ANNALES DE L'I. H. P.

F.P. CANTELLI

Considérations sur la convergence dans le Calcul des probabilités

Annales de l'I. H. P., tome 5, n° 1 (1935), p. 3-50
<http://www.numdam.org/item?id=AIHP_1935__5_1_3_0>

Considérations
sur la convergence dans le Calcul des probabilités

PAR

F. P. CANTELLI

I

1. — Un théorème, récemment démontré (1), m'a suggéré quelques remarques qui formeront le sujet de ces conférences.

2. Sur la notion de « variable éventuelle »

Rappelons d'abord une définition de la notion de variable éventuelle que j'ai donnée en 1916 (2) et qui nous servira de base.

Une variable éventuelle X, à une dimension, est une grandeur qui peut prendre une valeur quelconque comprise, soit dans l'intervalle x_0, x_1, *soit à l'extérieur de cet intervalle* x_0, x_1, *suivant qu'il s'agit de l'un ou l'autre des deux événements*

$$E(x_0,\ x_1), \qquad \overline{E}(x_0,\ x_1). \tag{1}$$

qui s'excluent mutuellement (l'intervalle $(x_0,\ x_1)$ comprend ou ne comprend pas ses limites ; ces dernières x_0, x_1, ont des valeurs quelconques, pourvu qu'elles soient réelles et que $x_1 > x_0$).

(1) V. GLIVENKO, *Sulla determinazione empirica delle leggi di probabilità*. Giornale dell'Istitut. Italiano degli Attuari, Gennaio, 1933, p. 92.
(2) *La tendenza ad un limite nel senso del Calcolo delle probabilità*. Rendiconti del Circolo matematico di Palermo, Tomo XLI, 1916.

On admet que l'un de ces deux événements. se produit nécessairement de sorte que leurs *probabilités* respectives,

$$p(x_0, x_1), \qquad \bar{p}(x_0, x_1) \tag{2}$$

satisfont en tout cas à la condition

$$p(x_0, x_1) + \bar{p}(x_0, x_1) = 1. \tag{3}$$

Remarquons que d'après la définition précédente on a

$$\lim_{\substack{x_0 \to -\infty \\ x_1 \to \infty}} p(x_0, x_1) \leqslant 1. \tag{4}$$

Il est facile de donner des exemples pour lesquels cette relation soit vraie avec le signe $<$ ou avec le signe $=$. Soit X, par exemple, une variable éventuelle qui prend les valeurs $1, \frac{1}{2}, \frac{1}{3}, 0$, avec les probabilités $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$. On aura dans ce cas

$$\lim_{\substack{x_0 \to -\infty \\ x_1 \to \infty}} p(x_0, x_1) = 1. \tag{5}$$

Par contre, prenons

$$Y = \frac{1}{X}. \tag{6}$$

Y peut être considérée comme satisfaisant à la définition donnée ; elle aura cependant une singularité au point $x = 0$. Pour cette nouvelle variable, il est évident que

$$\lim_{\substack{y_0 \to -\infty \\ y_1 \to \infty}} p(y_0, y_1) = \frac{3}{4} < 1. \tag{7}$$

Pour plus de simplicité nous nous bornerons dans ce qui va suivre uniquement à des variables satisfaisant à l'égalité (4) : on pourrait les appeler des variables éventuelles à *loi de répartition régulière.*

En posant

$$\lim_{x_0 \to -\infty} p(x_0, x) = V(x) \tag{8}$$

pour la *loi de probabilité totale* relative à la variable que l'on considère, on aura évidemment

$$\lim_{x \to -\infty} V(x) = 0, \qquad \lim_{x \to +\infty} V(x) = 1. \tag{9}$$

Dans mon travail déjà cité, j'ai considéré, pour la démonstration de certains théorèmes, une fonction de plusieurs *variables éventuelles* $X_1, X_2, \ldots, X_k$,

$$f(X_1, X_2, \cdots, X_k). \tag{10}$$

D'après la définition donnée, f représentera une *variable éventuelle* à une dimension si, en tenant compte des probabilités des X_i, des conditions d'interdépendance de ces variables et aussi des principes des probabilités totales et composées, on peut l'exprimer comme une variable Y qui prend soit une valeur de l'intervalle (y_0, y_1) soit une valeur extérieure à cet intervalle, avec les probabilités respectives

$$p(y_0, y_1), \qquad \overline{p}(y_0, y_1), \tag{11}$$

satisfaisant à la condition

$$p(y_0, y_1) + \overline{p}(y_0, y_1) = 1. \tag{12}$$

Ces considérations justifient les définitions des variables éventuelles *équivalentes* et celle de *forme canonique* d'une variable éventuelle, questions dont je me suis occupées dans un travail publié (1) en 1916.

En effet, du point de vue du Calcul des probabilités, $f(X_1, X_2, \ldots, X_k)$ et Y représentent une seule et même variable sous deux formes différentes, que nous pouvons appeler *équivalentes* ; on peut même dire que Y représente la fonction $f(X_1, X_2, \ldots, X_k)$ écrite sous *forme canonique*. En même temps, nous voyons clairement pourquoi nous pouvons appeler *équivalentes* des variables éventuelles qui admettent la même *forme canonique*.

Dans ce qui suit, nous considérerons exclusivement des variables éventuelles

$$f(X_1, X_2, \cdots, X_k)$$

auxquelles correspondent, dans leur forme canonique, des *lois de répartition régulière* ; plus brièvement, on peut dire que (10) représente une variable éventuelle à *loi de répartition régulière*.

3. Sur la loi des probabilités totales d'une variable éventuelle.

D'après la définition d'une variable éventuelle, $V(x)$ est une fonc-

(1) *Sulla legge dei grandi numeri*, Reale Accademia dei Lincei (Memorie), Serie V, vol. XI, fascicolo VII. Roma, 1916.

tion positive, non décroissante de la variable x et *bien définie en chaque point* x. Ses discontinuités, qui ne peuvent être que des discontinuités de première espèce, appartiennent, à un ensemble dénombrable de points, qui a, par conséquent, une mesure nulle. Dans chaque point de continuité on a

$$(13) \qquad \lim_{\varepsilon_1 \to 0} V(x_1 - \varepsilon_1) = \lim_{\varepsilon_2 \to 0} V(x_1 + \varepsilon_2).$$

Soit maintenant x_1 un point de discontinuité. Pour pouvoir définir sans ambiguité une variable éventuelle, nous devons indiquer explicitement si $V(x_1)$ *comprend ou non* le point x_1 de l'intervalle $(-\infty, x_1)$. Dans l'affirmative, nous devons poser

$$(14) \qquad V(x_1) = \lim_{\varepsilon \to 0} V(x_1 + \varepsilon).$$

Si l'on ne faisait pas cette convention, on serait conduit à faire correspondre aux intervalles

$$(-\infty, x_1), \qquad (-\infty, x_1 + 0),$$

(qui doivent être considérés comme *identiques*) deux probabilités différentes $V(x_1)$ et $V(x_1 + 0)$, ce qui est inadmissible. Donc, avec la définition de la variable éventuelle que nous avons adoptée, nous devrons considérer $V(x)$ comme une fonction continue à droite dans ses points de discontinuité.

Les considérations précédentes ne contredisent en aucune manière les hypothèses de M. P. Lévy (1) qui attribue à $V(x)$ en un point de discontinuité x_1, la valeur :

$$\frac{V(x_1 - 0) + V(x_1 + 0)}{2}.$$

M. P. Lévy est guidé en cette circonstance par des considérations de commodité analytique et poursuit d'ailleurs un but différent de celui que je me suis proposé d'atteindre dans le présent travail.

Si le point x_1 était exclu, nous devrions poser pour la probabilité de la variable éventuelle dans l'intervalle $(-\infty, x_1)$:

$$(15) \qquad V(x_1 - 0) = \lim_{\varepsilon \to 0} V(x_1 - \varepsilon).$$

(1) *Calcul des probabilités*, Paris, Gauthier-Villars, 1925.

Dans un point de discontinuité de $V(x)$ le saut à droite est nul, tandis que le saut à gauche est représenté par

$$s_g = V(x_1) - V(x_1 - 0) = V(x_1 + 0) - V(x_1 - 0). \tag{16}$$

Considérons tous les sauts de $V(x)$ dans l'intervalle $(-\infty, x)$, le point x inclus ; on sait que leur somme

$$\varphi(x) = \sum_{-\infty < x_i \leqslant x} s_g(x_i) \tag{17}$$

constitue unse série convergente. On peut aisément en déduire le résultat bien connu, que la fonction

$$\psi(x) = V(x) - \varphi(x) \tag{18}$$

est une fonction continue à droite et à gauche en chaque point x, et, au surplus, non décroissante dans l'intervalle $(0, \infty)$.

Il suit, que nous pouvons écrire,

$$V(x) = \varphi(x) + \psi(x), \tag{19}$$

où $\varphi(x)$ est la fonction des sauts, positive et non décroissante et $\psi(x)$ une fonction continue, également positive et non décroissante.

4. Un théorème de convergence

On peut imaginer, une suite de variables éventuelles

$$X_1, X_2, \cdots, X_n, \cdots \tag{20}$$

telle que la probabilité $V_n(x)$, pour que X_n prenne une valeur de l'intervalle $(-\infty, x)$, tende vers la probabilité $V(x)$ pour qu'une variable donnée X prenne une valeur du même intervalle $(-\infty, x)$, quel que soit d'ailleurs *le point fixé x* lequel peut être compris ou non dans l'intervalle considéré.

La convergence de $V_n(x)$ vers $V(x)$ est-elle uniforme quelle que soit la valeur de x ? En d'autres termes, en désignant par

$$m_n \mid V_n(x) - V(x) \mid \tag{21}$$

le plus grand écart en valeur absolue entre $V_n(x)$ et $V(x)$ dans l'intervalle $(-\infty, +\infty)$, peut-on écrire

$$\lim_{n \to \infty} m_n \mid V_n(x) - V(x) \mid = 0\ ? \tag{22}$$

La réponse est affirmative. Il est utile d'en donner une démonstration détaillée qui permette de compléter en même temps d'autres démonstrations connues [1].

Je démontrerai que si $V_n(x)$ tend vers $V(x)$ *en chaque point donné x*, on peut déterminer un entier $\bar{n}$ tel que, pour tout $n > \bar{n}$ et *quelle que soit la valeur de x*, on ait

$$|V_n(x) - V(x)| < \eta, \tag{23}$$

η étant un nombre positif, donné à l'avance, et aussi petit qu'on le voudra.

On connaît la manière de ranger les sauts de la fonction $V(x)$ dans l'intervalle $(-\infty, +\infty)$ par ordre de grandeur décroissante et on sait que leur somme forme une série convergente. Si ω_i désigne le saut de rang i, on aura la série

$$\omega_1 + \omega_2 + \cdots + \omega_k + \omega_{k+1} + \cdots \tag{24}$$

et l'on pourra déterminer un entier k tel que

$$\omega_{k+1} + \omega_{k+2} + \cdots < \frac{\eta}{4}. \tag{25}$$

Au moyen de r points $(r \geqslant k)$ divisons l'intervalle $(-\infty, +\infty)$ en intervalles partiels :

$$\begin{aligned}(-\infty, \nu_1 + 0), (\nu_1 + 0, \nu_2 - 0), (\nu_2 - 0, \nu_2 + 0), (\nu_2 + 0, \nu_3 - 0), \cdots \\ \cdots, (\nu_r - 0, \nu_r + 0), (\nu_r + 0, \infty),\end{aligned} \tag{26}$$

dans le but d'analyser le comportement de la fonction $V(x)$ aux extrémités de chacun de ces intervalles.

Les points $\nu_1, \nu_2, \ldots, \nu_r$, $(\nu_r \geqslant k)$ sont choisis de manière à satisfaire aux conditions suivantes :

a) les inégalités suivantes doivent être satisfaites

$$\begin{aligned} V(\nu_1 + 0) &< \frac{\eta}{2}, \\ 1 - V(\nu_r + 0) &< \frac{\eta}{2}; \end{aligned} \tag{27}$$

b) les k sauts $\omega_1, \omega_2, \ldots, \omega_k (r > k)$ doivent correspondre à des points de la suite $\nu_1, \nu_2, \ldots, \nu_r$;

(1) Cf. p. ex. P. Lévy, *loc. cit.* (4), p. 192.

c) dans chaque intervalle du type

$$(\nu_i + 0,\ \nu_{i+1} - 0) \tag{28}$$

la variation de $V(x)$ doit être inférieure à $\eta/2$, c'est-à-dire

$$V(\nu_{i+1} - 0) - V(\nu_i + 0) < \frac{\eta}{2}; \tag{29}$$

ce qu'on peut obtenir aisément en augmentant convenablement r, car dans chacun de ces intervalles la somme des sauts est certainement inférieure à $\eta/4$. Dans l'un des intervalles (28), la variation de $V(x)$ pourrait avoir une valeur supérieure à $\eta/4$ par suite de la variation de la fonction *continue* $\psi(x)$, qui, à son tour, contribue à former $V(x)$, mais cette dernière variation peut être rendue aussi petite que l'on veut, dans tous les intervalles $(\nu_i + 0,\ \nu_{i+1} - 0)$, en augmentant convenablement r.

De ce qui précède, et en tenant compte de l'hypothèse que la suite $V_n(x)$ tend en chaque point vers $V(x)$, on peut déduire qu'il est possible de déterminer un entier $\bar{n}$ tel que, pour tout $n > \bar{n}$, on ait

$$| V_n(x_i) - V(x_i) | < \frac{\eta}{2} \tag{30}$$

pour tous les points x_i qui coïncident avec les extrémités des intervalles (26) considérés ci-dessus. Il s'ensuit alors aisément que, pour les mêmes valeurs de $\bar{n}$ et de η et *pour tout point* x de l'intervalle $(-\infty, +\infty)$, l'inégalité (23) sera également vraie.

Il suffira, évidemment, de le démontrer pour des valeurs de x comprises dans un intervalle du type

$$(\nu_i + 0,\ \nu_{i+1} - 0). \tag{31}$$

Par rapport à un de ces intervalles, on a, en effet,

$$V(\nu_{i+1} - 0) - V(\nu_i + 0) < \frac{\eta}{2}, \tag{32}$$

$$| V_n(\nu_{i+1} - 0) - V(\nu_{i+1} - 0) | < \frac{\eta}{2}, \tag{33}$$

$$| V_n(\nu_i + 0) - V(\nu_i + 0) | < \frac{\eta}{2}. \tag{34}$$

Ces inégalités, jointes aux deux suivantes, variables en chaque point x de l'intervalle (31)

$$V_n(\nu_{i+1} - 0) \geqslant V_n(x) \geqslant V_n(\nu_i + 0) \tag{35}$$

$$V(\nu_{i+1} - 0) \geqslant V(x) \geqslant V(\nu_i + 0) \tag{36}$$

permettent de démontrer aisément le résultat annoncé.

5. Sur un théorème fondamental de statistique mathématique

Considérons une variable éventuelle X, avec une loi de probabilités totales $V(x)$.

Effectuons n épreuves sur la variable X et considérons le nombre de fois $V_n(x)$ où une valeur non supérieure à x se présente dans ces n épreuves. La fréquence relative qui correspond à la probabilité $V(x)$ est donnée par

$$V_n(x) = \frac{\nu_n(x)}{n}. \tag{37}$$

Si x est un point de discontinuité de $V(x)$, et si l'on veut inclure ce point, il faudra prendre pour $V(x)$, la valeur $V(x + o)$; au contraire, si on veut l'exclure, on devra écrire $V(x - o)$.

$V_n(x)$ peut être considérée comme une loi de probabilités totales relative à une variable éventuelle X_n.

Si nous considérons $V(x)$ comme une probabilité relative à un événement E_x, $V_n(x)$ représentera la fréquence avec laquelle cet événement se produit au cours de n épreuves. On peut alors, d'une manière précise, déduire d'un cas particulier de la loi *uniforme* des grands nombres (1), le résultat suivant : étant donnés deux nombres positifs ε_1, ε_2 arbitrairement petits, on peut déterminer un entier $\bar{n}$ tel que, pour tout $n > \bar{n}$, et pour chaque *point donné x*, on ait

$$| V_n(x) - V(x) | < \varepsilon_1, \tag{38}$$

avec une probabilité supérieure à $1 - \varepsilon_2$, et cela que le point x soit un point de continuité ou non ; si x est un point de discontinuité, il faut préciser s'il est exclu ou non.

(1) F. P. CANTELLI, *Sulla probabilità come limite della frequenza*. Rendiconti della R. Accademia dei Lincei, vol. XXVI, serie V, gennaio 1917, p. 39-45.

On peut facilement démontrer que $V_n(x)$ tend vers $V(x)$ uniformément, pour tout point x de l'intervalle $(-\infty, +\infty)$.

Divisons d'abord l'intervalle $(-\infty, +\infty)$ en intervalles du type (26), ainsi que nous l'avons indiqué. En appliquant un théorème bien connu de BOOLE, on parvient immédiatement à la conclusion suivante : ε, η étant arbitrairement petits, on peut déterminer un nombre $\bar{n}$ tel que, simultanément, pour tout $n > \bar{n}$ et pour les extrémités des intervalles considérés, on ait

$$| V_n(x) - V(x) | < \frac{\eta}{2} \tag{39}$$

avec une probabilité supérieure à $1 - \varepsilon$. Il s'ensuit, d'après ce que nous avons dit précédemment, qu'on peut même affirmer avec une probabilité supérieure à $1 - \varepsilon$, qu'il existe un entier $\bar{n}$ tel que, simultanément, pour tout $n > \bar{n}$ et pour toute valeur de x de l'intervalle $(-\infty, +\infty)$, on ait

$$| V_n(x) - V(x) | < \eta. \tag{40}$$

Considérons maintenant une suite de nombres positifs ε_i, tendant vers zéro et tels que la série

$$\alpha_n = \sum_{i=n}^{\infty} \varepsilon_i \tag{41}$$

soit convergente ; considérons encore une autre suite de nombres positifs η_i qui tendent également vers zéro. A chaque valeur de i on pourra faire correspondre un entier $\bar{n}_i$ tel que, simultanément, pour tout $n > \bar{n}_i$ et pour toutes les valeurs de x de l'intervalle $(-\infty, +\infty)$, on ait

$$| V_n(x) - V(x) | < \eta_i. \tag{42}$$

avec une probabilité supérieure à $1 - \varepsilon_i$.

En appliquant alors une extension du théorème de BOOLE (1), on déduit que $V_n(x)$ tend uniformément vers $V(x)$, et satisfait aux conditions (42) $(i = n, n + 1, \ldots)$, avec une probabilité supérieure à

$$1 - \alpha_n. \tag{43}$$

(1) Cf. *loc. cit.* (6) et F. P. CANTELLI, *Su due applicazioni di un teorema di G. Boole alla Statistica matematica*. Rendiconti della R. Accademia dei Lincei, vol. XXVI, série V, 1917, p. 295-302.

En employant la notation (21) nous pourrons dire plus brièvement que la probabilité pour que l'on ait

$$m_n \mid V_n(x) - V(x) \mid \to 0 \tag{44}$$

diffère de l'unité aussi peu qu'on le voudra.

C'est à cette même conclusion qu'est parvenu M. GLIVENKO en utilisant une méthode différente de celle qui précède, mais toujours basée sur la loi uniforme des grands nombres.

6. Extension du théorème précédent à la valeur moyenne d'une variable éventuelle.

Soit

$$f(X_1, X_2, \cdots, X_k) \tag{45}$$

une variable éventuelle. D'après la définition, la probabilité pour que f prenne une valeur inférieure ou égale à y sera donnée par une fonction positive non décroissante

$$V(y).$$

La valeur moyenne de la variable $f(X_1, X_2, \ldots, X_k)$, correspondant à l'ensemble des valeurs de f comprises dans l'intervalle $(-\infty, y)$, sera donnée par l'intégrale

$$\mathcal{M}(f \leqslant y) = \int_{-\infty}^{y} z \cdot dV(z). \tag{46}$$

A cette valeur moyenne théorique, qui existe par hypothèse, correspond une valeur moyenne empirique : on l'obtient en faisant la somme des valeurs inférieures ou égales à y qui se seront présentées au cours de n épreuves concernant la valeur (45) et en divisant cette somme par n. Désignons par $\mathcal{M}_n(f \leqslant y)$ la moyenne empirique correspondant à la valeur théorique (46), de manière qu'on puisse écrire

$$\mathcal{M}_n(f \leq y) = \int_{-\infty}^{y} z \cdot dV_n(z). \tag{47}$$

Les discontinuités de la fonction $\mathcal{M}(f \leqslant y)$ coïncident avec les discontinuités de $V(z)$; $\mathcal{M}(f \leqslant y)$ doit être considérée comme continue à droite dans ses points de discontinuité, qui sont tous de première espèce.

A chaque ε, arbitrairement petit, on pourra faire correspondre un nombre positif α tel que :

$$(48) \qquad \left| \int_{-\infty}^{-\alpha} z \cdot dV(z) \right| + \int_{\alpha}^{\infty} z \cdot dV(z) < \varepsilon.$$

Désignons par

$$(50) \qquad m_n \mid \mathcal{M}_n(f \leq y) - \mathcal{M}(f \leq y) \mid$$

la plus grande différence en valeur absolue entre $\mathcal{M}_n$ et $\mathcal{M}$, pour les valeurs de y comprises dans l'intervalle $(-\infty, +\infty)$.

En appliquant la loi uniforme des grands nombres, relative aux moyennes des variables éventuelles, on peut démontrer, pour des conditions suffisamment étendues, que la probabilité pour que l'on ait

$$(51) \qquad \lim_{n \to \infty} m_n \mid \mathcal{M}_n(f \leq y) - \mathcal{M}(f \leq y) \mid = 0$$

peut se rapprocher de l'unité autant qu'on le voudra.

Pour abréger, je ne donnerai pas la démonstration du théorème précédent.

II

7. Quelques remarques critiques sur les considérations précédentes.

Il est naturel de se demander maintenant si les considérations des deux paragraphes précédents qui font appel à la loi uniforme des grands nombres, ne peuvent pas être simplifiées. En effet, un grand nombre d'entre elles seraient parfaitement superflues si la loi uniforme des grands nombres pouvait s'écrire

$$(54) \qquad \lim_{n \to \infty} V_n(z) = V(z)$$

au sens ordinaire de l'analyse, ou bien si l'on pouvait déduire directement la relation (54) du Calcul classique des probabilités.

On est donc naturellement conduit à poser les questions suivantes :

a) Considérons un événement de probabilité constante p ; supposons qu'on puisse effectuer une infinité d'épreuves successives sur cet événement et considérons ses fréquences relatives d'apparition

$$\frac{\nu_n}{n}, \qquad n = 1, 2, 3, \ldots$$

Cela étant, *la loi uniforme des grands nombres nous autorise-t-elle à écrire*

$$\lim_{n \to \infty} \frac{\nu_n}{n} = p, \tag{55}$$

au sens ordinaire de l'analyse ?

b) Peut-on justifier l'égalité (55), en se basant uniquement sur le Calcul classique des probabilités, ou sur une théorie abstraite des probabilités, *fondée également sur les principes des probabilités totales et des probabilités composées ?*

Ces questions nous conduisent à examiner à nouveau certains de nos travaux déjà anciens et auxquels se rattachent des considérations très intéressantes et plus récentes dues à M. von MISES.

Dans le but de justifier certaines opinions que j'ai soutenues dans mes précédents travaux, je désire tout d'abord mettre en évidence les raisons pour lesquelles, dès les années 1917-1918, j'ai donné une réponse négative aux deux questions indiquées ci-dessus. Je n'exclus en aucune manière la possibilité de réfuter les arguments que je donne, mais je dois avouer que je n'en ai pas encore trouvé le moyen. Ce fait constitue, à mon avis, une raison de plus pour que j'en parle, car si l'on pouvait mettre mon argumentation en échec on serait à même, je le crois, de faire faire d'importants progrès au Calcul des probabilités.

Je commencerai en rappelant les fondements d'une théorie abstraite du Calcul des probabilités dont je me suis occupé récemment [1]. Je mettrai ensuite en évidence le passage de cette théorie abstraite au Calcul des probabilités classique par l'intermédiaire d'un postulat empirique et j'ajouterai quelques considérations critiques à ce sujet. Tout cela justifiera, dans son ensemble, ma réponse négative aux deux questions indiquées plus haut.

(1) *Una teoria astratta del calcolo delle probabilità*, Giornale dell'Istituto Italiano degli Attuari, Anno III, nº 2. Roma, 1932.

8. Sur une théorie abstraite du Calcul des probabilités.

Pour qu'une « théorie abstraite du Calcul des probabilités », mérite réellement son nom, elle ne doit faire appel à aucune notion ayant un caractère empirique : elle ne doit donc pas utiliser les notions de cas possibles, d'événements, de probabilités, de dépendances ou d'indépendances d'événements, etc.

Pour réaliser une telle théorie, partons d'un ensemble mesurable de points, ayant mesure finie et non nulle que nous prendrons comme unité. L'ensemble considéré pourrait être quelconque ; mais pour plus de *simplicité*, je le supposerai formé par les points d'une surface de mesure égale à l'unité (surface de base).

J'appellerai *aires partielles* des aires *d'un seul tenant ou non*, formées par des points qui appartiennent à la surface considérée. Je désigne par E une aire partielle et par $\overline{E}$ l'aire complémentaire.

Soient E_1 et E_2 deux aires partielles ; $(E_1 + E_2)$ désignera l'ensemble des points qui appartiennent au moins à l'une des deux aires E_1, E_2 ; $(E_1 E_2)$, l'ensemble des points communs à ces deux aires ; enfin $m(E)$ signifie la mesure de l'ensemble E.

Deux aires partielles E_1 et E_2 peuvent avoir des *aires communes*. On a alors, ainsi qu'il est bien connu :

$$0 \leq m(E) \leq 1 \tag{56}$$

$$m(E) + m(\overline{E}) = 1 \tag{57}$$

$$m(E_1 + E_2) = m(E_1) + m(E_2) - m(E_1E_2), \tag{58}$$

$$0 \leq \frac{m(E_1E_2)}{m(E_i)} \leq 1, \qquad i = 1,2. \tag{59}$$

Dans le cas où les deux aires partielles E_1, E_2 n'ont pas d'aire commune, le dernier terme du second membre disparaît dans (58).

Considérons *n aires partielles*

$$E_1, E_2, \cdots, E_n; \tag{60}$$

on a

$$m(E_1E_2 \cdots E_n) + m(\overline{E}_1 + \overline{E}_2 + \cdots + \overline{E}_n) = 1. \tag{61}$$

Puisque

$$m(\overline{E}_1 + \overline{E}_2 + \cdots + \overline{E}_n) \leq m(\overline{E}_1) + \cdots + m(\overline{E}_n) \tag{62}$$

on aura aussi

$$(63)\qquad m(E_1 E_2 \cdots E_n) \geqslant 1 - \sum_{i=1}^{n} m(\overline{E}_i).$$

Cette formule correspond à une relation bien connue de BOOLE que j'ai mise en évidence en 1917 (1).

Je considère maintenant une suite illimitée d'aires partielles

$$(64)\qquad E_n,\ E_{n+1}, \cdots,\ E_{n+r}, \cdots;$$

je peux écrire

$$(65)\qquad m(E_n E_{n+1} \cdots E_{n+r}) \geqslant 1 - \sum_{i=n}^{n+r} m(\overline{E}_i).$$

Lorsque r augmente indéfiniment, le premier membre de (65) admet une limite et, si la série qui en résulte au second membre est convergente, on pourra écrire (1)

$$(66)\qquad m(E_n E_{n+1} \cdots) \geqslant 1 - \sum_{i=n}^{\infty} m(\overline{E}_i).$$

Cette inégalité constitue une extension du théorème de BOOLE.

Introduisons maintenant la notion d'*aires multipliables*. Je dirai que deux *aires partielles* E_1, E_2 sont *multipliables* si elles sont disposées de manière que l'on ait

$$(67)\qquad m(E_1 E_2) = m(E_1) \times m(E_2).$$

En remarquant que

$$(68)\qquad m(\overline{E}_1 E_2) + m(E_1 E_2) = m(E_2)$$

et en tenant compte de (67), on peut écrire

$$(69)\qquad m(\overline{E}_1 E_2) = [1 - m(E_1)] m(E_2) = m(\overline{E}_1) \times m(E_2)\,;$$

enfin de (67) et (68) on tire

$$(70)\qquad \frac{m(\overline{E}_1 E_2)}{m(\overline{E}_1)} = \frac{m(E_1 E_2)}{m(E_1)} = m(E_2).$$

(1) Cf. *loc. cit.* (7).
(2) Cf. *loc. cit.* (7).

Les *n aires partielles* (60) peuvent être multipliables deux à deux, trois à trois, etc., ou enfin *multipliables entre elles* : dans ce dernier cas on a, par définition :

$$m(E_1 E_2 \cdots E_n) = m(E_1) \cdot m(E_2) \cdots m(E_n). \tag{71}$$

On pourra dire que n aires partielles sont *complètement multipliables* si elles sont multipliables deux à deux, trois par trois, etc.

Si n aires partielles sont multipliables entre elles, on peut écrire

$$m(E_1 \cdot E_2 \cdots E_n) = \prod_{i=1}^{n} [1 - m(\overline{E}_i)] \tag{72}$$

et si la série

$$\sum_{i=1}^{\infty} m(\overline{E}_i)$$

est divergente, on a

$$m(E_1 E_2 \cdots) = 0, \tag{73}$$

et, par une extension de (61),

$$m(\overline{E}_1 + \overline{E}_2 + \cdots) = 1. \tag{74}$$

Les égalités correspondant à (73) et à (74) ont été données, la première fois (1), par M. BOREL.

Passons à un autre ordre d'idées.

Je crois que la définition de la *variable éventuelle* que j'ai donnée en 1916, et dont j'ai parlé précédemment, est suffisamment précise.

Introduisons maintenant la notion de *variable pondérée*, en indiquant quelle est sa signification et en examinant un certain nombre d'opérations que l'on peut effectuer sur ces variables ; ensuite, nous envisagerons les généralisations possibles de cette notion.

J'imagine notre *surface de base, totalement couverte* par les h aires partielles

$$E_1, E_2, \cdots, E_h, \tag{75}$$

qui n'ont pas de parties communes ; on a donc

$$\sum_{i=1}^{h} m(E_i) = 1. \tag{76}$$

(1) *Sur les probabilités dénombrables et leurs applications arithmétiques.* Rendiconti di Circolo Matematico di Palermo, t. XXVII (1er semestre 1909).

J'appelle *variable pondérée* la fonction $X = f(x)$ qui prend la valeur x_i, $(i = 1, 2, \ldots, h)$, en chaque point de l'aire partielle E_i.

D'une façon analogue, j'imagine la même surface totalement recouverte par k autres aires partielles *distinctes*

$$E_1', E_2', \cdots, E_k', \tag{77}$$

pour lesquelles on ait

$$\sum_{j=1}^{k} m(E_j') = 1. \tag{78}$$

Soit alors $Y = \varphi(y)$ la variable pondérée qui prend la valeur y_j en chaque point de l'aire partielle $E_j (j = 1, 2, \ldots, k)$.

On voit immédiatement qu'on a les relations très importantes suivantes [1] :

$$\begin{cases} m(E_i) = \sum_{j=1}^{k} m(E_i E_j'), & i = 1, 2, \cdots, h \\ m(E_j') = \sum_{i=1}^{h} m(E_i E_j'), & j = 1, 2, \cdots, k. \end{cases} \tag{79}$$

A chaque point de l'ensemble $(E_i\ E'_j)$ correspond maintenant la valeur x_i de X et la valeur y_i de Y ; f désignant le symbole « fonction », à chaque point de l'ensemble précédent on fait donc correspondre la valeur $f(x_i, y_j)$.

Il est clair que $f(x_i, y_j)$ constitue encore une *variable pondérée* sur la surface de base. En effet, toutes les aires partielles

$$(E_i E_j'), \qquad (i = 1, 2, \cdots, h\,;\ j = 1, 2, \cdots, k) \tag{80}$$

n'ont pas de parties communes et remplissent toutes la surface de base, comme on peut aisément le démontrer. En outre, à chaque point d'une aire partielle (80) correspond une valeur bien définie $f(x_i, y_j)$.

La fonction $f(x_i, y_j)$ constitue donc une variable pondérée, fonction des deux variables pondérées X, Y. En particulier, si l'on se donne

(1) Cf. F. P. CANTELLI, *Elementi di Matematica attuariale.* Sinossi della Cultura Universale e Pratica, vol. II, Società Editrice libraria, Milano, 1909.

deux variables pondérées sur une surface de base on pourra considérer la variable pondérée somme, produit, etc., sur la même surface.

J'appelle *valeur moyenne d'une variable pondérée,* la somme des produits des valeurs prises par la variable en chaque point des aires partielles considérées, par les mesures des aires correspondantes. Si $\mathcal{M}$ est le symbole de la valeur moyenne, j'écris

$$\mathcal{M}(X) = \sum_{i=1}^{h} x_i \cdot m(E_i). \tag{81}$$

Avec la définition de la variable pondérée, si l'on pose

$$\mathcal{M}(X) = m, \tag{82}$$

on déduit facilement que

$$\mathcal{M}(X - m) = 0. \tag{83}$$

Si l'on a plusieurs variables pondérées sur la même surface de base

$$X_1, \quad X_2, \cdots, X_n \tag{84}$$

on pourra définir, comme plus haut, les variables pondérées somme $X_1 + X_2$, et, par conséquent $X_1 + X_2 + X_3, \ldots$; $X_1 + X_2 + \ldots + X_n$.

D'une manière analogue on peut définir les variables pondérées produits X_1X_2 ; $X_1X_2X_3, \ldots$; $X_1X_2 \ldots X_n$.

En tenant compte des relations (79) on déduit

$$\mathcal{M}(X_1 + X_2) = \mathcal{M}(X_1) + \mathcal{M}(X_2) \tag{85}$$

et plus généralement

$$\mathcal{M}(X_1 + X_2 + \cdots + X_n) = \mathcal{M}(X_1) + \mathcal{M}(X_2) + \cdots + \mathcal{M}(X_n). \tag{86}$$

Si, en outre, les aires partielles (80) sont multipliables, on déduit très facilement que

$$\mathcal{M}(X_1X_2) = \mathcal{M}(X_1) \cdot \mathcal{M}(X_2) ; \tag{87}$$

nous pouvons dire, dans ce cas, que les deux variables pondérées X_1 et X_2 sont *multipliables.*

Si chacune des variables (84) est multipliable par le produit des variables pondérées précédentes, nous pourrons écrire plus généralement

(88) $$\mathcal{M}(X_1 X_2 \cdots X_n) = \mathcal{M}(X_1) \cdot \mathcal{M}(X_2) \cdots \mathcal{M}(X_n).$$

Considérons maintenant les variables pondérées

(89) $$X_1, X_2, \cdots, X_n,$$

qui appartiennent à la même surface de base, et dont les valeurs moyennes soient respectivement

(90) $$m_1, m_2, \cdots, m_n.$$

Les variables suivantes

(91) $$X_1 - m_1, \qquad X_2 - m_2 . \cdots, X_n - m_n$$

représenteront naturellement des variables pondérées qui appartiennent à la même surface de base et qui, en outre, seront multipliables si les variables (89) elles-mêmes le sont. En considérant alors la nouvelle variable pondérée

(92) $$X_{(n)} - \sum_{i=1}^{n} m_i = \sum_{i=1}^{n} (X_i - m_i),$$

on aura

(93) $$\mathcal{M}\left[X_{(n)} - \sum_{i=1}^{n} m_i\right] = \sum_{i=1}^{n} \mathcal{M}(X_i - m_i) = 0.$$

En considérant encore une autre variable pondérée

(94) $$\left\{\begin{aligned} &\left[X_{(n)} - \sum_{i=1}^{n} m_i\right]^2 = \left[\sum_{i=1}^{n} (X_i - m_i)\right]^2 = \\ &= \sum_{i=1}^{n} (X_i - m_i)^2 + 2\sum_{r,s} (X_r - m_r)(X_s - m_s) \end{aligned}\right.$$

et en supposant que les variables (89) soient multipliables, on déduit

(95) $$\mathcal{M}\left[X_{(n)} - \sum_{i=1}^{n} m_i\right]^2 = \sum_{i=1}^{n} \mathcal{M}(X_i - m_i)^2.$$

Les considérations précédentes sont fondées sur des résultats bien connus, concernant à la mesure des ensembles et les fonctions mesurables (BOREL, LEBESGUE) et tiennent également compte des travaux de MM. DE LA VALLÉE-POUSSIN, FRÉCHET, LÉVY.

Cependant, même en tenant compte d'un autre travail très intéressant de M. F. M. URBAN, je ne crois pas que nous soyons aujourd'hui en mesure d'affirmer que nous avons réussi de développer, d'une façon systématique, une théorie abstraite des variables pondérées, théorie qui fasse abstraction de toute notion empirique et qui possède les caractéristiques que nous avons indiquées au début.

9. Sur quelques théorèmes fondamentaux de la théorie des variables pondérées

Soient E_1, $\overline{E}_1$ deux aires partielles complémentaires, données sur la surface de base, et soient p, q, $(q = 1 - p)$ leurs mesures respectives.

Désignons par X_1 la variable pondérée qui prend la valeur 1 dans l'aire partielle E_1 et la valeur *zéro* dans l'aire partielle $\overline{E}_1$. D'une manière analogue, soient E_2, $\overline{E}_2$ deux autres aires partielles complémentaires, de mesures p et q, et considérons la variable pondérée X_2, qui prend la valeur 1 dans l'aire partielle E_2 et la valeur *zéro* dans l'aire partielle $\overline{E}_2$. *Je fixe* les aires E_2, $\overline{E}_2$ de façon que chacune d'elles soit multipliable par chacune des aires E_1, $\overline{E}_1$; les deux variables X_1 et X_2 seront donc multipliables.

On peut considérer alors sur la surface de base la variable pondérée

$$\frac{\nu_2}{2} = \frac{X_1 + X_2}{2} \tag{96}$$

qui prend les valeurs 1, $\frac{1}{2}$, $\frac{1}{2}$, 0 respectivement, dans les aires partielles

$$(E_1E_2), \qquad (E_1\overline{E}_2), \qquad (\overline{E}_1E_2), \qquad (\overline{E}_1\overline{E}_2) \tag{97}$$

dont les mesures sont respectivement

$$p^2, \qquad pq, \qquad qp, \qquad q^2. \tag{98}$$

(1) *Grundlagen der Wahrscheinlichkeitsrechnung und der Theorie der Beobachtungsfehler*, B. G. Teubner, Berlin, 1923.

D'une manière analogue, on peut considérer les variables X_3, X_4,..., X_n, qui ont la valeur 1 dans les aires partielles E_3, E_4,..., E_n, (appartenant toutes à la surface de base considérée et ayant toutes la mesure p), et la valeur zéro dans les aires partielles $\overline{E}_2$, $\overline{E}_2$,..., $\overline{E}_n$ complémentaires à E_3, E_4,..., E_n et toutes de mesure $q = 1 - p$. Les aires partielles E_3, $\overline{E}_3$ sont fixées de manière qu'elles soient multipliables par les aires partielles (97) ; les aires E_4, $\overline{E}_4$ sont déterminées de façon qu'elles soient, à leur tour, multipliables par chacune de celles qu'on a obtenues précédemment, etc. En d'autres termes, chacune des variables de la suite

$$X_1, \quad X_2, \cdots, X_n \tag{99}$$

est multipliable par le produit des variables qui la précèdent. En posant

$$\frac{\nu_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

on forme ainsi une suite de variables pondérées sur la surface de base. Une feuille de papier, de surface unité se trouvera ainsi divisée en 2^n aires partielles. La variable pondérée prendra 2^n valeurs (dont quelques-unes coïncideront), correspondant aux 2^n aires partielles envisagées.

Il n'y a aucune difficulté à établir un *procédé uniforme* de division de la surface de base pour pouvoir passer de la variable $\frac{\nu_n}{n}$ à la variable suivante $\frac{\nu_{n+1}}{n+1}$; on peut réaliser matériellement la division de la feuille de papier considérée jusqu'à un rang n qui peut être assez élevé. Si on réalise effectivement cette division on constatera que les régions dans lesquelles $\frac{\nu_n}{n}$ prend des valeurs voisines de p, occuperont des portions de plus en plus grandes de la surface de base, au fur et à mesure que n augmentera.

Naturellement, on ne peut pas continuer à subdiviser la surface de base à l'infini. Mais, ayant fixé le procédé de passage de la variable pondérée $\frac{\nu_n}{n}$ à la suivante $\frac{\nu_{n+1}}{n+1}$, nous pouvons bien nous demander à quel résultat l'on parviendrait si n augmentait au delà de toute limite.

En nous appuyant sur les considérations précédentes, et en ayant présente à l'esprit la démonstration du célèbre théorème de BERNOULLI, on comprend facilement qu'on puisse arriver au résultat suivant :

ε étant un nombre positif donné aussi petit qu'on le voudra, la mesure totale des aires dans lesquelles $\frac{\nu_n}{n}$ prend des valeurs comprises dans l'intervalle $p - \varepsilon$, $p + \varepsilon$ tend vers l'unité lorsque n augmente indéfiniment.

Rappelons-nous bien que cette mesure *tend* vers l'unité, mais qu'elle n'est jamais égale à l'unité. En effet, quel que soit n, la variable ν_n prend sur la surface de base toutes les valeurs $0, 1, 2, \ldots, n$ et la variable $\frac{\nu_n}{n}$ toutes les valeurs $0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$. Si grand que soit n il restera toujours une portion de la surface de base, de mesure *non nulle*, sur laquelle la variable $\frac{\nu_n}{n}$ prendra des valeurs différant de p par une quantité plus grande que ε.

Je passe maintenant à un autre théorème, qui correspond à un cas particulier de la loi uniforme des grands nombres.

Considérons sur la surface de base les *variables pondérées*

$$\frac{\nu_n}{n}, \quad \frac{\nu_{n+1}}{n+1}, \ldots \tag{100}$$

Considérons aussi la suite illimitée des ensembles de points I_n, $I_{n+1}, \ldots$ pour lesquels les inégalités suivantes

$$\left\{\begin{array}{l} \left|\frac{\nu_n}{n} - p\right| < \varepsilon_n, \\ \left|\frac{\nu_{n+1}}{n+1} - p\right| < \varepsilon_{n+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array}\right. \tag{101}$$

s ont satisfaites, les ε_i $(i = n, n + 1, \ldots)$ étant une suite de nombres positifs tendant vers zéro. Cela étant, on démontre que la suite ε_i peut être donnée de façon que la mesure de l'ensemble commun aux ensembles I_n, $I_{n+1}, \ldots$ *diffère de l'unité d'une quantité aussi petite qu'on le voudra, pourvu que n soit suffisamment grand.* En d'autres termes, on peut trouver une suite ε_i, $(i = n, n + 1,...)$, dont les éléments tendent vers zéro, telle que

$$m(I_n I_{n+1} I_{n+2} \cdots) = 1 - \eta(n), \tag{102}$$

avec

$$\lim_{n \to \infty} \eta(n) = 0. \tag{103}$$

Bornons-nous à un cas particulier du théorème précédent.

Soit ε un nombre donné, positif et indépendant de n ; on peut alors choisir n tel que $\varepsilon \geqslant \varepsilon_{n+r}$, $(r = 0, 1, 2,...)$. Il s'ensuit qu'en désignant par I'_n, I'_{n+1},... les ensembles pour lesquels les inégalités suivantes

$$\left\{\begin{array}{l} \left|\frac{\nu_n}{n} - p\right| < \varepsilon, \\ \left|\frac{\nu_{n+1}}{n+1} - p\right| < \varepsilon, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right. \tag{104}$$

sont satisfaites, on doit avoir

$$m(I'_n I'_{n+1} \cdots) > m(I_n I_{n+1} \cdots) = 1 - \eta(n). \tag{105}$$

Un cas particulier est le théorème correspondant à celui de Bernoulli puisque l'on a

$$m(I'_n) > m(I'_n I'_{n+1} \cdots) = 1 - \eta(n). \tag{106}$$

Faisons une remarque au sujet du théorème correspondant à (104) et (105), qui nous sera utile par la suite.

Je veux déterminer la mesure de l'ensemble des points de la surface de base pour lesquels, pour la *première fois*, et à partir du rang $n + r$, les inégalités suivantes, en nombre infini,

$$\left\{\begin{array}{l} \left|\frac{\nu_{n+r}}{n+r} - p\right| < \varepsilon, \\ \left|\frac{\nu_{n+r+1}}{n+r+1} - p\right| < \varepsilon, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right. \tag{107}$$

sont satisfaites. On peut raisonner de la manière suivante.

La mesure de l'ensemble commun

$$(108) \qquad m(I'_{n+i} \cdot I'_{n+i+1} \cdots),$$

ne décroît évidemment pas lorsque i augmente.

La mesure de l'ensemble $(I'_{n+r-1}\ I'_{n+r}\ldots)$ correspond à l'ensemble des points pour lesquels on a simultanément

$$(109) \qquad \left\{ \begin{array}{l} \left| \dfrac{\nu_{n+r-1}}{n+r-1} - p \right| < \varepsilon, \\ \left| \dfrac{\nu_{n+r}}{n+r} - p \right| < \varepsilon, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array} \right.$$

De même, la mesure de l'ensemble $(I'_{n+r}\ I'_{n+r+1}\ldots)$ correspond à l'ensemble des points pour lesquels on a simultanément

$$(110) \qquad \left\{ \begin{array}{l} \left| \dfrac{\nu_{n+r}}{n+r} - p \right| < \varepsilon, \\ \left| \dfrac{\nu_{n+r+1}}{n+r+1} - p \right| < \varepsilon, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array} \right.$$

Le passage de (109) à (110) introduit un ensemble de points qui, pour la *première fois*, entrent en jeu pour satisfaire aux inégalités en nombre infini (107).

La mesure de l'ensemble de points cherché est évidemment

$$(111) \qquad m(I'_{n+r}I'_{n+r+1}\cdots) - m(I'_{n+r-1}I'_{n+r}\cdots);$$

désignons-la par

$$(112) \qquad p'_{n+r}.$$

Si l'on écrit alors la relation

$$(113) \qquad m(I'_{n+r}I'_{n+r+1}\cdots) = m(I'_{n+r-1}I'_{n+r}\cdots) + p'_{n+r},$$

pour les valeurs de r, $r-1$, $r-2$,..., 1 et si l'on fait la somme, on aura

$$(114) \qquad m(I'_{n+r}I'_{n+r+1}\cdots) = m(I'_{n}I'_{n+1}\cdots) + \sum_{i=1}^{r} p'_{n+i}.$$

Puisqu'en augmentant r le premier membre de l'égalité précédente tend vers l'unité, tout en restant cependant inférieur à cette valeur, il est évident que nous pouvons écrire

$$m(\mathrm{I}'_n\mathrm{I}'_{n+1}\cdots) + \sum_{r=1}^{\infty} p'_{n+i} = 1 \tag{115}$$

L'égalité précédente est intéressante, car elle peut expliquer l'origine de certaines interprétations erronées qui se présentent quand on fait appel aux notions empiriques du Calcul des probabilités, ainsi que nous le verrons par la suite.

10. Considérations critiques sur la théorie précédente

Dans la théorie abstraite que nous venons de développer, il n'a pas été question ni de probabilités, ni d'événements, ni de cas également possibles; nous n'avons utilisé aucune notion empirique. La théorie est exclusivement analytique et les suites d'ensembles et des fonctions qu'elle utilise ont été bien définies. Cette théorie est fondée sur les opérations élémentaires relatives aux mesures d'ensemble de points, sur la définition d'ensembles multipliables et des variables pondérées, et sur rien d'autre.

Comment se comporteront les valeurs $\frac{\nu_n}{n}$, c'est-à-dire les fréquences des variables pondérées considérées ci-dessus ? Evidemment, de la façon indiquée par la théorie. On ne peut rien affirmer *a priori*, qui soit indépendant de la théorie : on risquerait d'en recevoir un démenti.

Puis-je, par exemple, affirmer à propos des inégalités (104), qu'il existe une valeur finie $\bar{n}$ telle que pour tout $n > \bar{n}$ on puisse écrire

$$\left| \frac{\nu_n}{n} - p \right| < \varepsilon. \tag{116}$$

pour *tous les points de la surface de base* ?

Lorsqu'on tient compte de (115), la réponse est négative.

En effet, les ensembles de points pour lesquels les inégalités

$$\left\{\begin{array}{l} \left| \frac{\nu_k}{k} - p \right| < \varepsilon, \\ \left| \frac{\nu_{k+1}}{k+1} - p \right| < \varepsilon, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right. \tag{117}$$

en nombre infini, sont satisfaites à partir de $k = n, n + 1, \ldots, n + r, \ldots$, ont respectivement pour mesures

$$(118) \qquad \left\{ \begin{array}{l} m(I'_n I'_{n+1} \cdots), \\ m(I'_n I'_{n+1} \cdots) + p'_{n+1}, \\ m(I'_n I'_{n+1} \cdots) + p'_{n+1} + p'_{n+2}, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ m(I'_n I'_{n+1} \cdots) + p'_{n+1} + p'_{n+2} + \cdots + p'_{n+r}. \end{array} \right.$$

Il est bien vrai que (115) est valable pour $r \to \infty$, mais on ne peut jamais fixer une valeur de $\overline{n}$ telle que les inégalités (117) soient valables sur *toute la surface de base* : si grand que soit $\overline{n}$ il reste toujours un ensemble résiduel sur cette surface, ainsi que l'indiquent les valeurs (118).

Dans la théorie abstraite dont je viens de donner un résumé succinct, on n'a donc pas le droit de dire que

$$(119) \qquad \lim_{n \to \infty} \frac{\nu_n}{n} = p,$$

dans le sens ordinaire de l'analyse. L'allure des fréquences $\frac{\nu_n}{n}$ ne peut être que celle requise par la théorie.

Revenons, pour un instant, au cas simple qui correspond au théorème de BERNOULLI.

En fixant un rang n, soit $1 - \gamma_n$ la mesure de l'ensemble des points pour lesquels on a

$$(120) \qquad \left| \frac{\nu_n}{n} - p \right| < \varepsilon.$$

Puisqu'il s'agit d'aires multipliables il existe, sur la surface de base, des points pour lesquels

$$(121) \qquad \frac{\nu_n}{n} = 1$$

et dont la mesure est

$$(122) \qquad p^n.$$

En augmentant n, la mesure précédente tend vers zéro, mais *il y a toujours* des points qui satisfont à (121), quoique à la limite ils forment un ensemble de mesure nulle.

Cela est inévitable. Un Calcul des probabilités basé sur les principes des probabilités totales et des probabilités composées, comporte nécessairement la considération d'événements de probabilité nulle, mais qui ne sont pas théoriquement impossibles.

Il me semble illogique d'affirmer que ces événements n'existent pas.

11. Sur le calcul classique des probabilités.

« On ne peut guère donner une définition satisfaisante de la proba-
« bilité. On dit ordinairement : la probabilité d'un événement est le
« rapport du nombre des cas favorables à cet événement, au nombre
« total des cas possibles. »

« La définition précédente est incomplète ; il faut ajouter : « à
« condition que tous les cas soient également vraisemblables. »

« La définition complète de la probabilité est donc une sorte de
« pétition de principe : comment reconnaître que tous les cas sont
« également probables ? Une définition mathématique n'est pas
« possible : nous devons, dans chaque application, faire des *conven-*
« *tions*, dire que nous considérons tel et tel cas comme également pro-
« bables. Ces conventions ne sont pas tout à fait arbitraires, mais
« elles échappent à l'esprit du mathématicien qui n'aura pas à les
« examiner, une fois qu'elles seront admises. »

« Ainsi tout problème de probabilités offre deux périodes d'étude :
« la première, métaphysique pour ainsi dire, qui légitime telle ou
« telle convention ; la seconde, mathématique, qui applique à ces
« conventions les règles du calcul. »

A ces considérations du grand mathématicien H. POINCARÉ, je crois qu'on ne peut pas, même aujourd'hui, ajouter ou enlever un seul mot.

J'ai voulu rappeler ces considérations parce que ce que je dirai me semble s'accorder parfaitement avec les idées du grand mathématicien.

La probabilité est une notion empirique, elle n'est nullement une notion abstraite. De même que toutes les notions empiriques (c'est une question de mesure), elle ne se prête pas à une définition rigoureuse.

Je mets dans une urne b boules blanches et r boules d'une couleur différente : ce sont des sphères de même diamètre et du même poids, dans la limite des possibilités d'exécution matérielle. Ayant fermé les yeux et mêlé les boules dans l'urne, j'en extrais une. J'observe sa couleur et je la remets dans l'urne. Les yeux toujours fermés, je les mélange de nouveau et je tire encore une boule. J'observe sa couleur et je la remets dans l'urne. Je répète n fois la même opération.

Soit ν_n le nombre de fois que j'ai tiré une boule blanche ; $\frac{\nu_n}{n}$ sera la fréquence relative de ces événement pour n épreuves.

Dans les conditions matérielles dans lesquelles je me suis placé, il arrive que si n est suffisamment grand, $\frac{\nu_n}{n}$ a une valeur voisine du rapport bien connu $\frac{b}{b+r}$.

En outre, si les boules blanches de l'urne considérée sont numérotées de 1 à b et celles de couleur différente de $b+1$ à $b+r$, il arrive que dans un grand nombre d'épreuves, les fréquences de sortie d'une boule portant un nombre déterminé sont à peu près égales au rapport

$$\frac{1}{b+r}. \tag{123}$$

Il s'agit ici d'une constatation objective dont chacun peut se rendre compte en répétant l'expérience.

Mais il n'est pas possible d'effectuer une infinité d'épreuves du type considéré. Toutefois, l'expérience ayant été faite et confirmée un très grand nombre de fois, *naturellement pour n fini*, on peut énoncer la *proposition suivante, qui a, elle, un caractère physique* :

Le rapport $\frac{b}{b+r}$ et le rapport $\frac{1}{b+r}$, qu'on appelle respectivement *probabilité d'extraire dans une épreuve une boule blanche* et *probabilité d'extraire dans une épreuve une boule déterminée*, constituent une prévision de la fréquence relative avec laquelle on extraira une boule blanche ou une boule déterminée, dans un grand nombre d'épreuves faites dans les conditions précisées ci-dessus.

J'appellerai les cas qui se rapportent à l'extraction d'une boule déterminée des *cas également possibles*, pour indiquer que les probabilités respectives, égales, permettent de prévoir des fréquences à peu près égales.

Il s'agit là cependant d'une affirmation de caractère expérimental, physique, qui n'exclut pas la possibilité de trouver, pour une valeur de n très élevée, un écart considérable entre ν_n/n et le rapport $b/(b+r)$.

Considérons, deux urnes au lieu d'une seule : l'une contenant b boules blanches et r boules d'une couleur différente ; l'autre b_1 boules blanches et r_1 boules d'une couleur différente. Les boules sont sphériques, ont même diamètre et même poids.

J'effectue n d'extractions doubles, de la même manière que plus haut. Je suppose que dans les n extractions une paire de boules blanches soit sortie α_n fois et qu'une boule blanche ait été extraite de la première urne ν_n fois. Comme on peut écrire identiquement

$$\frac{\alpha_n}{n} = \frac{\nu_n}{n} \cdot \frac{\alpha_n}{\nu_n}, \tag{124}$$

il devra résulter (et il résulte en effet d'après les constatations de caractère expérimental précédentes), que pour n suffisamment grand

$$\left\{ \begin{array}{l} \dfrac{b}{b+r} \cong \dfrac{\nu_n}{n}, \\[2ex] \dfrac{b_1}{b_1+r_1} \cong \dfrac{\alpha_n}{\nu_n}, \end{array} \right. \tag{125}$$

et par conséquent

$$\frac{b}{b+r} \cdot \frac{b_1}{b_1+r_1} \cong \frac{\alpha_n}{n}. \tag{126}$$

Le principe des probabilités composées est ainsi justifié par l'expérience et dans la mesure permise par les observations, à savoir : pour des événements indépendants la probabilité composée est une prévision de la fréquence relative avec laquelle l'événement composé apparaît dans un grand nombre d'épreuves identiques.

On peut justifier tout aussi bien le principe des probabilités composées et le principe des probabilités totales pour des événements dépendants. Finalement, on conclut que :

Pour des schémas donnés, la définition expérimentale de la probabilité, considérée comme prévision de la fréquence relative d'un événement donné dans un grand nombre d'épreuves, justifie l'application des principes des probabilités totales et des probabilités composées.

Après avoir ainsi analysé le caractère expérimental de la notion de probabilité, il s'agit maintenant d'en établir la théorie correcte.

Nous pouvons représenter un événement E_1, de probabilité donnée p, sur la surface de base dont j'ai parlé dans la théorie abstraite, par une aire partielle E_1 de mesure p_1 ; l'événement contraire $\overline{E_1}$ sera représenté par l'aire partielle complémentaire, de mesure q_1. Un autre événement E_2 de probabilité p_2 sera représenté par une autre aire partielle E_2, de mesure p_2 ; la probabilité de l'événement composé $(E_1 E_2)$, si les deux événements sont compatibles, sera donnée par

$$(127) \qquad m(E_1E_2)$$

et, si les deux événements sont indépendants, E_2 devra être placée de façon que

$$(128) \qquad m(E_1E_2) = m(E_1) \cdot m(E_2) = p_1 \cdot p_2,$$

etc.

En nous appuyant de cette façon sur les principes des probabilités totales et composées, nous allons construire une théorie du calcul des probabilités, qui correspondra exactement à la théorie abstraite précédemment établie.

Le développement du calcul classique des probabilités comporte trois étapes :

1° Recherche de la signification *expérimentale* de la notion de probabilité et de celle des cas également possibles ; justification *expérimentale* des principes des probabilités totales et composées, appliqués à une classe donnée d'événements.

2° Elaboration de la théorie abstraite fondée sur ces deux principes, appliqués à la classe d'événements que l'on considère. La théorie ainsi obtenue est indépendante de la notion physique de probabilité comme cela résulte des développements précédents ; elle repose uniquement sur des conventions relatives à l'application des principes des probabilités totales et composées.

3° Si l'application de ces principes est conforme aux faits expérimentaux, les probabilités déduites de la théorie peuvent servir effectivement pour prévoir les fréquences relatives des événements considérés.

Dans les pages précédentes, nous avons toujours considéré uniquement un certain schéma bien déterminé, le schéma des urnes. Pour appliquer le calcul des probabilités à d'autres schémas, différents de ceux que nous avons utilisés jusqu'ici, il faudrait nécessairement justifier *a priori* la légitimité de cette généralisation. Or, on a tenté de nombreuses expériences dans ce sens, et les résultats obtenus ont été en parfait accord avec les précédents. La théorie abstraite du calcul des probabilités peut donc s'appliquer même à ces autres schémas, et par extension, à d'autres schémas qui n'ont pas encore été soumis au contrôle expérimental.

En cas de doute sur l'existence des probabilités ou sur la légitimité de l'application des principes des probabilités totales et composées à la classe d'événements que l'on considère, on devra comparer naturellement les résultats de la théorie avec l'expérience et décider d'après les résultats de cette comparaison ; ce procédé est commun à toutes les sciences d'observation.

12. Considérations critiques sur le calcul des probabilités.

Les grandes difficultés auxquelles on fait habituellement allusion lorsqu'on parle du calcul des probabilités et qui ont trait à la définition de cas *également possibles*, me semblent à l'heure actuelle complètement surmontées, au moins pour certains schémas particuliers. Pour le schéma des urnes, par exemple, on peut considérer qu'elles sont éliminées si l'on admet que l'ouvrier qui a tourné les boules est suffisamment habile pour les rendre parfaitement identiques et si l'expérimentateur ferme les yeux avant d'en tirer une de l'urne.

Il est certain que la définition physique de la probabilité n'est pas satisfaisante, mais ce défaut est partagé par toutes les définitions à caractère physique. La théorie n'a cependant rien à faire avec la définition physique de la probabilité ; elle raisonne sur des *conventions* concernant l'application des principes des probabilités totales et des probabilités composées à une classe particulière d'événements. Les conséquences de la théorie seront en accord avec l'expérience si les conventions faites sont justifiées par l'observation. S'il n'en est pas ainsi, la théorie élaborée reste quand même irréprochable : mais alors il peut se faire que ses conséquences ne soient plus vérifiées par l'expérience.

Or, supposons pour un instant que, pour une classe particulière d'événements, on choisisse une définition de la probabilité, *fondée sur des considérations de fréquence*. Admettons même que l'on considère légitime *l'application réitérée* des principes des probabilités totales et composées.

On pourra alors développer une théorie, qui à son tour permettra de calculer, par exemple, le comportement particulier des fréquences autour des probabilités. Supposons que ce comportement diffère de celui qui a été pris comme base lorsqu'on a choisi arbitrairement et en dehors de toute théorie la définition de la probabilité. Que doit-on en conclure ?

Je crois qu'il faut tout simplement conclure que la définition de la probabilité a été mal choisie, qu'elle n'est pas compatible avec l'application réitérée des principes mentionnés plus haut et qui constituent la base de la théorie. Et si l'application réitérée des deux principes indiqués à la classe d'événements que l'on examine est considérée comme légitime, il faut bien dire que cette définition est en contradiction avec la théorie.

Considérons une urne renfermant des boules blanches et des boules de différentes couleurs. Soit p le rapport du nombre des boules blanches au nombre total des boules et choisissons p comme probabilité d'extraire une boule blanche dans une épreuve quelconque. On peut affirmer que si n est assez grand, on aura

$$\frac{\nu_n}{n} \cong p \tag{129}$$

sauf en des cas tout à fait exceptionnels.

Imaginons qu'on fasse une infinité d'extractions, et soit

$$\frac{\nu_n}{n}, \frac{\nu_{n+1}}{n+1}, \frac{\nu_{n+2}}{n+2}, \ldots \tag{130}$$

la suite illimitée de fréquences qu'on est ainsi amené à considérer. La valeur de chaque fréquence dépend évidemment des valeurs prises par les fréquences précédentes.

Cela étant, peut-on affirmer que

$$\lim_{n \to \infty} \frac{\nu_n}{n} = p. \tag{131}$$

dans le sens ordinaire de l'analyse ?

La réponse à cette question a été négative lorsque nous nous sommes placés au point de vue de la théorie abstraite du calcul des probabilités ; il convient cependant d'examiner ce même problème du point de vue physique.

Si (131) est légitime, on pourra toujours trouver en entier $\bar{n}$ tel que pour tout $n > \bar{n}$, on ait

$$\left|\frac{\nu_n}{n} - p\right| < \varepsilon, \tag{132}$$

ε étant un nombre positif choisi à l'avance.

Il est évident qu'on ne peut pas démontrer cette proposition en s'appuyant sur le théorème de Bernoulli ; il faut faire appel à un autre procédé de démonstration.

Je considère les inégalités en nombre infini

$$\left\{\begin{array}{l} \left|\frac{\nu_n}{n} - p\right| < \varepsilon, \\ \left|\frac{\nu_{n+1}}{n+1} - p\right| < \varepsilon, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \left|\frac{\nu_{n+r}}{n+r} - p\right| < \varepsilon, \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right. \tag{133}$$

Soit p_n la probabilité pour que ces inégalités (133) se vérifient à partir du rang n, sans tenir compte de ce qui se passe pour les inégalités d'un rang inférieur à n ; soit p'_{n+1} la probabilité pour qu'elles se vérifient *seulement* à partir de l'indice $n + 1$;... soit p'_{n+r} la probabilité pour qu'elles se vérifient seulement à partir de l'indice n $n + r$; etc.

Pour des événements incompatibles, on peut écrire, en vertu du principe des probabilités totales

$$p_{n+r} = p_{n+r-1} + p'_{n+r} \tag{134}$$

et par conséquent

$$p_{n+r} = p_n + p'_{n+1} + p'_{n+2} + \cdots + p'_{n+r}. \tag{135}$$

Puisque l'on a pour un cas particulier de la loi uniforme des grands nombres

$$\lim_{r \to \infty} p_{n+r} = 1, \tag{136}$$

on aura aussi

$$p_n + p'_{n+1} + p'_{n+2} + \cdots + p'_{n+r} + \cdots = 1. \tag{137}$$

Il en résulte, donc, que la probabilité pour que les inégalités (133) soient simultanément vérifiées à partir d'un des indices

(138) $$n, \quad n+1, \quad n+2, \cdots$$

est égale à l'unité.

Le théorème exprimé par (137) est le même que celui énoncé par (115). Mais l'égalité (137), considérée du point de vue physique, peut facilement suggérer une conclusion erronée. En effet, on pourrait croire que dans la suite (138) se trouve certainement un nombre $\bar{n}$ tel que pour tout $n > \bar{n}$ on ait

(139) $$\left|\frac{\nu_n}{n} - p\right| < \varepsilon.$$

Cette déduction est fausse : car, si grand que l'on choisisse $\bar{n}$, l'égalité (137) nous dit que l'existence d'un tel indice est seulement *probable* ; (137) nous montre qu'il pourrait aussi bien être $\bar{n} + r$, avec une probabilité p'_{n+r}, au lieu de $\bar{n}$, etc.

La loi uniforme des grands nombres ne peut donc non plus justifier l'égalité (131). Cela est tout naturel : *l'application réitérée* des principes des probabilités totales et composées, nous oblige inuévitablement à affirmer que, dans une suite infinie d'extractions d'une urne pour laquelle *la probabilité d'extraire une boule blanche dans chaque épreuve est constamment égale à p*, il peut très bien arriver que la suite des fréquences d'extraction ne tende pas vers p.

Une question se pose alors : Pourquoi ce fait ne résulte-t-il pas des observations ?

La réponse me semble simple : il faudrait un temps infini pour effectuer une infinité d'épreuves et arriver à cette conclusion. Cela rend donc impossible toute vérification expérimentale directe. Cependant on peut justifier l'affirmation précédente par une série de considérations assez simples.

Soit 1/2 la probabilité d'extraire à chaque épreuve une boule blanche d'une urne déterminée. La probabilité d'extraire n boules blanches en n épreuves est

(140) $$\frac{1}{2^n}.$$

Faisons maintenant 2^{2n} séries indépendantes de n extractions chacune. Par rapport à ces 2^{2n} séries de n extractions, le nombre moyen de fois que boule blanche est extraite n fois en n épreuves est

$$2^{2n} \cdot \frac{1}{2^n} = 2^n. \tag{141}$$

Le calcul des probabilités nous apprend que la valeur moyenne théorique d'une variable éventuelle est une prévision de la moyenne arithmétique des valeurs prises par la variable dans un grand nombre d'épreuves.

Supposons maintenant qu'on répète k fois le groupe de 2^{2n} séries de n extractions.

Soient $v_1, v_2, \ldots, v_k$ le nombre de cas où une boule blanche aura été extraite n fois dans n épreuves, respectivement dans le premier groupe de 2^{2n} extractions, dans le deuxième groupe, etc... etc., dans le $k^{ème}$ groupe.

Si k est suffisamment grand on aura

$$\frac{\nu_1 + \nu_2 + \cdots + \nu_k}{k} \cong 2^{2n} \frac{1}{2^n} = 2^n. \tag{142}$$

Il s'ensuit que dans les k groupes de 2^{2n} séries d'extractions, la boule blanche se sera présentée un grand nombre de fois, successivement n fois en n épreuves. Si l'on avait le temps, tout cela pourrait être soumis au contrôle expérimental. Si l'on faitait tendre n vers l'infini le nombre des suites (130) où les fréquences au lieu de tendre vers $1/2$ prennent toujours la valeur 1, deviendrait lui-même infini. Cela n'est naturellement pas contrôlable par l'expérience, mais le calcul des probabilités classique nous montre clairement qu'il n'est pas impossible qu'une suite de fréquences (130), correspondant à l'existence d'une probabilité constante p, ne tende pas vers p.

Il est certain qu'on ne peut pas trouver dans le calcul des probabilités tout le déterminisme qu'on souhaite volontiers ! Pour réaliser une suite de fréquences dans le schéma bernoullien, il faudrait faire une infinité d'extractions : l'allure des fréquences reste subordonnée à l'action d'une infinité de causes sur lesquelles on ne peut dire que très peu de chose ; l'expérience nous apprend que, *pour un nombre fini d'observations*, il y a généralement compensation des écarts.

Lorsqu'on admet que

$$\lim_{n \to \infty} \frac{\nu_n}{n} = p, \tag{143}$$

on néglige des cas dont la probabilité a une valeur très petite mais n'est jamais nulle. En calculant la valeur moyenne du nombre de fois où un événement de probabilité p, peut se présenter en N épreuves, c'est-à-dire Np, on peut toujours, aussi petit que soit p, choisir un N suffisamment grand pour que Np lui-même soit aussi grand que l'on veut : le calcul des probabilités ne refuse pas le droit d'existence aux événements de probabilité infinitésimale, pourvu que l'on augmente convenablement le nombre des épreuves.

L'égalité (143) n'est donc pas compatible avec l'application réitérée du principe des probabilités composées. Cette application nous oblige en effet à prendre en considération des cas de probabilité aussi petite que l'on veut, cas que cette égalité néglige.

III

1. La convergence au sens du calcul des probabilités. Historique

On n'affirme rien de nouveau en constatant que certaines recherches, considérées comme particulièrement difficiles à une époque déterminée, apparaissent ensuite à une époque postérieure tout à fait simples et élémentaires. En mettant en œuvre des résultats déjà obtenus et en améliorant progressivement ses méthodes, la science arrive en effet à des généralisations, découvre des analogies, simplifie, synthétise un grand nombre de résultats, dont chacun n'a été obtenu qu'au prix d'un gros effort, ce qui permet en fin de compte de présenter sous une forme très simple des notions qui autrefois apparaissaient difficiles à concevoir.

Qu'il me soit permis de faire ici un bref historique de l'évolution de nos idées sur une question toute particulière, à savoir, *sur la manière dont les fréquences d'un événement convergent vers une limite* et sur le sens qu'on doit attribuer à cette affirmation : cela justifiera en grande partie les conceptions qui sont, d'après moi, à la base du calcul des probabilités.

En 1908, M. R. de MONTESSUS publiait ses intéressantes *Leçons élémentaires sur le calcul des probabilités* (1). Ces leçons donnèrent lieu en Italie à de très intéressantes discussions. M. de MONTESSUS s'oc-

(1) Gauthiers-Villars, éditeur, Paris, 1908.

cupe, dans le volume indiqué, de la « définition logique du hasard » qu'il avait donnée au Congrès de philosophie de Genève en 1904.

Transcrivons ici quelques passages :

« Le signe distinctif immédiat d'événements dûs au hasard est « que : de tels événements étant partagés en classes et ces classes « en catégories, le rapport du nombre total d'événements de la classe, « au nombre total d'événements de l'une quelconque des catégories « tend irrégulièrement vers une limite déterminée, quand le nombre « d'événements considérés devient de plus en plus grand. »

« Qu'est-ce que le hasard ?

« La réponse est prévue : Etant donné que certains événements « ont un caractère commun et, pour cette raison, constituent une « classe, mais diffèrent à certains points de vue, ce qui me permet « de les ranger en catégories bien définies, le hasard consiste dans « l'absence de relations bien définies entre les causes, rangeant tel « événement de telle classe dans telle catégorie, et les caractères dis- « tinctifs de cette catégorie.

« Le *Calcul des probabilités* a pour objet l'étude de la *fréquence* « *relative* des événements dont l'arrivée dépend du hasard. »

« En donnant des exemples intéressants, de caractèreexpé rimental « l'auteur poursuit : « En général et sur un grand nombre d'épreuves, le « rapport du nombre d'arrivées d'un événement au nombre d'épreu- « ves, tend, à mesure que croît le nombre d'épreuves, vers le rapport « du nombre des cas favorables au nombre des cas possibles. »

« La proposition que je viens d'énoncer est connue sous le nom « de *loi de Bernoulli*. Elle est purement expérimentale ; non pas qu'on « en ait poussé très loin la vérification (GAUSS a cependant compté « le nombre d'as que le sort lui attribuait dans ses parties de « whist quotidiennes), mais plutôt que ses conséquences en font une « loi vérifiée à chaque instant. Telles sont les lois de la *Mécanique* « *rationelle*.

« Sur cette loi, une fois admise, sont fondés la définition de la « *Probabilité mathématique* et le calcul des probabilités tout entier.

« J'ajoute que, *comme il est naturel*, la loi de Bernoulli devient une « conséquence de la définition de la probabilité mathématique et « qu'on peut la démontrer en partant de cette définition.

« Définition de la probabilité mathématique : On exprime le fait « d'expérience révélé par la loi de Bernoulli en définissant la *proba-*

« *bilité mathématique* comme le rapport du nombre des cas favorables « au nombre des cas possibles, étant admis que les cas envisagés « sont tous également possibles. Ce mot *probabilité* indique seule- « ment que, sur un grand nombre d'épreuves, le rapport du nombre « d'arrivées d'un événement au nombre d'*expériences* tendra préci- » sément vers le rapport du nombre des cas favorables au nombre « des cas possibles : et rien de plus. »

Je ne commenterai pas les passages précédents extraits du volume de M. R. de MONTESSUS. Toutefois, il me faut signaler le fait que les opinions précédentes conduisirent certains auteurs italiens qui les avaient adoptées à la conviction que la probabilité devait être considérée comme la limite des *fréquences relatives*.

En effet, M. Tullio BAGNI, homme d'une grande valeur, titulaire d'une chaire de mathématique au « Regio Istituto Superiore di Scienze Economiche e Commerciali » de Rome, publiait en 1915 un travail (1) résumant ses leçons, dans l'avant-propos duquel on lit les lignes suivantes :

« Je n'emploie pas la notion de « probabilité d'un événement ». « MM. BRUNS et De MONTESSUS sont, je crois, les seuls auteurs qui « ont précisément affirmé qu'ils voulaient fonder le calcul des pro- « babilités sur une proposition de caractère expérimental.

« .....Or, en quoi consiste la loi des grands nombres sinon dans « l'affirmation de l'existence de limites déterminées pour les fréquences « de certains événements, lorsqu'on imagine que les épreuves et les « observations sont répétées indéfiniment sur les mêmes événements ?

« *La mesure de la probabilité*, ou *la probabilité* tout court, *d'un événement normal*, est le rapport entre le nombre des cas favorables « à l'événement et le nombre des cas possibles, ou mieux : la fréquence « des cas favorables à l'événement dans la classe des cas possibles, « ou enfin : la limite vers laquelle tend la fréquence de l'événement « dans une suite de s épreuves sur le fait accidentel, normal d'où « l'événement peut dériver, lorsqu'on augmente indéfiniment le « nombre s.

Ces idées soulevèrent de nombreuses critiques parmi les probabilistes italiens. On posa les problèmes suivants :

(1) *Teoria matematica dei fenomeni collettivi*. Firenze, G. Barbera, 1915.

1º Le théorème de Bernoulli, dans sa forme connue, tel qu'il résulte de la théorie, est-il compatible ou non, avec la formule.

$$\lim_{n \to \infty} \frac{\nu_n}{n} = p\,?. \tag{144}$$

2º Peut-on démontrer, dans le cas général, même en renonçant à écrire la formule précédente, qu'il est possible de trouver une suite de nombres positifs ε_n, $\varepsilon_{n+1}, \ldots$, tendant vers zéro, et tels que la probabilité de la coexistence des inégalités, en nombre infini

$$\left\{\begin{array}{l} \left|\frac{\nu_n}{n} - p\right| < \varepsilon_n, \\ \left|\frac{\nu_{n-1}}{n+1} - p\right| < \varepsilon_{n+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array}\right. \tag{145}$$

dans lesquelles ν_n/n, $\nu_{n+1}/(n+1), \ldots$ représentent des fréquences dépendantes, soit aussi proche de l'unité qu'on le désire ?

Ne peut-on pas imaginer des cas pour lesquels cette probabilité soit nulle ?

3º Le cas plus général du théorème de Bernoulli, la loi des grands nombres de Laplace-Tchébycheff (dont les conditions de validité ont été données par M. Lindeberg, par M. Lévy et par moi-même) qui se rapporte à des moyennes de variables éventuelles indépendantes, est-il compatible ou non avec l'assertion que les moyennes observées *convergent* vers un nombre déterminé ?

4º Cette dernière assertion peut-elle être justifiée, au moins avec une grande probabilité, et sous quelles conditions ?

Ces problèmes donnèrent lieu à des discussions très intéressantes entre les probabilistes italiens : j'aime citer parmi eux MM. G. Castelnuovo et P. Medolaghi.

Les questions seraient peut-être restées sans réponse si l'on n'avait pas procédé à de nouvelles recherches. On les entreprit, et M. Castelnuovo dans une conférence au « Seminario matematico » de l'Université de Rome le 27 mai 1916, pouvait démontrer le résultat que le théorème de Bernoulli ne permettait en aucune façon d'écrire

$$\lim_{n \to \infty} \frac{\nu_n}{n} = p. \tag{146}$$

En 1918, ainsi que je l'indiquerai plus loin, toutes les questions posées reçurent une réponse précise.

En 1919, M. CASTELNUOVO (1) publiait un traité de calcul des Probabilités, qui tient compte des résultats des nouvelles études entreprises et qui représente un gros effort pour dissiper les équivoques qui existaient à cette époque. Je tiens à citer quelques passages de ce traité à l'appui de ce que je viens de dire.

« *Loi empirique du cas*. Dans une série d'épreuves, répétées un grand « nombre de fois dans les mêmes conditions, chaque événement « possible se manifeste avec une fréquence (relative) qui est à peu » près égale à sa probabilité. L'approximation croît ordinairement « lorsqu'on augmente le nombre des épreuves.

« On dit aussi souvent, en employant un langage impropre qui « doit être expliqué pour éviter les malentendus : la fréquence d'un « événement éventuel, dans une série d'expériences faites dans les « mêmes conditions, tend, lorsqu'on augmente le nombre de celles-ci, « vers une limite qui est égale à la probabilité de l'événement.

« La locution *tend vers une limite*, ne doit pas être interprétée dans « le sens précis qu'on lui attribue en analyse, mais dans le sens vague « qu'elle a dans les sciences expérimentales, lorsqu'on dit, par exemple « que « le mélange de deux gaz ou de deux liquides *tend* à devenir « uniforme avec le temps, ou que les températures de deux corps « voisins *tendent* à s'égaler ».

« L'emploi du mot *limite* dans son sens empirique serait contre-in- « diqué si le postulat précédent devait servir à la construction logique « du calcul des probabilités. Mais cela n'est pas le cas : son seul rôle « est simplement d'établir un lien entre la théorie et les applications, « dans lesquelles le mot limite est ou devrait être employé dans le « sens indiqué.

« A propos de ces lois des grands nombres, une question va se pré- « senter qui a déjà un intérêt remarquable, *même dans le cas plus* « *simple du théorème de Bernoulli*, auquel nous nous rapportons. « D'une urne de composition donnée on fait une série illimitée d'extrac- « tions, en remettant tour à tour la boule extraite dans l'urne. En « divisant, après chaque épreuve, le nombre des boules blanches « extraites par le nombre des extractions faites, on a une suite

(1) G. CASTELNUOVO. *Calcolo delle probabilitá*, Albrighi e Segati, Milano-Roma-Napoli, 1919.

« illimitée de fréquences

(147) $$f_1, f_2, \cdots, f_n, f_{n+1}, \cdots$$

« qui doivent être comparées à la probabilité p de l'événement.

« Le théorème de Bernoulli affirme que, pour ε donné, la probabi-« lité de l'inégalité $|f_n - p| < \varepsilon$ se rapprochera autant qu'on veut « de l'unité pourvu que l'on prenne n suffisamment grand. Or, la pro-« babilité de la coexistence des inégalités, en nombre infini

(148) $$|f_n - p| < \varepsilon, \qquad |f_{n+1} - p| < \varepsilon, \cdots,$$

« est certainement inférieure à la probabilité considérée plus haut ; « et pourtant le théorème de Bernoulli ne peut nous fournir aucune « indication sur la valeur de cette probabilité : on peut même envi-« sager la possibilité qu'elle soit nulle. En réalité, cela n'est pas vrai « car on peut démontrer que, ε étant donné *arbitrairement*, la proba-« bilité pour que toutes les fréquences depuis la $n^{ème}$, soient simul-« tanément comprises entre $p - \varepsilon$ et $p + \varepsilon$, peut s'approcher de « l'unité autant qu'on le voudra, pourvu que l'on prenne n suffisam-« ment grand. Le théorème a été, démontré dans l'hypothèse que les « intervalles successifs dans lesquels doivent être comprises les fré-« quences diminuent graduellement et tendent vers zéro. »

Dans le premier volume de la deuxième édition du *Calcul des Probabilités* (1), M. CASTELNUOVA ajoute en se référant au théorème sus-indiqué : « C'est seulement après l'avoir établi et précisé conve-« nablement qu'on peut affirmer qu'en augmentant le nombre des « épreuves, la fréquence d'un événement *tend*, avec une *certitude* « *pratique*, vers la valeur de la probabilité ».

2. La convergence vers une limite au sens du Calcul des probabilités et la convergence d'une suite de variables pondérées vers un nombre.

Les considérations historiques que je me suis permis de faire expliquent suffisamment, je pense, les motifs qui m'ont poussé à publier ma deuxième note de 1916, déjà mentionnée,.

Il fallait clarifier le fait que la convergence vers une limite, telle qu'elle résulte du théorème de Bernoulli ou de la loi des grands

(1) ZANICHELLI, Bologna, 1925.

nombres, est différente de la convergence ordinaire telle que l'envisage l'analyse.

On pouvait le faire en généralisant la notion de « convergence vers une limite » suggérée par *les deux cas particuliers* cités, en démontrant que la « convergence » au sens ordinaire de l'analyse pouvait être considérée *comme un cas particulier de la précédente* et en remarquant jusqu'à quel point les théorèmes sur la « convergence ordinaire » étaient valables pour l'autre « convergence ».

J'avais déjà essayé de préciser la notion abstraite de *grandeur éventuelle*, dans mes travaux de 1909 et de 1910 (1) ; mais ce fut seulement dans le deuxième de mes travaux de 1916 que j'en donnais la définition précise indiquée plus haut et que je l'appelais *variable aléatoire* (en italien : *variable causale*). Cette définition rencontra la faveur du public scientifique, avec la dénomination qui est maintenant universellement employée.

Dans le deuxième de mes travaux de 1916, j'écrivis les lignes suivantes :

« Une suite de variables éventuelles

$$X_1, X_2, \cdots, X_n, \cdots \tag{149}$$

« tend, au sens du calcul des probabilités, vers un nombre fixe N, si,
» en désignant par η un nombre positif quelconque fixé d'avance,
« la probabilité P_n pour que la variable éventuelle X_n prenne une
« valeur de l'intervalle $(N - \eta, N + \eta)$, c'est-à-dire la probabilité
« P_n des inégalités

$$N - \eta \leqslant X_n \leqslant N + \eta, \tag{150}$$

« est telle que

$$\lim_{n \to \infty} P_n = 1. \tag{151}$$

« Si au lieu du symbole *lim* de la « convergence ordinaire » vers une
« limite, nous employons le symbole *Lim* pour désigner la convergence

(1) *Elementi di Matematica attuariale.* Sinossi della Cultura universale e pratica, vol. II, Società Editrice Libraria, Milano, 9109.

Intorno ad un teorema fondamentale della teoria del rischio. Bollettino dell'Associazione degli Attuari Italiani, nº 24, Milano, 1910.

Intorno ad un teorema di Calcolo delle probabilitá. Giornale di Matematiche di Battaglini, vol. XLIX.

« vers une limite au sens du Calcul des probabilités, nous pourrons « écrire, au lieu de (151),

$$\lim_{n \to \infty} X_n = N. \tag{152}$$

« On trouve des exemples de suites de variables éventuelles « satisfaisant à (152), dans l'étude de la *loi des grands nombres*.

« Supposons maintenant que, comme cas particulier, la suite de « variables éventuelles soit remplacée par une suite de nombres

$$a_1, a_2, \cdots, a_n, \cdots \tag{153}$$

« auxquels correspondent des probabilités constamment égales à « l'unité, et rappelons-nous que, selon la définition des variables « éventuelles, l'unité représente la certitude tandis que *zéro* « représente l'impossibilité. Dans ce cas particulier, P_n ne peut « prendre que la valeur *zéro* ou 1 et pourtant si (152) est valable, « il existe un entier $\bar{n}$ tel que pour tout $n > \bar{n}$ les inégalités, « correspondantes aux (150),

$$N - \eta \leqslant a_n \leqslant N + \eta \tag{154}$$

« seront satisfaites.

« La « convergence » vers un nombre d'une suite de variables éven- « tuelles se réduit donc à la « convergence ordinaire », lorsque la suite « indiquée plus haut est identique à une suite de nombres.

« On peut affirmer cependant que les théorèmes valables pour la « convergence vers une limite au sens du Calcul des probabilités et « *qui dépendent seulement de la définition donnée*, seront *également* « *valables* pour la convergence ordinaire vers une limite, au sens de « l'analyse, tandis qu'on ne peut rien dire au sujet de la proposition « réciproque.

« Cela nous incite à examiner, en vue des applications, jusqu'à « quel point les théorèmes relatifs à la convergence ordinaire, sont « valables pour la convergence au sens du Calcul de probabilité. »

Le programme de travail était donc précis.

Avant toutes choses, je voudrais faire quelques observations pour dissiper tout malentendu.

1° Dans ma note indiquée plus haut je me suis placé à un point de vue abstrait, bien que le langage adopté fût celui des applications. De même qu'on a l'habitude de parler d'une suite de fonctions, on peut parler d'une suite de variables éventuelles ou aléatoires, ou, si l'on veut, de variables pondérées, que l'on peut construire et représenter sur la *surface de base*, dont j'ai déjà parlé dans la théorie abstraite.

2° Je fis remarquer dans ma note que l'idée de convergence vers une limite au sens du Calcul des probabilités est suggérée précisément par la loi des grands nombres, que j'ai toujours citée. Cependant j'étais et je reste convaincu qu'il s'agit au fond d'une chose tout à fait distincte, contrairement à l'opinion d'un certain nombre d'auteurs. En effet :

a) La loi des grands nombres concerne seulement une suite particulière de variables qui tendent vers zéro au sens du Calcul des probabilités ;

b) Les suites de variables éventuelles considérées par la loi des grands nombres, ou d'autres suites analogues se rapportant aux applications, impliquent l'existence des valeurs moyennes des variables que l'on considère, tandis que la définition que j'ai donnée en fait abstraction, attendu qu'on peut construire des suites de variables tendant vers une limite au sens du Calcul des probabilités sans que leurs valeurs moyennes existent ;

c) Les suites de variables éventuelles considérées par la loi des grands nombres ont des caractéristiques fondamentales dont la définition que j'ai donnée fait abstraction. Par exemple, pour un théorème que j'ai démontré en 1923 [1], les valeurs prises par les variables éventuelles que l'on considère dans la loi des grands nombres, tendent vers zéro en effectuant un nombre infini d'oscillations autour de cette valeur. Les variables éventuelles considérées dans ma définition font totalement abstraction de cette particularité. Peut-être ai-je mal fait de parler de « convergence d'une suite de variables au sens du Calcul des probabilités » car, en vérité, il s'agit d'une généralisation qui se passe de beaucoup des caractéristiques du hasard. Mais, si j'ai employé d'une manière impropre l'expression de *convergence vers une limite au sens du Calcul des probabilités*, je l'ai fait uniquement pour mieux mettre en évidence qu'on ne devait pas confondre cette convergence avec la convergence ordinaire qui en découle comme cas particulier.

(1) *Sulla oscillazione delle frequenze intorno alla probabilità.* « Rivista internazionale di statistica » Metron, Padova, 1923.

Anna MEZZANOTTE, *Estensione di un teorema sulla oscillazione delle frequenze intorno alla probabilità*, ecc. Rendiconti del Circolo Matematico di Palermo, Tomo LII, 1928.

Ayant défini la notion de variable pondérée sur une surface de base et ayant construit une suite de variables pondérées, on peut écrire

$$\operatorname*{Lim}_{n \to \infty} X_n = N,$$

lorsque la mesure de l'ensemble des points pour lesquels

$$(155) \qquad | X_n - N | < \varepsilon,$$

tend vers l'unité, si petit que soit ε.

Dans ma note, j'ai donné précisément des théorèmes qui se rapportent à la convergence d'une suite de variables pondérées vers un nombre, indépendamment de *l'existence des valeurs moyennes de ces variables.*

Parmi ces théorèmes, je vais en rappeler un, qui peut être généralisé et transposé directement dans la théorie des variables pondérées.

Deux variables éventuelles X, Y *forment un système,* si l'on peut écrire les relations nécessaires d'après la théorie formelle des probabilités, pour que la probabilité $p(x_0, x_1 ; y_0, y_1)$ soit bien définie. $p(x_0, x_1 ; y_0, y_1)$ est la probabilité pour que, simultanément, X prenne une valeur de l'intervalle (x_0, x_1) et Y une valeur de l'intervalle (y_0, y_1) et cela quelles que soient les valeurs x_0, x_1 $(x_1 > x_0)$ et y_0, y_1 $(y_1 > y_0)$; ces probabilités $p(x_0,x_1 ; y_0,y_1)$ devront satisfaire à toutes les relations requises par la théorie formelle.

Considérons maintenant deux suites de variables éventuelles

$$(156) \qquad \begin{cases} X_1, X_2, \cdots, X_n, \cdots, \\ Y_1, Y_2, \cdots, Y_n, \cdots \end{cases}$$

telles qu'on ait :

$$(157) \qquad \lim_{n \to \infty} X_n = N, \qquad \lim_{n \to \infty} Y_n = M$$

et telles que les couples de variables X_r, Y_s, correspondant aux indices r et s, forment un système. Comme on le voit, les variables X_r, Y_s, *peuvent être dépendantes.*

Or, si $f(x,y)$ indique une fonction de x, y continue au point $x = N$, $y = M$, et si la fonction $f(X_r, Y_s)$ représente une variable éventuelle à une dimension, il résulte que :

$$(158) \qquad \lim_{\substack{r \to \infty \\ s \to \infty}} f(X_r, Y_s) = f(N, M).$$

3. Convergence uniforme au sens du Calcul des probabilités et convergence uniforme d'une suite de variables pondérées vers un nombre.

Dans mon second travail de 1916 j'ai généralisé la notion de convergence vers une limite, suggérée par la loi des grands nombres.

Ces considérations peuvent paraître aujourd'hui extrêmement simples : elles ne l'étaient pas à l'époque et, à mon avis, elles donnent lieu encore aujourd'hui à de grandes difficultés lorsqu'on veut les utiliser pour l'établissement d'une théorie générale des variables pondérées.

La théorie des variables pondérées devrait être une généralisation de la théorie des fonctions de variables réelles.

Laissons cependant de côté ce problème, considérons la probabilité concernant l'affirmation que la fameuse loi des grands nombres correspondrait à la « convergence vers une limite » au sens de l'analyse, et demandons-nous si, par hasard, cette probabilité ne pourrait pas être nulle. Le problème peut être posé de la manière suivante : Considérons une suite de variables éventuelles indépendantes

$$X_1, X_2, \cdots, X_n, \cdots \tag{159}$$

ayant, respectivement, les valeurs moyennes :

$$m_1, m_2, \cdots, m_n, \cdots \tag{160}$$

et admettons qu'on ait :

$$\lim_{n \to \infty} \frac{m_1 + m_2 + \cdots + m_n}{n} = m. \tag{161}$$

Est-il possible de fixer alors une suite de nombres positifs ε_i, qui tendent vers zéro, et tels qu'à l'existence simultanée des inégalités, en nombre infini

$$\left\{\begin{array}{l} \left|\dfrac{X_1 + X_2 + \cdots + X_n}{n} - m\right| < \varepsilon_n, \\ \left|\dfrac{X_1 + X_2 + \cdots + X_{n+1}}{n+1} - m\right| < \varepsilon_{n+1}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right. \tag{162}$$

corresponde une probabilité

$$P_n = 1 - \eta_n, \tag{163}$$

avec

$$\lim_{n \to \infty} \eta_n = 0\ ?$$

La réponse est affirmative, pourvu que les variables de la suite (159) satisfassent à des conditions très peu restrictives, ainsi que je l'ai montré rigoureusement dans une de mes notes [1] de 1917.

J'ai cherché à déterminer *exactement* une borne inférieure P'_n de P_n, répondant au problème concret examiné et telle qu'elle tendît vers l'unité lorsqu'on faisait croître n indéfiniment.

J'ai dû recourir pour cela au théorème général de Boole, dont j'ai parlé, et à des considérations déjà anciennes [2] au sujet d'un théorème bien connu de Bienaymé (mes raisonnements ont été retrouvés pour des cas particuliers, par d'autres auteurs quelques dizaines d'années plus tard) ; l'application de l'intégrale de Gauss-Laplace [2], ainsi que je l'ai indiqué à la fin de mon travail de 1917.

Du théorème démontré découle naturellement une réponse affirmative à la question posée, même pour les cas particuliers qui se rapportent au théorème de Bernoulli et au cas célèbre des probabilités variables d'une épreuve à l'autre.

A ces considérations se rattachent encore des recherches intéressantes et bien connues de M. KHINTCHINE (1924), de M. KOLMOGOROFF (1928) et de M. LÉVY.

Ces recherches ont trait aux inégalités, en nombre infini,

$$\left| \frac{X_1 + X_2 + \cdots + X_i}{i} - m_{(i)} \right| \leqslant \varepsilon_i, \qquad i = n, n+1, \cdots \tag{164}$$

avec

$$m_{(i)} = \frac{m_1 + m_2 + \cdots m_i}{i}.$$

Les autres précédents cherchent les valeurs minima des ε_i telles

(1) Cf. *loc. cit.* (6).

(2) Cf. *loc. cit.* (18) et Maurice FRÉCHET, *Le generalizzazioni della ineguaglianza di Bienaymé*, Giornale dell'Istituto Italiano degli Attuari, Anno II, nº 1, gennaio, 1931.

(3) Paul LÉVY, *Sulla legge forte dei grandi numeri*, Giornale dell'Istituto Italiano degli Attuari, Anno II, nº 1, gennaio, 1931.

que les inégalités (164) soient vérifiées avec une probabilité qui tende vers l'unité lorsque n croît i définiment.

Sous certaines conditions, ils ont été conduits à écrire

$$\varepsilon_i = c\frac{M_{(i)}}{\sqrt{i}}\sqrt{\frac{2\log\log M_{(i)}}{i}}, \qquad (c > 1)$$

avec

$$M^2_{(i)} = \sum_{r=1}^{i} M(X_r - m_r)^2.$$

Des valeurs plus satisfaisantes pour les ε_i ont été données par M. P. LÉVY dans un mémoire très remarquable, publié dans le *Bulletin des Sciences Mathématiques* (2e série, t. LV, 1931), pour le cas particulier du schéma Bernoullien. On voit facilement qu'en modifiant un peu ses raisonnements, on peut étendre les résultats obtenus sous certaines conditions aux cas des mouyennes de variables aléatoires.

Qu'il me soit permis toutefois d'attirer l'attention sur un problème d'un caractère moins théorique : étant donnée une suite de nombres positifs décroissants et tendant vers zéro, ε_i, déterminer une borne inférieure P'_n *suffisamment* voisine de la probabilité de coexistence des inégalités (164) ; et cela en commençant par une valeur peu élevée de n.

C'est ce point de vue que j'ai adopté ; il a été suivi par M. Ignazio MESSINA (1) qui a utilisé une formule de M. de la VALLÉE-POUSSIN. Aujourd'hui il serait facile d'améliorer les résultats obtenus.

Il faut dire toutefois que le théorème dont je me suis occupé se rapporte également à une suite particulière de variables éventuelles

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \tag{165}$$

dans laquelle les variables X_i satisfont à des conditions particulières.

D'accord avec des remarquables considérations de M. FRÉ-

(1) *Intorno a un nuovo teorema di Calcolo delle probabilità*, Giornale di Matematiche di Battaglini, vol. LVI (1918), Napoli.

Su un nuovo teorema di Calcolo delle probabilità, sul teorema di Bernoulli e sui postulati empirici per la loro applicazione, Supplemento al Bollettino del Lavoro e della Previdenza Sociale, Roma, 1920.

Un teorema sulla legge uniforme dei grandi numeri, Giornale dell'Istituto Italiano degli Attuari, Anno IV, nº 1, gennaio, 1933, XI.

CHET (1), et en tenant aussi compte des recherches (2) de M. LÉVY, je vais donner maintenant la définition suivante, en empruntant à M. FRÉCHET le terme suggére par lui-même de *convergence uniforme* qui se justifie à tous les points de vue.

Une suite de variables pondérées

$$Y_1, Y_2, \cdots, Y_n, \cdots \tag{166}$$

converge uniformément vers une limite m si l'on peut déterminer une suite de nombres positifs, tendant vers zéro,

$$\varepsilon_n, \varepsilon_{n+1}, \cdots \tag{167}$$

et tels que, sur la surface de base, la mesure P_n de l'ensemble commun aux ensembles de points pour lesquels les inégalités

$$\begin{aligned} |Y_n - m| &< \varepsilon_n, \\ |Y_{n+1} - m| &< \varepsilon_{n+1}, \end{aligned} \tag{168}$$

sont respectivement valables, converge vers l'unité lorsque n augmente indéfiniment.

Cette définition est-elle équivalente à celle qui dérive du théorème déjà démontré ?

Elle ne l'est pas pour les mêmes raisons que j'ai déjà indiquées et pour lesquelles j'ai toujours évité d'employer le terme de convergence *stocastique*, introduit par M. SLUTSKY, au lieu de *convergence au sens du Calcul des probabilités* et qui n'est qu'*une convergence en mesure*, comme l'a montré M. FRÉCHET.

D'après la définition donnée, on pourra écrire

$$\operatorname*{Lim_u}_{n \to \infty} Y_n = m. \tag{169}$$

Je souhaite fortement que les très intéressantes études commencées dans ce domaine par MM. FRÉCHET et LÉVY soient poursuivies, et nous permettent de complèter dans un bref délai la théorie des variables pondérées.

(1) *Sur la convergence en probabilité*, Metron, vol. VII, nº 4, Roma, 1930.

(2) *Sur les séries dont les termes sont des variables éventuelles indépendantes*, Studia mathematica, Tom. III, Lwow, 1931.