éléments de nature au contraire déterminée, mais auxquels on n'avait pas songé jusqu'à une époque récente en donnant ces définitions et ces propriétés.

La notion de *distance* qui nous est si familière n'a concerné pendant longtemps que les points de la géométrie classique. On l'a étendue plus tard aux points de géométries plus compliquées dont nous nous contenterons de donner les noms, géométrie riemannienne, géométrie projective, etc... On a ensuite trouvé qu'elle rendait des services en l'étendant aux *espaces fonctionnels*. On peut appeler espace fonctionnel *distancié*, une certaine catégorie de fonctions pour lesquelles une distance est définie.

Pour éviter ou reculer le plus possible toute technicité mathématique un peu spéciale, nous reporterons à la fin, des exemples précis de définitions de la distance de deux fonctions, définitions variant avec la famille de fonctions considérées. Mais nous pouvons dire que ces définitions sont assujetties aux conditions suivantes :

On suppose qu'à tout couple de fonctions $f(x)$, $g(x)$ appartenant à une certaine famille Φ , on fait correspondre un nombre, représenté par (f, g) , appelé distance de f et de g et satisfaisant aux conditions suivantes :

- 1^o $(f, g) = (g, f) \geq 0$;
- 2^o $(f, g) = 0$ si f et g ne sont pas distinctes et réciproquement;
- 3^o On a l'« inégalité triangulaire » $(f, g) \leq (f, h) + (h, g)$.

Nous rapprochant maintenant de notre but final, nous rappellerons que c'est cette notion de distance de deux fonctions qui nous a permis de définir une infinité de *vrais* indices de corrélation. Je n'entrerai pas dans les détails de cette question qui ont été fort bien exposés devant vous par notre jeune collègue M. Féron, dont la conférence paraîtra prochainement dans le *Journal de la Société de Statistique* (1).

Nous compléterons cependant ces publications sur un point. Simaïka a attiré notre attention sur deux conditions à ajouter aux conditions qu'il convient d'imposer à un vrai indice de corrélation et que nous avions énoncées dans

(1) Voir aussi notre mémoire : *Anciens et nouveaux indices de corrélation. Leur application au calcul des retards économiques*. *Econometrica*, vol. 15, 1947, p. 1-30. Des erreurs dans les calculs numériques, n'affectant pas, du reste, les conclusions de l'article, seront rectifiées dans un prochain numéro.

notre mémoire d'Econometrica : A) un indice de corrélation doit pouvoir prendre toutes les valeurs comprises entre 0 et 1; B) il doit croître avec le degré de dépendance.

En ce qui concerne la condition A), Féron nous a communiqué ce résultat : que la condition A est vérifiée par les indices de corrélation r , η , g , j , d mentionnés dans notre article.

Il utilise à cet effet les tableaux de corrélation à deux lignes et deux colonnes (fig. 1). En posant avec les notations habituelles.

y	x	x_1	x_2
y_1	n_{11}	n_{12}	n_{21}
y_2	n_{12}	n_{21}	n_{22}

Fig. 1.

$$n_{11} = a, \quad n_{12} = b, \quad n_{21} = c, \quad n_{22} = d,$$

on peut s'assurer que, dans le cas où l'on a

$$d = a, \quad c = b,$$

les cinq indices sont égaux à $\frac{|a-b|}{a+b}$. En faisant varier $\frac{a}{b}$ de 0 à 1, on voit que ce rapport passe aussi par toutes les valeurs rationnelles de 1 à 0 (1).

Quant à la condition B, il faudrait pour la vérifier donner un sens au degré de dépendance. Si l'on y arrivait, les cinq indices ne pourraient vérifier B) que s'ils variaient eux-mêmes dans le même sens.

Or Féron a donné des exemples théoriques où il n'en est pas ainsi. Et ce résultat est confirmé dans l'application numérique de ces cinq indices, que j'ai faite et mentionnée plus loin. Ainsi : ou bien quatre au moins des cinq indices considérés ne vérifient pas la condition B, ou bien ce qui est plus probable on ne peut donner un sens au degré de dépendance qu'en le représentant par un indice et ce degré s'exprimera différemment suivant l'indice choisi.

Éléments aléatoires quelconques. — L'emploi des indices de corrélation étant d'une nécessité extrême et d'un emploi constant dans les applications de la Statistique aux sciences humaines comme aux sciences de la nature, le recours à la notion de distance de deux fonctions dans la construction de ces indices suffirait à lui seul à justifier le titre de cette conférence.

Mais nous pouvons aller plus loin. Nous n'avons encore prolongé la notion de distance que du cas de deux points de la géométrie euclidienne au cas de deux fonctions. Or, nous allons pouvoir passer maintenant au cas de deux éléments de nature quelconque sans abandonner pour cela le champ des mathématiques utilisables dans les domaines pratiques.

Nous dirons qu'un ensemble D d'éléments de *nature quelconque* devient un *espace distancié* quand, à tout couple d'éléments f, g de D (f, g ne sont plus maintenant nécessairement des fonctions), correspond un nombre représenté

(1) Quant aux valeurs irrationnelles, on les obtiendrait pour la forme théorique du tableau de corrélation correspondant au tableau empirique ci-dessus, en remplaçant les entiers a, b, c, d , par les probabilités correspondantes.

par (f, g) , appelé distance de f et de g et vérifiant les trois conditions 1^o, 2^o, 3^o ci dessus.

La simple introduction de la notion de distance dans l'espace où sont choisis les éléments aléatoires permet d'étendre à ces éléments aléatoires un grand nombre de définitions et de propriétés des nombres aléatoires et fournit par conséquent une *énorme* extension de la théorie des nombres aléatoires.

Nous nous limiterons ici aux extensions des notions de dispersion, de valeurs typiques et de convergence stochastique.

La considération de cette dernière commence à peine à se répandre dans l'enseignement du calcul des probabilités; on ne la trouverait sans doute encore dans aucun traité de statistique. Pourtant sa signification n'est pas moins grande au point de vue pratique qu'au point de vue théorique.

On connaît l'importance du théorème de Bernoulli. Son énoncé sous la forme classique est assez difficile à saisir et on lui a souvent donné une signification plus étendue qu'il ne comporte. Soit un événement E de probabilité constante qui se répète R fois dans n épreuves. D'après le théorème de Bernoulli, nous n'avons pas le droit de dire que la fréquence $F = \frac{R}{n}$ de E tend vers la probabilité p de E quand n croît mais seulement que la probabilité P pour que F diffère de p de moins d'une quantité donnée E , tend vers l'unité. Cantelli a rattaché ce fait à une notion plus générale, celle de *convergence en probabilité*.

Nous la définirons d'une façon beaucoup plus générale que celle de Cantelli, en passant du cas qu'il considère, des nombres aléatoires, au cas des éléments aléatoires de nature quelconque. Si, à chaque épreuve, les éléments aléatoires X , X_1 , X_2 ..., sont choisis au hasard dans un espace distancié D , on dira que X_n converge en probabilité vers X si, pour tout nombre ϵ positif, la probabilité que la distance (X, X_n) soit inférieure à ϵ , converge vers l'unité quand n croît indéfiniment. Dans le cas où X et les X_n sont des nombres aléatoires, il suffit de prendre pour distance (X, X_n) la valeur absolue $|X - X_n|$ de la différence $X - X_n$. On voit que le théorème de Bernoulli s'énonce alors d'une façon plus brève et plus frappante : la fréquence $F = \frac{R}{n}$ d'un événement E de probabilité constante p converge en probabilité vers p quand le nombre n des épreuves croît indéfiniment.

Mais une découverte fondamentale (à mon avis, la plus importante en calcul des probabilités des cinquante dernières années), est venue apporter au théorème de Bernoulli, un complément d'information considérable et en même temps servir de point de départ à un très grand nombre des progrès ultérieurs de la théorie. Non seulement F converge en probabilité vers p mais encore, d'après M. Borel, l'événement consistant en ce que F converge au sens ordinaire vers p est un événement dont la probabilité est égale à l'unité.

Il introduisait ainsi du même coup la notion de convergence presque certaine qui a permis à beaucoup d'auteurs de donner une plus grande valeur à des énoncés antérieurs ne concernant que la convergence en probabilité. Car la convergence presque certaine implique la convergence en probabilité, mais non réciproquement.

Mais nous pouvons maintenant donner aussi une extension nouvelle et consi-

dérable à cette notion, en nous plaçant encore dans le cas d'éléments aléatoires X, X_n choisis au hasard dans un espace distancié D . Nous commençons d'abord par dire que : si x, x_n sont des éléments de D , x_n converge vers x quand la distance (x, x_n) tend vers 0. C'est là une extension toute naturelle et qui ne concerne pas encore le calcul de probabilités. Celui-ci intervient, au contraire, quand nous dirons que X_n converge *presque certainement* vers X si l'événement, consistant en ce que dans une épreuve, X_n converge au sens qui vient d'être rappelé vers X , est un événement de probabilité égale à 1.

Voilà donc étendues au cas d'éléments abstraits, deux modalités particulières de la convergence stochastique, notion générale où la convergence est moins stricte dans les ensembles d'épreuves de moindre probabilité.

Nous avons pu en définir de même une troisième modalité dans un espace distancié : la convergence en moyenne d'ordre r et même des convergences stochastiques plus générales encore. Nous dirons que dans un espace distancié D X_n converge vers X en moyenne d'ordre r si $\mathcal{M}(X_n, X)^r$ tend vers 0 (en désignant en général par $\mathcal{M} U$ la valeur moyenne d'un nombre aléatoire U). Or on démontre que si X, Y, Z sont trois éléments aléatoires appartenant à un espace distancié D et si $r \geq 1$, on a

$$(1) \quad \sqrt[r]{\mathcal{M}(X, Y)^r} \leq \sqrt[r]{\mathcal{M}(X, Z)^r} + \sqrt[r]{\mathcal{M}(Z, Y)^r}.$$

Éléments globaux. — On est alors conduit à considérer à côté de l'espace distancié D où chaque épreuve détermine les éléments X, Y, Z , un autre espace dont les éléments $[X], [Y], [Z]$ sont, par exemple, pour $[X]$, l'ensemble des déterminations de X et que nous pouvons pour cette raison appeler un élément *global*. Plus exactement $[X]$ est déterminé : 1^o par la fonction a dans la relation $X = a(R)$ qui fait correspondre X de D au résultat R d'une épreuve; 2^o par la loi de probabilité de R . Et on voit d'après (1) qu'on peut faire, de cet espace des éléments globaux, un espace distancié B , en prenant pour distance de deux éléments globaux $[X], [Y]$ l'expression

$$([X], [Y]) = \sqrt[r]{\mathcal{M}(X, Y)^r}$$

pourvu qu'on considère comme éléments distincts, deux éléments qui sont identiques sauf dans un ensemble d'épreuves ayant une probabilité nulle.

Cette conception des éléments globaux conduit alors à de nouvelles formes de convergence stochastique. Il suffit de supposer qu'on ait associé à chaque couple $[X], [Y]$ d'éléments globaux, une distance $([X]), ([Y])$, vérifiant les conditions 1^o, 2^o, 3^o ci-dessus et de dire que X_n converge stochastiquement vers X (relativement à cette définition de la distance) quand $([X_n]), ([X])$ tend vers 0. Par exemple, on peut définir plusieurs distances non équivalentes des éléments globaux, telles que les convergences stochastiques correspondantes soient toutes identiques avec la convergence en probabilité.

Il ne suffit pas d'étendre les définitions, il faut étendre les théorèmes et, à cet effet, leurs démonstrations. Nous renverrons pour ces précisions techniques à notre mémoire en impression dans les Annales de l'Institut Henri Poin-

(1) Notions dont les définitions se trouvent rappelées dans les deux derniers mémoires cités, comme dans notre livre sur « Les espaces abstraits ». Gauthier-Villars, 1928.

caré ou, pour ceux qui se contenteraient des résultats, à l'exposé qui en est fait sans démonstration dans un article qui paraîtra en anglais dans un nouveau périodique américain « Mathematics Magazine ».

Notons toutefois que si certaines démonstrations relatives aux nombres aléatoires s'étendent immédiatement aux éléments aléatoires de nature quelconque quand on remplace en général $(X - Y)$ par la distance (X, Y) , d'autres nécessitent des modifications essentielles. Celles-ci tiennent, *entre autres*, au fait que, si tout espace euclidien est en même temps « séparable et complet »(1), il n'en est pas de même de tous les espaces distanciés.

Dispersion et positions typiques. — Quand X est un élément aléatoire choisi au hasard dans un espace distancié D , comment peut-on lui étendre les définitions fondamentales usitées en statistique pour la moyenne, la valeur équitable, la dispersion d'un nombre aléatoire? Nous indiquerons deux moyens d'y arriver. Le plus simple et le plus général consiste à étendre non pas les définitions habituelles des valeurs typiques mais celles qui résultent d'une de leurs propriétés. On sait que si X est un nombre aléatoire, a un nombre certain et si la valeur moyenne de $|X|^r$, soit $\mathcal{M} |X|^r$ est finie, alors la quantité $\mathcal{M} |X - a|^r$ est une quantité dépendant de a , qui atteint son minimum, quand a varie, pour une valeur de a qui pour $r = 2$ est la valeur moyenne \bar{X} de X et pour $r = 1$ est une valeur équitable $\bar{\bar{X}}$ de X . De plus, on peut mesurer la dispersion de X par le minimum de $\sqrt[r]{\mathcal{M} |X - a|^r}$, minimum qui est l'écart quadratique moyen de X pour $r = 2$ ou son écart moyen pour $r = 1$. On peut appeler \bar{X} et $\bar{\bar{X}}$ les valeurs typiques d'ordres respectifs 1 et 2 de X . Dès lors, l'idée vient de généraliser ce théorème sous forme de définition, de la manière suivante.

Soient X un élément aléatoire, a un élément certain, appartenant tous deux à un espace distancié D . On appellera écart moyen d'ordre r de X la borne inférieure, finie ou non, de $\sqrt[r]{\mathcal{M} (X, a)^r}$ quand a varie. On pourra mesurer la dispersion de X par un tel écart moyen.

Quand X est borné en moyenne d'ordre r , c'est-à-dire quand son écart moyen d'ordre r est fini, nous appellerons position typique d'ordre r de X dans D , toute détermination de a , s'il en existe dans D , telle que $\mathcal{M} (X, a)^r$ y atteigne son minimum.

Quand X n'est pas borné en moyenne d'ordre r , nous avons montré dans le mémoire déjà cité des *Annales de l'Institut H. P.* comment on peut étendre la définition précédente. Il est intéressant de noter qu'en appliquant cette extension au cas des nombres aléatoires, on arrive à une définition de la valeur moyenne qui, dans le cas d'une dispersion infinie, est plus générale que la définition classique (1).

On peut d'ailleurs définir des *positions typiques plus générales* encore, en recourant à la notion de distance de deux éléments globaux. Quant l'écart moyen d'ordre r de X est fini, la position typique d'ordre r de X est l'élément certain, s'il en existe, dont la distance à l'élément global $[X]$ est la plus

(1) *Nouvelles définitions de la valeur moyenne et des valeurs équi-probables d'un nombre aléatoire.* Annales Univ. Lyon, 3^e série, Sciences, section A, 1946, p. 5-26.

petite quand on la compte dans l'espace B_r . De même si U est l'espace distancié obtenu en associant à l'ensemble des éléments globaux, une certaine distance $([X], [Y])$, une position typique de X (relative à U) sera un élément certain b , s'il en existe, dont la distance $([X], [b])$ à l'élément global $[X]$ est la plus petite quand on la compte dans l'espace U . Que b existe ou non, la dispersion de X (relativement à U) pourra être mesurée par la borne inférieure, finie ou non, de la distance $([X], [a])$ de l'élément global X à l'élément certain $'a$, quand a varie.

Bien entendu, les positions typiques et la dispersion dépendront de la définition adoptée pour la distance. Mais même dans le cas des nombres aléatoires, nous savons qu'il n'y a pas une définition des valeurs typiques mais un grand nombre, de telles définitions d'ailleurs non équivalentes. On distingue ainsi, valeur moyenne, valeur équiprobable, dominante, moyenne géométrique, moyenne harmonique, etc. Chacune d'elles a ses avantages et ses inconvénients et il y a lieu de choisir entre elles dans chaque cas particulier, celle qui est la plus représentative. L'essentiel pour les positions typiques et la dispersion sera de bien choisir la définition de la distance qui convient au cas considéré. C'est précisément à cet instant que devra être prise en considération la nature des éléments envisagés, laquelle avait pu être négligée dans la définition générale. Mais avant de considérer quelques-uns de ces cas particuliers, passons à la seconde définition que nous avons donné de l'élément moyen, c'est-à-dire de la position typique d'ordre deux.

C'est celle qui s'applique lorsque l'espace distancié D (où est choisi au hasard l'élément aléatoire considéré X) est en même temps « vectoriel », c'est-à-dire où cet espace D est un des espaces abstraits définis simultanément par Wiener et Banach dans lesquels à tout couple ordonné de deux éléments a, b de D est associé un « vecteur » abstrait ξ . Et ceci de sorte que ce vecteur abstrait jouisse de celles des propriétés des vecteurs géométriques qui peuvent s'énoncer sans intervention de la géométrie. Par exemple, on définit la somme de deux vecteurs, la multiplication d'un vecteur par un nombre, etc... (1).

Dans ce cas, il est possible de parler, d'une somme d'éléments de D , d'une combinaison linéaire de ces éléments; il est donc possible de généraliser l'intégrale classique $\int_{-\infty}^{+\infty} x d F(x)$ (où l'on a posé $F(x) = \text{Prob. } [X < x]$) qui représente la moyenne d'un nombre aléatoire X , en la considérant comme limite d'une somme.

Mais il y a assez loin cependant de cette intégrale numérique à l'intégrale abstraite qu'il s'agit d'introduire. Sans reproduire ici les raisonnements qui nous ont conduit à cette définition, indiquons brièvement celle-ci : soit R un élément aléatoire choisi au hasard dans un espace abstrait Δ quelconque et $X = \Phi(R)$, un élément correspondant à R et appartenant à un espace vectoriel distancié (ou, en abrégé, un espace $W-B$) et définissons l'intégrale :

$$(2) \quad \bar{X} = \int_{\Delta} \Phi(\xi) p(de)$$

qui va représenter la position moyenne de X . Dans cette intégrale, nous désignons par $p(h)$ la fonction de distribution de R (c'est-à-dire la probabilité

(1) Pour plus de détails sur les espaces de Wiener-Banach, voir, par exemple, les pages 123 à 147 de notre ouvrage « Les Espaces abstraits »,

pour que R appartienne à un ensemble h d'éléments de Δ et par ξ un élément du sous-ensemble d de Δ . L'intégrale (f) n'est qu'une notation pour représenter un élément certain de $W-B$, à savoir \bar{X} , défini comme limite (1), quand ϵ tend vers zéro, de certains éléments de $W-B$ qui sont des sommes :

$$s = \sum p(e_k) \cdot \Phi(\eta_k).$$

Dans cette relation, η_k est un élément quelconque de e_k et les e_k sont des sous-ensembles disjoints de Δ , dont la réunion est Δ et sur chacun desquels l'oscillation (2) de $\Phi(R)$ est inférieure au nombre $\epsilon > 0$.

Dans un mémoire (3), nous avons précisé des cas très généraux où une somme s a un sens — c'est-à-dire où la suite d'éléments de $W-B$ qu'elle représente est convergente — et où, quand ϵ tend vers 0, s tend vers une limite indépendante du choix des e_k et des η_k .

On voit que \bar{X} est une intégrale abstraite d'une fonction Φ abstraite d'un élément abstrait R , R appartenant à Δ , Φ et \bar{X} appartenant à l'espace vectoriel distancié $W-B$.

La distance dans les espaces fonctionnels. — Donnons les exemples de définitions annoncés plus haut. Considérons un ensemble E de fonctions numériques $f(x)$ définies sur un segment fixe S ($a \leq x \leq b$). On pourra faire de E un espace distancié en y associant une définition de la distance. Suivant la nature des fonctions de E , on aura intérêt à y adopter telle ou telle définition de la distance.

Par exemple, quand E est l'ensemble des fonctions continues sur s , on pose généralement $(f, g) = \text{maximum de } |f(x) - g(x)|$ quand x varie sur s .

Dans la théorie des moindres carrés, si $f(x)$ et $g(x)$, continues ou non, sont de carrés intégrables sur s , c'est-à-dire telles que :

$$\int_a^b f^2(x) dx \quad \text{et} \quad \int_a^b g^2(x) dx$$

soient finies, on pose :

$$(f, g) = \sqrt{\frac{1}{b-a} \int_a^b [f(x) - g(x)]^2 dx}.$$

Enfin, rappelons que le calcul des probabilités associe à chaque nombre aléatoire X , sa fonction de répartition :

$$F(x) = \text{Prob. } [X < x].$$

Or, une telle fonction n'appartient nécessairement à aucune des deux catégories précédentes. Toute fonction de répartition est définie de $-\infty$ à $+\infty$, elle est non décroissante et varie de 0 à 1. On a donc jugé utile de définir une distance spéciale à cette catégorie de fonction. Paul Lévy a donné une définition qui permet d'obvier à la difficulté présentée par la présence possible

(1) C'est-à-dire que

$$\lim_{\epsilon \rightarrow 0} (\bar{X}, s) = 0$$

(2) L'oscillation de $\Phi(R)$ sur un sous-ensemble h de Δ est la borne supérieure des distances $(\Phi(R), \Phi(R'))$ quand R et R' varient sur h .

(3) *L'intégrale abstraite d'une fonction abstraite d'une variable abstraite avec application à la moyenne d'un élément aléatoire de nature quelconque*, Revue Scientifique, 1945.

de discontinuités. Il prend pour distance (F, G) de deux lois de répartition $F(x), G(x)$, le plus grand des segments limités sur les parallèles à la seconde bissectrice par les courbes représentatives de $F(x)$ et $G(x)$. On aura eu soin de compléter ces courbes, pour les rendre continues, si elles ne le sont pas, par des segments perpendiculaires à l'axe des x et ayant pour abscisses celles des points de discontinuités respectifs de $F(x)$ et $G(x)$.

On démontre que ces trois définitions de la distance vérifient les trois conditions 1^o, 2^o, 3^o ci-dessus, pourvu qu'on considère comme non distinctes deux fonctions égales presque partout, c'est-à-dire égales sauf peut-être en un ensemble de points qu'on peut, pour tout $\epsilon < 0$ enfermer dans un ensemble dénombrable d'intervalles de longueur totale inférieure à ϵ .

Cette troisième définition de la distance est l'une de celles qui peuvent être employées dans notre définition de nouveaux indices de corrélation.

D'autre part, les deux premiers espaces sont des espaces de Wiener-Banach. (Il suffit de considérer comme somme de deux éléments f, g , la somme des fonctions correspondantes $f(x), g(x)$ et comme produit $c f$ le produit de $f(x)$ par la constante c). On peut donc représenter dans les deux premiers espaces, par une intégrale de la forme (f) , la moyenne d'une fonction aléatoire.

Comparaison des deux définitions de la moyenne. — La seconde définition a l'avantage sur la première de donner une expression générale et explicite de la moyenne \bar{X} alors que la première, tout en restant aussi précise, ne donne le moyen de déterminer la moyenne que dans chaque cas particulier. Si nous avons tenu à donner aussi la première, c'est d'une part, qu'elle généralise non seulement la valeur moyenne en prenant $r = 2$, mais aussi la ou les valeurs équiprobables en prenant $r = 1$. C'est surtout, d'autre part, parce qu'elle est applicable à des espaces dont on sait seulement qu'ils sont distanciés sans savoir s'ils sont vectoriels. La plupart des espaces fonctionnels usuels sont vectoriels distanciés. Mais, pour le moment, les espaces dont les éléments sont des courbes ou des surfaces, s'ils sont facilement distanciables, n'ont pas été jusqu'ici présentés comme des espaces vectoriels. Ce serait même là une question très importante à résoudre de savoir si cela est possible de façon naturelle. Une réponse positive serait utile dans bien des domaines et permettrait, en particulier, d'appliquer aux courbes ou aux surfaces la seconde comme la première définitions.

Revenons à la seconde définition de la moyenne.

On peut obtenir des résultats d'autant plus précis qu'on particularise plus la catégorie d'espace distancié considérée. Appelons espace G un espace vectoriel distancié qui est, soit un espace à un nombre finis de dimensions, soit l'espace considéré plus haut des fonctions de carré sommables sur un segment S (1). Et, pour tout point X de G , représentant par $\|X\|^2$, le carré de la distance du point X à l'origine dans le premier cas ou l'intégrale $\int_a^b X^2(t) dt$ dans le second cas. Alors, en posant $S = X_1 + \dots + X_n$ et en supposant les X_k indépendants, on pourra généraliser l'égalité de Bienaymé sous la forme :

$$\mathfrak{M} \|S - \bar{S}\|^2 = \sum_{k=1}^n \mathfrak{M} \|X_k - \bar{X}_k\|^2.$$

(1) Voir dans notre dernier mémoire cité, une définition plus générale de G .

On en déduit une extension de la loi des grands nombres à l'espace G sous la forme suivante :

Soit $\sigma_k^2 = \mathfrak{M} \|\bar{X}_k - \bar{X}_k\|^2$ la fluctuation de X_k et soit $V_n = \frac{X_1 + \dots + X_n}{n}$.

Si les X_k sont bornés en moyenne d'ordre deux, c'est-à-dire si les $\mathfrak{M} \|\bar{X}_k\|^2$ sont finis et si l'on suppose encore les X_k (de G) indépendants, alors :

1^o Si le rapport

$$\frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2} \text{ tend vers zéro avec } \frac{1}{n},$$

$V_n - \bar{V}_n$ tend vers 0 en probabilité.

2^o Si la condition un peu plus stricte suivante est réalisée, à savoir si la série

$$\sigma^2 + \dots + \frac{\sigma_n^2}{n^2} + \dots$$

est convergente, alors $V_n - \bar{V}_n$ tend vers 0 presque certainement.

Nous venons de résumer quelques-unes de nos recherches sur la théorie des éléments aléatoires. Mais tout cela n'est qu'un début; il y a énormément à faire.

Une des questions les plus délicates sera de déterminer (et ce n'est pas là une question exclusivement mathématique) quelle définition de la distance il convient d'adopter en vue d'une application déterminée. Une autre sera de trouver quelles sont les lois de probabilité les plus simples qu'on peut rencontrer dans les applications. C'est là un problème purement expérimental dont nous avons commencé l'étude simultanément avec nos recherches théoriques.

Nous étudions actuellement numériquement la loi de fréquence des formes d'un fil noué jeté sur une table. Est un peu plus avancée l'étude de la loi de fréquence des formes d'une coupe horizontale d'un crâne humain à hauteur du front. Ce cas est plus simple que celui du fil :

1^o parce que les formes assez régulières, assez simples et assez peu dissemblables de ces contours crâniens permettent de les représenter chacun par une seule équation en coordonnées polaires :

$$\rho = f(\omega)$$

(le centre étant par exemple le milieu de l'axe de symétrie (approximative du contour) où $f(\omega)$, étant de période 2π , peut être développé en série de Fourier.

$$\rho = A_0 + A_1 \cos \omega + B_1 \sin \omega + \dots + A_n \cos n \omega + B_n \sin n \omega + \dots$$

2^o on peut obtenir une approximation suffisante en général par une dizaine d'harmoniques. D'où 25 coefficients à calculer tandis que nous estimons à 120 le nombre des coefficients à calculer pour déterminer suffisamment les formes très compliquées des fils mentionnés ci-dessus.

Nous avons d'abord cherché à voir si les lois de fréquence des coefficients sont normales et si ces coefficients sont indépendants. Cela nous a donné l'occasion d'appliquer à cette recherche quatre indices de corrélation : le coefficient de corrélation r , le coefficient de connexion g de Gini, l'indice j de

Jordan et notre indice diagonal d . On constate que généralement r est voisin de 0 tandis que les trois autres indices prennent des valeurs assez petites, mais montant déjà jusque vers 0,2; 0,3; 0,4

Le coefficient r nous aurait fait conclure à l'indépendance des coefficients; les trois autres indices marquent une dépendance faible mais non négligeable.

De même que pour l'étude théorique, l'étude statistique est aussi à ses débuts et devra être non seulement renouvelée avec une précision accrue, mais étendue à d'autres questions. J'espère que de nouveaux chercheurs viendront apporter leur contribution à l'une ou l'autre de ces formes de recherches et collaborer ainsi au développement d'un nouveau et passionnant chapitre de la science qui nous réunit ici.

Maurice FRÉCHET.