

ANNALES DE L'I. H. P.

MAURICE FRÉCHET

**Les éléments aléatoires de nature quelconque
dans un espace distancié**

Annales de l'I. H. P., tome 10, n° 4 (1948), p. 215-310

[<http://www.numdam.org/item?id=AIHP_1948__10_4_215_0>](http://www.numdam.org/item?id=AIHP_1948__10_4_215_0)

© Gauthier-Villars, 1948, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P. » 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/>

Les éléments aléatoires de nature quelconque dans un espace distancié

par

Maurice FRÉCHET.

PRÉFACE.

On trouvera dès l'*Introduction* qui suit, la signification du titre de ce Mémoire, dont une Table des matières figure à la page 309. Nous nous contenterons ici de dire en deux mots qu'il constitue la mise au point de certains des cours que nous avons faits à la Sorbonne en 1945 et en 1946. Ce travail n'est qu'un des résultats des grands efforts que nous avons accomplis, *tant du côté statistique et expérimental*, que du côté théorique, pour jeter les bases d'un Chapitre entièrement nouveau, et que nous croyons assuré d'un grand avenir, du Calcul des Probabilités. Dans le présent Mémoire, nous croyons avoir montré quelle formidable extension de ce Calcul résulte de la simple introduction de la notion de distance de deux éléments aléatoires.

Dans un prochain article, nous étudierons de même l'intégration et la différenciation « stochastiques » et à cette fin il nous sera nécessaire d'adjoindre à cette notion de distance, celle du vecteur abstrait correspondant.

On pourrait nous reprocher d'avoir limité nos applications dans ce qui suit aux cas déjà abordés des nombres aléatoires, des vecteurs aléatoires et des fonctions aléatoires. Mais c'est qu'il s'agissait avant tout de montrer que notre théorie générale était susceptible d'application

quelle qu'elle soit; que d'autre part elle permet, même dans ces cas particuliers d'obtenir du nouveau, comme notre généralisation (p. 239) de la notion de valeur moyenne d'un *nombre* aléatoire; qu'enfin et surtout, tout ceci n'est qu'un début sur lequel nous comptons revenir et être suivi. Par ailleurs, cette limitation nous a permis de rester élémentaire et l'on observera que cet article peut être lu par un lecteur n'ayant que des notions très succinctes de Calcul des Probabilités. De même, ce lecteur trouvera rappelées ici-même, les quelques définitions que nous aurons à emprunter à la Théorie des espaces abstraits.

J'insiste encore avant de terminer sur l'importance des notions introduites ici dans les applications les plus concrètes, les plus pratiques, bien que je n'en ait dit ici que quelques mots (p. 217 et 221).

INTRODUCTION.

Éléments aléatoires nouveaux. — Le Calcul des probabilités a été implicitement ou explicitement, jusqu'à une époque récente, l'étude des *nombres* aléatoires et des points aléatoires dans un espace à une, deux ou trois dimensions (probabilités géométriques).

Depuis peu, on a souvent cherché à étendre les résultats obtenus aux séries aléatoires, aux vecteurs aléatoires et aux fonctions numériques aléatoires de variables numériques certaines.

Mais la nature, la science et la technique offrent de nombreux exemples d'éléments aléatoires qui ne sont, ni des nombres, ni des séries, ni des vecteurs, ni des fonctions.

Telles sont par exemple, la forme d'un fil jeté au hasard sur une table, la forme d'un œuf pris au hasard dans un panier d'œufs.

On a ainsi une courbe aléatoire, une surface aléatoire. On peut aussi considérer d'autres éléments mathématiques aléatoires : par exemple des transformations aléatoires de courbe en courbe. On peut aussi rencontrer des éléments qui n'ont pas jusqu'ici été décrits mathématiquement.

On étudiait les *nombres* aléatoires obtenus en choisissant au hasard une ville dans un pays donné et en notant un nombre relatif à cette ville, comme sa population ⁽¹⁾, le nombre de ses maisons, etc. Mais on peut aussi considérer des éléments attachés à une ville choisie au hasard et qui ne peuvent se décrire par l'intermédiaire d'une des notions mathématiques usuelles : nombre, fonction, courbe, etc. Par exemple, la moralité de sa population, son état d'esprit politique, l'impression de beauté qu'elle donne, etc. C'est une catégorie nouvelle d'éléments aléatoires. Sans aller jusque-là, il paraît certain que l'urbanisme conduira à étudier des éléments aléatoires tels que la forme d'une ville prise au

⁽¹⁾ On a ainsi observé que la distribution des populations des villes vérifie la loi de probabilité dite de Pareto.

hasard, vérifier ainsi par exemple, d'une manière scientifique l'hypothèse de la tendance au développement des villes vers l'Ouest, etc.

Nous voyons ainsi que si l'on s'est occupé surtout jusqu'ici en Calcul des Probabilités et en Statistique Mathématique, des *nombres* aléatoires, *ce n'est pas* parce que l'étude d'éléments aléatoires de natures très diverses, ne s'impose pas aussi bien dans les applications. C'est surtout parce que, comme dans toutes les sciences, on s'est attaché d'abord à ce qui était plus simple ou plus facile.

Étude simultanée. — Admettant donc maintenant qu'il fallait bien, un jour ou l'autre, aborder l'étude des diverses natures d'éléments aléatoires (et constatant d'ailleurs comme plus haut que cette étude a déjà commencé), une question se pose. Devons-nous traiter les éléments aléatoires en étudiant séparément et successivement chacune de leurs catégories : vecteurs, puis fonctions, puis courbes, puis surfaces, etc.? Assurément, il y aura des différences dans les méthodes à employer et dans les résultats. Mais devons-nous imiter ce qui s'est passé quand on a étudié successivement forces, vitesses, accélérations, rotations, tourbillons, etc., pour s'apercevoir seulement assez tard qu'à côté des résultats différents obtenus, il y avait un grand nombre de notions et de propriétés communes qu'on a fini par réunir dans une théorie commune, l'Analyse vectorielle, au lieu de les répéter plusieurs fois sous des noms différents?

Cette expérience nous a conduit à nous demander s'il serait possible d'étendre certaines notions et certains résultats concernant les nombres aléatoires à des éléments aléatoires de nature plus générale. Or, nous allons voir que cela est possible, non seulement pour une liste déterminée d'éléments aléatoires, de natures diverses mais spécifiées, mais même pour des éléments aléatoires abstraits.

Les éléments abstraits. — Nous n'entendons pas d'ailleurs sous la dénomination d'élément abstrait, une entité métaphysique, mais tout uniment un élément dont la nature nous est inconnue ou même plus simplement un élément de nature connue mais n'entrant pas en ligne de compte. Loin d'être une conception nouvelle en principe, c'est déjà celle qui règne dans l'arithmétique la plus élémentaire, où l'addition des nombres 2 et 3 d'éléments de deux collections garde le même sens qu'il s'agisse d'ajouter 2 et 3 pommes ou 2 et 3 strophes poétiques.

La théorie des espaces abstraits donne le moyen d'aborder cette théorie des éléments aléatoires abstraits. Nous n'en supposons d'ailleurs pas la connaissance, et donnerons au fur et à mesure les définitions nécessaires.

Loi de probabilité. — Parmi les premières notions à généraliser figurent celles de loi de probabilité, de valeurs typiques (moyenne, équiprobable, etc.), de dispersion, de convergence stochastique, que nous allons d'abord examiner.

En ce qui concerne la loi de probabilité, s'il y a des difficultés d'application, il n'y a aucune difficulté de principe. Soit X un élément aléatoire choisi au hasard dans un ensemble $\mathcal{E}$ d'éléments. Soit maintenant e un sous-ensemble quelconque d'éléments de $\mathcal{E}$. Nous dirons que e est « probabilisable », relativement à X , s'il y a une probabilité déterminée que X appartienne à e . Cette probabilité, déterminée par e , est une fonction de e , que nous pouvons représenter par $p(e)$ et appeler la *fonction de distribution* de X . Connaître la *loi de probabilité* de X , c'est connaître sa fonction de distribution, c'est-à-dire la correspondance entre les ensembles e probabilisables relativement à X et les probabilités $p(e)$ pour que X appartienne à e . Si, par exemple, X est un nombre aléatoire compris entre 0 et 1 et dont la loi de probabilité est dite uniforme, les ensembles probabilisables relativement à X sont les ensembles, dits mesurables ⁽¹⁾, de nombres entre 0 et 1 et la fonction de distribution $p(e)$ est le nombre, dit mesure de e , qui se réduit, quand e est un segment, à la longueur de ce segment.

Il est clair que la fonction $p(e)$ ne peut être quelconque. On a d'abord $0 \leq p(e) \leq 1$. Et naturellement $p(\mathcal{E}) = 1$. Et d'autre part si $e_1, e_2, \dots, e_n$ sont des sous-ensembles probabilisables de E qui sont disjoints, c'est-à-dire sans éléments communs, on a en vertu du principe des probabilités totales

$$(1) \quad p(e_1) + p(e_2) + \dots + p(e_n) = p(e_1 + \dots + e_n),$$

(en désignant par $e_1 + \dots + e_n$ la réunion des éléments de e_1 , de e_2 , ... et de e_n), ceci quand $e_1 + \dots + e_n$ est aussi probabilisable.

⁽¹⁾ C'est la restriction dans ce cas simple aux ensembles mesurables qui justifie la limitation dans le cas général aux ensembles « probabilisables ».

Pour éviter des difficultés d'application de cette égalité, nous supposons que la famille des ensembles probabilisables est additive au sens restreint, c'est-à-dire telle que si $e_1, e_2, \dots, e_n$ sont des ensembles probabilisables disjoints, $e_1 + \dots + e_n$ est aussi probabilisable. C'est là une précaution de principe, car dans la pratique tous les événements qu'on rencontre ont chacun une probabilité déterminée.

Nous voyons, d'autre part, qu'en vertu du principe des probabilités totales, toute fonction de distribution est additive au sens restreint, c'est-à-dire vérifie (1).

En fait, pour éviter d'autres difficultés théoriques et sans qu'il en résulte aucune restriction dans la pratique, nous supposons même que la famille $\mathcal{X}$ des ensembles probabilisables et la fonction de distribution de X sont additives non seulement au sens restreint ci-dessus mais au sens complet, consistant en ce que la validité des définitions précédentes est étendue au cas où au lieu d'un nombre fini d'ensembles e_i , on en a une suite infinie mais dénombrable (c'est-à-dire qu'on peut numéroter par la suite infinie des entiers) $e_1, e_2, \dots, e_n, e_{n+1}, \dots$.

Quand on veut passer au calcul de $p(e)$, on se heurte à des difficultés, même dans le cas simple où X est un *nombre*. De sorte que pour déterminer la loi de probabilité, on se contente généralement dans ce cas de donner les valeurs de $p(e)$ pour des ensembles de nombres particuliers, à savoir ceux qui sont inférieurs à un nombre donné x arbitraire. On introduit ainsi la « fonction de répartition de X » : $F(x) = \text{Prob}\{X < x\}$, qui n'est autre que $p(e_x)$ où e_x est l'ensemble des nombres $X < x$ ⁽¹⁾.

De même, dans la théorie des fonctions numériques aléatoires $X(t)$ d'un nombre numérique certain t , on définit généralement la loi de probabilité par l'ensemble des fonctions

$$F_n(x_1, x_2, \dots, x_n | t_1, t_2, \dots, t_n) = \text{Prob}[X(t_1) < x_1, \dots, X(t_n) < x_n],$$

où sont variables $n, t_1, \dots, t_n, x_1, \dots, x_n$.

Mais ce sont là des restrictions d'ordre pratique et l'on ne peut considérer la loi de probabilité d'un élément aléatoire comme complètement déterminée que si l'on connaît $p(e)$ pour tout ensemble probabilisable e .

(1) Ici, comme dans toutes nos publications, nous représentons les éléments aléatoires par de grandes lettres et les éléments certains par des minuscules. On évite ainsi nombre de confusions qui se présentent dans les mémoires négligeant cette distinction typographique.

D'ailleurs, on sait bien, par exemple dans le cas où X est un nombre, que la connaissance de sa fonction de répartition $F(x)$ entraîne celle de sa fonction de distribution $p(e)$.

Nous indiquerons plus loin en Note additionnelle, page 306, un moyen très général d'étendre la notion de fonction de répartition à des éléments aléatoires de nature quelconque.

Détermination statistique de la loi de probabilité. — Comme l'étude des éléments aléatoires de nature nouvelle en est encore à ses débuts, il est d'une importance capitale de ne pas attendre d'en avoir avancé la théorie pour en entreprendre l'étude expérimentale. Car la théorie aura besoin de considérer, pour commencer, des lois simples de probabilité et il serait illusoire de commencer par celles dont l'expression mathématique est la plus simple, si elles ne se rencontraient pas dans les cas concrets les plus simples. C'est pourquoi nous avons déjà entrepris en même temps que l'étude théorique des éléments aléatoires une étude expérimentale de quelques exemples particuliers. Ceux-ci sont actuellement : la forme aléatoire d'un fil jeté sur une table, la forme aléatoire d'une section horizontale (prise à un niveau anatomique déterminé) d'un crâne humain choisi au hasard.

Pour ce second exemple, notre choix s'est porté sur deux collections de contours obtenus au « conformateur » d'une part sur des personnes vivantes (mais anonymes), contours fournis gracieusement par la grande maison parisienne de chapellerie Sools, d'autre part sur une collection ethnographique de crânes d'individus morts, collection mise aimablement à notre disposition par M. Lester, Directeur du Laboratoire d'Ethnographie.

En même temps, un de nos auditeurs du Caire, M. Rifaat, a entrepris, sous la direction du Docteur Simaïka, l'étude statistique d'une collection de contours crâniens préhistoriques qui lui ont été prêtés par le professeur Morand, de Londres.

Pour chacune de ces quatre collections, les fonctions représentatives sont actuellement développées en série de Fourier et l'on procèdera à une analyse statistique de ces séries.

En particulier, on déterminera les lois de fréquence des premiers coefficients pris isolément et l'on étudiera d'autre part leur corrélation

au moyen de divers indices de corrélation et, en particulier, de l'indice diagonal que nous avons défini récemment ⁽¹⁾ et qui n'est pas restreint comme le coefficient de corrélation, r , au repérage des corrélations *linéaires*.

Ce sont là des opérations longues et difficiles auxquelles ont bien voulu coopérer, chacun de leur côté (avec le Laboratoire de Calcul de l'Institut Henri Poincaré) deux physiciens bien connus qui se sont occupés avec succès des phénomènes périodiques, MM. Labrouste et Duffieux, assistés de leurs aides, opérant respectivement sur deux analyseurs harmoniques.

Pour les fils, environ 120 coefficients ! pour *chaque* forme de fil, paraissent nécessaires pour la représenter convenablement, et nous disposons de 100 de ces formes, correspondants à 100 jets. Pour les contours crâniens dont les formes sont infiniment moins variées, nous nous sommes contentés de 12 harmoniques soient 25 coefficients pour chaque crâne.

Les calculs sont en cours, mais sans être encore à même d'en tirer des conclusions définitives, nous nous sommes déjà posé le problème suivant. Si l'on représente par

$$\varphi = a_0 + a_1 \cos \omega + b_1 \sin \omega + \dots + a_n \cos n\omega + b_n \sin n\omega + \dots,$$

le développement en série de Fourier de l'équation en coordonnées polaires ρ, ω , d'un contour crânien, on pourrait assez naturellement être conduit à supposer comme premier cas simple que chaque coefficient a_k ou b_k obéit à la loi de Laplace et que les coefficients $a_0, a_1, b_1, a_2, b_2, \dots$, sont indépendants ⁽²⁾.

Les premiers calculs faits semblent révéler que *les coefficients, s'ils ne sont pas indépendants*, ont au moins une dépendance assez lâche et qu'ils suivent, quoique assez grossièrement, *la loi de Laplace*.

En attendant d'avoir terminé cette étude statistique des contours crâniens, M^{lle} Duhamel a bien voulu entreprendre, sur notre demande, une analyse statistique basée sur les tableaux de coefficients de Fourier aimablement fournis par M. Ozil. Celui-ci, guidé par de tout autres

⁽¹⁾ *Anciens et nouveaux indices de corrélation...* (*Econometrica*, vol. 15, 1947, p. 1-30).

⁽²⁾ Nous appliquons ici, dans un cas particulier, l'idée de définir la loi de probabilité d'un élément aléatoire par celle d'une suite de nombres définissant cet élément, idée précisée dans la Note additionnelle de la page 306.

motifs que notre étude des éléments aléatoires de nature quelconque a, dès 1938, émis l'idée qu'on pourrait aider à la classification des races au moyen de l'analyse harmonique des profils humains et a procédé depuis, à cette analyse et à sa discussion grâce aux moyens de recherche et de publication mise à sa disposition par M. Canac, Directeur du Centre de Recherches scientifiques industrielles et maritimes de Marseille. Les tableaux de M. Ozil ne correspondent pour des formes en fait assez compliquées qu'à 9 harmoniques et représentent seulement la moitié expressive du profil, c'est-à-dire une fonction *non* périodique. C'est pourquoi, nous ne considérons leur étude statistique que comme pouvant seulement nous donner de premières indications qui pourront nous guider dans la poursuite de l'étude des trois catégories de courbes aléatoires mentionnées plus haut et de celles qui pourront être entreprises plus tard. On pourra sans doute assez vite aborder l'étude statistique des surfaces aléatoires, puis plus tard celle d'éléments aléatoires de natures plus complexes.

Retour à la théorie. Moyennes. — Dans un premier essai ⁽¹⁾ de généralisation de la notion de moyenne d'un nombre aléatoire, nous avons d'abord défini la moyenne d'un élément aléatoire appartenant à un « espace abstrait *vectoriel* distancié » ⁽¹⁾. Ce dernier espace a été obtenu simultanément en 1920 par Wiener et Banach en combinant la notion d'espace abstrait vectoriel défini autrefois par Grassmann et la notion d'espace abstrait distancié que nous avons introduite beaucoup plus tard en 1904 et dont la définition sera rappelée page 226.

En imposant à la notion de moyenne quelques conditions intuitives simples, nous avons montré ⁽¹⁾ que si X reste dans toute épreuve dans un même espace de Wiener-Banach, la moyenne de X peut s'exprimer au moyen d'une généralisation convenable de la notion d'intégrale. Nous renverrons à ce sujet à notre Mémoire cité ⁽¹⁾.

Cette extension convenait bien au cas où l'élément aléatoire X est une *fonction* numérique aléatoire d'un nombre certain, cas où l'on peut considérer ces fonctions comme appartenant à un espace vectoriel distancié. On peut, en effet, y définir la distance de diverses manières

⁽¹⁾ *L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque* (Revue Scientifique, 82^e Année, 1944, p. 483-512).

comme il sera expliqué pages 244, 247, et d'autre part, les fonctions jouissent de plusieurs propriétés des vecteurs; par exemple, on sait en quoi consiste la somme de deux fonctions et le produit d'une fonction par un nombre.

Mais il y a des cas où, même si une telle analogie avec les vecteurs est possible à établir, on ne voit pas bien actuellement comment la réaliser, d'une façon conforme à notre intuition des notions de somme, de produit, etc. (Voir cependant le dernier alinéa de la *Note additionnelle*, p. 308.)

C'est pourquoi, dans le même Mémoire (à la page 512) nous avons indiqué brièvement une autre extension de la notion de moyenne, s'appliquant cette fois à un espace distancié qui *ne serait pas* supposé vectoriel. Toutefois, en appliquant notre définition au cas particulier où l'élément aléatoire X est un nombre, notre définition ne couvrirait que le cas où la dispersion de X est finie. Bien que ce soit le seul cas se présentant en statistique numérique, il n'en est pas de même en statistique théorique et il restait à étendre notre définition au cas d'un élément abstrait X de façon qu'elle couvre encore quand X est un nombre, le cas où l'intégrale classique

$$\mathcal{M}X = \int_{-\infty}^{+\infty} x dF(x).$$

est absolument convergente (que la dispersion de X soit finie ou non). C'est ce que nous ferons plus loin, pages 234, 236, 237, après avoir rappelé page 233, la définition convenant au cas d'une dispersion finie.

Puisque, par le même procédé, on peut généraliser la notion de valeur équiprobable, nous commencerons par celle-ci qui donne lieu à des écritures plus simples et nous étendrons ensuite les définitions obtenues à des valeurs typiques plus générales. Indiquons d'abord l'idée à la base de ces extensions.

Éléments typiques. — La statistique a constamment besoin de donner une idée de ce que sont, en gros, les éléments d'une collection complexe, difficile à saisir dans son ensemble, en les assimilant à un seul élément choisi aussi ressemblant que possible à tous les éléments de la collection et qui, pour cette raison, soit appelé élément typique de cette collection. Si cette collection est formée des nombres d'un tableau, un élément

typique, pour leur ressembler, devra d'abord être aussi un nombre; il devra en outre en différer aussi peu que possible. Supposons que par un moyen quelconque, on puisse définir un « écart » entre un nombre quelconque a et l'ensemble $y_1, \dots, y_n$ du tableau, c'est-à-dire un nombre qui repère d'une façon satisfaisante l'ordre de grandeur de l'ensemble des valeurs absolues des différences entre a et les y . Alors toute valeur de a pour laquelle cet écart est minimum pourra être naturellement considérée comme un nombre typique du tableau.

Le plus naturel (et en même temps le plus simple au point de vue des calculs numériques) est de prendre pour cet écart de a et des y , la moyenne arithmétique des écarts de a avec les y , soit

$$\frac{|y_1 - a| + \dots + |y_n - a|}{n},$$

on sait qu'alors cet écart atteint son maximum quand a coïncide avec une médiane des y .

Une définition beaucoup moins naturelle et conduisant à des calculs numériques moins simples (mais qui a de considérables avantages mathématiques) de l'écart de a avec les y consiste à prendre pour cet écart

$$\sqrt{\frac{(y_1 - a)^2 + \dots + (y_n - a)^2}{n}}.$$

On sait qu'alors cet écart atteint son minimum quand a est égal à la moyenne arithmétique des y .

En correspondance avec ces définitions statistiques, quand on passe au calcul des probabilités, toute valeur certaine $\bar{X}$ telle que l'écart moyen, $\mathcal{M}|X - a|$, d'un nombre aléatoire X avec un nombre certain a atteigne son minimum, quand a varie, pour $a = \bar{X}$, est une valeur équiprobable de X . De même toute valeur certaine $\bar{X}$ telle que l'écart quadratique moyen $\sqrt{\mathcal{M}(X - a)^2}$ du nombre aléatoire X avec le nombre certain a , atteigne son minimum quand a varie, pour $a = \bar{X}$ est une valeur moyenne de X .

On observe que $\mathcal{M}|X - a|$, (l'écart moyen de X avec a) est la moyenne de l'écart aléatoire de X avec a . On est donc conduit naturellement, quand on est accoutumé à l'idée d'un espace abstrait où un écart est défini, à généraliser la définition de la valeur équiprobable de la manière suivante.

Définition de l'écart dans un espace abstrait. — Supposons que X et a soient des éléments (l'un aléatoire, l'autre certain) pris dans un ensemble $\mathcal{E}$ d'éléments *de nature quelconque* où un écart est défini. C'est-à-dire, supposons qu'à tout couple d'éléments a, b de $\mathcal{E}$ soit associé un nombre $(a, b) = (b, a) \geq 0$, appelé écart de a et de b et tel que (a, b) ne soit nul que si a et b ne sont pas distincts et réciproquement. Alors $\mathcal{M}(X, a)$ est une sorte d'écart moyen de X et de a . Si $\mathcal{M}(X, a)$ est fini pour au moins un élément certain a , $\mathcal{M}(X, a)$ a une borne inférieure finie $m \geq 0$, quand a varie sur $\mathcal{E}$. Si, en outre, il existe un élément c de $\mathcal{E}$ pour lequel l'écart moyen $\mathcal{M}(X, a)$ atteint son minimum, cet élément c pourra être considéré comme une position typique de X .

Par analogie avec le cas où X est un nombre, on pourrait aussi substituer dans ce qui précède à l'écart moyen μ de X et de a , leur écart quadratique moyen $\sigma_a = \sqrt{\mathcal{M}(X, a)^2}$ qui constitue aussi une sorte d'indice de dissemblance entre a et X .

Si le minimum de σ_a quand a varie est atteint pour une position d de a , on pourra aussi considérer d comme une position typique de a .

Mais observons que si (a, b) est un écart, $(a, b)^2$ en est aussi un, étant symétrique, ≥ 0 et nul seulement si $b \equiv a$. Si l'on applique la première définition en remplaçant (a, b) par $(a, b)^2$, et (X, a) par $(X, a)^2$, on aura à chercher le minimum de $\mathcal{M}(X, a)^2$ au titre de la première définition. Or il ne se distingue pas du minimum de σ_a^2 qui intervient dans la seconde. $(X, a)^2$ comme (X, a) jouent le rôle de $|X - a|$, on ne voit pas finalement pourquoi la première définition conduirait comme on serait d'abord tenté de le penser, à une généralisation de la valeur équiprobable et la seconde à une extension de la notion de valeur moyenne.

Espaces distanciés. — Il n'en est plus de même si la notation (a, b) désigne non seulement un écart, mais une *distance*, c'est-à-dire si cet écart vérifie quels que soient a, b, c de $\mathcal{E}$, « l'inégalité triangulaire »

$$(2) \quad (a, b) \leq (a, c) + (c, b).$$

Alors, il n'y a aucune raison pour qu'on ait nécessairement aussi

$$(a, b)^2 \leq (a, c)^2 + (c, b)^2 \quad \text{quels que soient } a, b, c \text{ de } \mathcal{E},$$

de sorte que le carré d'une distance ne peut généralement pas être considéré lui-même comme une distance.

Quand à tout couple d'éléments a, b d'un ensemble E , on peut associer une distance entendue au sens précédent, on dira que ces éléments appartiennent à un *espace distancié* $\mathcal{O}$.

Dans la suite, nous supposons toujours que les éléments aléatoires $X, Y, \dots$, déterminés simultanément par une même épreuve, sont définis sur la même catégorie d'épreuves et que ces éléments peuvent être considérés comme des points abstraits appartenant à un même espace distancié $\mathcal{O}$.

Dès lors, si nous pouvons continuer à dire, quand (a, b) est un écart que si $\mathcal{M}(X, a)$ atteint son minimum en un élément certain $a = c$ de $\mathcal{E}$, c est une position typique de X , c'est seulement *quand la notation* (a, b) *représente une distance*, dans un espace distancié $\mathcal{O}$ que nous dirons :

I. S'il existe un élément certain $\bar{X}$ de $\mathcal{O}$ tel que $\mathcal{M}(X, a)$ atteigne pour $a \equiv \bar{X}$ un minimum fini, $\bar{X}$ est une position équiprobable de X .

II. S'il existe un élément certain $\bar{X}$ de $\mathcal{O}$ tel que $\mathcal{M}(X, a)^2$ atteigne pour $a \equiv \bar{X}$ un minimum fini, $\bar{X}$ est une position moyenne de X .

Rien n'indique dans ces définitions que s'il existe une position moyenne, il n'en existe qu'une seule, bien qu'il en soit ainsi quand X est un nombre aléatoire. Cela ne doit pas nous surprendre puisque, même dans ce cas simple, la définition analogue d'une valeur équiprobable peut, comme on sait, nous donner une infinité de valeurs équiprobables prises arbitrairement dans un segment probable.

D'autre part, si $\mathcal{M}(X, a)$ et $\mathcal{M}(X, a)^2$ sont finis, chacun pour au moins une position de l'élément certain, a de $\mathcal{O}$, leurs bornes inférieures sont finies mais ne sont pas nécessairement atteintes. C'est ce qu'on verrait en prenant par exemple pour X un élément appartenant à l'espace distancié $\mathcal{O}$ constitué par les nombres réels entre -1 et $+1$, *sauf* zéro, en prenant $|a - b|$ pour distance (a, b) et en supposant la loi de probabilité uniforme. Alors

$$\mathcal{M}(X, a)^2 = \frac{1}{2} \int_{-1}^{+1} (x - a)^2 dx = a^2 + \frac{1}{2} \int_{-1}^{+1} x^2 dx,$$

a une borne inférieure finie $\frac{1}{2} \int_{-1}^{+1} x^2 dx$, mais cette borne n'est pas atteinte puisque a n'atteint pas la valeur zéro sur $\mathcal{O}$.

Dispersion. — Cependant on pourra toujours caractériser la *dispersion* de X au moyen de ces bornes inférieures.

Nous appellerons écart moyen de X , la borne inférieure μ de $\mathfrak{M}(X, a)$ quand a varie sur $\mathcal{O}$ et écart quadratique moyen de X , la borne inférieure σ de $\sqrt{\mathfrak{M}(X, a)^2}$ quand a varie sur $\mathcal{O}$. D'après l'inégalité de Schwarz, on a

$$\mu^2 \leq [\mathfrak{M}(X, a)]^2 \leq \mathfrak{M}(X, a)^2$$

et par suite

$$\mu^2 \leq \sigma^2, \quad \mu \leq \sigma.$$

Nous venons de rappeler des définitions déjà données dans notre Mémoire cité plus haut. Les extensions qui suivent au cas d'un ordre k quelconque et surtout au cas d'une dispersion infinie sont inédites.

Remarque. — Observons que l'on a

$$(3) \quad (X, a) \leq (X, b) + (b, a),$$

si donc il existe un élément b de $\mathcal{O}$ tel que $\mathfrak{M}(X, b)$ soit fini, alors $\mathfrak{M}(X, a)$ est aussi fini quel que soit a (sans être nécessairement borné quand a varie).

Dès lors, $\mathfrak{M}(X, a)$ est fini quel que soit a , ou infini quel que soit a .

L'inégalité

$$(4) \quad (X, a)^2 \leq 2(X, b)^2 + 2(a, b)^2,$$

déduite de (3), montre de même que $\mathfrak{M}(X, a)^2$ est fini quel que soit a ou infini quel que soit a . D'ailleurs, l'inégalité de Schwarz montre que si $\mathfrak{M}(X, a)^2$ est fini, il en est de même de $\mathfrak{M}(X, a)$; la réciproque n'ayant pas lieu.

Ordre. — Plus généralement, l'inégalité,

$$(5) \quad (X, a)^k \leq 2^{k-1}[(X, b)^k + (b, a)^k] \quad \text{pour } k \geq 1,$$

facile à établir à partir de l'inégalité

$$(1 + \lambda)^k \leq 2^{k-1}(1 + \lambda^k) \quad (\lambda \geq 0),$$

permet de conclure que si $\mathfrak{M}(X, a)^k$ est (pour $k \geq 1$) fini pour au moins une position b de a , dans $\mathcal{O}$, cette quantité est finie pour tout élément certain a de $\mathcal{O}$.

Nous dirons alors que X est *borné en moyenne d'ordre k* .

Dans le cas où X est un nombre aléatoire, on appelle écart moyen de X avec a la quantité $\mathfrak{M}|X - a|$ et écart quadratique moyen de X avec a la quantité $\sqrt{\mathfrak{M}[X - a]^2}$. On peut maintenant appeler plus généralement *écart moyen d'ordre k* d'un élément aléatoire X choisi au hasard dans un espace distancié $\mathcal{O}$ avec un élément certain a de $\mathcal{O}$, la quantité $\sqrt[k]{\mathfrak{M}(X, a)^k}$.

Nous pourrions appeler dispersion d'ordre k ou *écart moyen d'ordre k* d'un élément aléatoire X borné en moyenne d'ordre k , la borne inférieure (nécessairement finie et ≥ 0) de l'écart moyen d'ordre k de X avec un élément certain a , quand a varie sur $\mathcal{O}$.

Une inégalité utile. — On a, dans $\mathcal{O}$

$$(6) \quad (X, Y) \leq (X, Z) + (Z, Y),$$

d'où

$$(7) \quad \mathfrak{M}(X, Y) \leq \mathfrak{M}(X, Z) + \mathfrak{M}(Z, Y).$$

Étendons cette propriété des écarts moyens aux écarts quadratiques moyens. On a, d'après (6),

$$(X, Y)^2 \leq (X, Z)^2 + (Y, Z)^2 + 2(X, Z)(Y, Z),$$

d'où

$$\begin{aligned} \mathfrak{M}(X, Y)^2 &\leq \mathfrak{M}(X, Z)^2 + \mathfrak{M}(Y, Z)^2 + 2\mathfrak{M}(X, Z)\mathfrak{M}(Y, Z) \\ &\leq \mathfrak{M}(X, Z)^2 + \mathfrak{M}(Y, Z)^2 + 2\sqrt{\mathfrak{M}(X, Z)^2\mathfrak{M}(Y, Z)^2}, \end{aligned}$$

en vertu de l'inégalité de Schwarz. Par conséquent

$$(8) \quad \sqrt{\mathfrak{M}(X, Y)^2} \leq \sqrt{\mathfrak{M}(X, Z)^2} + \sqrt{\mathfrak{M}(Y, Z)^2}.$$

Ainsi, on a établi la relation

$$(9) \quad \sqrt[k]{\mathfrak{M}(X, Y)^k} \leq \sqrt[k]{\mathfrak{M}(X, Z)^k} + \sqrt[k]{\mathfrak{M}(Z, Y)^k},$$

pour $k = 1, 2$. Mais *cette inégalité subsiste* pour les autres valeurs de $k \geq 1$ comme nous allons le voir.

On peut d'abord démontrer de façon élémentaire, lorsque k est un entier pair, $k = 2r$, une inégalité un peu plus faible. Soit

$$U = (X, Y), \quad T = (Z, X), \quad W = (Z, Y).$$

On a

$$U \leq T + W,$$

d'où

$$\begin{aligned} \mathfrak{M} U^{2r} &\leq \mathfrak{M} (T + W)^{2r} = \sum_s C_{2r}^s \mathfrak{M} T^s W^{2r-s} \\ &\leq \sum_s C_{2r}^s \sqrt{\mathfrak{M} T^{2s} \mathfrak{M} W^{4r-2s}}. \end{aligned}$$

Or on sait que

$$\sqrt[2s]{\mathfrak{M} T^{2s}} \leq \sqrt[4r]{\mathfrak{M} T^{4r}} = \alpha, \quad \sqrt[4r-2s]{\mathfrak{M} W^{4r-2s}} \leq \sqrt[4r]{\mathfrak{M} W^{4r}} = \beta,$$

d'où

$$\mathfrak{M} U^{2r} \leq \sum_s C_{2r}^s \sqrt{\alpha^{2s} \beta^{4r-2s}} = \sum_s C_{2r}^s \alpha^s \beta^{2r-s} = (\alpha + \beta)^{2r}.$$

D'où enfin

$$\begin{aligned} \sqrt[2r]{\mathfrak{M} U^{2r}} &\leq \sqrt[4r]{\mathfrak{M} T^{4r}} + \sqrt[4r]{\mathfrak{M} W^{4r}} \\ &\leq \sqrt[2r]{\mathfrak{M} T^{4r}} + \sqrt[2r]{\mathfrak{M} W^{4r}}. \end{aligned}$$

Mais cette inégalité est moins forte que (9) puisque

$$\mathfrak{M} T^{2r} \leq \sqrt{\mathfrak{M} T^{4r}}, \quad \mathfrak{M} W^{2r} \leq \sqrt{\mathfrak{M} W^{4r}}.$$

Pour démontrer l'inégalité (9), il suffit évidemment de démontrer que pour deux nombres aléatoires quelconques V, W définis à la fois dans chaque épreuve, on a

$$(10). \quad \sqrt[k]{\mathfrak{M} |V+W|^k} \leq \sqrt[k]{\mathfrak{M} |V|^k} + \sqrt[k]{\mathfrak{M} |W|^k} \quad \text{pour } k \geq 1.$$

On peut déduire cette inégalité d'une inégalité dite de Hölder concernant les nombres certains. Au contraire, nous retrouverons celle-ci comme un cas particulier de (10) que nous allons établir directement en suivant exactement un raisonnement de F. Riesz, donné pour établir deux inégalités du Calcul intégral (lui-même déclarait d'ailleurs son raisonnement extensible à des cas plus généraux).

Il montre d'abord qu'on a

$$(11) \quad x^\alpha \leq \alpha x + 1 - \alpha \quad \text{pour } x \geq 0, 0 < \alpha < 1.$$

en notant que la dérivée $\alpha(x^{2-1} - 1)$ de $x^2 - \alpha x - 1 + \alpha$ par rapport à x est > 0 pour $0 < x < 1$, < 0 pour $x > 1$. En remplaçant dans (11), x par $\frac{A}{B}$, ($A \geq 0$, $B > 0$), on a

$$(12) \quad A^\alpha B^{1-\alpha} \leq \alpha A + (1-\alpha)B,$$

qui reste vrai pour $B = 0$.

Soient alors deux nombres aléatoires Y et Z toujours ≥ 0 et tels que

$$\mu = \sqrt[k]{\mathfrak{M} Y^k} \quad \text{et} \quad \nu = \sqrt[k-1]{\mathfrak{M} Z^{\frac{k}{k-1}}}$$

soient finis. Remplaçons dans (12), A par $\left(\frac{Y}{\mu}\right)^k$, B par $\left(\frac{Z}{\nu}\right)^{\frac{k}{k-1}}$, α par $\frac{1}{k}$; nous aurons

$$\frac{YZ}{\mu\nu} \leq \frac{1}{k} \frac{Y^k}{\mu^k} + \left(1 - \frac{1}{k}\right) \frac{Z^{\frac{k}{k-1}}}{\nu^{\frac{k}{k-1}}}.$$

La valeur moyenne du second membre est, par hypothèse, finie et même égale à

$$\frac{1}{k} \frac{\mathfrak{M} Y^k}{\mu^k} + \left(1 - \frac{1}{k}\right) \frac{\mathfrak{M} Z^{\frac{k}{k-1}}}{\nu^{\frac{k}{k-1}}} = \frac{1}{k} + 1 - \frac{1}{k} = 1.$$

Donc la valeur moyenne du premier membre est aussi finie : $\mathfrak{M} YZ$ existe et l'on a $\frac{\mathfrak{M} YZ}{\mu\nu} \leq 1$, c'est-à-dire une première inégalité intéressante qui généralise l'inégalité de Schwarz (obtenue pour $k = 2$)

$$(13) \quad \mathfrak{M} YZ \leq \sqrt[k]{\mathfrak{M} Y^k} \cdot \sqrt[k-1]{\mathfrak{M} Z^{\frac{k}{k-1}}},$$

d'où

$$(14) \quad |\mathfrak{M} VW| \leq \sqrt[k]{\mathfrak{M} |V|^k} \sqrt[k-1]{\mathfrak{M} |W|^{\frac{k}{k-1}}} \quad \text{pour } k > 1.$$

On en déduit immédiatement l'inégalité (9) en appliquant (14) aux deux termes du second membre de l'inégalité évidente

$$\mathfrak{M} |V + W|^k \leq \mathfrak{M} |V| |V + W|^{k-1} + \mathfrak{M} |W| |V + W|^{k-1},$$

ce qui donne

$$\mathfrak{M} |V + W|^k \leq \sqrt[k]{\mathfrak{M} |V|^k} \sqrt[k-1]{\mathfrak{M} |V + W|^k} + \sqrt[k]{\mathfrak{M} |W|^k} \sqrt[k-1]{\mathfrak{M} |V + W|^k},$$

d'où, en divisant par le facteur commun du second membre (1), l'inégalité (9) qui se trouve établie pour $k > 1$. Elle est d'ailleurs évidente pour $k = 1$.

Remarques. — I. L'inégalité (14) peut s'écrire sous forme plus symétrique

$$\mathfrak{M}|VW| \leq \sqrt[k]{\mathfrak{M}|V|^k} \sqrt[h]{\mathfrak{M}|W|^h},$$

en posant $\frac{1}{k} + \frac{1}{h} = 1$, avec $k > 1$.

Quand V, W sont des nombres aléatoires prenant respectivement avec des probabilités égales les valeurs $a_1, \dots, a_n; b_1, \dots, b_n$, cette inégalité se réduit à celle de Hölder

$$\left| \sum_{r=1}^n a_r b_r \right| \leq \sqrt[k]{\sum_{r=1}^n |a_r|^k} \sqrt[h]{\sum_{r=1}^n |b_r|^h}.$$

II. Dans ce qui suit, nous avons démontré un certain nombre de propriétés relatives à l'ordre k en supposant implicitement $k \geq 1$ et faisant usage de (9). Beaucoup de ces propriétés pourraient être étendues au cas où $0 < k < 1$ en utilisant une formule à laquelle nous n'avons pensé qu'après envoi de notre manuscrit, à savoir

$$(9^{bis}) \quad \mathfrak{M}(X, Y)^k \leq \mathfrak{M}(X, Z)^k + \mathfrak{M}(Z, Y)^k \quad \text{pour } 0 < k \leq 1.$$

Cette dernière relation se déduit immédiatement de l'inégalité facile à prouver

$$|a + b|^k \leq |a|^k + |b|^k \quad \text{pour } 0 < k \leq 1,$$

quels que soient les nombres a et b .

(1) Si ce facteur commun était nul, l'inégalité (9) s'ensuivrait sans démonstration.

CHAPITRE I.

LES POSITIONS TYPIQUES.

Première définition d'une position typique d'ordre k . — En nous plaçant au point de vue exposé page 227, on pourra appeler *position typique d'ordre k* d'un élément aléatoire X borné en moyenne d'ordre k , tout élément certain, c , de $\mathcal{O}$, s'il en existe, tel que l'écart moyen d'ordre $k: \sqrt[k]{\mathfrak{M}(X, a)^k}$, de X avec un élément certain a , atteigne pour $a \equiv c$, son minimum ⁽¹⁾ quand a varie sur $\mathcal{O}$ ⁽²⁾.

Une position équiprobable et une position moyenne ne seront autres alors que des positions typiques d'ordres respectifs 1 et 2.

Il s'agit maintenant de définir des positions typiques d'ordre k de X , de sorte que la définition ait un sens même quand X n'est pas borné en moyenne d'ordre k . Il faut, de plus, qu'elle se réduise à la précédente quand X est borné en moyenne d'ordre k et que, dans le cas où X est un nombre, elle concorde avec les définitions classiques quand celles-ci sont applicables.

Position typique généralisée. — Pour arriver à cette définition, un moyen classique se présente : remplacer approximativement X par un élément aléatoire X_n borné en moyenne d'ordre k et assujettir la position typique d'ordre k de X à jouer approximativement le rôle d'une position typique d'ordre k de X_n . On peut appliquer cette idée de diverses manières. Nous adopterons la suivante. Pour l'énoncer nous ferons intervenir un élément arbitraire b de $\mathcal{O}$ dont l'introduction paraît

⁽¹⁾ Il est clair que l'on pourrait remplacer dans cet énoncé l'écart moyen par $\mathfrak{M}(X, a)^k$. La substitution se fera dans une des généralisations ultérieures (par exemple, p. 240).

⁽²⁾ Depuis la rédaction de ce Mémoire, Shafik Doss a présenté une troisième définition de la position moyenne en parlant d'une idée différente [*Sur la moyenne d'un élément aléatoire abstrait* (*C. R. Acad. Sci.*, t. 226, 1948, p. 1418-1419)].

On peut définir d'autres positions typiques en partant d'une idée analogue (*Sur une nouvelle définition des positions typiques*, par Maurice Fréchet, mêmes *C. R. Acad. Sci.*, 226, 1948, p. 1419-1420. Cette Note est développée dans deux mémoires en cours d'impression.

compliquer les choses, mais est due à ce qu'il n'y a, pour un espace distancié quelconque, aucun élément privilégié ne jouant le rôle du zéro.

Soient a, b deux éléments certains de $\mathcal{O}$ l'un, a , qu'on fera varier, l'autre, b , fixe. Désignons par X_n^b un élément identique à X quand $(b, X) \leq n$ et à b dans le cas contraire. On a

$$(20) \quad (X_n^b, a) = \begin{cases} (X, a) & \text{dans le premier cas,} \\ (a, b) & \text{dans le second cas.} \end{cases}$$

D'ailleurs dans le premier cas

$$(X, a) \leq (X, b) + (a, b) \leq n + (a, b)$$

Donc, dans toute épreuve, on a

$$(X_n^b, a) \leq n + (a, b),$$

d'où

$$(21) \quad \sqrt[k]{\mathfrak{M}(X_n^b, a)^k} \leq n + (a, b).$$

Ainsi l'écart moyen d'ordre k de X_n^b avec a

$$\varphi_{n,b}(a) = \sqrt[k]{\mathfrak{M}(X_n^b, a)^k},$$

est une fonction de a finie quels que soient a et b , il a donc une borne inférieure finie $m_{n,b,k} \geq 0$ quand a varie sur $\mathcal{O}$.

Deuxième définition. — Nous appellerons position typique d'ordre k de X tout élément certain δ de $\mathcal{O}$, indépendant de b et de n , tel que

$$\lim_{n \rightarrow \infty} [\mathfrak{M}(X_n^b, \delta)^k - (m_{n,b,k})^k] = 0.$$

On peut se demander quelles sont les relations de cette deuxième définition avec la première. Nous allons démontrer que la première est un cas particulier de la seconde. La comparaison ne peut se faire que quand la première définition a un sens, c'est-à-dire quand X est borné en moyenne d'ordre k . Notons d'abord que dans ce cas $\varphi_{n,b}(a)$ converge uniformément vers $\varphi(a) = \sqrt[k]{\mathfrak{M}(X, a)^k}$ lorsque a varie sur $\mathcal{O}$. Car, on a

$$(22) \quad |\varphi(a) - \varphi_{n,b}(a)| = |\sqrt[k]{\mathfrak{M}(X, a)^k} - \sqrt[k]{\mathfrak{M}(X_n^b, a)^k}| \leq \sqrt[k]{\mathfrak{M}(X, X_n^b)^k},$$

en vertu de (9).

Or si $g(x)$ est la fonction de répartition de la variable aléatoire $(X, b)^k$, puisque

$$(X, X_n^b) = 0 \quad \text{si} \quad (X, b)^k < n^k$$

et $(X, X_n^b)^k = (X, b)^k$ dans le cas contraire, on a

$$\mathfrak{M}(X, X_n^b)^k = \int_{n^k}^{+\infty} x \, dg(x),$$

et le second membre tend vers zéro avec $\frac{1}{n}$, puisque dans le cas actuel nous supposons fini $\int_0^{+\infty} x \, dg(x) = \mathfrak{M}(X, b)^k$.

Comme d'ailleurs le second membre de (22) est indépendant de α , on voit bien que la convergence de $\varphi_{n,b}(\alpha)$ vers $\varphi(\alpha)$ est uniforme. Il en résulte que $m_{n,b,k}$ tend vers la borne inférieure m_k de $\varphi(\alpha)$. C'est une conséquence connue de la convergence uniforme. Rappelons sa démonstration : on vient de prouver que

$$|\varphi(\alpha) - \varphi_{n,b}(\alpha)| \leq \varepsilon_{nb},$$

où $\varepsilon_{n,b}$ indépendant de α , tend vers zéro avec $\frac{1}{n}$. On a donc

$$m_k \leq \varphi(\alpha) \leq \varphi_{n,b}(\alpha) + \varepsilon_{nb},$$

d'où

$$m_k \leq m_{n,b,k} + \varepsilon_{nb};$$

de même

$$m_{n,b,k} \leq m_k + \varepsilon_{nb},$$

d'où

$$m_k = \lim_{n \rightarrow \infty} m_{n,b,k}.$$

D'après ce qui précède, on a, quel que soit l'élément certain γ de $\mathcal{D}$,

$$\lim_{n \rightarrow \infty} |[\varphi_{n,b}(\gamma)]^k - (m_{n,b,k})^k| = |[\varphi(\gamma)]^k - (m_k)^k|.$$

Donc si γ est une position typique d'ordre k de X , 1° au sens de la première définition, le second membre est nul, donc aussi le premier et γ est aussi position typique d'ordre k au sens de la deuxième; 2° au sens de la deuxième, le premier est nul, donc aussi le second, et γ est aussi position typique d'ordre k au sens de la première définition.

En résumé, dans le cas où X est borné en moyenne d'ordre k , les deux ensembles de ses positions typiques d'ordre k aux sens respectifs

de la première et de la deuxième définition, sont identiques. En particulier ils sont vides en même temps. La première définition est bien exactement équivalente, lorsqu'elle est applicable, à la deuxième dont elle est un cas particulier.

Autre forme de la seconde définition. — La meilleure manière d'étendre la première définition au cas où X n'est pas borné en moyenne d'ordre k serait sans doute de définir un nouvel écart moyen de deux éléments aléatoires X, Y tel qu'il reste fini non seulement quand l'écart moyen d'ordre k est fini, mais encore dans certains cas plus généraux et d'appliquer la première définition à cet écart. Si l'on prend pour Y un élément certain a de $\mathcal{O}$, le nouvel écart de X et de a serait une fonction $h(a)$ de a et une position typique de X serait un élément certain γ tel que $h(\gamma)$ soit le minimum de $h(a)$. Pour que $h(a)$ ait bien le caractère d'un écart il faudrait qu'il se réduise à (c, a) dans le cas particulier où X se réduit à un élément certain, c , de $\mathcal{O}$. Pour généraliser la première définition, il faudrait choisir cet écart de sorte que dans le cas où X est borné en moyenne d'ordre k , les deux définitions coïncident. Il serait intéressant de chercher un tel écart ou simplement de chercher s'il est possible qu'il existe un tel écart.

Comme première tentative dans cette voie, on peut définir une telle fonction, quand on néglige la condition posée plus haut pour le cas de $X = c$, en donnant une autre forme à la seconde définition.

Considérons la différence

$$h_{nb}(a) = \mathfrak{N}(X_n^b, a)^k - (m_{nkb})^k,$$

Cette fonction est évidemment ≥ 0 . Nous appellerons $h_b(a)$ la plus petite des limites de $h_{nb}(a)$ quand $n \rightarrow \infty$. Pour tout a de $\mathcal{O}$, $h_b(a)$ sera finie ou infinie mais ≥ 0 .

S'il existe un élément certain $\gamma^{(k)}$ de $\mathcal{O}$ indépendant de b et tel que $h_b(\gamma^{(k)}) = 0$, nous l'appellerons position typique d'ordre k de X . Il est clair que cet élément réalise le minimum d'une fonction $h_b(a)$, qui, sans être exactement de la nature d'un écart entre X et a , reflète assez bien (sinon dans sa valeur absolue, du moins dans ses variations avec a) l'idée intuitive qu'on peut se faire d'un tel écart. D'autre part, il est bien évident que cette définition est entièrement équivalente à la seconde définition (donnée p. 234).

Remarque. — On aurait pu penser plus naturel de faire intervenir dans les deux formes de la seconde définition l'écart moyen d'ordre k , $\sqrt[k]{\mathfrak{M}(X_n^b, a)^k}$ et non $\mathfrak{M}(X_n^b, a)^k$. Nous verrons, page 240, qu'une telle substitution ne peut être admise.

Définition auxiliaire. — Une interprétation de notre exposé intuitif de la page 233 qui aurait pu paraître plus naturelle que celle qui a conduit à la seconde définition, est celle qui conduit à ce que nous appellerons une définition auxiliaire d'une position typique d'ordre k de X : c'est un élément certain $\gamma^{(k)}$ de $\mathcal{O}$ indépendant de b et qui est la limite ⁽¹⁾ quand $n \rightarrow \infty$ d'une position typique d'ordre k , $\gamma_n^{(k)}$, de X_n^b .

Cette définition est même celle que nous avons adoptée de préférence dans un mémoire ⁽²⁾ consacré à la traduction des résultats actuels dans le cas où X est un nombre. Car, dans ce cas, X_n^b qui est borné dans l'ensemble des épreuves, a toujours au moins une valeur typique d'ordre k . Mais quand X se déplace dans un espace $\mathcal{O}$ dont on sait seulement qu'il est distancié, l'existence des $\gamma_n^{(k)}$ n'est pas assurée, comme nous l'avons vu, page 227. De sorte que, non seulement, on ne peut dire que la définition auxiliaire soit équivalente à la seconde définition, mais on ne peut même pas dire que ce soit strictement une généralisation de la première définition.

Cependant, si X est borné en moyenne d'ordre k et s'il existe une position typique d'ordre k , $\gamma^{(k)}$ de X au sens de la définition auxiliaire, il en existera aussi au moins une au sens de la première définition, à savoir précisément $\gamma^{(k)}$.

En effet on a, d'après (9),

$$|\sqrt[k]{\mathfrak{M}(X_n^b, \gamma^{(k)})^k} - m_{nk}| = |\sqrt[k]{\mathfrak{M}(X_n^b, \gamma_n^{(k)})^k} - \sqrt[k]{\mathfrak{M}(X_n^b, \gamma_n^{(k)})^k}| \leq (\gamma^{(k)}, \gamma_n^{(k)}) \rightarrow 0.$$

La limite du premier membre est donc nulle; or, d'après la page 235, c'est aussi

$$\sqrt[k]{\mathfrak{M}(X, \gamma^{(k)})^k} - m_k.$$

Donc $\gamma^{(k)}$ réalise bien le minimum $(m_k)^k$ de $\mathfrak{M}(X, a)^k$ quand a varie sur $\mathcal{O}$.

⁽¹⁾ Bien entendu, on dit que $\gamma_n^{(k)} \rightarrow \gamma^{(k)}$ si la distance $(\gamma_n^{(k)}, \gamma^{(k)}) \rightarrow 0$ avec $\frac{1}{n}$.

⁽²⁾ *Nouvelles définitions de la valeur moyenne et des valeurs équiprobables d'un nombre aléatoire* (Ann. Scient. Université de Lyon, 3^e série, 1946, section A, p. 5-26).

La réciproque n'est pas vraie, d'abord parce que nous ne savons pas si les $\gamma_n^{(k)}$ existent, ni ensuite, quand ils existent, s'ils convergent. Même si X a au moins une position typique au sens de la définition auxiliaire et est borné en moyenne d'ordre k , on peut seulement dire que l'ensemble des positions typiques d'ordre k de X au sens auxiliaire est *compris* dans l'ensemble de celles au sens de la première définition.

LES POSITIONS TYPIQUES D'ÉLÉMENTS ALÉATOIRES DE DIVERSES NATURES PARTICULIÈRES.

I. *Cas où X est un nombre.* — Pour que les considérations précédentes aient une base solide, il faut qu'elles se raccordent de façon satisfaisante dans le cas où X est un nombre aléatoire [et bien entendu en prenant dans ce cas $(X, Y) = |X - Y|$] avec la théorie classique des nombres aléatoires.

Valeurs équiprobables. — Nous avons montré ailleurs ⁽¹⁾ que si X est un nombre aléatoire, sa valeur équiprobable ou son segment équiprobable au sens des deux formes de la seconde définition des valeurs typiques d'ordre 1 le sont aussi au sens classique et inversement, autrement dit que les deux définitions sont alors équivalentes. Elles ne sont, au contraire, équivalentes à la définition auxiliaire que lorsqu'il n'y a qu'une valeur équiprobable. Dans le cas où, de plus, l'écart moyen de X est fini (X borné en moyenne), on sait déjà depuis longtemps que la définition d'une valeur équiprobable au sens classique est équivalente à la première définition de la page 233. Donc, dans ce dernier cas, les quatre définitions sont équivalentes.

Valeur moyenne. — Dans le même cas où X est un nombre aléatoire, nous avons montré ⁽¹⁾ que s'il existe une position moyenne $\mathcal{M}X$, au sens classique, c'est-à-dire si, en posant $F(x) = \text{Prob. } [X < x]$, l'intégrale $\int_{-\infty}^{+\infty} x dF(x)$ est absolument convergente, alors il existe une position moyenne au sens de notre seconde définition, comme au sens

⁽¹⁾ Voir la note ⁽²⁾, p. 237.

de la définition auxiliaire, il n'en existe qu'une seule et elle est égale à $\mathfrak{M} X$.

Quand on ne sait pas s'il existe une position moyenne $\bar{X}$ au sens classique, il faut pour qu'il en existe au sens de la seconde définition ou de la définition auxiliaire qui lui est, ici, équivalente, que la valeur principale, au sens de Cauchy de $\int_{-\infty}^{+\infty} x dF(x)$ existe (c'est-à-dire que $\lim_{\lambda \rightarrow \infty} \int_{-\lambda}^{+\lambda} x dF(x)$ existe et soit unique) et alors $\bar{X}$ sera égal à cette valeur principale, de sorte qu'il n'existe qu'une seule telle valeur $\bar{X}$. Mais il ne suffit pas que la valeur principale existe pour que la position moyenne au sens de la seconde définition existe.

En effet, on a

$$\begin{aligned} \psi_{n,b}(a) = \mathfrak{M} |X_n^b - a|^2 &= \int_{b-n}^{b+n} (x - a)^2 dF(x) + \int_{-\infty}^{b-n} (b - a)^2 dF(x) \\ &+ \int_{b+n}^{+\infty} (b - a)^2 dF(x) = (a - R_n)^2 + m_{n,b}^2, \end{aligned}$$

avec

$$R_n = \int_{b-n}^{b+n} x dF(x) + \int_{-\infty}^{b-n} b dF(x) + \int_{b+n}^{+\infty} b dF(x),$$

et, pour qu'il existe $\bar{X}$ tel que

$$0 = \lim_{n \rightarrow \infty} [\psi_{n,b}(\bar{X}) - m_{n,b}^2] = \lim_{n \rightarrow \infty} [\bar{X} - R_n]^2,$$

il faut que R_n tende vers une limite indépendante de b quand $n \rightarrow \infty$ — et alors cette limite sera $\bar{X}$.

De sorte que pour que la définition de la moyenne à notre nouveau sens coïncide avec la définition de la moyenne par l'intégrale $\int_{-\infty}^{+\infty} x dF(x)$, il faut et il suffit que cette intégrale soit définie, comme plus haut, comme la limite (indépendante de b) de R_n , ou plus simplement (puisque les deux derniers termes de R_n tendent vers zéro) de l'intégrale

$$J_n = \int_{b-n}^{b+n} x dF(x).$$

Dans le mémoire cité plus haut, nous avons donné deux exemples montrant que cette nouvelle définition de la moyenne est effectivement

plus stricte que la définition de l'intégrale $\int_{-\infty}^{+\infty} x dF(x)$ par sa valeur principale et moins stricte que la définition classique qui exige sa convergence absolue. On peut dire que la valeur moyenne de X à notre nouveau sens est la valeur *superprincipale* de $\int_{-\infty}^{+\infty} x dF(x)$ en entendant ainsi la limite (supposée unique et indépendante de b) de $\int_{b-n}^{b+n} x dF(x)$.

On pourrait avoir l'idée de modifier la seconde définition en exigeant que soit infiniment petit avec $\frac{1}{n}$ non pas

$$\mathfrak{N}(X_n^b - a)^2 - (m_{nb})^2 \quad \text{mais} \quad \sqrt{\mathfrak{N}(X_n^b - a)^2} - m_{nb}.$$

Dans le même mémoire nous avons montré que si X n'est pas borné en moyenne d'ordre 2 et a pourtant une moyenne au sens classique ou même au sens de notre première définition, la modification en question de la seconde définition conduirait à considérer *tout* nombre certain a comme une valeur moyenne à ce nouveau sens. Cette modification ne peut donc être retenue.

Remarquons enfin que la fonction

$$h_{n,b}(a) = \mathfrak{N}(X_n^b, a)^k - (m_{nkb})^k,$$

considérée page 236, a ici pour $k=2$, la valeur $(a - R_n)^2$. De sorte que lorsque X a une valeur moyenne $\bar{X}$ au sens de notre seconde définition, R_n tendant vers $\bar{X}$, $h_{n,b}(a)$ a bien une limite

$$h(a) = (a - \bar{X})^2.$$

Ainsi, même quand X n'est pas borné en moyenne quadratique (c'est-à-dire quand $\int_{-\infty}^{+\infty} x^2 dF(x) = +\infty$), on peut dire que $\bar{X}$ réalise le minimum d'une certaine fonction de a , $h(a)$, jouant presque le rôle du carré d'un écart entre X et a ou, tout au moins, le rôle de la « partie finie » de l'intégrale $\int_{-\infty}^{+\infty} (x - a)^2 dF(x)$ dans le langage de M. Hadamard.

II. CAS DE L'ESPACE CARTÉSIEN. — Un point aléatoire X de l'espace cartésien à r dimensions est défini par ses r coordonnées $X_1, \dots, X_r$; avec

des notations évidentes, la distance y est définie par

$$(X, Y) = \sqrt{[X_1 - Y_1]^2 + \dots + [X_r - Y_r]^2}.$$

1. Position moyenne. — Pour définir la position moyenne de X , on considère d'abord la quantité

$$\mathfrak{M}(X, a)^2 = \mathfrak{M}(X_1 - a_1)^2 + \dots + \mathfrak{M}(X_r - a_r)^2,$$

où $a_1, \dots, a_r$ sont les coordonnées d'un point certain a , dans le cas où X est quadratiquement borné en moyenne, c'est-à-dire où $\mathfrak{M}(X, b)^2$ est fini pour tout point certain b . En prenant pour b l'origine, on voit que c'est le cas où $\mathfrak{M}X_1^2, \dots, \mathfrak{M}X_r^2$ sont finis. Dans ce cas, $\mathfrak{M}X_1, \dots, \mathfrak{M}X_r$ existent au sens classique et l'on a

$$\mathfrak{M}(X, a)^2 = (a_1 - \mathfrak{M}X_1)^2 + \dots + (a_r - \mathfrak{M}X_r)^2 + m,$$

où m est indépendant de a .

Alors $\mathfrak{M}(X, a)^2$ atteint son minimum pour $a \equiv \bar{X}$ où $\bar{X}$ a les coordonnées

$$a_1 = \mathfrak{M}X_1, \dots, a_r = \mathfrak{M}X_r,$$

et pour ce point seulement : le point moyen est le point de coordonnées moyennes.

Passons au cas où X n'est pas quadratiquement borné en moyenne. Avec les notations déjà employées, $X^{n,b}$ (où b est un point certain de coordonnées $b_1 \dots b_r$) est un point aléatoire pour lequel $\mathfrak{M}(X^{n,b}, a)^2$ est fini. Par suite cette quantité a quand a varie un minimum fini $m_{n,b}^2$. Il s'agit d'étudier un cas où notre définition générale s'applique, c'est-à-dire, il s'agit de chercher s'il existe un point certain γ , $(\gamma_1, \dots, \gamma_r)$, indépendant de b et de n et tel que

$$\mathfrak{M}(X^{n,b}, \gamma)^2 - m_{n,b}^2 = (\gamma_1 - \mathfrak{M}X_1^{n,b})^2 + \dots + (\gamma_r - \mathfrak{M}X_r^{n,b})^2,$$

tende vers zéro avec $\frac{1}{n}$. Il faut donc que

$$\mathfrak{M}X_1^{n,b}, \dots, \mathfrak{M}X_r^{n,b},$$

tendent vers des limites respectives, $\gamma_1, \dots, \gamma_r$ indépendantes de b .

Si donc, il y a à notre sens une position moyenne $\bar{X} = \gamma$, de coordonnées $\gamma_1, \dots, \gamma_r$, *il n'y en a qu'une* et elle est donnée par

$$\gamma_1 = \lim_{n \rightarrow \infty} \mathfrak{M}X_1^{n,b}, \dots, \gamma_r = \lim_{n \rightarrow \infty} \mathfrak{M}X_r^{n,b},$$

Et réciproquement si ces dernières limites existent et sont indépendantes de b , il y a une position moyenne de X à notre sens et il n'y en a qu'une dont les coordonnées sont égales à ces limites.

Notons qu'ici b est un point et non un nombre, de sorte qu'on ne peut encore dire que ces limites sont $\mathfrak{M}X_1, \dots, \mathfrak{M}X_r$.

La définition de $X_k^{n,b}$, qui fait intervenir aussi les valeurs de $X_2, \dots, X_n$ par la considération de (b, X) , est complexe. Pour simplifier, considérons le nombre aléatoire $Y_k^{n,b}$ égal à X_k quand $|X_k - b_k| \leq n$ et à b_k dans le cas contraire ⁽¹⁾. Alors $|X_k^{n,b} - Y_k^{n,b}|$ ne sera différent de zéro que si l'on a à la fois

$$|X_k - b_k| \leq n, \quad (b, X) > n.$$

ou à la fois

$$n < |X_k - b_k| \quad \text{et} \quad (b, X) \leq n.$$

Le dernier cas ne peut se présenter puisque $|X_k - b_k| \leq (b, X)$. On voit alors qu'on a

$$|X_k^{n,b} - Y_k^{n,b}| \leq Z_k,$$

où Z_k n'est différent de zéro que si $(b, X) > n$, cas où l'on peut le prendre égal à (b, X) .

Or supposons que $X_1, \dots, X_r$ aient chacun une moyenne $\overline{X}_1, \dots, \overline{X}_r$ au sens classique. Alors $|X_1|, \dots, |X_r|$ ou aussi $|X_1 - b_1|, \dots, |X_r - b_r|$ auront aussi une moyenne finie et par suite aussi, aura une moyenne finie la quantité

$$(b, X) \leq |X_1 - b_1| + \dots + |X_r - b_r|.$$

Soit

$$g(t) = \text{Prob}[(b, X) < t].$$

On aura

$$(30) \quad |\mathfrak{M}X_k^{n,b} - \mathfrak{M}Y_k^{n,b}| \leq \mathfrak{M}|X_k^{n,b} - Y_k^{n,b}| \leq \mathfrak{M}Z_k = \int_n^{+\infty} t dg(t).$$

Or on vient de voir que dans le cas actuel

$$\int_0^{+\infty} t dg(t) = \mathfrak{M}(X, b),$$

est fini. Donc le dernier membre de (30) tend vers zéro avec $\frac{1}{n}$. Or dans

⁽¹⁾ C'est $Y_k^{n,b}$ et non $X_k^{b,n}$, que la définition de la page 234 ferait intervenir pour la détermination de $\overline{X}_k$.

le cas actuel si X_k a une valeur moyenne à notre sens,

$$\mathfrak{M} Y_k^{n,b} \rightarrow \bar{X}_k.$$

Donc $\mathfrak{M} X_k^{n,b}$ tend aussi vers $\bar{X}_k$.

En résumé quand les coordonnées $X_1, \dots, X_r$ du point aléatoire X ont respectivement des moyennes $\bar{X}_1, \dots, \bar{X}_r$ au sens classique, le point aléatoire X a une position moyenne et une seule, $\bar{X}$, à notre sens et les coordonnées de $\bar{X}$ sont les valeurs moyennes, aussi à notre sens, des coordonnées de X .

Remarque. — Il est évidemment plus rapide de poser *a priori* et par définition comme on le fait d'habitude que la position moyenne d'un point aléatoire est le point dont les coordonnées sont les moyennes des coordonnées du point aléatoire.

On trouvera peut-être, néanmoins, qu'il était intéressant de justifier cette définition analytique en la déduisant d'un principe général applicable non seulement aux points aléatoires mais aux éléments de nature quelconque. Et ceci d'autant mieux que ces deux façons de procéder appliquées à d'autres éléments de nature importante peuvent donner des résultats discordants, comme on le montrera plus loin (p. 246).

2. Positions équiprobables. — Nous n'en dirons que quelques mots en les considérant comme cas particuliers ($k = 1$) des positions moyennes d'ordre k .

Nous avons déjà montré par l'exemple de la valeur équiprobable d'un nombre aléatoire qu'il peut y avoir dans un espace cartésien, plus d'une position moyenne d'ordre k . Démontrons qu'il y a toujours au moins une position moyenne d'ordre k pour un point aléatoire X d'un espace cartésien, au moins quand ce point est borné en moyenne d'ordre k , c'est-à-dire quand $\varphi(a) = \sqrt[k]{\mathfrak{M}(X, a)^k}$ est fini pour chaque point certain a .

En effet, notons d'abord que $\varphi(a)$ est continu. Cela résulte de ce que d'après (9)

$$|\varphi(a) - \varphi(b)| \leq (a, b).$$

Mais on a de même

$$\begin{aligned} |(a, b) - \sqrt[k]{\mathfrak{M}(X, a)^k}| &\leq \sqrt[k]{\mathfrak{M}(X, b)^k}, \\ \varphi(a) &\geq (a, b) - \varphi(b). \end{aligned}$$

Si donc, on prend a assez loin pour que

$$(a, b) \geq 2\varphi(b),$$

on aura $\varphi(a) \geq \varphi(b)$, c'est-à-dire que si S est la sphère de centre b et de rayon $2\varphi(b)$, la borne inférieure de $\varphi(a)$ dans tout l'espace sera la même que dans S seulement. Or une fonction qui comme $\varphi(a)$ est continue dans la sphère S y atteint certainement son minimum dans S (qui se trouve être aussi son minimum dans tout l'espace) en au moins un point $\gamma^{(k)}$. Il y a bien au moins un point certain, à savoir $\gamma^{(k)}$ qui est position moyenne d'ordre k de X .

III. CAS OU L'ÉLÉMENT ALÉATOIRE EST UNE FONCTION. Deux notations. — Lorsque les éléments aléatoires sont pris dans une famille $\mathcal{F}$ de fonctions numériques $x(t)$ d'une variable numérique t ($\alpha \leq t \leq \beta$), on a à distinguer deux sortes de notations. On peut désigner par $\overline{X}(t)$ la valeur moyenne du nombre aléatoire $X(t)$ pour t donné et par $[\overline{X(t)}]$ la position moyenne de l'élément aléatoire $[X(t)]$ défini par la correspondance entre l'ensemble des valeurs de t et l'ensemble des valeurs de $X(t)$ (Pour abrégé, nous supprimerons souvent les crochets).

Dans la littérature actuelle, on considère comme évident que la définition de la position moyenne de $X(t)$ doit être la suivante : c'est la fonction égale pour chaque valeur de t à la valeur moyenne $\overline{X}(t)$ du nombre $X(t)$. Mais il n'y a aucune raison pour qu'il en soit nécessairement ainsi. Nous allons en donner un exemple.

1. Cas des fonctions continues. — Si la famille $\mathcal{F}$ est constituée par les fonctions continues sur un segment fixe (α, β) , la définition usuelle et d'ailleurs la plus naturelle, de la distance, est, avec des notations évidentes,

$$(31) \quad (x, y) = \text{Maximum de } |x(t) - y(t)|,$$

quand t varie sur (α, β) .

Exemple particulier où les deux notations ne sont pas équivalentes. — Considérons le cas où $X(t)$, pris au hasard dans cette famille $\mathcal{F}$, ne peut prendre que les déterminations $x_1(t)$, $x_2(t)$ avec les probabilités respectives p et q . Alors, si $a(t)$ est une fonction certaine

continue sur (α, β) , on a

$$\varphi(\alpha) = \mathfrak{M}(X, \alpha)^2 = p(x_1, \alpha)^2 + q(x_2, \alpha)^2.$$

Posons

$$x_2(t) - x_1(t) = x(t);$$

on aura en posant $b(t) = \alpha(t) - x_1(t)$

$$\begin{aligned}\varphi(\alpha) &= p[\text{Max}|b(t)|]^2 + q[\text{Max}|x(t) - b(t)|]^2 \\ &= p \text{Max} b^2(t) + q \text{Max}[x(t) - b(t)]^2.\end{aligned}$$

Cherchons si $\varphi(\alpha)$ atteint sa borne inférieure. On a en appelant m et M le minimum et le maximum de $|x(t)|$ et μ le maximum de $|b(t)|$

$$\text{Max}|x(t) - b(t)| \geq \text{Max}[|x(t)| - |b(t)|] \geq M - \mu \quad \text{pour } \mu \leq M.$$

Lorsque $\mu \leq m (\leq M)$, on a donc

$$(32) \quad \varphi(\alpha) \geq p\mu^2 + q(M - \mu)^2 = \mu^2 - 2\mu qM + qM^2 = (\mu - qM)^2 + pqM^2.$$

Donc $\varphi(\alpha) \geq pqM^2$.

Quand $\mu \geq m$, on peut écrire

$$\varphi(\alpha) \geq p\mu^2 \geq pm^2,$$

et pour que le second membre soit encore $\geq pqM^2$, il suffit que

$$pm^2 \geq pqM^2,$$

ou

$$(33) \quad q \leq \frac{m^2}{M^2},$$

cas où nous nous placerons maintenant.

Donc, lorsque cette condition est satisfaite, on a toujours

$$(34) \quad \varphi(\alpha) \geq pqM^2.$$

Si l'on suppose $x(t)$ toujours ≥ 0 , on a même dans (34) l'égalité pour un choix convenable de α , c'est-à-dire de b . Il suffit de prendre $b(t) = qM$. Car alors

$$x(t) - b(t) = x(t) - qM \geq m - qM \geq m - \frac{m^2}{M} = m\left(1 - \frac{m}{M}\right) \geq 0,$$

donc

$$|x - b| \equiv x(t) - b(t),$$

dont le maximum est $M - qM = pM$, d'où, puisqu'alors $a = x_1 + qM$,

$$\varphi(x_1 + qM) = pq^2M^2 + qp^2M^2 = pqM^2.$$

Ainsi une position moyenne de X existe, qui est $\overline{X}(t) = x_1(t) + qM^{(1)}$.

D'autre part pour chaque valeur de t

$$\overline{X}(t) = px_1(t) + qx_2(t) = x_1(t) + qx(t).$$

On voit alors qu'on a

$$x_1(t) \leq \overline{X}(t) \leq \overline{X}(t) \leq x_2(t),$$

car en retranchant partout $x_1(t)$, cela revient à

$$0 \leq qx(t) \leq qM \leq m \frac{m}{M} \leq m \leq x(t).$$

De plus, si $x_1(t)$ n'est pas identique à $x_2(t)$, on a

$$\overline{X}(t) - \overline{X}(t) = q[M - x(t)] \geq 0,$$

de sorte que $\overline{X}(t)$ n'est identique à $\overline{X}(t)$ que dans le cas où $x(t)$ est une constante.

La dispersion sera, au nouveau sens, la racine carrée de

$$\sigma^2 = \mathfrak{N}(X, x_1 + qM)^2 = \varphi(x_1 + qM) = pqM^2,$$

soit $\sigma = M\sqrt{pq}$. Tandis que la dispersion de $X(t)$ pour t donnée sera $\sigma(t)$ où

$$\begin{aligned} \sigma^2(t) &= \mathfrak{N}[X(t) - \overline{X}(t)]^2 = p[x_1 - (px_1 + qx_2)]^2 + q[x_2 - (px_1 + qx_2)]^2 \\ &= pq^2(x_1 - x_2)^2 + qp^2(x_2 - x_1)^2 = pqx^2(t). \end{aligned}$$

De sorte que $\sigma(t) < \sigma$ sauf pour la ou les valeurs de t pour lesquelles $x(t) = M$. Ici encore on n'aura $\sigma(t) \equiv \sigma$ que lorsque $x(t)$ est lui-même constant.

Par contre, avec cette même définition de la distance, la position équiprobable $\overline{X}(t)$ de $X(t)$ est identique, dans l'exemple actuel, à la fonction $\overline{X}(t)$ qui pour chaque valeur de t est la valeur équiprobable de $X(t)$.

Car pour $q < \frac{1}{2}$ par exemple, cette dernière fonction est $x_1(t)$. Or pour trouver $\overline{X}(t)$, formons

$$\begin{aligned} \psi(a) &= \mathfrak{N}(X, a) = p \text{Max}|x_1 - a| + q \text{Max}|x_2 - a| \\ &= p \text{Max}|b(t) - a| + q \text{Max}|x(t) - b(t)|. \end{aligned}$$

(1) On pourrait montrer que c'est la seule, mais, pour notre but, ce n'est pas nécessaire.

On a

$$x(t) - b(t) \geq m - \mu.$$

Si donc

$$\mu \leq m, \quad \text{Max}[x(t) - b(t)] = \text{Max}[x(t) - b(t)] \geq M - \mu.$$

d'où

$$(35) \quad \psi(a) \geq p\mu + q(M - \mu) = (p - q)\mu + qM \geq qM.$$

Si $\mu > m$, $\psi(a) \geq p\mu > pm > qM$ pourvu que $q < \frac{m}{m+M}$. Si donc q est assez petit, $\psi(a)$ sera $\geq qM$ quel que soit a . Or pour $a(t) \equiv x_1(t)$, on a

$$\psi(a) = q \text{Max}(x_2 - x_1) = qM.$$

Donc $\psi(a)$ atteint son minimum pour $a \equiv x_1$. Il ne l'atteint pas pour une autre fonction a . Car si $\mu > m$, $\psi(a) > qM$ et si pour $\mu \leq m$, on avait $\psi(a) = qM$, on aurait, d'après (35),

$$qM \geq (p - q)\mu + qM \geq qM.$$

Donc

$$0 = \mu = \text{Max}[a(t) - x_1(t)] \quad \text{c'est-à-dire} \quad a(t) = x_1(t).$$

2. Cas des fonctions de carré intégrable. — Pour la commodité du calcul (et non pas parce que la définition de la distance de deux fonctions que nous allons étudier maintenant est meilleure en soi), nous prendrons comme on le fait souvent pour distance de deux fonctions $X(t)$, $Y(t)$ définies sur un segment (α, β) , l'expression

$$(36) \quad (X, Y) = \sqrt{\frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} [X(t) - Y(t)]^2 dt}.$$

Si l'on se bornait aux fonctions continues sur (α, β) , on aurait avec cette définition de la distance un espace qui ne serait pas métriquement complet ⁽¹⁾. Nous supposerons donc seulement que les fonctions $X(t)$, $Y(t)$ sont de carrés intégrables sur α, β , étant entendu qu'on considère comme équivalentes deux fonctions égales *presque partout* ⁽²⁾, sur (α, β) .

⁽¹⁾ Voir la définition, p. 265.

⁽²⁾ C'est-à-dire égales sauf peut-être sur un ensemble de valeurs de t de mesure nulle (Ensemble qu'on peut enfermer dans une suite dénombrable d'intervalles dont la longueur totale est aussi petite que l'on veut).

Ceci étant, cherchons la position moyenne de $X(t)$. Si $X(t)$ est quadratiquement borné, cette position moyenne $\gamma(t)$ est une fonction, de carré intégrable, telle que $\mathfrak{M}(X, a)^2$ soit minimum quand la fonction certaine $a(t)$ est égale à $\gamma(t)$.

Or

$$(37) \quad \mathfrak{M}(X, a)^2 = \frac{1}{\beta - \alpha} \mathfrak{M} \int_{\alpha}^{\beta} [X(t) - a(t)]^2 dt,$$

d'où ⁽¹⁾

$$(38) \quad \mathfrak{M}(X, a)^2 = \frac{1}{\beta - \alpha} \left\{ \int_{\alpha}^{\beta} a^2(t) dt - 2 \mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt + \mathfrak{M} \int_{\alpha}^{\beta} X^2(t) dt \right\},$$

dont les trois termes sont finis ⁽¹⁾. Pour ne pas interrompre l'argument, admettons un instant : 1° que pour chaque valeur de t sur (α, β) , le nombre aléatoire $X(t)$ ait une valeur moyenne finie $\bar{X}(t)$, sauf peut-être pour un ensemble de valeurs de t de mesure nulle; 2° que la suite des nombres $\bar{X}(t)$ quand t varie, définisse une fonction de t de carré intégrable sur (α, β) ; 3° que l'on ait

$$\mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt = \int_{\alpha}^{\beta} \mathfrak{M}[X(t) a(t)] dt.$$

Alors

$$\mathfrak{M}(X, a)^2 = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} [a(t) - \bar{X}(t)]^2 dt + m,$$

où m est indépendant de $a(t)$. Le minimum de $\mathfrak{M}(X, a)^2$ a lieu pour $a(t) = \bar{X}(t)$. Ainsi $\bar{X}(t)$ est une position moyenne de la fonction $X(t)$ et (puisque l'on est convenu de considérer deux fonctions égales presque partout comme équivalentes), c'est la seule position moyenne. Si l'on représente la position moyenne de la fonction $X(t)$ par $[\bar{X}(t)]$ pour bien marquer que chaque épreuve consiste à choisir au hasard la fonction

⁽¹⁾ On suppose $\mathfrak{M}(X, a)^2$ fini, donc la moyenne de l'intégrale dans (37) existe. En particulier en prenant $a \equiv 0$, $\mathfrak{M} \int_{\alpha}^{\beta} X^2(t) dt$ existe. Donc, dans (38) tous les termes existent et sont finis, sauf peut être $\mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt$. Ainsi, par différence, on voit que ce terme (où $\int_{\alpha}^{\beta} X(t) a(t) dt$ est fini) existe aussi et est fini

$X(t)$ toute entière et non successivement ses valeurs, $[\overline{X(t)}]$ et $\overline{X}(t)$ ont des origines et des significations différentes. Ici, et essentiellement parce que nous avons pris pour définition de la distance, l'expression (36), ces deux notions se trouvent coïncider et l'on a

$$(39) \quad [\overline{X(t)}] = \overline{X}(t),$$

c'est-à-dire que ces deux fonctions sont égales presque partout.

Revenons aux points laissés en suspens. Le point 1° n'est qu'une conséquence du point 2°.

En ce qui concerne l'hypothèse 2°, comme, d'après l'inégalité de Schwarz, on a $[\mathfrak{M}X]^2 \leq \mathfrak{M}X^2$, il suffit de prouver que $\mathfrak{M}X^2(t)$ est intégrable. Or, comme nous l'avons vu en note de la page précédente,

$$\mathfrak{M}(X, 0)^2 = \frac{1}{\beta - \alpha} \mathfrak{M} \int_{\alpha}^{\beta} X^2(t) dt,$$

est fini. Si l'on peut prouver qu'on peut permuter $\mathfrak{M}$ avec le signe d'intégration, non seulement le point 2° sera prouvé mais aussi le point 3°. Nous avons, dans les deux cas, à prouver la propriété (P)

$$(P) \quad \left\{ \begin{array}{l} \text{Si } \int_{\alpha}^{\beta} Y(t) dt \text{ [où } Y(t) \text{ est une fonction aléatoire]} \text{ est un nombre aléa-} \\ \text{toire } U \text{ fini qui possède une moyenne déterminée } \mathfrak{M}U, \text{ alors } Y(t) \text{ a une} \\ \text{intégrale finie} \end{array} \right. \quad (40) \quad \int_{\alpha}^{\beta} Y(t) dt = \mathfrak{M} \int_{\alpha}^{\beta} Y(t) dt.$$

Il y a un cas simple où cette propriété a évidemment lieu. C'est celui où $Y(t)$ ne peut prendre qu'un nombre fini de déterminations $y_1(t), \dots, y_r(t)$ chacune intégrable sur (α, β) et avec les probabilités respectives $p_1, \dots, p_r$.

Alors $W = \int_{\alpha}^{\beta} Y(t) dt$ est bien un nombre aléatoire qui prend les valeurs finies $\int_{\alpha}^{\beta} y_k(t) dt$, avec les probabilités p_k et qui a, par suite, une moyenne déterminée

$$\mathfrak{M}W = \sum_k p_k \int_{\alpha}^{\beta} y_k dt.$$

Or on a évidemment

$$\sum_{k=1}^r p_k \int_{\alpha}^{\beta} y_k(t) dt = \int_{\alpha}^{\beta} \sum_{k=1}^r p_k y_k(t) dt.$$

c'est-à-dire que $\mathcal{M}Y(t)$ est intégrable et qu'on a bien, dans ce cas, l'égalité (39).

Dans les points 2° et 3°, nous avons à supposer vérifiée la propriété P pour $Y = X^2(t)$ et pour $Y = X(t)a(t)$. On a supposé $X(t)$ et $a(t)$ de carrés intégrables sur (α, β) . Si $X(t)$ n'a qu'un nombre fini de déterminations $x_1(t), \dots, x_s(t)$, chacune de carré intégrable sur α, β , d'abord $\mathcal{M}X(t) = \sum p_k x_k(t)$ et X vérifie la propriété 1°. De plus $X^2(t)$ et $X(t)a(t)$, seront bien des fonctions n'ayant qu'un nombre fini de déterminations chacune intégrable sur (α, β) et vérifieront la propriété (P). Ainsi les points 1°, 2° et 3° sont vérifiés au moins quand $X(t)$ appartient à la classe m des fonctions aléatoires qui n'ont qu'un nombre fini de déterminations, chacune de carré intégrable.

Indiquons encore un autre cas où l'on démontre facilement que les conditions 1°, 2°, 3° sont vérifiées.

Il suffit évidemment pour vérifier 1° et 2° de supposer que $X(t)$ est choisi au hasard dans un champ $\mathcal{F}$ de fonctions $x(t)$ continues également bornées (c'est-à-dire telles qu'il existe un nombre certain K indépendant de t tel que $|x(t)| < K$ pour toutes les fonctions de $\mathcal{F}$). Pour vérifier de plus 3°, il nous suffira de supposer en outre que les fonctions $x(t)$ de $\mathcal{F}$ sont également continues (c'est-à-dire telles que pour tout $\omega > 0$, il existe $\varepsilon > 0$, indépendant de $x(t)$, pour lequel

$$|x(t) - x(t')| < \omega \quad \text{quand} \quad |t - t'| < \varepsilon.$$

En effet, divisons (α, β) en intervalles de longueurs $< \varepsilon$ par des points

$$\alpha_1 = \alpha < \dots < \alpha_r < \dots < \alpha_s = \beta.$$

On aura d'abord

$$\left[\int_{\alpha}^{\beta} |X(t)a(t)| dt \right]^2 \leq \int_{\alpha}^{\beta} X^2 dt \int_{\alpha}^{\beta} a^2(t) dt = K^2(\beta - \alpha) \int_{\alpha}^{\beta} a^2(t) dt.$$

Donc $\int_{\alpha}^{\beta} X(t)a(t) dt$ existe et c'est un nombre aléatoire borné; par

conséquent il a une moyenne déterminée. De plus

$$\left| \int_{\alpha}^{\beta} X(t) a(t) dt - \sum_r \int_{\alpha_r}^{\alpha_{r+1}} X(\alpha_r) a(t) dt \right| = \left| \int_{\alpha}^{\beta} [X(t) - Z(t)] a(t) dt \right| \\ \leq \sqrt{\int_{\alpha}^{\beta} [X(t) - Z(t)]^2 dt \int_{\alpha}^{\beta} a^2(t) dt} \leq \omega \sqrt{(\beta - \alpha) \int_{\alpha}^{\beta} a^2(t) dt} = \omega M.$$

où $Z(t) = X(\alpha_r)$ dans (α_r, α_{r+1}) et où M est un nombre indépendant de X . Ainsi

$$\int_{\alpha}^{\beta} X(t) a(t) dt = \sum_r \lambda_r X(\alpha_r) + \Theta \omega M,$$

où les $\lambda_r = \int_{\alpha_r}^{\alpha_{r+1}} a(t) dt$ sont indépendants de X et où Θ est un nombre aléatoire, en valeur absolue ≤ 1 . Les $X(\alpha_r)$ sont en valeur absolue, inférieurs à K . Donc le second membre, étant borné dans l'ensemble des épreuves, a une moyenne déterminée et par suite aussi le premier membre et l'on a

$$\mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt = \sum_r \lambda_r \mathfrak{M} X(\alpha_r) + \theta \omega M,$$

où θ est un nombre certain en module ≤ 1 . On peut donc écrire

$$\left| \mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt - \sum_r \int_{\alpha_r}^{\alpha_{r+1}} [\mathfrak{M} X(\alpha_r)] a(t) dt \right| \leq \omega M,$$

ou

$$(41) \quad \left| \mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt - \int_{\alpha}^{\beta} [\mathfrak{M} Z(t)] a(t) dt \right| < \omega M.$$

D'autre part, puisque $|X(t)| < K$, $X(t)$ a une moyenne déterminée $\bar{X}(t)$ qui est une fonction de t bornée donc intégrable et même de carré intégrable. Par suite, on peut écrire

$$(42) \quad \left| \int_{\alpha}^{\beta} [\mathfrak{M} X(t)] a(t) dt - \sum_r \int_{\alpha_r}^{\alpha_{r+1}} [\mathfrak{M} X(\alpha_r)] a(t) dt \right| \\ = \left| \int_{\alpha}^{\beta} \{\mathfrak{M}[X(t) - Z(t)]\} a(t) dt \right| \leq \omega M.$$

D'où, en combinant les deux inégalités (41), (42)

$$\left| \mathfrak{M} \int_{\alpha}^{\beta} X(t) a(t) dt - \int_{\alpha}^{\beta} [\mathfrak{M} X(t)] a(t) dt \right| \leq 2 \omega M.$$

Ceci ayant lieu pour tout $\omega > 0$, l'égalité constituant la condition 3° se trouve établie.

Dès lors, ayant précédemment établi la relation

$$\overline{X(t)} = \overline{X}(t),$$

pour toute famille $\mathcal{F}$ de fonctions aléatoires $X(t)$ de carrés intégrables sur $\alpha\beta$ vérifiant les conditions 1°, 2°, 3°, nous voyons qu'il suffit de raisonnements élémentaires pour prouver qu'on peut prendre pour $\mathcal{F}$ l'une ou l'autre des deux importantes familles particulières qui viennent d'être considérées. Puisqu'il s'agit seulement ici de donner un exemple d'application de notre théorie générale, nous ne chercherons pas à déterminer la famille $\mathcal{F}$ la plus générale, vérifiant 1°, 2°, 3°, recherche qui ferait intervenir des raisonnements assez délicats. Nous signalerons seulement qu'on y arriverait en utilisant une indication de Kolmogoroff donnée dans un cas analogue et consistant à étendre au cas actuel les théorèmes de Fubini et de Tonelli signalés sous les nos 5 et 6, à la page 262 de l'ouvrage *Théorie de l'intégrale*, par Saks.

LES POSITIONS TYPIQUES DANS LES ESPACES D'ÉLÉMENTS ALÉATOIRES GLOBAUX.

Premiers espaces d'éléments aléatoires globaux. — L'inégalité (9) peut, parmi d'autres motifs, conduire à l'idée de considérer des éléments aléatoires comme points d'un nouvel espace distancié.

Quand chaque épreuve (d'une certaine catégorie C d'épreuves) détermine par son résultat R certains éléments Y, Z, ..., on peut considérer ceux-ci comme des fonctions de R

$$Y = g(R), \quad Z = h(R), \dots$$

Mais on peut à leur tour envisager ces fonctions comme points d'un certain espace fonctionnel et par suite Y, Z, ... comme éléments d'un certain espace Δ correspondant.

Dans le cas particulier où Y, Z, ... sont des éléments aléatoires qui dans toute épreuve sont pris au hasard parmi les éléments d'un espace distancié $\mathcal{O}$, on peut distinguer parmi les éléments de Δ ceux qui sont bornés en moyenne d'ordre k ; ils appartiennent à un sous-espace $\mathcal{B}_k$ de Δ . L'inégalité (9) conduit à considérer à son tour $\mathcal{B}_k$ comme un

espace distancié, où l'on peut adopter pour distance de deux éléments Y, Z (considérés *globalement* dans l'ensemble des épreuves) l'expression

$$(50) \quad ([Y], [Z]) = \sqrt[k]{\mathfrak{N}(Y, Z)^k}.$$

Cette quantité est finie puisque, si l'on prend pour une des fonctions $g(R), h(R), \dots$ une fonction indépendante de R , c'est-à-dire un élément certain α , on aura pour $k \geq 1$,

$$\sqrt[k]{\mathfrak{N}(Y, Z)^k} \leq \sqrt[k]{\mathfrak{N}(Y, \alpha)^k} + \sqrt[k]{\mathfrak{N}(Z, \alpha)^k}$$

où, dans $\mathfrak{B}_k$, le second membre est fini par hypothèse. Notons en passant, la différence des notations (Y, Z) qui représente la distance des déterminations prises dans $\mathcal{O}$ par Y et Z dans *une* épreuve et $([Y], [Z])$ qui représente la distance de deux *fonctions* $Y = g(R), Z = h(R)$, c'est-à-dire de deux éléments aléatoires considérés *globalement* dans l'ensemble des épreuves. (Nous dirons que $[Y], [Z]$ représentent des *éléments aléatoires globaux*.) La première est une distance dans l'espace $\mathcal{O}$, la seconde dans l'espace $\mathfrak{B}_k$.

On voit qu'on a, pour cette distance globale,

$$(51) \quad \begin{cases} ([Y], [Z]) = ([Z], [Y]) \geq 0, \\ ([Y], [Y]) = 0, \\ ([Y], [Z]) \leq ([Y], [T]) + ([T], [Z]). \end{cases}$$

En ce qui concerne la réciproque de (51), si l'on a

$$0 = ([Y], [Z]),$$

on a

$$\mathfrak{N}(Y, Z)^k = 0,$$

ce qui exige seulement que Y soit identique à Z , *presque certainement* ⁽¹⁾. C'est donc seulement en considérant comme équivalentes deux éléments aléatoires de $\mathfrak{B}_k$ qui coïncident dans toute épreuve, sauf peut-être dans un cas de probabilité nulle, que l'on pourra considérer $\sqrt[k]{\mathfrak{N}(X, Y)^k}$ comme une distance et $\mathfrak{B}_k$ comme un espace distancié. Mais c'est là une restriction courante en analyse où, par exemple, l'espace des fonctions de carrés intégrables peut être distancié en considérant comme équivalentes deux fonctions qui ne sont égales qu'en dehors d'un ensemble de mesure nulle [voir note ⁽²⁾, p. 247].

(1) C'est-à-dire, identique, sauf peut-être dans un ensemble d'épreuves dont la probabilité est nulle.

Nouveaux espaces d'éléments aléatoires globaux. — Nous venons de montrer que si l'on considère les éléments aléatoires de $\mathcal{B}_k$, on peut y définir de manière naturelle, une distance, par la formule (50). Mais nous verrons plus loin, page 263, qu'il est parfois utile de définir d'autres distances. Celles-ci seront définies en supposant encore que les éléments aléatoires considérés sont pris au hasard à chaque épreuve dans un espace distancié déterminé $\mathcal{O}$.

Mais ce dernier point n'est pas essentiel. Supposons donc que l'on considère un ensemble d'éléments aléatoires $Y, Z, \dots$ tous à la fois déterminés à chaque épreuve d'une catégorie particulière C d'épreuves et fonctions : $Y = g(R), Z = h(R), \dots$ du résultat R de cette épreuve, fonction étant ici prise au sens général de transformation : $R, Y, Z, \dots$ peuvent n'être pas des nombres. Nous pouvons considérer $Y, Z, \dots$ comme appartenant à deux sortes d'espaces. L'un est un espace $\mathcal{E}$, où sont pris au hasard à chaque épreuve, les déterminations correspondantes de $Y, Z, \dots$. L'autre, Δ , est un espace dont les éléments sont en fait des *fonctions* certaines $g, h, \dots$, mais qu'on peut aussi appeler l'espace des éléments aléatoires globaux $Y, Z, \dots$, pris chacun dans l'ensemble des épreuves.

Par exemple, on choisit au hasard un individu R dans une population donnée; $Y, Z, T, U, V, \dots$ seront son poids, sa taille, son âge, la forme de son nez, l'idée qu'il a en tête au moment du choix, etc. Les trois premiers sont des éléments pris à chaque épreuve dans l'espace des nombres réels. Mais on peut considérer aussi l'espace Δ dont les éléments sont, l'un le poids aléatoire de R , le second sa taille aléatoire, le troisième son âge aléatoire, etc. Ceci étant, il se pourra qu'on puisse définir *dans* Δ une distance de deux éléments et ceci de façon à répondre à notre intuition de la distance. Par exemple, si les fonctions $g, h, \dots$ sont numériques, on pourrait prendre comme distance de deux éléments $Y = g(R), Z = h(R)$, le maximum de $|g(R) - h(R)|$ quand R varie sur $\mathcal{E}$ (sans qu'il soit nécessaire que R soit un nombre).

Quand il s'agira d'éléments aléatoires, il sera en général préférable de donner une définition de la distance, où intervient le hasard et où l'on tient moins compte des ensembles d'épreuves dont la probabilité est plus petite.

C'est par exemple ce qui a lieu dans $\mathcal{B}_k$ où si par exemple (X, Y)

n'a que deux valeurs α , β prises avec les probabilités p , q , on a

$$([X], [Y]) = \sqrt[k]{p\alpha^k + q\beta^k}.$$

On voit que si q est petit le premier membre dépendra surtout de α et peu de β .

C'est dans l'hypothèse générale, dont cet exemple n'est qu'un cas particulier, que nous allons nous placer dans la suite.

Produit scalaire. — On peut établir une inégalité qui nous sera utile par la suite

Soient Y , Z , T trois points d'un espace distancié Δ d'éléments aléatoires globaux. En posant

$$a = ([Y], [Z]), \quad b = ([Y], [T]), \quad c = ([Z], [T]),$$

on pourra considérer a , b , c comme les côtés d'un triangle. Dans ce triangle, on a

$$(52) \quad ab \cos \varphi = \frac{a^2 + b^2 - c^2}{2},$$

en appelant φ l'angle des côtés a , b . L'expression $ab \cos \varphi$ est le produit scalaire des côtés a , b . Nous pouvons donc appeler *produit scalaire généralisé* de YZ et YT et représenter par $YZ.YT$ la quantité

$$YZ.YT = \frac{([Y], [Z])^2 + ([Y], [T])^2 - ([Z], [T])^2}{2},$$

où le second membre a un sens et une valeur indépendants de l'interprétation géométrique précédente.

D'après (52), on a

$$\left| \frac{a^2 + b^2 - c^2}{2ab} \right| \leq 1,$$

[et l'égalité n'a lieu que si

$$(52) \quad c = |a \pm b|,$$

c'est-à-dire si le triangle est aplati, cas où ses trois sommets sont alignés]. On a donc l'inégalité que nous voulions établir

$$(53) \quad |YZ.ZT| \leq ([Y], [Z])([Y], [T]).$$

L'égalité ne peut avoir lieu que si comme dans (52)

$$(52 \text{ bis}) \quad ([Z], [T]) = ([Y], [Z]) \pm ([Y], [T]),$$

cas où nous pourrions dire que les trois éléments aléatoires $[Y]$, $[Z]$, $[T]$ sont alignés dans l'espace Δ des éléments aléatoires globaux.

Coefficient de linéarité. — On pourra donc appeler le rapport

$$(55) \quad \rho = \frac{YZ \cdot YT}{([Y], [Z])([Y], [T])},$$

coefficient d'alignement ou de linéarité de Y, Z, T dans l'espace des éléments aléatoires globaux.

Cas de $\mathcal{B}_2$. — Considérons le cas particulier où à chaque épreuve Y, Z, T appartiennent à $\mathcal{B}_2$. Dans ce cas, on peut préciser la signification géométrique de l'égalité $\rho = \pm 1$, non seulement dans $\mathcal{B}_2$ mais dans l'espace distancié $\mathcal{O}$. Quand $\rho = 0 = \pm 1$, (55) devient

$$(56) \quad \mathfrak{N}(Z, T)^2 = \mathfrak{N}(Y, Z)^2 + \mathfrak{N}(Y, T)^2 - 2\theta \sqrt{\mathfrak{N}(Y, Z)^2 \mathfrak{N}(Y, T)^2},$$

ou

$$(57) \quad \sqrt{\mathfrak{N}(Z, T)^2} = |\sqrt{\mathfrak{N}(Y, Z)^2} - \theta \sqrt{\mathfrak{N}(Y, T)^2}|.$$

Or

$$|(Y, Z) - (Y, T)| \leq (Z, T) \leq (Y, Z) + (Y, T),$$

d'où en élevant au carré et prenant les moyennes

$$(58) \quad -2\mathfrak{N}[(Y, Z)(Z, T)] \leq \mathfrak{N}(Y, Z)^2 + \mathfrak{N}(Y, T)^2 - \mathfrak{N}(Z, T)^2 \leq 2\mathfrak{N}[(Y, Z)(Z, T)]$$

et en combinant (56) et (58)

$$-2\mathfrak{N}[(Y, Z)(Z, T)] \leq 2\theta \sqrt{\mathfrak{N}(Y, Z)^2 \mathfrak{N}(Y, T)^2} \leq 2\mathfrak{N}[(Y, Z)(Z, T)],$$

donc

$$\sqrt{\mathfrak{N}(Y, Z)^2 \mathfrak{N}(Y, T)^2} \leq \mathfrak{N}[(Y, Z)(Y, T)].$$

Or d'après l'égalité de Schwarz

$$\mathfrak{N}[(Y, Z)(Y, T)] \leq \sqrt{\mathfrak{N}(Y, Z)^2 \mathfrak{N}(Y, T)^2}$$

donc

$$\mathfrak{N}[(Y, Z)(Y, T)] = \sqrt{\mathfrak{N}(Y, Z)^2 \mathfrak{N}(Y, T)^2}.$$

Mais on sait que l'inégalité de Schwarz

$$|\mathfrak{M} UV|^2 \leq \mathfrak{M} U^2 \mathfrak{M} V^2,$$

ne peut devenir une égalité que si les nombres aléatoires U, V sont tels qu'il existe deux nombres certains λ, μ non tous deux nuls, tels que l'égalité $\lambda U + \mu V = 0$ ait lieu presque certainement.

Donc ici

$$(59) \quad \lambda(Y, Z) + \mu(Y, T) = 0,$$

presque certainement.

Nous venons d'établir une conséquence de la relation symétrique

$$(60) \quad \sqrt{\mathfrak{M}(Z, T)^2} \pm \sqrt{\mathfrak{M}(Y, Z)^2} \pm \sqrt{\mathfrak{M}(Y, T)^2} = 0,$$

condition non seulement nécessaire mais aussi suffisante pour que $\rho = \pm 1$. On verrait donc de la même façon, en permutant Y, Z, T , qu'il existe aussi des nombres certains λ', μ' non tous nuls tels que

$$(61) \quad \lambda'(Z, Y) + \mu'(Z, T) = 0$$

presque certainement. L'événement **E** consistant en ce que l'une *ou* l'autre de ces relations n'ait pas lieu est donc aussi de probabilité nulle. Leur réalisation simultanée a lieu presque certainement.

L'égalité (60) a lieu évidemment quand une des trois racines carrées est nulle, cas où un couple déterminé des éléments Y, Z, T est formé de deux éléments aléatoires qui, dans chaque épreuve (sauf un cas de probabilité nulle) occupent la même position (aléatoire) dans l'espace distancié $\mathcal{O}$.

En dehors de ce cas, les trois racines de (60) sont $\neq 0$ et par suite, comme en vertu de (59) et (61), on a

$$\lambda^2 \mathfrak{M}(Y, Z)^2 = \mu^2 \mathfrak{M}(Y, T)^2, \quad \lambda'^2 \mathfrak{M}(Z, Y)^2 = \mu'^2 \mathfrak{M}(Z, T)^2$$

où les trois moyennes sont maintenant supposées $\neq 0$, les nombres certains $\lambda, \mu, \lambda', \mu'$ sont aussi tous $\neq 0$. Dès lors les égalités (59) et (61) peuvent s'écrire sous la forme

$$(62) \quad (Y, Z) = \gamma(Z, T), \quad (Y, T) = \delta(Z, T) \quad \text{avec} \quad \gamma \neq 0, \delta \neq 0,$$

d'où

$$(63) \quad \sqrt{\mathfrak{M}(Y, Z)^2} = \gamma \sqrt{\mathfrak{M}(Z, T)^2}, \quad \sqrt{\mathfrak{M}(Y, T)^2} = \delta \sqrt{\mathfrak{M}(Z, T)^2},$$

et en portant dans (57), on aura

$$(64) \quad 1 = |\gamma - \theta\delta|.$$

d'où en tenant compte de (62)

$$(65) \quad (Z, T) = |(Y, Z) - \theta(Y, T)|,$$

ou, presque certainement

$$(66) \quad (Z, T) = |(Y, Z) \mp (Y, T)|,$$

suivant que $\rho = +1$ ou -1 .

Réciproquement, si (62) et (66), ont lieu presque certainement, mais avec, pour (66), un signe $-$ ou $+$ non aléatoire, c'est-à-dire si (62) et (65) ont lieu en prenant pour θ un nombre *certain* égal à $+1$ ou à -1 , on a d'abord en vertu de (63),

$$YZ \cdot YT = \frac{(\gamma^2 + \delta^2 - 1)}{2} \mathcal{N}(Z, T)^2,$$

et

$$\sqrt{\mathcal{N}(Y, Z)^2 \mathcal{N}(Y, T)^2} = \gamma\delta \mathcal{N}(Z, T)^2;$$

puis (64) découle de (63) et (65).

Par suite, en tenant compte de (64),

$$\rho = \frac{\gamma^2 + \delta^2 - 1}{2\gamma\delta} = \frac{\gamma^2 + \delta^2 - (\gamma - \theta\delta)^2}{2\gamma\delta} = 0, \quad \text{donc } |\rho| = 1,$$

Ainsi pour que $|\rho| = 1$, il faut et il suffit que Z soit identique à T presque certainement, cas où $\rho = 1$ ou plus généralement qu'on ait presque certainement les relations (62) et (65) où θ est un nombre certain égal à ± 1 , et s'il en est ainsi, on aura non seulement $|\rho| = 1$, mais $\rho = 0$. On peut exprimer ces deux conditions dans un langage géométrique plus intuitif. Nous pouvons exprimer la relation (66) dans l'espace $\mathcal{O}$ comme plus haut la relation analogue dans l'espace $\mathcal{B}_2$, en disant d'abord que, sauf un cas de probabilité nulle, dans chaque épreuve Y, Z, T seront alignés dans $\mathcal{O}$. Il faut ensuite tenir compte de l'une des égalités (62), car l'autre sera alors conséquence de (64) en posant l'égalité (65). On peut exprimer l'une ou l'autre de ces égalités (62), en tenant compte de l'alignement de Y, Z, T en disant que, sauf un cas de probabilité nulle, les positions Y, Z, T déterminent dans chaque cas un segment rectiligne de longueur aléatoire, mais partagé

dans un rapport certain. De plus, en s'inspirant de (65), on pourra dire que si $\theta = 1$, Y est en dehors de ZT et si $\theta = -1$, que Y est entre Z et T .

Dans le cas où Y , Z , T seraient des points aléatoires dans un espace $\mathcal{O}$ qui serait euclidien et où l'on aurait pris pour distance la distance euclidienne, on n'aurait plus là de simples façons de s'exprimer, mais l'énoncé même des conditions pour que $\rho = \pm 1$.

Nouvelle généralisation des positions typiques d'un élément aléatoire. — Dans un espace distancié Δ d'éléments aléatoires globaux, nous dirons que l'un, Y , de ces éléments aléatoires, a une position typique $\bar{Y}$ (correspondant à la définition adoptée pour la distance dans Δ), si la distance $([Y], [a])$ de Y à un élément certain a , appartenant à Δ atteint son minimum quand a devient $\bar{Y}$. Il n'est certain, ni qu'un tel élément $\bar{Y}$ existe, ni qu'il soit unique. Qu'il existe ou non, on appellera *dispersion* de Y (relativement à la définition adoptée pour la distance dans Δ) la borne inférieure (≥ 0), qui existe toujours, de $([Y], [a])$ quand a varie dans Δ .

CHAPITRE II.

CONVERGENCES STOCHASTIQUES.

Notion générale. — La notion de distance de deux éléments aléatoires de nature quelconque permet d'étendre immédiatement à ces éléments les définitions des diverses convergences dites stochastiques.

Supposons que chaque épreuve d'une certaine catégorie C détermine à la fois une suite d'éléments aléatoires $Y, Y_1, Y_2, \dots, Y_n, \dots$ appartenant à un espace distancié $\mathcal{O}$. Nous appellerons *convergence stochastique de Y_n vers Y* (dans l'ensemble des épreuves de C) tout mode de convergence qui est d'autant moins strict sur un ensemble particulier e d'épreuves que la probabilité qu'une épreuve appartienne à e est plus petite. Sous cette désignation intuitive et encore vague, on range un certain nombre de modes de convergence, de définitions très précises et déjà profondément étudiées dans le cas des nombres aléatoires :

les convergences *légale*, *en probabilité*, *en moyenne d'ordre k*, (en particulier pour $k = 2$, *en moyenne quadratique*) et *presque certaine*. La notion de distance de deux éléments aléatoires *globaux*, introduite pages 253, 254, permet de considérer des convergences stochastiques plus variées encore. Dans un espace distancié, Δ , d'éléments aléatoires globaux, nous pouvons en effet dire que Y_n converge stochastiquement vers Y dans Δ (c'est-à-dire au sens correspondant à la définition adoptée pour la distance dans Δ), si $([Y_n], [Y])$ tend vers zéro avec $\frac{1}{n}$.

Limite des dispersions et des positions typiques. — Supposons que $([X], [X_n]) \rightarrow 0$ avec $\frac{1}{n}$ et soient D et D_n les dispersions de X , X_n . Par hypothèse, il y a deux éléments certains a_n , b_n tels que

$$D \leq ([X], [a_n]) < D + \frac{1}{n}, \quad D_n \leq ([X_n], [b_n]) < D_n + \frac{1}{n}.$$

D'où

$$D \leq ([X], [b_n]) \leq ([X], [X_n]) + ([X_n], [b_n]) = ([X], [X_n]) + D_n + \frac{1}{n},$$

et de même

$$D_n \leq ([X_n], [a_n]) \leq ([X_n], [X]) + ([X], [a_n]) = ([X], [X_n]) + D + \frac{1}{n}.$$

Donc

$$|D - D_n| \leq ([X], [X_n]) + \frac{1}{n},$$

ce qui montre que si X_n converge stochastiquement vers X , la dispersion de X_n tend vers celle de X (pourvu que les dispersions et cette convergence stochastique soient relatives à la même définition de la distance des deux éléments globaux). Par exemple, si X_n converge en moyenne quadratique vers X , l'écart quadratique moyen de X_n tend vers celui de X (que ce dernier soit fini ou non).

Supposons maintenant qu'il existe pour chaque X_n au moins une position typique $\bar{X}_n$. Alors

$$D \leq ([X], [\bar{X}_n]) \leq ([X], [X_n]) + ([X_n], [\bar{X}_n]),$$

ou

$$0 \leq ([X], [\bar{X}_n]) - D \leq ([X], [X_n]) + (D_n - D) \rightarrow 0.$$

Donc

$$(67) \quad D = \lim_{n \rightarrow \infty} ([X], [\bar{X}_n]).$$

Supposons en outre que les positions typiques $\bar{X}_n$ tendent vers un élément certain α ⁽¹⁾ ou même plus généralement qu'on puisse extraire de la suite des $\bar{X}_n$ une suite convergente et soit α sa limite. On aura alors

$$|([X], [\alpha]) - ([\bar{X}_n], [X])| \leq ([\alpha], [\bar{X}_n]).$$

Alors en faisant croître n (au besoin par valeurs convenables), le dernier terme tendra vers zéro ⁽¹⁾ et il en résultera d'après (67) que $([X], [\alpha]) = D$, c'est-à-dire que X a au moins une position typique et que la limite α des positions typiques $\bar{X}_n$ des X_n est une position typique $\bar{X} = \alpha$ de X .

Convergence en moyenne d'ordre k . — Quand des éléments aléatoires Y_n, Y sont choisis au hasard dans un espace distancié $\mathcal{O}$, on dit que $Y_n \rightarrow Y$ en moyenne d'ordre k quand $\mathfrak{M}(Y_n, Y)^k \rightarrow 0$ avec $\frac{1}{n}$. Il revient au même de dire que $\sqrt[k]{\mathfrak{M}(Y_n, Y)} \rightarrow 0$. On a vu que (pour $k \geq 1$) cette racine peut être considérée comme une distance

$$([Y_n], [Y]) = \sqrt[k]{\mathfrak{M}(Y_n, Y)^k},$$

dans un espace $\mathcal{B}_k$ d'éléments aléatoires bornés en moyenne d'ordre k .

Application. — Supposons qu'on prenne

$$([Z], [Y]) = \sqrt[k]{\mathfrak{M}(Z, Y)^k}.$$

Si Y et les Y_n sont chacun borné en moyenne d'ordre k et si Y_n a une position typique γ_n , d'ordre k alors d'après ce qui précède, quand Y_n converge vers Y en m. o. k . (en abrégé pour : en moyenne d'ordre k) et s'il existe une suite σ extraite des γ_n qui soit convergente : 1° Y aura aussi au moins une position typique d'ordre k ; 2° la limite de σ sera une position typique d'ordre k de Y .

L'intérêt de cette application concerne naturellement surtout les cas où $k = 1$ ou 2 .

Cas de l'espace cartésien. — Soient deux points Z, Y de l'espace cartésien à r dimensions; $Z_1, \dots, Z_r; Y_1, \dots, Y_r$ leurs coordonnées.

⁽¹⁾ A dire vrai, nous supposons ici, ce qui n'a pas encore été fait, qu'il a été donné une définition de la convergence des points certains dans Δ et que celle-ci est celle qu'on obtient en considérant les points certains comme des éléments globaux particuliers.

En posant $U_j = (Y_j - Z_j)^2$, on a

$$(Z, Y)^k = (\Sigma U_j)^{\frac{k}{2}} \geq U_j^{\frac{k}{2}},$$

d'où

$$\begin{aligned} \sqrt[k]{\mathfrak{M}(Y_j, Z_j)^k} &= \sqrt[k]{\mathfrak{M} U_j^{\frac{k}{2}}} \leq \sqrt[k]{\mathfrak{M}(Z, Y)^k} \\ &= \left[\sqrt[k]{\mathfrak{M}(\Sigma U_j)^{\frac{k}{2}}} \right]^{\frac{1}{2}} \leq \left[\Sigma \sqrt[k]{\mathfrak{M} U_j^{\frac{k}{2}}} \right]^{\frac{1}{2}} \leq \sqrt{\Sigma (\sqrt[k]{\mathfrak{M}(Y_j, Z_j)^k})^2}. \end{aligned}$$

Donc la condition *nécessaire et suffisante* pour que Z converge en m. o. k vers Y est que *chaque coordonnée* de Z converge en m. o. k vers la coordonnée correspondante de Y . Toutefois, l'avant-dernière inégalité, étant une conséquence de (9), n'est établie que pour $\frac{k}{2} \geq 1$. La conclusion ci-dessus n'est donc démontrée que pour $k \geq 2$.

Convergence en probabilité. — Quand Y_n et Y sont choisis au hasard dans un espace distancié $\mathcal{O}$, nous dirons que $Y_n \rightarrow Y$ en probabilité lorsque pour tout nombre $\varepsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \{ \text{Prob}[(Y_n, Y) > \varepsilon] \} = 0.$$

Ce mode de convergence a été inventé dans le cas où les Y_n et les Y sont des nombres pour généraliser le théorème de Bernoulli qui, grâce à lui peut s'énoncer d'une façon plus intuitive : la fréquence f_n d'un événement de probabilité constante p , au cours de n épreuves, converge en probabilité vers p .

Remarque. — Cette définition suppose que l'événement $(Y_n, Y) > \varepsilon$ a une probabilité déterminée. Dans la suite, il sera entendu que tous les événements dont les probabilités auront été considérées dans les raisonnements sont supposés avoir chacun une probabilité déterminée.

Nous avons déjà fait observer qu'en pratique tous les événements qu'on rencontre sont *probabilisables*, de sorte qu'il y aura là une précaution de rigueur sans importance pratique.

Cas de l'espace cartésien. — Cherchons la condition pour la convergence en probabilité du point aléatoire $X^{(n)}$ de coordonnées $X_1^{(n)}, \dots, X_r^{(n)}$ vers le point aléatoire X de coordonnées $X_1, \dots, X_r$. On a

$$|X_k - X_k^{(n)}| \leq (X, X^{(n)}).$$

d'où

$$\text{Prob}[|X_k - X_k^{(n)}| < \varepsilon] \geq \text{Prob}[(X, X^{(n)}) < \varepsilon].$$

Donc si $X^{(n)} \rightarrow X$ en probabilité, $X_k^{(n)}$ converge en probabilité vers X_k pour $k = 1, 2, \dots, r$.

D'autre part, si l'on a simultanément

$$|X_1 - X_1^{(n)}| < \varepsilon, \quad \dots, \quad |X_r - X_r^{(n)}| < \varepsilon,$$

on a

$$(X, X^{(n)}) < \varepsilon \sqrt{r}.$$

Donc, en utilisant la formule de Boole (rappelée page 266), on a

$$\begin{aligned} \text{Pr}[(X, X^{(n)}) < \varepsilon \sqrt{r}] &\geq \text{Pr}[|X_1 - X_1^{(n)}| < \varepsilon, \dots, |X_r - X_r^{(n)}| < \varepsilon] \\ &\geq \text{Pr}[|X_1 - X_1^{(n)}| < \varepsilon] + \dots + \text{Pr}[|X_r - X_r^{(n)}| < \varepsilon] - r + 1. \end{aligned}$$

Dès lors, si $X_1^{(n)}, \dots, X_r^{(n)}$ tendent respectivement en probabilité vers $X_1, \dots, X_r$, le dernier membre tend vers 1, donc aussi le premier et par suite $X^{(n)}$ converge en probabilité vers X ; en résumé : *la condition nécessaire et suffisante pour qu'un point aléatoire $X^{(n)}$ de l'espace euclidien converge en probabilité vers un point aléatoire X est que chaque coordonnée de $X^{(n)}$ converge en probabilité vers la coordonnée correspondante de X .*

La convergence en probabilité déduite d'une distance. — On peut montrer qu'on peut attacher à tout couple d'éléments aléatoires X, Y d'un espace distancié $\mathcal{O}$, une distance $([X], [Y])$ choisie de sorte que la condition nécessaire et suffisante pour que $Y_n \rightarrow Y$ en probabilité soit que $([Y_n], [Y]) \rightarrow 0$. On peut même le faire d'une infinité de façons.

On peut d'abord généraliser une définition que nous avons donnée en 1930 pour le cas des nombres aléatoires. On prendra pour $([X], [Y])$ la borne inférieure, quand le nombre certain positif ε varie, de la somme

$$\varepsilon + \text{Prob}[(X, Y) > \varepsilon].$$

J'avais fait observer que si cette définition atteignait le but principal, qui était d'établir la *possibilité* de définir la convergence en probabilité au moyen d'une distance, néanmoins, cette définition faisait intervenir la somme de deux quantités de dimensions différentes et qu'il serait au moins plus esthétique d'éviter cette rencontre. Ky Fan a montré qu'on pouvait le faire en faisant intervenir au lieu d'une somme, un quotient

pour lequel cette disparité de dimension n'a pas les mêmes inconvénients. En étendant sa définition des nombres aléatoires aux éléments aléatoires, $([X], [Y])$ se trouve défini comme la borne inférieure des valeurs de $\varepsilon (> 0)$, telle que

$$\frac{\text{Prob}[(X, Y) > \varepsilon]}{\varepsilon} < 1.$$

Pour chacune de ces deux définitions, on peut démontrer (exactement comme dans le cas des nombres aléatoires) : 1° qu'elle fournit une distance; 2° que la condition nécessaire et suffisante pour que la distance ainsi définie de $[Y_n]$ et de $[Y]$ tende vers zéro est que Y_n converge vers Y en probabilité.

Enfin, on peut aussi étendre aux éléments aléatoires une famille de définitions (correspondant aux différentes fonctions f employées) que j'avais données pour les nombres aléatoires. On appellera f -écart de Z et de Y , la quantité $\lambda \geq 0$ qui vérifie la relation

$$(70) \quad f(\lambda) = \mathfrak{M} f((Z, Y)),$$

où $f(x)$ est une fonction ≥ 0 , continue, croissante, définie pour $x \geq 0$, avec $f(0) = 0$.

Quand $f(x) \equiv x$, λ est l'écart moyen de Z et de Y ; quand $f(x) \equiv x^2$, λ est leur écart quadratique moyen. Mais il peut arriver que ces quantités soient infinies. Il y a, au contraire deux cas importants où λ est nécessairement fini, d'abord le cas où, dans toute épreuve, (Z, Y) est fini, ce qui a lieu en particulier quand il existe au moins un point certain c tel que les écarts (c, Z) et (c, Y) soient bornés dans la catégorie d'épreuves envisagée. Alors λ est fini pour chaque fonction f . Le second cas est celui où $f(x)$ est borné quand x varie.

Dans ce cas, on a, d'après l'inégalité généralisée de Bienaymé,

$$\{ \text{Prob}[f((Y_n, Y)) \geq f(\varepsilon)] \} \leq \frac{f(\lambda_n)}{f(\varepsilon)}$$

ou

$$\{ \text{Prob}[(Y_n, Y) \geq \varepsilon] \} \leq \frac{f(\lambda_n)}{f(\varepsilon)},$$

où λ_n est le f -écart de Y_n et de Y .

On voit alors que si $\lambda_n \rightarrow 0$, $Y_n \rightarrow Y$ en probabilité. Inversement, si $Y_n \rightarrow Y$ en probabilité, comme on a

$$\mathfrak{M} f((Y_n, Y)) \leq f(\varepsilon) + M \{ \text{Prob}[(Y_n, Y) \geq \varepsilon] \},$$

en appelant M la borne supérieure de $f(x)$, on voit que pour n assez grand, on aura

$$f(\lambda_n) = \mathfrak{M} f(Y_n, Y) < f(\varepsilon) + \varepsilon,$$

c'est-à-dire que $f(\lambda_n)$ et par suite λ_n tendent vers zéro avec $\frac{1}{n}$. Ainsi la condition nécessaire et suffisante pour que le f -écart de Y_n et Y tende vers zéro avec $\frac{1}{n}$ est une condition indépendante de la forme de la fonction $f(\lambda)$ (pourvu qu'elle possède les propriétés admises plus haut) et cette condition c'est que $Y_n \rightarrow Y$ en probabilité. De même que nous l'avons montré dans le cas des nombres aléatoires, on peut même énoncer la même propriété pour une f -distance si l'on renonce à cette condition pourtant bien naturelle que, dans le cas de deux éléments aléatoires Y, Z dont la distance (Y, Z) reste constante et égale à α , la f -distance soit égale à α . Il suffit, à cet effet, de prendre pour f une fonction telle que

$$f(\lambda + \mu) \leq f(\lambda) + f(\mu)$$

et de prendre pour f -distance de Y, Z non pas le nombre λ donné par (70), mais la quantité $f(\lambda)$ elle-même.

La notion d'espace complet. — La généralisation d'une importante proposition de Slutsky fournit un exemple de l'attention avec laquelle les présentes généralisations doivent être menées. Toutes celles que nous avons faites jusqu'ici s'appliquaient à un espace distancié quelconque. Maintenant, il va nous être nécessaire d'introduire et d'employer la notion d'espace *complet*. Une autre généralisation, page 279, nécessitera celle d'espace *séparable*, etc. Le risque d'erreur provient de ce qu'un espace euclidien étant à la fois complet, séparable, etc., notre habitude de raisonner sur l'espace euclidien peut nous faire négliger la nécessité de l'une de ces conditions.

Dans un espace distancié $\mathcal{O}$, on voit, grâce à l'inégalité triangulaire, que si une suite d'éléments $x_1, x_2, \dots, x_n, \dots$, est convergente, alors $(x_n, x_m) \rightarrow 0$ quand le plus petit des nombres n et $m \rightarrow \infty$, ce que nous exprimerons en disant que toute suite convergente dans $\mathcal{O}$ vérifie le critère de Cauchy.

Mais la réciproque n'est pas vraie. On dira qu'un espace distancié est (métriquement) *complet* quand elle est vraie.

Ceci étant, en généralisant la démonstration donnée ⁽¹⁾ quand l'espace distancié est celui des nombres (qui est complet), on pourra généraliser une proposition de Slutsky et dire :

Pour qu'une suite d'éléments aléatoires d'un espace distancié métriquement complet $\mathcal{O}$ soit convergente en probabilité, il faut et il suffit que pour tout $\varepsilon > 0$,

$$(71) \quad \text{Prob}[(X_m, X_n) < \varepsilon]$$

tende vers l'unité quand le plus petit des nombres n et m tend vers l'infini.

On peut exprimer cette proposition sous une forme différente et la démontrer ainsi indirectement en introduisant un espace distancié Δ d'éléments aléatoires globaux.

Condition de convergence en probabilité. — Considérons l'espace distancié Δ où chaque élément *global* $[X]$, est un élément aléatoire dont les positions appartiennent à un espace distancié $\mathcal{O}$ complet et où la distance est définie, comme plus haut, par $([X], [Y]) =$ borne inférieure quand ε positif varie de $\{\varepsilon + \text{Prob}[(X, Y) > \varepsilon]\}$.

Cet espace Δ est métriquement complet. Soit, en effet, une suite d'éléments aléatoires *globaux* $[X_n]$ de Δ vérifiant le critère généralisé de Cauchy et soit $\sum_p \varepsilon_p$ une série convergente de nombres > 0 . Il existe un nombre N_p , tel que

$$(72) \quad ([X_n], [X_m]) \leq \varepsilon_p \quad \text{pour } n \geq N_p \text{ et } m \geq N_p.$$

Et, par suite, on voit facilement qu'on a pour ces valeurs de n, m ,

$$\{\text{Prob}[(X_n, X_m) \leq \varepsilon_p]\} \geq 1 - \varepsilon_p.$$

En vertu de l'inégalité de Boole généralisée

$$\text{Prob}[H_1 \text{ et } H_2 \dots \text{ et } H_r, \dots] \geq 1 - [\text{Prob } C(H_1) + \dots + \text{Prob } C(H_r) + \dots]$$

où $C(H_r)$ est l'événement contraire à H_r , on aura ici

$$\text{Prob } e_p \geq 1 - (\varepsilon_p + \varepsilon_{p+1} + \dots)$$

⁽¹⁾ Voir, par exemple, nos *Recherches théoriques modernes sur le Calcul des Probabilités*, Gauthier-Villars, Livre I, p. 167-169 de la Première édition, 1928. (La seconde édition, en préparation, ne paraîtra qu'après le présent mémoire.) Nous désignerons dans la suite ce Livre par R. T., 1^{er}.

en désignant par e_p la réalisation simultanée des événements

$$(X_{N_p}, X_{N_{p+1}}) \leq \varepsilon_p, \quad (X_{N_{p+1}}, X_{N_{p+2}}) \leq \varepsilon_{p+1}, \quad \dots$$

qui entraîne

$$(X_{N_{p+r}}, X_{N_{p+q}}) < \omega_{p+h}$$

quels que soient r et q supérieurs à h en posant

$$\omega_p = \varepsilon_p + \varepsilon_{p+1} + \dots$$

Dès lors, quand e_p a lieu, les X_{N_p} vérifient le critère de Cauchy dans $\mathcal{D}$ et *puisque* $\mathcal{D}$ est complet, forment une suite convergente. La convergence des $Y_r = X_{N_r}$ a lieu au moins quand a lieu l'événement e_p de probabilité $> 1 - \omega_p$; sa probabilité est donc égale à 1. Ainsi, on voit d'abord qu'on peut extraire de la suite des X_n , une suite $Y_1, Y_2, \dots$, qui converge *presque certainement*, c'est-à-dire sauf un cas de probabilité nulle. Soit Y un élément aléatoire identique à la limite des Y_p quand les Y_p convergent et prenant une position arbitrairement choisie dans le cas contraire. On a, quand e_p a lieu,

$$(X_{N_p}, Y) \leq (X_{N_p}, X_{N_{p+q}}) + (X_{N_{p+q}}, Y) < \omega_p + (X_{N_{p+q}}, Y)$$

d'où, quand $q \rightarrow \infty$,

$$(X_{N_p}, Y) \leq \omega_p.$$

Or on a vu que la probabilité de e_p est $\geq 1 - \omega_p$; on a donc

$$(73) \quad ([X_{N_p}], [Y]) \leq \omega_p + \{ \text{Prob}[(X_{N_p}, Y) > \omega_p] \} \leq 2\omega_p$$

et par suite, on a pour $n > N_p$, d'après (72) et (73),

$$([X_n], [Y]) \leq ([X_n], [X_{N_p}]) + ([X_{N_p}], [Y]) \leq 3\omega_p.$$

Donc $([X_n], [Y])$ tend vers zéro avec $\frac{1}{n}$, X_n converge en probabilité vers Y . Ainsi l'espace distancié Δ est complet.

D'autre part, la démonstration précédente fournit un résultat intéressant. Nous avons vu plus haut que si une suite d'éléments d'un espace distancié $\mathcal{D}$ converge en probabilité, elle vérifie dans Δ le critère de Cauchy. Nous venons de voir aussi que si dans cet espace Δ , on prend la suite des X_n , pour cette suite, qu'on suppose convergente en probabilité vers Z , on peut en extraire une suite $\mathcal{Z}$ qui converge presque certainement vers un élément Y et Y est aussi limite en probabilité des X_n .

Dès lors, puisque

$$([Z], [Y]) \leq ([Z], [X_n]) + ([X_n], [Y]),$$

$([Z], [Y])$ est nul, c'est-à-dire que Y est égal à Z presque certainement et par conséquent la suite $\mathfrak{Z}$ converge aussi presque certainement vers Z . En résumé : de toute suite d'éléments aléatoires X_n (dont les positions appartiennent à un espace distancié complet $\mathcal{Q}$) qui converge vers l'élément aléatoire Z en probabilité, on peut extraire une suite qui converge presque certainement vers Z .

Remarque. — La condition de Slutsky (71) est évidemment nécessaire pour la convergence en probabilité. Elle est suffisante, car, si elle a lieu, on peut trouver pour ε_p donné un nombre N_p , tel que la condition (72) soit vérifiée quand $n > N_p$ et $m > N_p$. Or, on vient de démontrer qu'alors X_n converge en probabilité.

Condition de convergence en moyenne d'ordre k . — I. Il résulte de l'inégalité généralisée de Bienaymé

$$\{\text{Prob}[(Y_n, Y) > \varepsilon]\} \leq \frac{\mathfrak{M}(Y_n, Y)^k}{\varepsilon^k}$$

que si une suite d'éléments aléatoires Y_n converge vers Y en moyenne d'ordre k , cette suite converge aussi en probabilité. Mais la réciproque n'est pas vraie en général. On pourrait démontrer qu'elle est vraie quand les Y_n sont *également bornés*, mais nous la démontrerons plus loin, page 272, dans le cas plus général où les Y_n sont *également sommables* d'ordre k .

II. Si Y_n converge en m. o. k . vers Y , il résulte de l'inégalité (9) ou ici

$$\sqrt[k]{\mathfrak{M}(Y_n, Y_p)^k} \leq \sqrt[k]{\mathfrak{M}(Y_n, Y)^k} + \sqrt[k]{\mathfrak{M}(Y_p, Y)^k},$$

qu'à tout nombre $\omega > 0$, correspond N tel que

$$\sqrt[k]{\mathfrak{M}(Y_n, Y_p)^k} < \omega \quad \text{pour } n > N \text{ et } p > N$$

ou encore que

$$(74) \quad \lim_{\frac{1}{n} + \frac{1}{p} \rightarrow 0} \sqrt[k]{\mathfrak{M}(Y_n, Y_p)^k} = 0.$$

C'est une généralisation du critère de Cauchy, condition nécessaire

pour la convergence de Y_n en m. o. k . Montrons que cette condition est aussi suffisante. Quand cette condition est vérifiée, il résulte de l'inégalité de Bienaymé

$$\{\text{Prob}[(Y_n, Y_m) > \varepsilon]\} < \frac{\mathfrak{M}(Y_n, Y_m)^k}{\varepsilon^k}$$

que la condition de Slutsky est vérifiée et par suite que, si l'espace $\mathcal{O}$ est complet, Y_n converge en probabilité vers un élément aléatoire Y . Il suffit de prouver que sa convergence vers Y a lieu aussi en m. o. k . ⁽¹⁾ Cette proposition est déjà connue quand les Y_n sont des nombres. La démonstration donnée pour ce cas par Paul Lévy ⁽²⁾, peut même, avec des modifications convenables, se prêter à une triple généralisation d'un lemme préalable du même auteur sous la forme suivante.

Soient $U_1, U_2, \dots, U_p, \dots$, une suite d'éléments aléatoires convergeant en probabilité vers U et soit V un autre élément aléatoire. On a

$$(75) \quad \mathfrak{M}(U, V)^k \leq \lim_{p \rightarrow \infty} \mathfrak{M}(U_p, V)^k$$

(que les moyennes qui figurent dans cette inégalité soient finies ou non).

En effet, par définition, dans les deux cas, on a

$$\mathfrak{M}(U, V)^k = \int_0^{+\infty} z^k dF(z),$$

où

$$F(z) = \text{Prob}[(U, V) < z].$$

Si ω est un nombre positif arbitraire et si $\mathfrak{M}(U, V)^k$, fini ou non, est supérieur à Λ , on pourra trouver M tel que

$$\int_0^M z^k dF(z) > \Lambda - \omega,$$

c'est-à-dire

$$(76) \quad \mathfrak{M}(U', V)^k > \Lambda - \omega,$$

où

$$U' \begin{cases} \equiv U & \text{pour } (U, V) \leq M, \\ \equiv V & \text{pour } (U, V) > M. \end{cases}$$

⁽¹⁾ Si une suite $V_1, V_2, \dots$ converge en probabilité vers V ou même presque certainement (voir p. 284), vers V , elle ne converge pas nécessairement en m. o. k . On le voit en prenant par exemple pour V un nombre aléatoire fini dans chaque épreuve, mais non borné et même tel que $\mathfrak{M}|V|^k$ soit infini et $V_n = \frac{V}{n}$.

⁽²⁾ Voir sa *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, p. 56-57.

Soient $\varepsilon, \varepsilon'$ deux nombres positifs arbitraires. On peut aussi prendre un nombre q tel que

$$(77) \quad \{ \text{Prob}[(U_p, U) > \varepsilon] \} < \varepsilon' \quad \text{pour } p > q.$$

D'autre part, on a

$$(78) \quad (U_p, V_p) \leq \varepsilon$$

dans toute épreuve, si l'on prend

$$V_p \equiv \begin{cases} U & \text{pour } (U_p, U) \leq \varepsilon, \\ U_p & \text{pour } (U_p, U) > \varepsilon. \end{cases}$$

En comparant (V_p, V) à (U', V) on voit qu'on a $(V_p, V) \geq (U', V)$, sauf peut-être quand se produit le concours C des événements

$$(U, V) \leq M, \quad (U_p, U) > \varepsilon.$$

La probabilité de C est $< \varepsilon'$ d'après (77) et quand C se produit, on a

$$(U', V) = (U, V) \leq M.$$

Posons

$$N = (U', V), \quad N_p = (V_p, V),$$

et

$$P = 0 = P_p$$

sauf quand C se produit, cas où nous poserons

$$N = 0 = N_p, \\ P = (U', V) \quad \text{et} \quad P_p = (V_p, V).$$

On aura

$$(U', V) = P \quad \text{ou} \quad N$$

et

$$(V_p, V) = P_p \quad \text{ou} \quad N_p$$

suivant que C a lieu ou non. De plus,

$$N_p \geq N \quad \text{et} \quad P \leq M.$$

D'où

$$\mathfrak{N}(V_p, V)^k = \mathfrak{N} N_p^k + \mathfrak{N} P_p^k \geq \mathfrak{N} N^k$$

et par suite,

$$A - \omega < \mathfrak{N}(U', V)^k = \mathfrak{N} N^k + \mathfrak{N} P^k \leq \mathfrak{N} N^k + M^k \varepsilon' \leq \mathfrak{N}(V_p, V)^k + M^k \varepsilon'.$$

D'où

$$\mathfrak{N}(V_p, V)^k > A - \omega - M^k \varepsilon'.$$

ε' étant pris arbitrairement, on peut le prendre tel que $M^k \varepsilon' < \omega$, d'où

$$\mathfrak{M}(V_p, V)^k > A - 2\omega.$$

Mais d'après (78) et (9), on a

$$\sqrt[k]{\mathfrak{M}(U_p, V)^k} \geq \sqrt[k]{\mathfrak{M}(V_p, V)^k} - \varepsilon > \sqrt[k]{A - 2\omega} - \varepsilon$$

pour $p > q$, d'où

$$\lim_{p \rightarrow \infty} \mathfrak{M}(U_p, V)^k \geq [\sqrt[k]{A - 2\omega} - \varepsilon]^k$$

et ceci pour des ω et ε positifs aussi petits que l'on veut, donc

$$\lim_{p \rightarrow \infty} \mathfrak{M}(U_p, V)^k \geq A \quad \text{pour tout } A < \mathfrak{M}(U, V)^k.$$

D'où résulte enfin (75). Le lemme de Paul Lévy est ainsi étendu au cas plus général où, d'une part, on ne suppose pas que les U, V, U_p sont des nombres, où d'autre part on ne suppose pas vérifiée par les U_p , la condition de Cauchy qui intervient dans son raisonnement, où enfin k n'est plus supposé égal à 2 mais simplement ≥ 1 .

Appliquons l'inégalité (75) en prenant

$$V = Y_n, \quad U_p = Y_p, \quad U = Y,$$

on aura, pour n fixe,

$$\mathfrak{M}(Y, Y_n)^k \leq \lim_{p \rightarrow \infty} \mathfrak{M}(Y_p, Y_n)^k.$$

Or, si la condition (74) est vérifiée, à tout $\eta > 0$, correspond $\theta > 0$, tel que

$$\frac{1}{n} + \frac{1}{p} < \theta \quad \text{entraîne} \quad \mathfrak{M}(Y_p, Y_n)^k < \eta.$$

Pour $n > \frac{2}{\theta}$, cette condition sera remplie quel que soit $p > \frac{2}{\theta}$, donc, finalement

$$\mathfrak{M}(Y, Y_n)^k \leq \eta \quad \text{pour } n > \frac{2}{\theta}.$$

Ainsi, quand l'espace distancié $\mathcal{O}$ (où à chaque épreuve sont choisis au hasard les Y_n) est métriquement complet, l'espace distancié $\mathcal{B}_k$ où la distance de deux éléments aléatoires globaux $[X], [Y]$ est prise égale à $\sqrt[k]{\mathfrak{M}(X, Y)^k}$, est aussi métriquement complet.

Égale sommabilité. — Nous dirons que des éléments aléatoires X formant une famille $\mathcal{F}$ sont également sommables ⁽¹⁾ en moyenne d'ordre k lorsque, pour au moins un élément certain a , si l'on pose

$$(80) \quad U_n(a, X) = \begin{cases} (a, X)^k & \text{pour } (a, X) \geq n, \\ 0 & \text{au cas contraire,} \end{cases}$$

$\mathfrak{M} U_n(a, X)$ converge *également* vers zéro, c'est-à-dire quand, pour tout $\varepsilon > 0$, il existe un nombre N indépendant du choix de l'élément X dans $\mathcal{F}$, tel que

$$(81) \quad \mathfrak{M} U_n(a, X) < \varepsilon \quad \text{pour } n > N.$$

C'est ce qui aura lieu en particulier quand les X de $\mathcal{F}$ sont *également* bornés, c'est-à-dire s'il existe un élément certain b et un nombre H tel que $(b, X) < H$ dans toute épreuve et pour tout X de $\mathcal{F}$. Car alors

$$(a, X) < (a, b) + H$$

et pour $n > (a, b) + H$, on aura

$$(a, X) < n, \quad \text{d'où} \quad U_n(a, X) = 0$$

et

$$\mathfrak{M} U_n(a, X) = 0 < \varepsilon.$$

D'ailleurs, s'il existe un élément certain a de $\mathcal{Q}$ jouissant de la propriété (81), tout autre élément b de $\mathcal{Q}$ jouira de la même propriété. En effet, on a

$$(b, X) \leq (a, X) + (a, b).$$

Si donc $(b, X) \geq p$, on aura $(a, X) \geq n$ en prenant $p - (a, b) \geq n$.

Or, dans ce cas, on a, d'après (80),

$$\sqrt[k]{U_p(b, X)} \leq \sqrt[k]{U_n(a, X)} + Z,$$

où $Z = (a, b)$ si $(b, X) \geq p$ et $Z = 0$ dans le cas contraire; d'où, en tenant compte de (9),

$$\sqrt[k]{\mathfrak{M} U_p(b, X)} \leq \sqrt[k]{\mathfrak{M} U_n(a, X)} + (a, b) \sqrt[k]{\Pi_p},$$

où Π_p est la probabilité pour que $(b, X) \geq p$. Or on a évidemment

$$\mathfrak{M} U_n(a, X) \leq n^k \Pi_p,$$

⁽¹⁾ Nous généralisons ici une expression introduite par Flament dans l'étude des ensembles d'intégrales de fonctions ordinaires.

d'où

$$\sqrt[k]{\mathfrak{M} U_p(b, X)} \leq \sqrt[k]{\mathfrak{M} U_n(a, X)} \left[1 + \frac{(a, b)}{n} \right] \leq \sqrt[k]{\varepsilon} \left[1 + \frac{(a, b)}{2N} \right]$$

pour $p \equiv (a, b) + 2N$, en prenant $n = 2N$. On voit qu'en prenant ε assez petit, le premier membre sera aussi petit que l'on veut pour p assez grand : b peut jouer le rôle de a .

Application I. — Supposons que $X_n \rightarrow X$ en probabilité et que X et X_n soient également sommables en moyenne d'ordre r . Il s'agit de montrer que $\mathfrak{M}(X, X_n)^r \rightarrow 0$.

Soit

$$\varepsilon > 0 \quad \text{et} \quad V_n = \begin{cases} (X, X_n)^r & \text{pour } (X, X_n) > \varepsilon, \\ 0 & \text{pour } (X, X_n) \leq \varepsilon. \end{cases}$$

On a

$$0 \leq (X, X_n)^r - V_n \leq \varepsilon^r.$$

donc

$$(82) \quad \mathfrak{M}(X, X_n)^r \leq \varepsilon^r + \mathfrak{M} V_n.$$

Puisque X et les X_n sont également sommables, il existe un élément certain, c , tel qu'en posant

$$U_p(c, X) = \begin{cases} (c, X) & \text{pour } (c, X) > p, \\ 0 & \text{pour } (c, X) \leq p, \end{cases}$$

on puisse, pour tout $\eta > 0$, choisir p de sorte que

$$\mathfrak{M} U_p(c, X) < \eta, \quad \mathfrak{M} U_p(c, X_n) < \eta.$$

Or on a

$$\sqrt[k]{\mathfrak{M} V_n} \leq (W_n + T_n),$$

où

$$\begin{array}{lll} \text{quand } (X, X_n) > \varepsilon, & \text{on prend } W_n = (c, X), & T_n = (c, X_n), \\ \text{» } (X, X_n) \leq \varepsilon, & \text{» } W_n = 0, & T_n = 0. \end{array}$$

Donc

$$\sqrt[k]{\mathfrak{M} V_n} \leq \sqrt[k]{\mathfrak{M}(W_n)^r} + \sqrt[k]{\mathfrak{M}(T_n)^r}.$$

Or quand $(X, X_n) \leq \varepsilon$, on a

$$W_n = 0;$$

quand $(X, X_n) > \varepsilon$, on a

$$W_n = (c, X);$$

alors, ou bien

$$(c, X) \leq p \quad \text{et} \quad W_n^r \leq p^r,$$

ou bien

$$(c, X) > p \quad \text{et} \quad W_n^r = (c, X)^r = U_p(c, X).$$

Donc

$$\begin{aligned} \mathfrak{M}(W_n)^r &\leq p^r \{ \text{Prob}[(X, X_n) > \varepsilon] \} + \mathfrak{M}[U_p(c, X)] \\ &\leq p^r \{ \text{Prob}[(X, X_n) > \varepsilon] \} + \eta. \end{aligned}$$

Soit $\omega > 0$. Pour n assez grand, $n > N$, $\{ \text{Prob}[(X, X_n) > \varepsilon] \} < \omega$, d'où

$$\mathfrak{M}(W_n)^r \leq p^r \omega + \eta < 2\eta \quad \text{pour } n > N,$$

où, η étant arbitraire, p est alors choisi et où l'on a pris $\omega < \frac{\eta}{p^r}$.

Donc $\mathfrak{M}(W_n)^r \rightarrow 0$. On montrerait de même que $\mathfrak{M}(T_n)^r \rightarrow 0$, donc aussi $\mathfrak{M}V_n$. En prenant n assez grand, $n > N'$, on aura $\mathfrak{M}V_n < \varepsilon^r$, d'où, par (82), $\mathfrak{M}(X, X_n)^r < 2\varepsilon^r$. Finalement

$$\mathfrak{M}(X, X_n)^r \rightarrow 0 \quad \text{quand } n \rightarrow \infty.$$

Donc quand X et les X_n sont également sommables en moyenne d'ordre r , il y a équivalence entre la convergence en probabilité et la convergence en moyenne d'ordre r .

Application II. — Quand $X_n \rightarrow X$ en probabilité, que deviennent les positions moyennes $\bar{X}_n$ et équiprobables $\bar{\bar{X}}_n$, s'il en existe, de X_n ? Et quelles sont leurs relations avec $\bar{X}$ et $\bar{\bar{X}}$ si celles-ci existent?

On peut donner des exemples dès le cas le plus simple, celui où les X_n et X sont des nombres aléatoires, où $X_n \rightarrow X$ en probabilité, où X_n et X ont des moyennes, $\bar{X}_n$ et $\bar{X}$ et où $\bar{X}_n$ a une limite sans que cette limite soit égale à $\bar{X}$ ⁽¹⁾.

Mais supposons que X et les X_n soient également bornées ou plus généralement qu'elles soient également sommables en moyenne d'ordre r ($r = 1$ ou 2). Dans le premier cas : $r = 1$, on a vu, page 261, que si les positions équiprobables $\bar{\bar{X}}_n$ existent et si l'on peut en extraire une suite σ convergente, alors $\bar{\bar{X}}$ existe aussi et σ tend vers une position équiprobable de X . Dans le cas où $r = 2$, on sait que les X et X_n seront aussi également sommables d'ordre 1, donc non seulement le résultat précédent s'étend aux positions moyennes, mais il subsiste pour les positions équiprobables.

(1) Voir un exemple dû à Slutsky, p. 176, R. T., 1^{er}.

Convergence légale. — Nous dirons que Y_n converge *légalement* vers Y lorsque la loi de probabilité de Y_n converge vers celle de Y ou plus précisément quand la fonction de distribution $p_n(e)$ de Y_n converge vers la fonction de distribution $p(e)$ de Y ⁽¹⁾.

Déjà, dans le cas des nombres aléatoires, nous avons fait observer ⁽²⁾ que l'utilité de cette convergence légale consiste à fournir, quand on connaît $p(e)$, une expression asymptotique de $p_n(e)$, mais que pour une épreuve déterminée, elle ne fournit aucun renseignement sur Y_n connaissant Y .

C'est dans ce sens que l'un des premiers problèmes à résoudre concernant les autres modes de convergence stochastique de Y_n vers Y consiste à chercher si cette convergence entraîne la convergence légale.

Remarque. — Il nous sera utile, pour la suite, de rappeler deux formules ⁽³⁾.

Observons d'abord que si des événements E_n sont tels que ⁽⁴⁾

$$(83) \quad E_1 \subset E_2 \subset \dots \subset E_n \dots$$

(c'est-à-dire que E_1 a pour conséquence E_2 , etc.), on a

$$(84) \quad \text{Prob}[E_1 \text{ ou } E_2 \dots \text{ ou } E_n \dots] = \lim_{n \rightarrow \infty} \text{Prob } E_n.$$

Et si l'on a

$$(85) \quad F_1 \supset F_2 \supset \dots \supset F_{n-1} \supset F_n \dots,$$

on a

$$(86) \quad \text{Prob}[F_1 \text{ et } F_2 \dots \text{ et } F_n \dots] = \lim_{n \rightarrow \infty} \text{Prob } F_n.$$

La convergence en probabilité est légale. — Soit X_n convergeant en probabilité vers X .

La loi de probabilité de X peut être exprimée par la probabilité que X appartienne à un ensemble donné E , soit $P(E)$; de même, soit

$$P_n(E) = \text{Prob}(X_n \in E).$$

⁽¹⁾ Toutefois nous verrons qu'on est amené naturellement à élargir légèrement cette définition en n'imposant pas la convergence quand e appartient à une certaine catégorie exceptionnelle d'ensembles précisée plus loin, page 277, en particulier note ⁽²⁾.

⁽²⁾ Voir la page 172 de notre livre, R. T., 1^{re}.

⁽³⁾ Voir leur démonstration à peu près évidente à la page 24 de mon ouvrage R. T., 1^{re}.

⁽⁴⁾ La notation $x \in E$ signifie que x est un élément de l'ensemble E ; la notation $F \subset E$ signifie que l'ensemble F appartient à l'ensemble E .

Nous allons comparer $P(E)$ aux limites de $P_n(E)$ quand $n \rightarrow \infty$. Soit i l'intérieur de E , j son extérieur, c'est-à-dire soit i l'ensemble des points x de E dont chacun est le centre d'une sphéroïde S_x de rayon non nul ε_x ⁽¹⁾ appartenant à E et soit j l'intérieur de l'ensemble complémentaire de E . On a

$$\Pr(X_n \in E) \geq \Pr[X \in i \text{ et } X_n \in S_X].$$

D'où, en supposant $\varepsilon_x = 1$ quand x n'appartient pas à i ,

$$P_n(E) \geq \Pr[X \in i] + \Pr[X_n \in S_X] - 1.$$

Or soit ε positif arbitraire. On a

$$\begin{aligned} \Pr[X_n \in S_X] &= \Pr[(X, X_n) < \varepsilon_X] \geq \Pr[(X, X_n) < \varepsilon \text{ et } \varepsilon_X > \varepsilon] \\ &\geq \Pr[(X, X_n) < \varepsilon] + \Pr[\varepsilon_X > \varepsilon] - 1. \end{aligned}$$

Il est clair que $\Pr[\varepsilon_X > \varepsilon]$ ne décroît pas quand ε décroît et, en vertu de (84), converge vers $\Pr[\varepsilon_X > 0] = 1$. Dès lors, pour η , positif arbitraire donné, il existe une valeur ε_0 de ε telle que

$$\Pr[\varepsilon_X > \varepsilon_0] > 1 - \eta,$$

d'où

$$\Pr[X_n \in S_X] > \Pr[(X, X_n) < \varepsilon_0] - 1 + \eta.$$

On peut enfin, ε_0 , η étant fixés, prendre p , tel que

$$\Pr[(X, X_n) < \varepsilon_0] > 1 - \eta \quad \text{pour } n > p.$$

D'où

$$P_n(E) \geq P(i) - 2\eta \quad \text{pour } n > p.$$

Mais le même raisonnement peut s'appliquer au complémentaire de E et à l'intérieur j de ce complémentaire. On a donc

$$1 - P_n(E) \geq P(j) - 2\eta \quad \text{pour } n > p'$$

d'où

$$P(i) - 2\eta \leq P_n(E) \leq 1 - P(j) + 2\eta$$

et par suite,

$$P(i) \leq \text{toute limite (quand } n \rightarrow \infty) \text{ de } P_n(E) \leq 1 - P(j).$$

Dans le cas où X , X_n sont des nombres aléatoires dont les fonctions de répartition sont $F(x)$, $F_n(x)$, et où l'on prend pour E l'ensemble

⁽¹⁾ Ce sphéroïde est l'ensemble des points y de ω tels que $(x, y) < \varepsilon_x$.

des points X d'abscisse $< x$, i sera $\equiv E$ et j sera l'ensemble des points d'abscisses $> x$, de sorte qu'on aura

$$F(x) \leq \text{toute limite de } F_n(x) \leq F(x + 0).$$

On retrouve ainsi cette proposition connue que $F_n(x)$ tend vers $F(x)$ en tout point x où $F(x)$ est continu. On voit, en outre, que, dans ce cas, la continuité de $P(h)$ pour $h = E$ est une conséquence de la condition $P(i) + P(j) = 1$ ou encore de $P(f) = 0$ en appelant f la frontière de E , c'est-à-dire $f = \mathcal{O} - i - j$. Généralisons cette conclusion comme une définition. Dans un espace distancié [où l'on a encore $P(i) \leq P(E) \leq 1 - P(j)$, puisque $i \subset E \subset (\text{complémentaire de } j) \equiv C(j)$], la fonction $P(h)$ sera dite *continue pour* $h = E$ quand $P(i) = P(E) = P(C(j))$ ⁽¹⁾. Nous aurons alors le résultat suivant :

Dans tout espace distancié, si l'élément aléatoire X_n converge en probabilité vers X , la fonction de distribution de X_n soit $P_n(h) = \text{Prob}[X_n \in h]$, converge vers la fonction de distribution de X , soit $P(h) = \text{Prob}(X \in h)$, pour tout ensemble E pour lequel $P(h)$ est continu ⁽²⁾.

On sait que, même pour le cas simple où X et X_n sont des nombres, la réciproque n'est pas vraie.

Remarque. — Ce théorème généralise un théorème connu, concernant les nombres aléatoires, au moyen d'une généralisation convenable de la notion de fonction continue. Mais on peut se demander si une généralisation différente de cette notion ne pourrait étendre le champ de validité du théorème obtenu.

Nous allons voir qu'il n'en est pas ainsi. Soit, en effet, dans l'espace distancié $\mathcal{O}$ un ensemble E et sa frontière f supposée non vide et supposons qu'une autre généralisation de la notion de continuité nous conduise à déclarer continue pour $h = E$, une fonction de distribu-

⁽¹⁾ On peut dire que si $P(i) = P(E)$, $P(h)$ est continue par l'intérieur pour $h = E$ et que si $P(C(j)) = P(E)$, $P(h)$ est continue par l'extérieur pour $h = E$, expressions correspondant dans l'exemple traité ci-dessus aux continuités à gauche et à droite de $F(x)$.

⁽²⁾ On pourra dire que la convergence en probabilité de X_n vers X entraîne sa convergence légale si l'on élargit la définition de la convergence légale en n'imposant la convergence de $P_n(E)$ vers $P(E)$ que pour E tel que $P(h)$ soit continu pour $h = E$.

tion $P(h)$ d'un élément aléatoire X telle que $p(f) \neq 0$. Appelons alors A l'ensemble commun à E et f et $B = f - A$. On a

$$P(A) + P(B) = P(f) > 0,$$

donc l'un au moins des nombres $P(A)$, $P(B)$ n'est pas nul. Supposons $P(B) > 0$; prenons alors $X_n = X$ quand X n'appartient pas à B ; quand X appartient à B , comme il y a au moins un point γ de E à distance de X inférieure à $\frac{1}{n}$, on peut prendre X_n dans l'ensemble des points tels que γ . Ainsi, que X appartienne à B ou non, on a $(X, X_n) < \frac{1}{n}$. Donc $\left\{ \text{Prob} \left[(X, X_n) < \frac{1}{n} \right] \right\} = 1$ et X_n converge en probabilité vers X . Or la probabilité pour que X_n appartienne à E est égale à $P(E) + P(B)$, donc $P_n(E) - P(E) = P(B) \neq 0$, de sorte que $P_n(E)$ ne converge pas vers $P(E)$.

Lorsque $P(B) = 0$, on a $P(A) = P(f) > 0$. Prenons alors $X_n = X$ quand X n'appartient pas à A ; quand X appartient à A , on peut prendre et l'on prend pour X_n un point de l'ensemble $C(E)$ complémentaire de E , à distance de X inférieure à $\frac{1}{n}$. Alors on a toujours $(X_n, X) < \frac{1}{n}$, donc X_n converge en probabilité vers X . D'autre part

$$P_n(C(E)) = P(A) + P(C(E)) \quad \text{ou} \quad 1 - P_n(E) = P(A) + 1 - P(E),$$

d'où

$$P(E) - P_n(E) = P(A) > 0.$$

Donc $P_n(E)$ ne converge pas non plus vers $P(E)$.

Ainsi le théorème obtenu plus haut cesse d'être vrai pour toute définition de la continuité de $P(E)$ qui n'implique pas celle que nous avons adoptée plus haut.

Il est toujours précieux de pouvoir traduire une propriété où interviennent des probabilités comme la convergence en probabilité, par une propriété où n'interviennent que des éléments certains comme les propriétés des fonctions de distribution $P_n(E)$, $P(E)$. C'est pourquoi il va être utile, pour la rendre suffisante, de modifier la condition nécessaire qui vient d'être obtenue. On y arrivera en étendant au cas d'un espace distancié une remarque et des propositions de Kozakiewicz concernant les nombres aléatoires, mais en modifiant considérablement les démonstrations; en particulier, nous employons des fonctions de distributions au lieu de fonctions de répartition.

La convergence en probabilité de X_n vers X est une propriété qui, en principe, fait intervenir évidemment non seulement les lois de probabilité de X_n et de X pris séparément, mais la loi de distribution du couple X_n, X , soit

$$p_n(E, e) = \text{Prob}[X_n \in E \text{ et } X \in e].$$

La condition cherchée va faire intervenir $p_n(E, e)$ ou plus simplement $p_n(E, E)$.

Condition de convergence en probabilité de X_n vers X . — I. On a, avec les notations précédentes,

$$\begin{aligned} p_n(E, E) &\geq \text{Prob}[X_n \in E \text{ et } X \in i] \geq \text{Prob}[(X, X_n) < \varepsilon_X \text{ et } X \in i] \\ &\geq \text{Prob}[(X, X_n) < \varepsilon_X] + P(i) - 1. \end{aligned}$$

Quand $X_n \rightarrow X$ en probabilité, on voit comme plus haut que, pour η arbitraire, > 0 , il existe q indépendant de X et tel que

$$\text{Prob}[(X, X_n) < \varepsilon_X] > 1 - 2\eta \quad \text{pour } n > q.$$

Donc

$$p_n(E, E) > P(i) - 2\eta \quad \text{pour } n > q.$$

Mais il est clair que

$$\text{Prob}[X \in E] \geq \text{Prob}[X \in E \text{ et } X_n \in E].$$

On a donc

$$P(E) \geq p_n(E, E) > P(i) - 2\eta \quad \text{pour } n > q.$$

Quand on suppose que $P(h)$ est continu pour $h = E$, on en conclut

$$(87) \quad \lim_{n \rightarrow \infty} p_n(E, E) = P(E).$$

II. Pour démontrer la réciproque, nous commencerons par deux observations qui seraient inutiles si nous nous occupions seulement des nombres aléatoires.

Nous nous limiterons au cas où l'espace distancié $\mathcal{O}$ est *séparable*, c'est-à-dire où il existe une suite N dénombrable d'éléments $a_1, a_2, \dots, a_r, \dots$, de $\mathcal{O}$ tel que tout élément x de $\mathcal{O}$ soit limite d'une suite $a_{n_1}, a_{n_2}, \dots$, d'éléments, distincts ou non, convenablement extraite de la suite N . La plupart des espaces distanciés importants

(mais non tous) sont séparables et en particulier l'espace des nombres réels.

D'autre part, considérons un élément fixe b de $\mathcal{O}$ et le sphéroïde S_ρ de centre b et de rayon ρ . La probabilité $P(S_\rho)$ ⁽¹⁾ est évidemment, pour b fixe, une fonction non décroissante de ρ . Dès lors, elle est continue, sauf peut-être pour un ensemble dénombrable de valeurs de ρ , soient $\rho_1, \rho_2, \dots$. Pour ρ' différent des ρ_k , $P(S_\rho)$ est continu pour $\rho = \rho'$. C'est-à-dire que

$$P(S_{\rho'}) = \lim_{\varepsilon \rightarrow 0} P(S_{\rho' - \varepsilon}) = \lim_{\varepsilon \rightarrow 0} P(S_{\rho' + \varepsilon}) \quad (0 < \varepsilon < \rho').$$

Or l'intérieur i de $S_{\rho'}$ comprend au moins $S_{\rho' - \varepsilon}$. Donc

$$P(S_{\rho' - \varepsilon}) \leq P(i) \leq P(S_{\rho'})$$

et quand on fait tendre ε vers zéro, on voit qu'on a

$$P(i) = P(S_{\rho'}).$$

Dès lors $P(h)$ est, pour $h = S_{\rho'}$, continu par l'intérieur; de même par l'extérieur. Finalement on voit que $P(h)$ est continu pour $h = S_\rho$ pour toute valeur positive de ρ sauf $\rho_1, \rho_2, \dots$.

On voit que $P(h)$ sera continu en prenant pour h un sphéroïde de centre a_r et de rayon ρ , sauf peut-être pour une suite dénombrable de nombres $\rho_1^{(r)}, \rho_2^{(r)}, \dots$. L'ensemble T des $\rho_s^{(r)}$ est un ensemble dénombrable de nombres qu'on peut désigner par $\rho'_1, \rho'_2, \dots$. Dès lors, quel que soit q , il y a au moins un nombre ε_q différent de tous les nombres ρ'_k et $< \frac{1}{q}$. Et $P(h)$ sera continu en prenant pour h chacun des sphéroïdes U_1^q de centre a_r , de rayon ε_q . Ces préliminaires posés, on a évidemment

$$\begin{aligned} & \text{Prob}[(X, X_n) \leq 2\varepsilon_q] \\ & \geq \text{Prob} \{ [X \in U_1^q \text{ et } X_n \in U_1^q] \quad \text{ou} \quad \dots \quad \text{ou} \quad [X \in U_r^q \text{ et } X_n \in U_r^q] \}. \end{aligned}$$

Posons $V_1^q = U_1^q$ et, pour $r > 1$,

$$V_r^q = U_r^q - (U_1^q + \dots + U_{r-1}^q);$$

(1) Nous supposons donc ici que tout sphéroïde est un ensemble *probabilisable* pour X .

les ensembles $V_1^q, V_2^q, \dots$, étant disjoints, on pourra écrire l'inégalité précédente sous la forme

$$\begin{aligned} & \{ \text{Prob}[(X, X_n) \leq 2\varepsilon_q] \} \\ & \geq \text{Prob} \{ [X \in V_1^q \text{ et } X_n \in V_1^q] \quad \text{ou} \quad \dots \quad \text{ou} \quad [X \in V_r^q \text{ et } X_n \in V_r^q] \} \\ & = p_n(V_1^q, V_1^q) + \dots + p_n(V_r^q, V_r^q). \end{aligned}$$

Si nous pouvons démontrer que $P(h)$ est continu pour $h = \text{l'un quelconque des } V_r^q$, nous aurons, d'après l'hypothèse à la base de l'énoncé de la réciproque,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \{ \text{Prob}[(X, X_n) \leq 2\varepsilon_q] \} \\ & \geq P(V_1^q) + \dots + P(V_r^q) = P(V_1^q + \dots + V_r^q) = P(U_1^q + \dots + U_r^q). \end{aligned}$$

Le dernier terme tend vers 1 quand $r \rightarrow \infty$. Car tout élément x de $\mathcal{O}$ étant limite d'une suite des a_k , il y a au moins un a_k tel que $(x, a_k) \leq \varepsilon_q$. Par suite, x appartient à $U_1^q + \dots + U_k^q + U_{k+1}^q + \dots$. Cette dernière suite n'est donc autre que $\mathcal{O}$; par conséquent, en vertu de (86),

$$\lim_{r \rightarrow \infty} P(U_1^q + \dots + U_r^q) = P(\mathcal{O}) = 1.$$

On a donc

$$1 \geq \overline{\lim}_{n \rightarrow \infty} \dots \geq \lim_{n \rightarrow \infty} \{ \text{Prob}[(X, X_n) \geq 2\varepsilon_q] \} \geq 1.$$

Donc

$$\lim_{n \rightarrow \infty} \{ \text{Prob}(X, X_n) \leq 2\varepsilon_q \} = 1.$$

Pour $\omega > 0$ quelconque, il existe un $\varepsilon_q < \frac{\omega}{2}$. Donc

$$\text{Prob}[(X, X_n) < \omega] \geq \text{Prob}[(X, X_n) \leq 2\varepsilon_q].$$

Finalement, pour tout $\omega > 0$,

$$\lim_{n \rightarrow \infty} \text{Prob}[(X, X_n) < \omega] = 1,$$

c'est-à-dire que $X_n \rightarrow X$ en probabilité.

Démontrons maintenant la propriété admise chemin faisant. Comme V_r^q peut s'obtenir en formant successivement

$$W_1 = U_1^q, \quad W_2 = W_1 - U_2^q, \quad W_3 = W_2 - U_3^q, \quad \dots \quad \text{d'où} \quad V_r^q = W_r,$$

il suffit de prouver que si $P(h)$ est continu pour $h = U$ et $h = U_1$, il est continu pour $h = U - U_1$. Or, soient f et f_1 les frontières de U et U_1 ; on doit avoir $P(f) = P(f_1) = 0$. Mais tout point de la

frontière f_2 de $U - U_1$ appartient à f ou à f_1 ⁽¹⁾. On a donc

$$\begin{aligned} 0 \leq P(f_2) &\leq P(f + f_1) = P(f + [f_1 - f]) \\ &= P(f) + P(f_1 - f) \leq P(f) + P(f_1) = 0. \end{aligned}$$

Ainsi $P(f_2) = 0$ et par suite $P(h)$ est continu sur $U - U_1$.

En résumé, si X et X_n sont pris au hasard dans un espace *distancié séparable*, la condition *nécessaire et suffisante* pour que $X_n \rightarrow X$ en probabilité est que, pour tout ensemble E pour lequel $P(h)$ est continu, on ait $P(E) = \lim_{n \rightarrow \infty} p_n(E, E)$, c'est-à-dire

$$(87^{bis}) \quad \{ \text{Prob}[X \in E] \} = \lim_{n \rightarrow \infty} \text{Prob} \{ X \in E \text{ et } X_n \in E \}.$$

Condition de convergence en probabilité de X_n . — Les raisonnements précédents s'adaptent facilement au cas où l'on ne considère que la suite des X_n sans faire intervenir leur limite éventuelle X .

On a d'abord

$$\begin{aligned} \text{Prob}[X_n \in E] &\geq \text{Prob}[X_n \in E, X_m \in E] \geq \text{Prob}[X_n \in i, X_m \in E] \\ &\geq \text{Prob}[X_n \in i, (X_n, X_m) < \varepsilon_{X_n}] \\ &\geq \text{Prob}[X_n \in i] + \text{Prob}[(X_n, X_m) < \varepsilon] + \{ \text{Prob}[\varepsilon < \varepsilon_{X_n}] \} - 2. \end{aligned}$$

Or si X_n converge en probabilité vers X et si $P(h)$ est continu pour $h = E$, le premier membre et les trois premiers termes du dernier tendent respectivement vers $P(E)$, $P(E)$, 1, 1, quand $\frac{1}{n} + \frac{1}{m} \rightarrow 0$.

On a donc

$$\lim_{\frac{1}{n} + \frac{1}{m} \rightarrow 0} \{ \text{Prob}[X_n \in E, X_m \in E] \} = \text{Prob}[X \in E].$$

Puisqu'on ne connaît pas X , on ne connaît pas $P(E)$, mais on voit que, si X_n converge en probabilité, alors : 1° la fonction certaine

$$\text{Prob}[X_n \in E, X_m \in E]$$

doit, quand $\frac{1}{n} + \frac{1}{m} \rightarrow 0$, tendre vers une limite $\Pi(E)$ quand E n'appar-

(1) On voit facilement que pour qu'un point x appartienne à la frontière d'un ensemble E , il faut et il suffit que tout sphéroïde de centre x contienne un élément de E et un élément de $C(E)$. Dès lors, tout élément de f_2 est centre d'un sphéroïde contenant un élément de $U - U_1$ et un élément de $C(U - U_1)$, donc un élément commun à U et $C(U_1)$ et un élément appartenant à $C(U)$ ou à U_1 , donc il possède un élément de U et un élément de $C(U)$ ou un élément de $C(U_1)$ et de U_1 , donc il appartient à f ou à f_1 .

tient pas à une certaine famille exceptionnelle Φ ; 2° cette limite $\Pi(E)$ doit être une fonction de distribution, donc une fonction toujours ≥ 0 , additive au sens complet et telle que $\Pi(\mathcal{O}) = 1$; 3° la famille Φ , si elle existe, appartient à celle des ensembles E pour lesquels $\Pi(h)$ est non définie ou discontinue, s'il y en a.

Remarquons d'ailleurs que s'il existe un élément aléatoire Y dont $\Pi(E)$ soit une fonction de distribution, il n'en résulte nullement que Y soit limite en probabilité de X_n . On sait, en effet, déjà dans le cas où X_n est un nombre aléatoire, qu'il y a une infinité d'éléments aléatoires qui ont la même fonction de distribution.

Pour notre démonstration de la réciproque, nous devons supposer l'espace $\mathcal{O}$ à la fois séparable et complet.

On a, avec les notations précédentes,

$$\begin{aligned}
 (88) \quad & \left\{ \text{Prob}[(X_n, X_m) \leq 2\varepsilon_q] \right\} \\
 & \geq \text{Prob} \left\{ [X_n \in V_1^q \text{ et } X_m \in V_1^q] \text{ ou } \dots \text{ ou } [X_n \in V_r^q \text{ et } X_m \in V_r^q] \right\} \\
 & = \text{Prob}[X_n \in V_1^q \text{ et } X_m \in V_1^q] + \dots + \text{Prob}[X_n \in V_r^q \text{ et } X_m \in V_r^q].
 \end{aligned}$$

Nous supposons que chacun des derniers termes tend vers $\Pi(V_1^q), \dots, \Pi(V_r^q)$ respectivement, les $V_1^q, \dots, V_r^q$ étant définis comme précédemment à partir des U_r^q , mais les U_r^q étant définis à partir, non de $P(E)$, mais de la fonction $\Pi(E) = \lim_{\frac{1}{n} + \frac{1}{m} \rightarrow 0} \text{Prob}[X_n \in E, X_m \in E]$ (pour les E non exceptionnels). Donc

$$1 \geq \overline{\lim}_{\frac{1}{n} + \frac{1}{m} \rightarrow 0} \dots \geq \underline{\lim}_{\frac{1}{n} + \frac{1}{m} \rightarrow 0} \text{Prob}[(X_n, X_m) \leq 2\varepsilon_q] \geq \Pi(U_1^q + \dots + U_r^q)$$

quel que soit r . Et comme le dernier terme tend vers $\Pi(\mathcal{O}) = 1$, quand $r \rightarrow \infty$, on voit qu'on a

$$\lim_{\frac{1}{n} + \frac{1}{m} \rightarrow 0} \text{Prob}[(X_n, X_m) \leq 2\varepsilon_q] = 1$$

pour toute valeur de q et, par suite,

$$\lim_{\frac{1}{n} + \frac{1}{m} \rightarrow 0} \text{Prob}[(X_n, X_m) < \omega] = 1$$

pour tout $\omega > 0$.

La condition de Slutsky étant satisfaite, il en résulte que X_n converge en probabilité vers un élément aléatoire X .

Ainsi, nous avons démontré que : si un élément aléatoire X_n est choisi au hasard dans un espace $\mathcal{O}$ distancié, séparable et métriquement complet et si chaque épreuve détermine simultanément $X_1, X_2, \dots$, la condition *nécessaire et suffisante* pour que X_n converge *en probabilité* est qu'il existe une fonction d'ensembles E d'éléments de $\mathcal{O}$, soit $\Pi(E)$ qui soit ≥ 0 , additive au sens complet et pour laquelle $\Pi(\mathcal{O}) = 1$ et telle que

$$\text{Prob}[X_n \in E \text{ et } X_m \in E]$$

converge vers $\Pi(E)$ quand $\frac{1}{n} + \frac{1}{m} \rightarrow 0$ pourvu que $\Pi(h)$ soit continu pour $h = E$ ⁽¹⁾.

On sait qu'alors, d'après la proposition directe, $\text{Prob}[X_n \in E, X_m \in E]$ tend vers la fonction de distribution $P(E)$ de X (limite en probabilité de X_n), si $P(h)$ est continu pour $h = E$. Il en résulte que $P(E) = \Pi(E)$ pour tout ensemble E tel que $P(h)$ et $\Pi(h)$ soient continus pour $h = E$ et aussi qu'on peut prendre $P(E)$ pour $\Pi(E)$.

Convergence presque certaine. — Si X_n et X sont deux éléments aléatoires, définis simultanément sur chaque épreuve d'une catégorie déterminée, on définit la convergence presque certaine de X_n vers X par la condition que la probabilité de l'événement C consistant dans la convergence de X_n vers X soit $= 1$.

Cette définition, contrairement à celle de la convergence en probabilité, suppose donc que (quand C a lieu) on a donné un sens à l'expression : X_n converge vers X . Pour satisfaire à cette condition, nous pouvons ne pas nous en tenir au cas où X_n et X sont des nombres et considérer le cas où ils sont choisis au hasard (à chaque épreuve) dans un espace distancié $\mathcal{O}$.

Nous allons exprimer de plusieurs manières, ne faisant intervenir que des fonctions certaines, une condition nécessaire et suffisante pour la convergence presque certaine.

Mais auparavant, nous allons prouver ce fait, qui, intuitivement, est évident : si $X_n \rightarrow X$ presque certainement, il converge aussi vers X en probabilité.

Car soit $\varepsilon > 0$ arbitraire et E_n l'événement consistant en ce que,

⁽¹⁾ On suppose que $\Pi(E)$ est définie sur une famille (additive au sens complet) d'ensembles probabilisables pour tous les X_n et comprenant les sphéroides.

simultanément

$$(X, X_n) < \varepsilon, \quad (X, X_{n+1}) < \varepsilon, \quad \dots$$

Il est clair que

$$E_1 \subset E_2 \subset E_3 \dots$$

D'où, par (84),

$$\{\text{Prob}[X_n \rightarrow X]\} \leq \{\text{Prob}[E_1 \text{ ou } E_2 \text{ ou } \dots]\} = \lim_{n \rightarrow \infty} \text{Prob } E_n \leq 1.$$

Si donc $X_n \rightarrow X$ presque certainement, on a

$$\lim_{n \rightarrow \infty} \{\text{Prob } E_n\} = 1$$

et comme

$$\text{Prob } E_n \leq \text{Prob}[(X_n, X) < \varepsilon] \leq 1,$$

on a bien, pour chaque $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \{\text{Prob}[(X_n, X) < \varepsilon]\} = 1.$$

D'ailleurs, on sait bien que la convergence en probabilité n'entraîne pas inversement la convergence presque certaine. Et même, on peut citer des cas où $X_n \rightarrow X$ en probabilité alors que X_n ne converge vers X dans aucune épreuve.

Cas de l'espace cartésien. — Quand l'espace distancié est un espace cartésien à un nombre entier r de dimensions, $X^{(n)}$ a r coordonnées $X_1^{(n)}, \dots, X_r^{(n)}$. Si l'on prend, comme plus haut, avec des notations évidentes $[X, X^{(n)}] = \sqrt{[X_1 - X_1^{(n)}]^2 + \dots + [X_r - X_r^{(n)}]^2}$, la condition nécessaire et suffisante pour que *dans une épreuve* $X^{(n)} \rightarrow X$, est évidemment que chacune des coordonnées $X^{(n)}$ converge vers la coordonnée correspondante de X .

Dès lors, la condition nécessaire et suffisante pour que $X^{(n)} \rightarrow X$ presque certainement est que *chacune des coordonnées* de $X^{(n)}$ converge presque certainement vers la coordonnée correspondante de X .

Formule de Kolmogoroff généralisée. — Chaque épreuve déterminant encore à la fois $X_1, X_2, \dots$, dans un espace distancié $\mathcal{O}$, soit C l'événement consistant dans la convergence de cette série. Pour calculer la probabilité de C , appelons $C(\varepsilon)$ l'événement consistant en ce qu'il existe un entier n tel que

$$(X_q, X_m) < \varepsilon \quad \text{pour } q \text{ et } m > n.$$

Si l'espace distancié $\mathcal{O}$ où se meuvent les X_i est *métriquement complet*, C est le concours des $C(\varepsilon)$ où ε prend toutes les valeurs positives. Il est clair que si $\varepsilon > \varepsilon'$, on a $C(\varepsilon') \subset C(\varepsilon)$ et alors, d'après (86),

$$\text{Prob } C = \lim_{\varepsilon \rightarrow 0} \text{Prob } C(\varepsilon).$$

Mais $C(\varepsilon)$ est la réunion des événements $C_n(\varepsilon)$ où $C(\varepsilon)$ a lieu, n étant, cette fois, donné. Puisque $C_n(\varepsilon) \subset C_{n+1}(\varepsilon)$, on a, d'après (84),

$$\text{Prob } C(\varepsilon) = \lim_{n \rightarrow \infty} \text{Prob } C_n(\varepsilon).$$

De même, $C_n(\varepsilon)$ est le concours de $C_{n,n+1}(\varepsilon)$, $C_{n,n+2}(\varepsilon)$, ..., $C_{n,m}(\varepsilon)$, ..., où $C_{n,m}(\varepsilon)$ est l'événement consistant dans la réalisation des inégalités $(X_q, X_r) < \varepsilon$ où q, r prennent indépendamment les valeurs $n, n+1, \dots, m$. On a donc

$$\text{Prob } C_n(\varepsilon) = \lim_{m \rightarrow \infty} \text{Prob } C_{n,m}(\varepsilon)$$

et par suite

$$(89) \quad \text{Prob } C = \lim_{\varepsilon \rightarrow 0} \left\{ \lim_{n \rightarrow \infty} \left[\lim_{m \rightarrow \infty} \text{Prob } C_{n,m}(\varepsilon) \right] \right\}.$$

Il est parfois plus commode de faire intervenir les événements $H_{n,m}(\varepsilon)$ consistant dans la réalisation simultanée des inégalités $(X_n, X_r) < \varepsilon$ pour $r = n+1, n+2, \dots, m$. On a évidemment

$$C_{n,m}(\varepsilon) \subset H_{n,m}(\varepsilon) \subset C_{n,m}(2\varepsilon).$$

Donc

$$\text{Pr } C_{n,m}(\varepsilon) \leq \text{Pr } H_{n,m}(\varepsilon) \leq \text{Pr } C_{n,m}(2\varepsilon)$$

et, par suite,

$$\text{Prob } C(\varepsilon) \leq \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} H_{n,m}(\varepsilon) \leq \text{Prob } C(2\varepsilon)$$

et

$$(89 \text{ bis}) \quad \text{Pr } C = \lim_{\varepsilon \rightarrow 0} \left\{ \lim_{n \rightarrow \infty} \left[\lim_{m \rightarrow \infty} \text{Pr } H_{n,m}(\varepsilon) \right] \right\}$$

qui, dans le cas des nombres aléatoires, est la forme donnée à sa formule par Kolmogoroff.

Application. — Si X_n converge presque certainement,

$$\text{Pr } C = 1; \quad \text{d'où} \quad 1 = \lim_{\varepsilon \rightarrow 0} \text{Pr } C(\varepsilon),$$

où $\text{Pr } C(\varepsilon)$ ne peut croître quand ε décroît. Par conséquent,

$$1 \leq \text{Pr } C(\varepsilon) \leq 1.$$

Donc, quel que soit ε , $\text{Pr } C(\varepsilon)$ doit être égal à 1. Dès lors on doit avoir simplement

$$(90) \quad \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \text{Pr } H_{n,m}(\varepsilon) = 1$$

pour chaque valeur positive de ε .

Inversement, si cette égalité a lieu pour au moins une valeur positive de ε , on aura $\text{Prob } C(2\varepsilon) = 1$, d'où

$$(91) \quad \text{Pr } C = \lim_{\varepsilon \rightarrow 0} \text{Pr } C(2\varepsilon) = 1.$$

Ainsi quand les X_n appartiennent à un espace distancié *métriquement complet*, la condition nécessaire et suffisante pour la convergence presque certaine de X_n est (comme l'a observé Kozakiewicz, dans le cas où les X_n sont des nombres), que pour au moins une valeur de $\varepsilon > 0$, on ait (90), c'est-à-dire

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \{ \text{Prob}[(X_n, X_{n+1}) < \varepsilon \text{ et } (X_n, X_{n+2}) < \varepsilon \text{ et } \dots \text{ et } (X_n, X_{n+m}) < \varepsilon] \} = 1.$$

Intervention de la limite. — La condition précédente est commode en ce qu'elle n'exige pas la connaissance de la limite presque certaine de X_n . Mais pour la même raison, on peut et l'on doit lui en préférer une autre quand on veut que cette limite soit un élément aléatoire déterminé X . A cet effet, il suffit de reprendre le raisonnement précédent en substituant aux distances (X_n, X_q) les distances (X, X_n) .

Appelons Γ , $\Gamma(\varepsilon)$, $\Gamma_n(\varepsilon)$, $\Gamma_{n,m}(\varepsilon)$ les probabilités respectives : pour que $X_n \rightarrow X$; pour qu'il existe un entier n , tel que

$$(X, X_q) < \varepsilon \quad \text{pour } q > n;$$

pour que ces inégalités aient lieu, n étant donné, enfin pour que simultanément

$$(X, X_n) < \varepsilon, \quad (X, X_{n+1}) < \varepsilon, \quad \dots, \quad (X, X_m) < \varepsilon.$$

Alors

$$\text{Pr } \Gamma = \lim_{\varepsilon \rightarrow 0} \text{Pr } \Gamma(\varepsilon), \quad \text{Pr } \Gamma(\varepsilon) = \lim_{n \rightarrow \infty} \text{Pr } \Gamma_n(\varepsilon);$$

$$\text{Pr } \Gamma_n(\varepsilon) = \lim_{m \rightarrow \infty} \text{Pr } \Gamma_{n,m}(\varepsilon).$$

d'où

$$(89 \text{ ter}) \quad \text{Pr } \Gamma = \lim_{\varepsilon \rightarrow 0} \left\{ \lim_{n \rightarrow \infty} \left[\lim_{m \rightarrow \infty} \text{Pr } \Gamma_{n,m}(\varepsilon) \right] \right\}.$$

Quand $X_n \rightarrow X$ presque certainement, on a

$$1 = \text{Pr}\Gamma = \lim_{\varepsilon \rightarrow 0} \text{Pr}\Gamma(\varepsilon).$$

Et puisque $\text{Pr}\Gamma(\varepsilon)$ ne peut croître quand ε décroît, on a

$$\text{Pr}\Gamma(\varepsilon) = 1 \quad \text{pour tout } \varepsilon.$$

Dès lors on doit avoir pour chaque $\varepsilon > 0$,

$$(92) \quad \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \text{Pr}\Gamma_{n,m}(\varepsilon) = 1.$$

Réciproquement si cette égalité a lieu, comme le premier membre est $\text{Pr}\Gamma(\varepsilon)$ et que $\text{Pr}\Gamma = \lim_{\varepsilon \rightarrow 0} \text{Pr}\Gamma(\varepsilon)$, on aura

$$\text{Pr}\Gamma = 1.$$

Ainsi la condition *nécessaire et suffisante* pour que $X_n \rightarrow X$ presque certainement est la condition (92), *c'est-à-dire*,

$$(93) \quad \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \text{Pr}[(X, X_n) < \varepsilon, (X, X_{n+1}) < \varepsilon, \dots, (X, X_m) < \varepsilon] = 1.$$

On observera qu'au contraire de la condition (91), cette condition vient d'être établie *sans supposer* que soit métriquement complet l'espace distancié où X et les X_n sont choisis au hasard.

Autre forme de condition. — Établissons une condition pour la convergence presque certaine de X_n vers X , exprimée au moyen d'une fonction certaine analogue à la fonction de distribution, savoir ici

$$(94) \quad \Phi_{n,n+r}(E) = \text{Prob}[X_n \in E, \dots, X_{n+r} \in E].$$

En désignant encore par i l'intérieur de E et par ε_x le rayon d'un sphéroïde S_x ayant pour centre l'élément x de i et appartenant à E et prenant $\varepsilon_x = 1$ quand x n'appartient pas à i , on a

$$\begin{aligned} P_n(E) &= \text{Prob}[X_n \in E] \geq \Phi_{n,n+r}(E) \\ &\geq \text{Prob}[X \in i, X_n \in S_x, \dots, X_{n+r} \in S_x], \end{aligned}$$

d'où

$$(95) \quad P_n(E) \geq \text{Prob}[X \in i] + \text{Prob}[(X_n, X) < \varepsilon_x, \dots, (X_{n+r}, X) < \varepsilon_x] \rightarrow 1.$$

Or comme on l'a vu plus haut page 276, étant donné $\eta > 0$, il existe un nombre certain $\varepsilon > 0$, tel que

$$\{\text{Prob}[\varepsilon_X < \varepsilon]\} < \eta.$$

D'autre part, en appelant A l'avant-dernier terme de (95),

$$\begin{aligned} A &\geq \text{Prob}[\varepsilon_X \geq \varepsilon, (X_n, X) < \varepsilon, \dots, (X_{n+r}, X) < \varepsilon] \\ &\geq \{\text{Prob}[\varepsilon_X \geq \varepsilon]\} + \text{Prob}\Gamma_{n,n+r}(\varepsilon) - 1 \geq \text{Prob}\Gamma_{n,n+r}(\varepsilon) - \eta \end{aligned}$$

et par suite

$$(96) \quad P_n(E) \geq \Phi_{n,n+r}(E) \geq P(i) + \text{Prob}\Gamma_{n,n+r}(\varepsilon) - 1 - \eta.$$

où $\Gamma_{n,n+r}(\varepsilon)$ a la signification donnée page 287. A cet endroit on a montré que $\text{Pr } \Gamma_{n,n+r}(\varepsilon) \geq \text{Pr } \Gamma_n(\varepsilon)$ et que si $X_n \rightarrow X$ presque certainement, $\text{Pr } \Gamma_n(\varepsilon) \rightarrow 1$, de sorte que pour n assez grand et indépendant de r , ($n > N$), on a

$$\begin{aligned} \text{Pr } \Gamma_{n,n+r}(\varepsilon) &\geq \text{Pr } \Gamma_n(\varepsilon) > 1 - \eta, \\ P_n(E) &\geq \Phi_{n,n+r}(E) \geq P(i) - 2\eta \quad \text{pour } n > N. \end{aligned}$$

Si $P(h)$ est continu pour $h = E$, $P_n(E) \rightarrow P(E) = P(i)$ et puisque η est arbitraire (> 0), on en conclut

$$(97) \quad \lim_{n \rightarrow \infty} \Phi_{n,n+r}(E) = P(E),$$

r , dans cette relation, étant fixe ou variable avec n . D'ailleurs, par sa définition même, $\Phi_{n,n+r}(E)$ ne peut croître quand r croît et reste ≥ 0 , donc $\Phi_{n,n+r}(E)$ a une limite quand, n restant fixe, $r \rightarrow \infty$. On peut donc aussi écrire

$$(98) \quad \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \Phi_{n,n+r}(E) = P(E).$$

c'est-à-dire

$$(99) \quad \text{Prob}[X \in E] = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \text{Prob}[X_n \in E, X_{n+1} \in E, \dots, X_{n+r} \in E],$$

pourvu que $P(h)$ soit continu pour $h = E$.

Pour la réciproque il sera plus commode de mettre la condition nécessaire sous une forme un peu différente, qui nécessite un petit changement dans la démonstration.

Soit

$$\Psi_{n,n+r}(E) = \text{Prob}[X \in E, X_n \in E, \dots, X_{n+r} \in E].$$

On a

$$P(E) \geq \Psi_{n,n+r}(E) \geq \text{Prob}[X \in E, X_n \in E, \dots, X_{n+r} \in E]$$

et, d'après ce qui précède, le dernier terme est $> P(i) - 2\eta$ pour $n > N$.
Dès lors, on voit comme plus haut que, si $P(h)$ est continu pour $h = E$,
on a

$$(100) \quad P(E) = \lim_{n \rightarrow \infty} \Psi_{n,n+r}(E) = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \Psi_{n,n+r}(E)$$

ou

$$(100 \text{ bis}) \quad \text{Prob}[X \in E] = \lim_{n \rightarrow \infty} \lim_{r \rightarrow \infty} \text{Prob}\{X \in E, X_n \in E, \dots, X_{n+r} \in E\}.$$

Pour démontrer la réciproque, nous allons, comme pour la réciproque de la page 279, nous limiter au cas où l'espace distancié $\mathcal{O}$ (où sont pris au hasard X et X_n) est aussi *séparable*.

Ceci étant, avec les notations précédentes, le concours S_k des événements

$$X \in V_k^q, \quad X_n \in V_k^q, \quad \dots, \quad X_{n+r} \in V_k^q$$

entraîne le concours de

$$X \in U_k^q, \quad X_n \in U_k^q, \quad \dots, \quad X_{n+r} \in U_k^q,$$

lequel entraîne $\Gamma_{n,n+r}(2\varepsilon_q)$. On a donc pour tout entier $\alpha > 0$,

$$\text{Prob}\Gamma_{n,n+r}(2\varepsilon_q) \geq \text{Prob}[S_1 \text{ ou } S_2 \text{ ou } \dots \text{ ou } S_\alpha]$$

et puisque $S_1, \dots, S_\alpha$ sont incompatibles,

$$\text{Prob}\Gamma_{n,n+r}(\varepsilon_q) \geq \text{Prob} S_1 + \dots + \text{Prob} S_\alpha = \sum_{k=1}^{\alpha} \Psi_{n,n+r}(V_k^q).$$

Mais $P(h)$ est, comme on l'a vu, continu pour $h = V_k^q$ et si la condition (100) est vérifiée

$$P(V_k^q) = \lim_{n \rightarrow \infty} \Psi_{n,n+r}(V_k^q), \quad \dots$$

Pour $\omega > 0$ donné, il existe donc N_k tel que pour $n > N_k$

$$\Psi_{n,n+r}(V_k^q) > P(V_k^q) - \frac{\omega}{\alpha},$$

d'où, pour $n > N = N_1 + \dots + N_\alpha$,

$$\begin{aligned} \text{Prob}\Gamma_{n,n+r}(\varepsilon_q) &\geq P(V_1^q) + \dots + P(V_\alpha^q) - \omega \\ &= P(V_1^q + \dots + V_\alpha^q) - \omega = P(U_1^q + \dots + U_\alpha^q) - \omega. \end{aligned}$$

α étant arbitraire, prenons-le de sorte que

$$P(U_1'' + \dots + U_\alpha'') > P(\mathcal{O}) - \omega = 1 - \omega.$$

On voit qu'on a pour r quelconque et $n > N$

$$1 \geq \text{Prob } \Gamma_{n,n+r}(2\varepsilon_q) > 1 - 2\omega.$$

D'où

$$\lim_{n \rightarrow \infty} \text{Prob } \Gamma_{n,n+r}(2\varepsilon_q) = 1.$$

Et si ε est un nombre positif arbitraire, en prenant $2\varepsilon_q < \varepsilon$, on aura

$$\text{Prob } \Gamma_{n,n+r}(\varepsilon) \geq \text{Prob } \Gamma_{n,n+r}(2\varepsilon_q).$$

d'où

$$\lim_{n \rightarrow \infty} \text{Prob } \Gamma_{n,n+r}(\varepsilon) = 1.$$

Dès lors, comme on l'a vu plus haut, $X_n \rightarrow X$ presque certainement.

Ainsi, quand X et X_n sont choisis au hasard dans un espace distancié *séparable*, la condition *nécessaire et suffisante* pour que $X_n \rightarrow X$ presque certainement est que l'égalité (100) [explicitée par (100 bis)] soit vérifiée *pour tout* ensemble E tel que $P(h)$ soit continu pour $h = E$.

Condition de convergence presque certaine. — On a vu que si X_n converge presque certainement : 1° $\Phi_{n,n+r}(E)$ a une limite quand, n croissant indéfiniment, E n'appartient pas à une certaine famille exceptionnelle $\mathcal{G}$; 2° la limite est une fonction $\varpi(E)$ additive au sens complet. ≥ 0 et telle que $\varpi(\mathcal{O}) = 1$; 3° on peut prendre pour $\mathcal{G}$ la famille des ensembles E tels que $\varpi(h)$ ne soit ni défini ni continu pour $h = E$ ⁽¹⁾.

Montrons que si $\mathcal{O}$ est séparable et complet la réciproque est vraie. On a

$$\begin{aligned} \{ \text{Prob}[X_n \in E] \} &\geq \text{Prob}[X_n \in E, X_{n+r} \in E] \\ &\geq \text{Prob}[X_n \in E, X_{n+1} \in E, \dots, X_{n+r} \in E] \end{aligned}$$

ou avec les notations précédentes,

$$(100 \text{ ter}) \quad P_n(E) \geq p_{n,n+r}(E) \geq \Psi_{n,n+r}(E).$$

On a supposé que $\Phi_{n,n+r}(E) \rightarrow \varpi(E)$ si $\varpi(h)$ est continu pour $h = E$,

⁽¹⁾ On suppose que les ensembles de points de $\mathcal{O}$ ne faisant pas partie de $\mathcal{G}$ forment une famille additive au sens complet et comprenant tout sphéroïde de $\mathcal{O}$.

quand $n \rightarrow \infty$, de quelque façon que varie r entier ≥ 0 . Pour $r = 0$, $\Phi_{n,n+r}(E)$ se réduit à $P_n(E)$. Donc dans (100 *ter*) les termes extrêmes tendent vers $\varpi(E)$ et par suite aussi $p_{n,n+r}(E)$. Dès lors on en conclut que X_n converge au moins en probabilité vers un élément aléatoire X . Mais on peut aller plus loin et montrer directement que X_n converge presque certainement. Il suffit de procéder comme aux pages 290, 291, mais en remplaçant $\Psi_{n,n+r}$ par $\Phi_{n,n+r}$ et $\Gamma_{n,n+r}(\varepsilon)$ par $C_{n,n+r}(\varepsilon)$.

Remarque I. — On pourrait prendre aussi pour condition nécessaire et suffisante l'égalité (99) où le second membre ne contenant pas λ est donc un peu plus simple. Mais pour que le raisonnement subsiste, il faudrait faire intervenir la condition (89 *bis*) au lieu de (89 *ter*) et par suite supposer que l'espace $\mathcal{O}$ est non seulement distancié et séparable, mais encore, comme précédemment, métriquement complet.

Remarque II. — On a vu comment la simple introduction de la notion de distance de deux éléments aléatoires (se substituant à la valeur absolue de la différence de deux nombres aléatoires) permet une *énorme* généralisation de la théorie des diverses convergences stochastiques des nombres aléatoires.

On pourrait se demander s'il est possible d'aller au delà. La notion de convergence en probabilité (par exemple) d'une suite de nombres aléatoires a été étendue page 262 au cas d'une suite d'éléments aléatoires d'une façon qui fait intervenir essentiellement l'idée de distance de ces éléments. Pourrait-on la définir dans des cas plus généraux? La réponse est donnée par le théorème de la page 282. Ce théorème pourrait fournir une définition de la convergence en probabilité par la formule (87 *bis*) qui ne fait pas intervenir explicitement la distance. Celle-ci paraît être implicitement utilisée dans la condition de continuité de $P(h)$, quand on impose la condition que $P(f) = 0$, f étant la frontière de E . Mais la frontière d'un ensemble peut se définir sans intervention de la distance quand l'ensemble appartient à un espace *topologique*. De même pour la condition de séparabilité qui, avec les notations de la page 279, pourrait s'exprimer en disant que $\mathcal{O} = \bar{N}$ où $\bar{N}$, est la fermeture de N .

Ainsi une extension est possible. Nous nous contentons de le signaler ici, car pour les applications les plus importantes, il suffit de se limiter au cas où les éléments aléatoires considérés restent dans un espace distancié.

CHAPITRE III.

FONCTIONS ALÉATOIRES DE NATURE QUELCONQUE.

Dans ce qui précède, nous venons d'étudier diverses formes de convergence stochastique d'une suite d'éléments aléatoires X_n de nature quelconque, mais dépendant d'un paramètre entier n , cette suite étant déterminée tout entière par chaque épreuve.

Fonctions aléatoires. — L'idée vient naturellement d'étendre ces définitions au cas où le paramètre certain dont dépend l'élément est, non plus un entier n , mais un nombre réel quelconque t . La suite ainsi obtenue sera alors celle des déterminations dans une même épreuve d'un élément aléatoire $X(t)$ dépendant d'une variable certaine t . En somme, chaque épreuve ⁽¹⁾ déterminera une *fonction aléatoire* $X(t)$, mais une fonction qui ne sera pas nécessairement numérique.

Par exemple $X(t)$ est la suite des formes prises successivement aux différents instants t par une vague considérée au hasard; c'est encore la suite des états de santé, aux différents instants t , d'un individu pris au hasard dans une population donnée.

On pourra dire que $Y = X(t)$ est une transformation aléatoire d'un nombre certain t en un élément aléatoire de nature quelconque.

Convergences stochastiques. — On définira les diverses convergences stochastiques de $X(t)$ vers un élément aléatoire X , quand $t \rightarrow +\infty$ comme pour les convergences stochastiques de X_n quand $n \rightarrow \infty$. Mais une seconde idée vient naturellement aussi à l'esprit, celle d'étendre ces définitions au cas où t tend vers un nombre réel quelconque t_0 , infini ou fini. Il conviendra à cet effet de supposer que, pour chaque valeur de t , $X(t)$ est choisi (ainsi que sa limite) dans un espace distancié $\mathcal{O}$ indépendant de t .

⁽¹⁾ Dans les études récentes sur les fonctions aléatoires numériques, les applications à l'aérodynamique, etc..., conduisent à déterminer $X(t)$ par une suite d'épreuves correspondant aux valeurs successives de t .

Les applications que nous avons mentionnées pages 217, 221 et ici, amènent au contraire à considérer la fonction $X(t)$ toute entière comme un seul élément choisi dans une seule épreuve dans une famille donnée de fonctions.

Nous dirons alors qu'une fonction aléatoire $X(t)$ tend quand $t \rightarrow t_0$, vers un élément aléatoire X ; I, légalement; II, en moyenne d'ordre k ; III, en probabilité; IV, presque certainement, selon que :

I. la fonction de distribution $P(E/t)$ de $X(t)$ tend vers celle, $P(E)$ de X quand $t \rightarrow t_0$ pourvu que $P(h)$ soit continu pour $h = E$;

II. $\mathfrak{M}(X(t), X)^k \rightarrow 0$ quand $t \rightarrow t_0$;

III. Prob. $[(X(t), X) > \varepsilon]$ tend vers zéro pour chaque valeur de $\varepsilon > 0$, quand $t \rightarrow t_0$;

IV. La probabilité que $(X(t), X) \rightarrow 0$ quand $t \rightarrow t_0$ est égale à l'unité.

Application. — On voit aussitôt qu'à chaque mode de convergence stochastique va correspondre une définition de la « continuité stochastique » de $X(t)$.

Il lui correspondra aussi une définition de la dérivation stochastique ou de l'intégration stochastique, quand celles-ci sont possibles dans l'espace distancié $\mathcal{O}$ où se déplace $X(t)$. Nous ferons voir ailleurs que cela est possible dans certains espaces distanciés qu'on peut appeler *distanciés vectoriels*, parce qu'on peut y définir l'addition de deux éléments ou la multiplication d'un élément par un nombre réel.

Nous allons maintenant étudier ces applications de la convergence stochastique à la théorie des fonctions aléatoires de nature quelconque.

Remarque. — La convergence en moyenne d'ordre k a une signification mathématique aussi claire, mais une signification intuitive beaucoup moins nette que la convergence en probabilité, que, d'ailleurs, elle entraîne.

Il en sera de même des définitions correspondantes de la continuité, de la dérivation et de l'intégration. Ce sont pourtant les définitions de ces opérations correspondant à la convergence en m. o. k , qui sont les plus généralement étudiées. Mais il convient de se rendre compte que c'est *en raison de la plus grande facilité* des raisonnements *mathématiques* concernant la continuité, etc., en moyenne d'ordre k (spécialement en moyenne d'ordre 2) que les définitions correspondantes sont adoptées de préférence *et non pour des raisons de principe*.

Continuités stochastiques. — Une fonction aléatoire $X(t)$ sera évidemment dite continue stochastiquement (légalement, etc.) pour $t = t_0$, lorsque $X(t)$ converge stochastiquement (légalement, etc.), vers $X(t_0)$

quand $t \rightarrow t_0$. Il est clair que si elle est continue en moyenne d'ordre k ou presque certainement, elle est aussi continue en probabilité.

Fonction certaine associée à une fonction aléatoire. — Comme nous l'avons fait observer plus haut, il y a toujours intérêt à ramener un problème de Calcul des probabilités à un problème d'Analyse et en particulier à exprimer la condition qu'une fonction aléatoire $X(t)$ soit stochastiquement continue par une condition portant sur des fonctions certaines associées à $X(t)$. C'est ce que nous allons faire.

Il sera toujours facile d'y arriver quand le mode de convergence stochastique correspondant au mode de continuité stochastique considéré peut se traduire par une distance dans un espace Δ d'éléments aléatoires globaux. Dans Δ , à tout couple d'éléments aléatoires globaux $[X], [Y]$ est associée une distance $([X], [Y])$ et la convergence stochastique considérée de $X(t)$ vers X quand $t \rightarrow t_0$ s'y traduit par la condition que

$$\lim_{t \rightarrow t_0} ([X(t)], [X]) = 0;$$

la continuité stochastique correspondante se traduira par

$$\lim_{t \rightarrow t_0} \gamma(t, t_0) = 0,$$

où l'on a posé

$$(101) \quad \gamma(t, t_0) = ([X(t)], [X(t_0)]).$$

Or $\gamma(t, t_0)$ est une fonction certaine de t et de t_0 qui est bien déterminée par la loi de probabilité du couple aléatoire $X(t), X(t_0)$. Et elle est évidemment nulle pour $t = t_0$. C'est elle qu'on pourra associer à $X(t)$.

Donc : dans l'espace Δ , la condition nécessaire et suffisante pour la continuité stochastique (correspondant à Δ) de $X(t)$ pour $t = t_0$ est que la fonction certaine $\gamma(t, t_0)$, associée par (101) à $X(t)$, soit continue pour $t = t_0$.

Par exemple, on pourra (voir p. 264) prendre en correspondance avec la continuité en probabilité

$$(102) \quad f(\gamma(t, t_0)) = \mathfrak{M}f((X(t), X(t_0)))$$

quand pour chaque épreuve et pour chaque valeur de t , $X(t)$ appartient à un espace distancié $\mathcal{O}$, f étant une fonction ≥ 0 continue, non décroissante, avec $f(0) = 0$.

Dé même, la *continuité en moyenne d'ordre k de $X(t)$ pour $t = t_0$* s'exprimera par la continuité de

$$(103) \quad \gamma(t, t_0) = \sqrt[k]{\mathfrak{M}(X, X_0)^k}$$

pour $t = t_0$ [en écrivant pour abrégé $X = X(t)$, $X_0 = X(t_0)$] ⁽¹⁾.

Cette condition ne suppose pas que $X(t_0)$, ni $X(t)$, soient bornées en m. o. k . (en abrégé pour : moyenne d'ordre k). Si toutefois $X(t_0)$ est bornée en m. o. k ., $X(t)$ ne pourra être continu en m. o. k . pour $t = t_0$ que si $X(t)$ est aussi borné en m. o. k . pour chaque valeur de t suffisamment voisine de t_0 . C'est ce qui résulte de l'inégalité triangulaire

$$(104) \quad \sqrt[k]{\mathfrak{M}(X(t), \xi)^k} \leq \sqrt[k]{\mathfrak{M}(X(t_0), \xi)^k} + \sqrt[k]{\mathfrak{M}(X(t), X(t_0))^k}$$

déduite de (9) et où ξ est un élément certain arbitrairement choisi dans $\mathcal{O}$.

Remarquons même que la borne pourra être prise uniformément au sens suivant. Prenons ω tel que pour $|t - t_0| < \omega$, le dernier terme de (104) soit inférieur à un nombre ε donné > 0 . On a donc pour $|t - t_0| < \omega$,

$$\sqrt[k]{\mathfrak{M}(X(t), \xi)^k} \leq \sqrt[k]{\mathfrak{M}(X(t_0), \xi)^k} + \varepsilon = A,$$

de sorte que le premier membre est non seulement fini mais inférieur à un nombre fixe indépendant de t pour t suffisamment voisin de t_0 .

Relations des continuités stochastiques avec la continuité certaine. —

Si $X(t)$ est continue stochastiquement pour $t = t_0$, on ne peut affirmer qu'elle soit continue au sens ordinaire dans chaque épreuve. Réciproquement, supposons qu'elle soit continue au sens ordinaire dans chaque épreuve pour $t = t_0$; elle est alors *a fortiori* continue presque certainement et donc aussi continue en probabilité pour $t = t_0$. Mais il faudra une condition supplémentaire pour affirmer qu'elle est continue en m. o. k . Supposons par exemple que, en prenant pour $\mathcal{O}$ l'ensemble des nombres réels, $X(t)$ soit une fonction numérique égale à $(t - t_0)U$ où U est un nombre aléatoire fini, ≥ 0 dans chaque épreuve, mais dont la moyenne d'ordre k est infinie. Alors

$$\sqrt[k]{\mathfrak{M}(X(t), X(t_0))^k} = |t - t_0| \sqrt[k]{\mathfrak{M}U^k}$$

(1) Cette condition peut aussi être exprimée par la continuité pour $t = t_0$ de $[\gamma(t, t_0)]^k$, c'est-à-dire de $\mathfrak{M}(X(t), X(t_0))^k$, quantité dont l'expression est souvent plus simple que celle de $\gamma(t, t_0)$.

sera infini quel que soit $t \neq t_0$, de sorte que $X(t)$ n'est pas continu en m.o.k., bien que $X(t)$ soit évidemment continu pour $t = t_0$ dans chaque épreuve. Et pourtant $X(t)$ est dans cet exemple, borné en m. o. k. pour $t = t_0$. Par contre, si les éléments aléatoires $X(t)$ correspondant aux valeurs de t voisines de t_0 sont non seulement bornés en m. o. k., mais aussi également sommables d'ordre k (voir définition p. 272), comme d'autre part si $X(t)$ est continu pour $t = t_0$ dans toute épreuve, $X(t)$ est continu en probabilité pour $t = t_0$, alors il résulte de la page précédente que $X(t)$ sera aussi continu en m. o. k. pour $t = t_0$.

Un autre genre de difficulté se présente quand on considère l'ensemble S des valeurs de t pour lequel $X(t)$ est défini : généralement S sera un segment ou toute la droite des t . Si $X(t)$ est continu presque certainement pour chaque valeur de t sur S , peut-on dire qu'il est presque certain que $X(t)$ soit continu en tout point de S ? La réponse est négative, comme le montre l'exemple suivant. On choisit au hasard un point sur le segment $(0, 1)$ avec loi de probabilité uniforme. A chaque épreuve $X(t)$ est défini ainsi : quand à cette épreuve, on a choisi le point d'abscisse x_0 , on prend $X(t) = 1$ si $t = x_0$ et zéro dans le cas contraire. Il est clair que cette fonction n'est pour aucune épreuve, continue pour toute valeur de t (de 0 à 1). C'est-à-dire que la continuité dans une épreuve, pour toute valeur de t est un événement dont la probabilité non seulement n'est pas égale à 1, mais est même nulle. Pourtant pour chaque valeur de t , $X(t)$ est presque certainement continue. En effet, pour les épreuves où le point choisi n'est pas dans le segment $t_0 - \varepsilon, t_0 + \varepsilon$, $X(t)$ est dans ce segment égal à zéro et par suite, y est continu. La continuité de $X(t)$ pour $t = t_0$ est un événement dont la probabilité est donc $\geq 1 - 2\varepsilon$ pour tout $\varepsilon > 0$; elle est donc bien égale à 1.

Enfin observons aussi que :

1° Dans une épreuve, la continuité d'une fonction aléatoire $X(t)$ (dont les déterminations appartiennent à un espace distancié $\mathcal{O}$) définie sur un segment fini S , entraîne dans cette épreuve la continuité uniforme. Car, dans le cas contraire, il existerait $\varepsilon_0 > 0$, tel que pour tout entier n , il existe t_n et t'_n pour lesquels

$$(X(t_n), X(t'_n)) > \varepsilon_0, \quad |t_n - t'_n| < \frac{1}{n}.$$

On pourrait extraire de la suite des t_n une suite t_{n_p} convergente, vers

un point t_0 . Alors

$$0 < \varepsilon_0 < (X(t_{n_p}), X(t'_{n_p})) \leq (X(t_{n_p}), X(t_0)) + (X(t_0), X(t'_{n_p})).$$

Le second membre tendra vers zéro avec $\frac{1}{p}$, d'où la contradiction.

2° La continuité stochastique en chaque point n'entraînera pas la continuité uniforme dans chaque épreuve, puisqu'elle n'entraîne même pas (dans chaque épreuve) la continuité en chaque point. Cependant, on peut garder quelque chose de la continuité uniforme classique, sous la forme qui suit.

Continuité stochastique uniforme. — Considérons une fonction aléatoire $Z(t)$, telle que pour chaque t l'élément aléatoire global $[X(t)]$ (c'est-à-dire considéré dans l'ensemble des épreuves) appartienne à un espace distancié Δ . Nous dirons que la continuité stochastique de $X(t)$ correspondant à la définition de la distance $([X], [Y])$ dans Δ est uniforme sur un segment S , fini ou infini, de valeurs de t si à tout $\omega > 0$ correspond $\varepsilon > 0$, tel que sur S

$$|t - t'| < \varepsilon \quad \text{entraîne} \quad ([X(t)], [X(t')]) < \omega.$$

Une fonction aléatoire $X(t)$ continue uniformément en ce sens sur S est évidemment continue (au sens correspondant à Δ) en chaque point t de S . La réciproque, qui n'a pas lieu quand S est infini, est vraie quand il est fini.

En effet, s'il n'y avait pas continuité uniforme sur S ($\alpha \leq t \leq \beta$), il existerait $\omega_0 > 0$ tel que pour tout entier n , il existe deux points t_n, t'_n de S pour lesquels

$$|t_n - t'_n| < \frac{1}{n}, \quad ([X(t_n)], [X(t'_n)]) > \omega_0.$$

Or, en vertu du théorème de Weierstrass-Bolzano, on pourrait extraire de la suite des t_n , une suite convergente $t_{n_1}, t_{n_2}, \dots$. Soit u sa limite; c'est aussi celle de $t'_{n_1}, t'_{n_2}, \dots$, et l'on aurait

$$0 < \omega_0 < ([X(t_{n_r})], [X(t'_{n_r})]) \leq ([X(t_{n_r})], [X(u)]) + ([X(u)], [X(t'_{n_r})]).$$

Par hypothèse, les deux termes du second membre tendent vers zéro avec $\frac{1}{r}$, d'où la contradiction.

Ceci s'applique à la continuité en m. o. k . en posant

$$([X(t)], [X(t')]) = \sqrt[k]{\mathfrak{M}(X(t), X(t'))^k}.$$

Ceci peut s'appliquer aussi à la continuité en probabilité en prenant pour distance une des expressions considérées pages 263, 264. En prenant en particulier la première, on aura une valeur $\theta > 0$ telle que

$$\theta + \frac{1}{2} \text{Prob}[(X(t), X(t')) > \theta] < \omega,$$

d'où

$$(105) \quad \frac{1}{2} \text{Prob}[(X(t), X(t')) > \omega] < \omega \quad \text{pour } |t - t'| < \varepsilon.$$

Ainsi, quand $X(t)$ est continu en probabilité en chaque point t de S , à tout $\omega > 0$, correspond un $\varepsilon (> 0)$, *indépendant de t et t'* , pour lequel (105) a lieu, la probabilité étant calculée pour t et t' donnés.

Convergence stochastique d'une suite de fonctions aléatoires. — Soit $X^{(1)}(t), \dots, X^{(n)}(t), \dots$ une suite de fonctions aléatoires définie dans chaque épreuve en même temps que $X(t)$ dans le même espace distancié Δ d'éléments aléatoires globaux. Quand $X^{(n)}(t)$ converge stochastiquement vers $X(t)$ au sens de la convergence stochastique dans l'espace Δ , pour chaque valeur de t , que devient

$$\gamma^{(n)}(t, t_0) = ([X^{(n)}], [X_0^{(n)}])?$$

On déduit de l'inégalité triangulaire

$$|\gamma^{(n)}(t, t_0) - \gamma(t, t_0)| \leq ([X^{(n)}], [X]) + ([X_0^{(n)}], [X_0]).$$

Donc

$$\gamma^{(n)}(t, t_0) \rightarrow \gamma(t, t_0).$$

Par exemple, si $X^{(n)}(t)$ converge en probabilité vers $X(t)$ pour chaque valeur de t , alors $\gamma^{(n)}(t, t_0) \rightarrow \gamma(t, t_0)$ quand on définit γ par la formule (102) et $\gamma^{(n)}$ par une formule analogue.

De même, si $X^{(n)}(t)$ converge en m. o. k . vers $X(t)$ pour chaque valeur de t , la fonction certaine associée à $X^{(n)}(t)$, soit

$$\gamma^{(n)}(t, t_0) = \sqrt[k]{\mathfrak{M}(X^{(n)}(t), X^{(n)}(t_0))^k},$$

converge vers la fonction certaine $\gamma(t, t_0)$ associée à $X(t)$ par (103).

Mais on ne pourrait sans restriction remplacer ici les écarts moyens d'ordre k par leurs puissances k . On peut le faire cependant si $X(t)$

est borné en m. o. k . pour chaque valeur de t . Car alors si ξ est un élément certain de $\mathcal{O}$

$$\gamma(t, t_0) \leq \sqrt[k]{\mathfrak{M}(X(t), \xi)^k} + \sqrt[k]{\mathfrak{M}(X(t_0), \xi)^k},$$

donc $\gamma(t, t_0)$ est fini. Or pour $\omega > 0$ arbitraire, il existe N tel que

$$|\gamma^{(n)}(t, t_0) - \gamma(t, t_0)| < \omega \quad \text{pour } n > N,$$

donc, pour t fixé, il existe un nombre B pour lequel $\gamma(t, t_0) < B$ et $\gamma^{(n)}(t, t_0) < B$ indépendant de n pour $n > N$. D'où

$$\mathfrak{M}(X^{(n)}(t), X^{(n)}(t_0))^k - \mathfrak{M}(X(t), X(t_0))^k \leq |\gamma^{(n)}(t, t_0) - \gamma(t, t_0)| \cdot kB^{k-1}.$$

Donc si $X(t)$ est borné en m. o. k . pour toute valeur de t et si $X^{(n)}(t)$ converge vers $X(t)$ en m. o. k ., alors $\mathfrak{M}(X^{(n)}(t), X^{(n)}(t_0))^k$ tend vers $\mathfrak{M}(X(t), X(t_0))^k$.

Autre fonction associée. — Dans le cas où $X(t)$ est une fonction numérique, il y a intérêt, quand on étudie la dérivation en moyenne quadratique, à faire jouer le rôle de $\gamma(t, t_0)$ au produit moyen ⁽¹⁾ de $X(t)$, $X(t_0)$. Cette substitution est non seulement inutile, mais source de complication en ce qui concerne la continuité en m. q. (abrégé de moyenne quadratique).

Cependant, non seulement on a le droit de faire cette substitution, mais on peut même l'étendre au cas des éléments aléatoires de nature quelconque en remplaçant le produit moyen par le produit scalaire moyen introduit (p. 255),

$$(106) \quad \Gamma(t, t_0) = \xi X, \xi X_0 = \frac{1}{2} [\mathfrak{M}(\xi, X)^2 + \mathfrak{M}(\xi, X_0)^2 - \mathfrak{M}(X, X_0)^2],$$

où ξ est un élément certain de $\mathcal{O}$ choisi arbitrairement, mais une fois pour toutes.

Bien qu'il n'y ait aucun avantage actuel à remplacer $\gamma(t, t_0)$ par $\Gamma(t, t_0)$, montrons qu'on peut le faire afin de pouvoir employer une même fonction certaine pour la continuité en m. q. comme pour la dérivation en m. q. Mais il faudra supposer $X(t_0)$ borné en m. q. Or on a vu (p. 296) qu'alors si $X(t)$ est continu en m. q. pour $t = t_0$, $X(t)$ est borné en m. q. pour t voisin de t_0 .

Supposons donc $X(t)$ borné en moyenne quadratique pour t voisin

de t_0 , alors pour tout élément certain ξ de $\mathcal{O}$, le produit scalaire moyen ⁽¹⁾

$$\begin{aligned}\Gamma(t_0, t) &= \Gamma(t, t_0) = \xi X(t_0) \cdot \xi X(t) \\ &= \frac{1}{2} \{ \mathfrak{N}(\xi, X(t))^2 + \mathfrak{N}(\xi, X(t_0))^2 - \mathfrak{N}(X(t), X(t_0))^2 \}\end{aligned}$$

aura une valeur finie pour t voisin de t_0 . On a

$$\Gamma(t_0, t_0) = \mathfrak{N}(\xi, X(t_0))^2,$$

d'où

$$(107) \quad \mathfrak{N}(X(t), X(t_0))^2 = \Gamma(t, t) + \Gamma(t_0, t_0) - 2\Gamma(t, t_0).$$

D'après cette formule, si $\Gamma(t, t')$ est une fonction de t, t' continue pour $t = t_0 = t'$, ou, moins strictement, si $\Gamma(t, t_0)$ et $\Gamma(t, t)$ sont deux fonctions continues de t pour $t = t_0$, $X(t)$ sera continu en moyenne quadratique pour $t = t_0$.

Inversement, si $X(t)$ est continue en m. q. pour $t = t_0$, alors écrivant pour abréger X, X_0 pour $X(t), X(t_0)$, posons

$$(X, \xi) = (X_0, \xi) + \Theta(X, X_0).$$

En vertu de l'inégalité triangulaire, on pourra supposer $|\Theta| \leq 1$. Dès lors

$$2\Gamma(t, t_0) - 2\Gamma(t_0, t_0) = \mathfrak{N}[2\Theta(X_0, \xi)(X, X_0)] + \mathfrak{N}\Theta^2(X, X_0)^2 - \mathfrak{N}(X, X_0)^2.$$

D'où, en vertu de l'inégalité de Schwarz,

$$2|\Gamma(t, t_0) - \Gamma(t_0, t_0)| \leq 2\sqrt{\mathfrak{N}(X_0, \xi)^2 \mathfrak{N}(X, X_0)^2} + \mathfrak{N}(X, X_0)^2.$$

Quand $t \rightarrow t_0$, $\mathfrak{N}(X, X_0)^2$ tend vers zéro et par suite $\Gamma(t, t_0) \rightarrow \Gamma(t_0, t_0)$ si $X(t_0)$ est borné en m. q. De même

$$\begin{aligned}2|\Gamma(t, t) - \Gamma(t, t_0)| &= |\mathfrak{N}[2\Theta(X_0, \xi)(X, X_0)] + 2\mathfrak{N}(X, X_0)^2| \\ &\leq 2\sqrt{\mathfrak{N}(X_0, \xi)^2 \mathfrak{N}(X, X_0)^2} + 2\mathfrak{N}(X, X_0)^2\end{aligned}$$

et quand $t \rightarrow t_0$, $\Gamma(t, t) \rightarrow \Gamma(t_0, t_0)$.

Ainsi, quand $X(t_0)$ est borné en m. q., la condition nécessaire et suffisante pour que $X(t)$ soit continue en m. q. pour $t = t_0$ est que $\Gamma(t, t_0)$ et $\Gamma(t, t)$ soient continues au sens ordinaire pour $t = t_0$.

⁽¹⁾ On a d'abord employé dans le cas des nombres aléatoires, le coefficient de corrélation. Dans ce même cas des nombres, Dedebrand et Wehrle ainsi que Loève ont trouvé plus commode d'utiliser le produit moyen, que nous avons généralisé ici.

On va de même montrer que la convergence en m. q. entraîne la convergence de la fonction associée.

Suite de fonctions aléatoires. — Soit $X^{(n)}(t)$ une autre fonction aléatoire et $\Gamma^{(n)}(t, t_0)$ son produit scalaire moyen

$$\Gamma^{(n)}(t, t_0) = \frac{1}{2} [\mathfrak{M}(X^{(n)}, \xi)^2 + \mathfrak{M}(X_0^{(n)}, \xi)^2 - \mathfrak{M}(X^{(n)}, X_0^{(n)})^2].$$

Que devient $\Gamma_n(t, t_0)$ quand $X^{(n)}(t)$ converge en m. q. vers $X(t)$ pour toute valeur de t ? La question a été posée et résolue par Loève dans le cas des nombres. Nous allons étendre sinon sa démonstration, du moins son énoncé au cas d'éléments aléatoires pris dans un même espace distancié $\mathcal{O}$.

Supposons que $X^{(n)}(t)$ converge vers $X(t)$ en m. q. quand $n \rightarrow \infty$ et soit

$$\Gamma^{(n)}(t, t_0) - \Gamma(t, t_0) = \frac{1}{2} \mathfrak{M} U^{(n)},$$

avec

$$U^{(n)} = (X^{(n)}, \xi)^2 + (X_0^{(n)}, \xi)^2 - (X^{(n)}, X_0^{(n)})^2 \\ - (X, \xi)^2 - (X_0, \xi)^2 + (X, X_0)^2.$$

On a

$$(X_0^{(n)}, \xi) = (X_0, \xi) + \theta'_n(X_0, X_0^{(n)}), \\ (X^{(n)}, \xi) = (X, \xi) + \theta_n(X, X^{(n)}), \\ (X^{(n)}, X_0^{(n)}) = (X, X_0) - \Omega_n[(X, X^{(n)}) + (X_0, X_0^{(n)})],$$

avec

$$|\theta_n| \leq 1, \quad |\theta'_n| \leq 1, \quad |\Omega_n| \leq 1.$$

D'où

$$U^{(n)} = 2\theta_n(X, X^{(n)})(X, \xi) + \theta_n^2(X, X^{(n)})^2 \\ + 2\theta'_n(X_0, X_0^{(n)})(X_0, \xi) + \theta_n'^2(X_0, X_0^{(n)})^2 \\ + 2\Omega_n(X, X_0)[(X, X^{(n)}) + (X_0, X_0^{(n)})] - \Omega_n^2[(X, X^{(n)}) + (X_0, X_0^{(n)})]^2.$$

Par suite

$$|\mathfrak{M} U^{(n)}| \leq 2\sqrt{\mathfrak{M}(X, \xi)^2 \mathfrak{M}(X, X^{(n)})^2} + \mathfrak{M}(X, X^{(n)})^2 \\ + 2\sqrt{\mathfrak{M}(X_0, \xi)^2 \mathfrak{M}(X_0, X_0^{(n)})^2} + \mathfrak{M}(X_0, X_0^{(n)})^2 \\ + 2\sqrt{\mathfrak{M}(X, X_0)^2 [\sqrt{\mathfrak{M}(X, X^{(n)})^2} + \sqrt{\mathfrak{M}(X_0, X_0^{(n)})^2}]^2}.$$

Dès lors si $\mathfrak{M}(X, \xi)^2$, $\mathfrak{M}(X_0, \xi)^2$ et par suite $\mathfrak{M}(X, X_0)^2$ sont finis, on voit que $\Gamma^{(n)}(t, t_0) \rightarrow \Gamma(t, t_0)$.

Ainsi quand $X(t)$ est borné en m. q. pour chaque valeur de t , alors, si $X^{(n)}(t)$ converge en m. q. vers $X(t)$ quel que soit t , $\Gamma^{(n)}(t, t_0)$ converge au sens ordinaire vers $\Gamma(t, t_0)$, quel que soit t .

Autre forme de condition. — On peut donner une autre condition pour la continuité en probabilité en utilisant le critère de convergence de la page 282, mais par suite, en supposant non seulement distancié mais séparable, l'espace où $X(t)$ est choisi au hasard pour chaque valeur de t .

Soit $p(E, t)$ la fonction de distribution de $X(t)$. Posons

$$(108) \quad p(E, e, t, t_0) = \text{Prob}[X(t) \in E, X(t_0) \in e].$$

Pour que $X(t)$ soit continu en probabilité pour $t = t_0$, il faut et il suffit que

$$(109) \quad \lim_{t \rightarrow t_0} p(E, E, t, t_0) = p(E, t_0)$$

pour tout ensemble E tel que $p(h, t_0)$ soit continu en h pour $h = E$.

Il est d'ailleurs clair que

$$(110) \quad p(E, E, t_0, t_0) = p(E, t_0).$$

La condition s'exprime donc par la *continuité de* $p(E, E, t, t_0)$ en t pour $t = t_0$ pour tout ensemble E tel que $p(h, h, t_0, t_0)$ soit continu en h pour $h = E$.

Continuité presque certaine. — Nous avons montré ailleurs ⁽¹⁾ que, même dans le cas simple des nombres aléatoires, la convergence presque certaine ne peut s'exprimer par l'intermédiaire d'une distance de deux éléments aléatoires globaux. Par conséquent, pour exprimer la continuité presque certaine de $X(t)$ pour $t = t_0$, on ne pourra plus choisir une distance globale, comme fonction certaine associée à X . Mais, en se référant aux pages 288, 291, on peut employer des fonctions analogues aux fonctions de distribution.

Nous allons auparavant donner une formule analogue à la formule de Kolmogoroff, en passant des suites discontinues aux suites continues.

En employant les mêmes raisonnements qu'aux pages 286 à 292, on verra facilement que :

1° Si l'on appelle Γ la convergence, dans une épreuve, de $X(t)$ vers l'élément ξ de $\mathcal{D}$ quand $t \rightarrow t_0$ par valeurs quelconques, on a

$$(111) \quad \text{Prob } \Gamma = \lim_{\varepsilon > 0} \lim_{\tau > 0} \text{Prob}[(X(t), \xi) < \varepsilon \text{ pour } |t - t_0| < \tau].$$

(1) Voir page 235 de R. T. 1^{er}.

2° Si l'on appelle C la convergence dans une épreuve, de $X(t)$ quand $t \rightarrow t_0$, on a, quand l'espace $\mathcal{O}$ est métriquement *complet*,

$$(112) \quad \text{Prob C} = \lim_{\varepsilon \rightarrow 0} \lim_{\tau \rightarrow 0} \left\{ \text{Prob} \left[(X(t), X(t')) < \varepsilon \left\{ \begin{array}{l} |t - t_0| \leq \tau \\ \text{et} \\ |t' - t_0| \leq \tau \end{array} \right\} \right] \right\},$$

où $0 < \tau$.

Il en résulte par le même raisonnement qu'à la page 287, que la *condition nécessaire et suffisante* :

I. Pour que $X(t)$ converge *presque certainement* vers ξ de $\mathcal{O}$ quand $t \rightarrow t_0$, est que

$$(113) \quad \lim_{\tau \rightarrow 0} \left\{ \text{Prob} [(X(t), X(t_0)) < \varepsilon \text{ pour } |t - t_0| < \tau] \right\} = 1.$$

II. Pour que $X(t)$ converge *presque certainement* dans un espace $\mathcal{O}$ métriquement *complet*, quand $t \rightarrow t_0$ par valeurs distinctes de t_0 , est que

$$(114) \quad \lim_{\tau \rightarrow 0} \lim_{\tau' \rightarrow 0} \left\{ \text{Prob} [(X(t), X(t')) < \varepsilon \text{ pour } \tau \leq \frac{t - t_0}{t' - t_0} \leq \tau'] \right\} = 1.$$

De I, on déduit :

I bis. Pour que $X(t)$ soit *presque certainement continu* pour $t = t_0$, il faut et il suffit que, pour tout $\varepsilon > 0$, on ait

$$(113 \text{ bis}) \quad \lim_{\tau \rightarrow 0} \left\{ \text{Prob} [(X(t), X(t_0)) < \varepsilon \text{ pour } |t - t_0| < \tau] \right\} = 1.$$

On peut aussi exprimer la condition I en faisant intervenir les fonctions de distribution.

Soit pour $\omega > 0$ et en posant $X_0 \equiv X(t_0)$

$$p(E, \omega) = \text{Prob}[X(t) \in E \text{ pour } |t - t_0| < \omega] \quad \text{et} \quad P(E) = \text{Prob}[X_0 \in E].$$

Étudions le comportement de $p(E, \omega)$ quand $X(t)$ est continue presque certainement pour $t = t_0$. On a

$$\begin{aligned} P(E) &\geq p(E, \omega) \geq \text{Prob}[X_0 \in i \text{ et } (X(t), X_0) < \varepsilon_{X_0} \text{ pour } |t - t_0| < \omega] \\ &\geq \text{Prob}\{X_0 \in i \text{ et } \varepsilon < \varepsilon_{X_0} \text{ et } [(X(t), X_0) < \varepsilon \text{ pour } |t - t_0| < \omega]\} \\ &\geq P(i) + \text{Prob}[\varepsilon < \varepsilon_{X_0}] + \left\{ \text{Prob}[(X(t), X(t_0)) < \varepsilon \text{ pour } |t - t_0| < \omega] \right\} - 2. \end{aligned}$$

On a déjà vu que le second terme du dernier membre $\rightarrow 1$ quand $\varepsilon \rightarrow 0$ et d'après (113 bis), le dernier terme aussi quand ε étant fixé $\omega \rightarrow 0$. Donc, si $P(h)$ est continu pour $h = E$, on aura

$$(115) \quad P(E) = \lim_{\omega \rightarrow 0} p(E, \omega).$$

C'est une généralisation du résultat de la page 277.

Pour la réciproque, nous aurons à supposer comme à la page 279 que l'espace $\mathcal{D}$ est *séparable*. Alors supposons vérifiée (115) quand $P(h)$ est continu pour $h = E$.

En donnant à ε_q , V_1^q , V_2^q , ... les mêmes significations que page 280, pour le cas où

$$P(E) = \text{Prob}[X_0 \in E],$$

on aura

$$\begin{aligned} 1 &\geq \text{Prob}[(X(t), X_0) \leq 2\varepsilon_q \text{ pour } |t - t_0| < \omega] \\ &\geq \text{Prob}\{[X(t) \in V_1^q \text{ pour } |t - t_0| < \omega] \text{ ou } \dots \text{ ou } [X(t) \in V_r^q \text{ pour } |t - t_0| < \omega]\} \\ &\geq \{ \text{Prob}[X(t) \in V_1^q \text{ pour } |t - t_0| < \omega] \} + \dots \\ &\quad + \{ \text{Prob}[X(t) \in V_r^q \text{ pour } |t - t_0| < \omega] \} \\ &= p(V_1^q, \omega) + \dots + p(V_r^q, \omega). \end{aligned}$$

On suppose $P(h)$ continu pour h identique à l'un quelconque des V_s^q , le dernier membre tend donc vers

$$P(V_1^q) + \dots + P(V_r^q) = P(U_1^q + \dots + U_r^q)$$

quand $\omega \rightarrow 0$. Comme le dernier terme est pour r assez grand aussi voisin de 1 que l'on veut, on voit finalement qu'on a

$$\lim_{\omega \rightarrow 0} \{ \text{Prob}[(X(t), X(t_0)) \leq 2\varepsilon_q \text{ pour } |t - t_0| < \omega] \} = 1,$$

et en prenant $2\varepsilon_q < \varepsilon$

$$\lim_{\omega \rightarrow 0} \{ \text{Prob}[(X(t), X(t_0)) \leq \varepsilon \text{ pour } |t - t_0| < \omega] \} = 1,$$

c'est-à-dire d'après (113 bis) que $X(t)$ est presque certainement continue pour $t = t_0$.

En résumé, si $X(t)$ est, pour chaque valeur de t , choisi au hasard dans un espace distancié séparable $\mathcal{D}$ et si dans chaque épreuve $X(t)$ est déterminé pour toutes les valeurs de t appartenant à un voisinage fixe de t_0 , alors pour que $X(t)$ soit *presque certainement* continu pour $t = t_0$, il faut et il suffit que l'on ait

$$(116) \quad \{ \text{Prob}[X(t_0) \in E] \} = \lim_{\omega \rightarrow 0} \{ \text{Prob}[X(t) \in E, \text{ à la fois pour toutes les valeurs de } t \text{ telles que } |t - t_0| < \omega] \}$$

pourvu que le premier membre soit une fonction d'ensemble $P(h)$ continue pour $h = E$.

Généralisation du paramètre. — Dans ce qui précède, la fonction aléatoire $X(t)$ dépendait d'une variable t qui était un nombre certain.

L'essentiel ici était de supposer t certain, sans quoi tout devrait être profondément modifié. Par contre on peut garder les énoncés et les résultats précédents, sauf à y faire quelques modifications de forme (dont la teneur serait évidente), en supposant seulement que t est un élément certain appartenant à un espace distancié [en général distinct de celui où se meut $X(t)$].

NOTE ADDITIONNELLE.

Coordonnées dans un espace distancié séparable. — Nous avons montré il y a longtemps qu'on peut caractériser tout point d'un espace distancié « séparable » $\mathcal{S}$ de la façon suivante. $\mathcal{S}$ étant séparable, on peut en extraire une suite dénombrable N d'éléments

$$a_0, a_1, a_2, \dots,$$

tels que tout élément de $\mathcal{S}$ appartienne à N ou en soit un élément d'accumulation. Soit alors X un élément de $\mathcal{S}$. On pourra appeler « coordonnées » de X relativement à N , la suite ordonnée dénombrable des nombres

$$X_k = (X, a_k) = (a_0, a_k) \quad (k = 1, 2, \dots).$$

On peut prouver que la distance de X et de Y (de coordonnées $Y_1, Y_2, \dots$) est égale à la borne supérieure des nombres $|X_k - Y_k|$.

Quand X est un élément aléatoire choisi au hasard dans $\mathcal{S}$, sa loi de probabilité déterminera non seulement la loi de probabilité de chacune de ses coordonnées X_k , mais encore la loi de probabilité de toute suite ordonnée d'un nombre fini de ces coordonnées.

Il sera souvent plus commode de déterminer la loi de probabilité d'une telle suite que la loi de probabilité de X .

En particulier, posons

$$\begin{aligned} F_n(x_1, x_2, \dots, x_n) &= \text{Prob}[X_1 < x_1, \dots, X_n < x_n], \\ F(x_1, x_2, \dots, x_n, x_{n+1}, \dots) &= \text{Prob}[X_1 < x_1, X_2 < x_2, \dots, X_n < x_n, X_{n+1} < x_{n+1}, \dots]. \end{aligned}$$

La connaissance de la suite des fonctions F_n est équivalente à celle de la seule fonction F . D'une part, on a

$$F_n(x_1, \dots, x_n) = F(x_1, x_2, \dots, x_n, +\infty, +\infty, \dots).$$

Et d'autre part, en vertu des définitions de F_n et F et de (86), on a

$$F(x_1, x_2, \dots, x_p, x_{p+1}, \dots) = \lim_{n \rightarrow \infty} F_n(x_1, x_2, \dots, x_n).$$

Enfin connaissant les F_n , nous pourrions en déduire

$$\begin{aligned} G(x_1, x_2, \dots, y_1, y_2, \dots) &= \text{Prob}[x_1 \leq X_1 < y_1, x_2 \leq X_2 < y_2, \dots] \\ &= \lim_{n \rightarrow \infty} \text{Prob}[x_1 \leq X_1 < y_1, \dots, x_n \leq X_n < y_n] \\ &= \lim_{n \rightarrow \infty} \Delta_n F_n(x_1, \dots, x_n), \end{aligned}$$

où Δ_n est la « différence $n^{\text{ième}}$ » de $F_n(x_1, \dots, x_n)$ quand on donne aux x_k les accroissements $y_k - x_k$.

Dès lors, connaissant les F_n , nous connaissons ainsi la fonction de distribution $p(e)$ de X pour tous les ensembles e particulièrement simples qui sont des sortes de prismatoïdes semi-ouverts

$$x_1 \leq X_1 < y_1, \quad \dots, \quad x_n \leq X_n < y_n, \quad \dots$$

Et nous pourrions en déduire au moyen du théorème des probabilités totales la valeur de $p(e)$ pour les ensembles e qui peuvent se déduire de la famille de ces prismatoïdes par une suite dénombrable d'additions ou de soustractions d'ensembles. Si, par exemple, l'espace $\mathcal{S}$ considéré se réduisait à la ligne droite et par suite $F_n(x_1, x_2, \dots, X_n)$ à $F(x_1)$, la connaissance de $F(x_1)$ suffirait à connaître $p(e)$ pour tous les ensembles linéaires, e , mesurables au sens de Borel.

Remarque. — Nous avons ramené la connaissance d'une fonction de distribution $p(e)$, étendue à un espace distancié séparable quelconque $\mathcal{S}$ (espace appartenant ainsi à une famille d'espaces qui comprend, outre les espaces euclidiens à n dimensions et l'espace de Hilbert, tous les espaces que la pratique a conduit à considérer) à la connaissance d'une suite de fonctions d'un nombre fini croissant de variables. Mais la signification géométrique des coordonnées $X_1, X_2, \dots$ d'un élément X de $\mathcal{S}$ introduites en vue de cette réduction n'est pas très claire. On peut utiliser pour y remédier la forme donnée par Urysohn (puis complétée par Banach) à un théorème qui précise le théorème nous servant de point de départ. Ce dernier théorème peut s'exprimer en disant qu'il existe un espace distancié D_ω universel, c'est-à-dire indépendant de $\mathcal{S}$, tel que tout espace distancié séparable $\mathcal{S}$ soit isométrique à un sous-ensemble $\mathcal{S}_1$ de D_ω (c'est-à-dire qu'il existe une transformation ponctuelle biunivoque et conservant les distances, de $\mathcal{S}$ en $\mathcal{S}_1$). Urysohn a

montré qu'on pouvait remplacer D_ω par un espace universel non seulement distancié, mais lui-même séparable.

Enfin Banach a montré ⁽¹⁾ qu'on pouvait prendre pour cet espace universel, l'espace $\mathcal{C}$ dont chaque élément est une fonction numérique $f(x)$, continue sur $(0, 1)$ et où, par convention, on définit la limite d'une suite d'éléments $f_1, f_2, \dots, f_n, \dots$ de $\mathcal{C}$ comme la limite uniforme de $f_n(x)$ sur $(0, 1)$. Il suffit alors de prendre pour coordonnées $x_1, x_2, \dots$ de X de $\mathcal{S}$, des coordonnées de la fonction continue $f(x)$ correspondant à X . On peut, par exemple, en appelant $r_1, r_2, \dots$ la suite des nombres rationnels sur $(0, 1)$, poser

$$x_n = f(r_n).$$

Il est clair que si Y correspond à $g(x)$ et si

$$y_n = g(r_n),$$

on aura

$$(X, Y) = (f, g) = \max \text{ quand } n \text{ varie de } |f(r_n) - g(r_n)|,$$

d'où

$$(X, Y) = \max \text{ quand } n \text{ varie de } |x_n - y_n|.$$

On a la même égalité que plus haut, mais cette fois, nous savons que les x_n sont les valeurs d'une fonction continue aux points d'abscisse rationnelle et de même pour les y_n .

Les résultats rappelés ci-dessus pourraient être utilisés pour tenter de résoudre la difficulté soulevée page 224. Nous venons de voir qu'on peut toujours considérer un espace distancié séparable $\mathcal{S}$ comme isométrique d'un sous-ensemble $\mathcal{S}_1$ d'un espace *vectoriel* distancié. Il y aura, sans doute, des cas où en revenant de $\mathcal{S}_1$ à $\mathcal{S}$, on pourra faire correspondre aux vecteurs de $\mathcal{S}_1$ de nouveaux éléments considérés comme vecteurs de $\mathcal{S}$ et faire ainsi de $\mathcal{S}$ un espace *vectoriel* distancié séparable [de sorte qu'on pourra alors appliquer dans $\mathcal{S}$ notre première définition ⁽²⁾ de la moyenne, représentée ainsi par une intégrale]. Mais il ne sera pas toujours possible de compléter $\mathcal{S}$ ainsi. (Par exemple, il faut évidemment que $\mathcal{S}$ soit d'un seul tenant.) Toutefois, il vaudrait la peine de chercher si cela est possible pour les espaces $\mathcal{S}$ les plus simples pour lesquels la question se pose.

⁽¹⁾ Pages 187-188 de la *Théorie des opérations linéaires*, Tome I des *Monographies mathématiques*, Varsovie, 1932.

⁽²⁾ Voir note ⁽¹⁾, p. 283.

TABLE DES MATIÈRES.

PRÉFACE	p. 215
INTRODUCTION	p. 217
Éléments aléatoires nouveaux, p. 217. Leur étude simultanée, p. 218. Les éléments abstraits, p. 218. Loi de probabilité, p. 219. Détermination statistique des lois de probabilité des éléments aléatoires nouveaux, p. 221. Retour à la théorie. Moyennes, p. 223. Éléments typiques, p. 224. Espaces distanciés, p. 226. Dispersion, p. 228. Ordre, p. 228. Une inégalité utile, p. 229. Remarques, p. 229.	

CHAPITRE I.

Les positions typiques.

Première définition d'une position typique d'ordre k , p. 232. Position typique généralisée, p. 232. Deuxième définition, p. 234. Autre forme de la seconde définition, p. 236. Définition auxiliaire. p. 237.

LES POSITIONS TYPIQUES D'ÉLÉMENTS ALÉATOIRES X DE DIVERSES NATURES PARTICULIÈRES.

I. — *Cas où X est un nombre.*

Valeurs équiprobables, p. 238. Valeur moyenne, p. 238.

II. — *Cas de l'espace cartésien.*

Position moyenne. p. 241. Remarque, p. 243. Positions équiprobables, p. 243.

III. — *Cas où l'élément aléatoire est une fonction.*

Deux notations : $[\overline{X(t)}]$, $\overline{X}(t)$, p. 244. 1° Cas des fonctions continues, p. 244. Exemple particulier où les deux notations ne sont pas équivalentes, p. 244. 2° Cas des fonctions de carré intégrable, p. 247.

LES POSITIONS TYPIQUES DANS LES ESPACES D'ÉLÉMENTS ALÉATOIRES GLOBAUX.

Premiers espaces d'éléments aléatoires globaux, p. 252. Nouveaux espaces d'éléments aléatoires globaux, p. 254. Produit scalaire, p. 255. Coefficient de linéarité, p. 256. Cas de B_2 , p. 256. Nouvelles généralisations des positions typiques d'un élément aléatoire, p. 259.

CHAPITRE II.

Convergences stochastiques.

Notion générale, p. 259. Limite des positions typiques, p. 260. Convergence en moyenne d'ordre k , p. 261. Application, p. 261. Cas de l'espace cartésien, p. 261. Convergence en probabilité, p. 262. Remarque, p. 262. Cas de l'espace cartésien, p. 262. La convergence en probabilité déduite d'une distance, p. 263. La notion d'espace complet, p. 265. Condition de convergence en probabilité, p. 266. Condition de convergence en moyenne d'ordre k , p. 268. Égale sommabilité, p. 272. Applications, p. 273. Convergence légale, p. 275. Remarque, p. 275. La convergence en probabilité est légale, p. 275. Condition de convergence en probabilité de X_n vers X , p. 279. Condition de convergence en probabilité d'une suite, p. 282. Convergence presque certaine, p. 284. Cas de l'espace cartésien, p. 285. Formule de Kolmogoroff généralisée, p. 285. Application, p. 286. Intervention de la limite, p. 287. Autre forme de condition, p. 288. Condition de convergence presque certaine, p. 291. Remarque, p. 292.

CHAPITRE III.

Fonctions aléatoires de nature quelconque.

Fonctions aléatoires, p. 293. Convergences stochastiques, p. 293. Application, p. 293. Remarque, p. 293. Continuités stochastiques, p. 294. Fonction certaine associée à une fonction aléatoire, p. 295. Relations des continuités stochastiques avec la continuité certaine, p. 296. Continuité stochastique uniforme, p. 298. Convergence stochastique d'une suite d'éléments aléatoires, p. 299. Autre fonction associée, p. 300. Suite de fonctions aléatoires, p. 302. Autre forme de condition, p. 303. Continuité presque certaine, p. 303. Généralisation du paramètre, p. 305.

NOTE ADDITIONNELLE.

Coordonnées dans un espace distancié séparable, p. 306. Remarques, p. 307.

TABLE DES MATIÈRES.....	p. 309
-------------------------	--------