

# ANNALES DE L'I. H. P.

PAUL LÉVY

## Sur quelques points de la théorie des probabilités dénombrables

*Annales de l'I. H. P.*, tome 6, n° 2 (1936), p. 153-184

[http://www.numdam.org/item?id=AIHP\\_1936\\_\\_6\\_2\\_153\\_0](http://www.numdam.org/item?id=AIHP_1936__6_2_153_0)

© Gauthier-Villars, 1936, 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/>

# Sur quelques points de la théorie des probabilités dénombrables

PAR

PAUL LÉVY

---

## CHAPITRE I

### Principes généraux

I. — C'est en 1909, dans un mémoire rapidement devenu classique (1), que M. Émile BOREL a indiqué les premiers principes de la théorie des probabilités dénombrables, qui a pris depuis quelques années un si grand développement. Bien avant cette date, on avait considéré des suites infinies de variables aléatoires

(I)  $x_1, x_2, \dots, x_n, \dots,$

et établi à leur sujet des théorèmes asymptotiques ; mais, qu'il s'agît du théorème de BERNOULLI ou du théorème plus profond établissant le rôle de la loi de LAPLACE-GAUSS, on commençait par considérer, pour une valeur finie de  $n$ , la probabilité d'un événement dépendant des  $n$  premières variables choisies, et l'on étudiait ensuite sa limite pour  $n$  infini. Dans la théorie des probabilités dénombrables, l'événement même dont on étudie la probabilité dépend de tous les  $x_n$  ; il peut s'agir de savoir si ces variables ont une limite, si elles sont bornées, si

(1) E. BOREL. — *Sur les probabilités dénombrables et leurs applications arithmétiques*. Rendiconti d. circ. mat. di Palermo, t. XXVI (1909), p. 247-271.

la série  $\Sigma x_n$  est convergente. Tout un nouveau champ de recherches était ainsi ouvert par le mémoire de M. BOREL.

2. — Nous supposons essentiellement que les  $x_n$  sont des variables réelles, et dépendant de lois qui peuvent n'être pas indépendantes les unes des autres, mais qui sont bien définies. Nous entendons par là que tout système d'inégalités de la forme

$$(2) \quad a_\nu < x_\nu < b_\nu \quad (a_\nu < b_\nu; \nu = 1, 2, \dots, n),$$

à une probabilité bien déterminée, ces probabilités devant vérifier le principe des probabilités totales ; si certaines inégalités sont remplacées par des égalités, la probabilité considérée est alors aussi bien déterminée <sup>(1)</sup>.

Introduisons maintenant la fonction

$$F_n(x) = \mathbb{F}_{n-1} \{ x_n < x \},$$

$\mathbb{F}_{n-1}$  désignant une probabilité évaluée en fonction de  $x_1, x_2, \dots, x_{n-1}$ , supposés connus ;  $F_n(x)$  définit donc les conditions de l'expérience qui doit déterminer  $x_n$ , et si ces conditions dépendent des variables antérieures, il ne faut pas confondre cette fonction avec  $\mathbb{F} \{ x_n < x \}$ ,  $\mathbb{F}$  désignant une probabilité *a priori* (c'est-à-dire évaluée avant le début des expériences ; on peut observer que  $\mathbb{F}$  est la valeur probable de  $\mathbb{F}_{n-1}$ ). Cette fonction  $F_n(x)$ , qui dépend de  $x_1, x_2, \dots, x_{n-1}$ , est naturellement bien définie, sauf peut-être dans des cas dont la probabilité totale est nulle.

Désignons par  $C_n$  la courbe  $y = F_n(x)$ , rendue continue (s'il y a lieu) par l'adjonction de segments parallèles à l'axe des  $y$  correspondant à chacun des points de discontinuité de  $F_n(x)$  ; la donnée de  $y$  suffit à définir un point de  $C_n$ , et par suite  $x$ , sauf peut-être pour une infinité dénombrable de valeurs de  $y$ . Si d'ailleurs on prend pour  $y$  une variable aléatoire  $y_n$ , choisie entre 0 et 1 avec une probabilité uniformément répartie dans cet intervalle, la valeur  $x_n$  de  $x$  que l'on en déduit, bien définie sauf dans des cas dont la probabilité totale est nulle (et dans ces cas la valeur choisie importe peu), est bien la variable aléatoire qu'il s'agissait de définir, dont la fonction des probabilités totales est  $F_n(x)$ .

(1) Ainsi

$$\mathbb{F} \{ a \leq x < b \} = \lim_{\varepsilon \rightarrow 0} \mathbb{F} \{ a - \varepsilon < x < b \}.$$

En opérant ainsi successivement pour  $n = 1, 2, \dots$ , nous voyons que : dans le cas le plus général, on peut, sans rien changer à la loi de probabilité dont dépend l'ensemble des  $x_n$ , les considérer comme des fonctions de variables aléatoires auxiliaires  $y_1, y_2, \dots, y_n, \dots$ , indépendantes les unes des autres, chacune d'elles étant choisie entre 0 et 1 avec une répartition uniforme de la probabilité dans cet intervalle.

Si les  $x_n$  sont indépendants les uns des autres, chaque  $x_n$  ne dépend que de  $y_n$  ; dans le cas général  $x_n$  dépendra en outre de  $y_n, y_{n-1}, \dots, y_{n-1}$ .

Observons que la réduction des  $x_n$  à des variables telles que les  $y_n$ , choisies au hasard entre 0 et 1 et indépendamment les unes des autres, est évidemment possible d'une infinité de manières ; celle que nous venons de définir est caractérisée par le fait que  $x_n$  est une fonction non décroissante de  $y_n$ . On peut plus généralement prendre pour  $x_n$  n'importe quelle fonction mesurable de  $y_1, y_2, \dots, y_n$ , et obtenir ainsi une suite de variables aléatoires  $x_n$  réductible à celle des  $y_n$  par un procédé qui en général ne sera pas celui décrit ci-dessus.

### 3. — LA NOTION DE PARTITION DANS LES ENSEMBLES ABSTRAITS

Nous considérons les  $x_n$  comme les coordonnées d'un point  $x$  dans un espace à une infinité de dimensions, et dans cet espace nous considérerons spécialement le cube  $Q_\omega$  obtenu en faisant varier chacun des  $x_n$  de 0 à 1. L'hypothèse que les  $x_n$  sont des variables aléatoires indépendantes les unes des autres et que pour chacune d'elles la probabilité est uniformément répartie dans l'intervalle (0, 1) correspond à l'idée que la probabilité est répartie d'une manière homogène dans  $Q_\omega$  ; s'il en est ainsi la probabilité que  $x$  appartienne à une région  $\mathcal{R}$  est la mesure de cette région. L'étude de ce cas est d'autant plus importante que, comme nous venons de le voir, elle comprend tous les problèmes de probabilités dénombrables. Mais il y a avantage à commencer par quelques remarques qui s'appliquent, non seulement à l'étude d'une répartition quelconque de la probabilité dans  $Q_\omega$ , mais au cas d'un ensemble abstrait  $E$ , considéré comme une collection d'éléments  $e$  de nature absolument quelconque.

La notion de *partition* comprend deux idées. La première est celle de la division de  $E$  en sous-ensembles disjoints  $E_1, E_1', \dots$ , dont chacun sera divisé en sous-ensemble ( $E_2, E_2', \dots$ ), et ainsi de suite indéfiniment ; à chaque opération, on ne distinguera qu'un nombre fini de sous-en-

sembles <sup>(1)</sup>. La seconde idée est celle de la répartition de la probabilité entre les sous-ensembles ainsi distingués ; cette répartition doit être conforme au principe des probabilités totales, c'est-à-dire que le *poids* (ou probabilité) attribué à un sous-ensemble sera définitif et sera, à l'opération suivante, partagé exactement entre les sous-ensembles entre lesquels on le divise. En somme *la partition est une division de E en ensembles partiels, suivie d'une subdivision poussée jusqu'à l'infini, et comportant une répartition déterminée de la probabilité entre les sous-ensembles ainsi distingués.*

Nous appellerons *chaîne* une suite de sous-ensembles  $E_1, E_2, \dots$ , successivement distingués, et dont chacun est une partie du précédent ; une chaîne sera essentiellement supposée prolongée jusqu'à l'infini sauf dans le cas, évidemment possible, où elle aboutit après un nombre fini d'opérations à un  $E_n$  ne contenant qu'un élément ; il peut exister au plus une infinité dénombrable de chaînes finies.

Nous dirons qu'une partition est *parfaite* si toute chaîne infinie aboutit à *un élément et un seul*, c'est-à-dire s'il y a un  $e$  et un seul commun à tous les  $E_n$  qui constituent la chaîne.

L'importance de cette double restriction est évidente. Pour qu'une partition apparaisse comme satisfaisante, il faut d'abord que deux éléments distincts finissent toujours par être distingués, c'est-à-dire qu'aucune chaîne ne comprenne plus d'un élément. Ainsi une partition d'une aire plane divisée en régions toujours séparées par des parallèles à une même droite apparaîtrait comme une partition mal faite.

D'autre part il ne doit pas y avoir de chaîne *vide*. Ainsi, si l'on essaye de définir une partition de l'ensemble des nombres rationnels compris entre 0 et 1 en divisant cet intervalle en intervalles partiels, la plupart des chaînes aboutiraient à des nombres irrationnels étrangers à l'ensemble étudié ; elles ne contiendraient aucun élément de cet ensemble. Une partition de l'ensemble E comprenant ainsi des chaînes vides ne convient pas à l'étude de la répartition de la probabilité dans E <sup>(2)</sup>.

Mais il y a parfois intérêt à se contenter d'une *partition presque*

(1) On peut d'ailleurs sans introduire de difficulté essentielle supposer qu'on en distingue chaque fois une infinité dénombrable.

(2) Il est naturellement facile, pour tout ensemble dénombrable, d'obtenir une partition parfaite. Il suffit, à chaque opération, d'isoler un élément de l'ensemble des autres et de s'arranger pour que tous les éléments de l'ensemble, sauf un élément  $e_0$ , soient ainsi isolés l'un après l'autre ; il y aura ainsi une seule chaîne infinie, aboutissant à  $e_0$ . L'exemple indiqué est donc celui d'une partition mal faite dans un ensemble où une partition parfaite est possible.

*parfaite*, se distinguant de la partition parfaite par l'existence de chaînes vides, ayant une probabilité totale nulle. On aura bien ainsi une répartition de toute la probabilité entre les chaînes non vides, c'est-à-dire entre les éléments de  $E$  <sup>(1)</sup>.

Tel est le cas de la partition obtenue en divisant un intervalle  $I$  en intervalles partiels  $I_n$  indéfiniment subdivisés en intervalles de plus en plus petits. Chaque point séparant deux intervalles ainsi distingués,  $I_n'$  et  $I_n''$ , devant être attribué (ainsi que sa probabilité, peut-être positive) à l'un d'eux, par exemple  $I_n'$ , la chaîne contenant  $I_n''$  et aboutissant géométriquement à ce point sera vide, et aura une probabilité nulle ; comme il n'y a qu'une infinité dénombrable de tels points, on aura obtenu une partition presque parfaite, et ce procédé permet de définir, avec la même division de  $I$  en intervalles partiels, n'importe quelle répartition de la probabilité dans cet intervalle. Des remarques analogues s'appliquent à la division d'une aire plane en aires partielles très petites, par exemple par un quadrillage, avec toutefois cette différence qu'il y aura ici une infinité non dénombrable de chaînes vides.

On obtient évidemment une partition parfaite, liée à la division d'un intervalle  $I$  en intervalles partiels  $I_n$  de longueurs tendant vers zéro pour  $n$  infini, dans tout ensemble  $E$  tel qu'on puisse établir une correspondance bi-univoque entre ses éléments et les points de  $I$ , à cela près que les points séparant deux des intervalles  $I_n$  devront correspondre chacun à deux éléments distincts de  $E$ . Comme cela est possible pour tous les ensembles ayant la puissance du continu, on voit que : *dans tout ensemble ayant la puissance du continu, on peut définir une partition parfaite n'ayant que des chaînes infinies.*

Dans tout ensemble fini ou dénombrable, une partition parfaite est possible, mais comprend nécessairement des chaînes finies (et même, dans le premier cas, ne comprend que de telles chaînes).

J'ai établi en 1924 <sup>(2)</sup> des résultats réciproques du précédent ; une

(1) Si la répartition de la probabilité dans  $E$  est définie *a priori*, cette condition est nécessairement vérifiée. Mais si l'on cherche à définir une loi de probabilité en se donnant arbitrairement la répartition de la probabilité entre les  $E_n$  distingués successivement, on risque d'attribuer une probabilité positive à l'ensemble des chaînes vides.

(2) Conférence faite en janvier 1924 au séminaire de M. HADAMARD, et rédigée en un mémoire publié en 1925 dans la *Revue de Métaphysique et de Morale* et réimprimé à la fin de mon *Calcul des Probabilités*. Cet article contient, à la section V, une erreur, qui m'a été signalée par M. STEINHAUS ; il y a aussi une faute de rédaction à la fin de la section VI, où j'ai oublié de préciser qu'il s'agissait de fonctions bornées.

Mais l'ensemble de mon exposé est indépendant de ces erreurs, et, je crois avoir le droit d'at-

*partition parfaite* (et même plus généralement une partition où deux éléments distincts finissent toujours par être distingués) *n'est possible que dans les ensembles ayant au plus la puissance du continu. Une partition parfaite n'ayant que des chaînes infinies n'est possible que dans les ensembles ayant exactement la puissance du continu.*

Cela résulte immédiatement de ce que *la notion même de partition implique la possibilité d'une représentation linéaire.* En effet on peut faire correspondre l'ensemble  $E$  à un intervalle  $I$ , et à chaque opération de la partition divisant un sous-ensemble  $E_n$  en parties  $E_{n+1}$ ,  $E'_{n+1}$ , ..., faire correspondre à ces parties des intervalles  $I_{n+1}$ ,  $I'_{n+1}$ , ..., dont la réunion constitue l'intervalle  $I_n$  déjà choisi pour représenter  $E_n$ ; on peut de plus supposer que la longueur du plus grand des  $I_n$  tende vers zéro pour  $n$  infini. Chaque chaîne infinie aboutit donc à un point, chaque chaîne finie à un intervalle  $i$ , et tout point de  $I$  est obtenu une fois et une seule, soit comme extrémité d'une chaîne, soit comme élément d'un  $i$ , à l'exception des points séparant deux intervalles  $I_n$ , qui seront obtenus deux fois. Chaque chaîne correspondant à un élément  $e$  et un seul, les résultats énoncés deviennent évidents.

Il importe de remarquer que, si l'on a ainsi une représentation linéaire de la partition, considérée (indépendamment des poids) comme étant un mode de subdivision de  $E$ , il suffit d'attribuer à chaque  $I_n$  le même poids qu'à l'ensemble  $E_n$  qu'il représente, pour avoir une représentation linéaire de la loi de probabilité considérée. On peut d'ailleurs, en partant d'un intervalle  $I$  de longueur unité, s'arranger pour donner à chaque  $I_n$  une longueur égale à son poids. La probabilité d'un *sous-ensemble mesurable* de  $E$  est ainsi égale à la mesure de l'ensemble qui lui correspond dans la représentation linéaire <sup>(1)</sup>, la notion de sous-ensemble mesurable de  $E$  pouvant d'ailleurs précisément être définie par cette relation. Bien entendu, dans ces conditions, s'il existe au

tirer l'attention sur le fait que ce travail, malgré son caractère assez concis dans l'exposé de principes généraux qui me semblaient assez évidents, contenait dès 1925 un exposé d'ensemble de questions développées depuis, dans le cas particulier du cube à une infinité de dimensions, à partir de 1929, par M. JESSEN, qui ne connaissait évidemment pas mon mémoire, et en 1930 par M. M. STEINHAUS, qui à cette époque devait connaître déjà mon mémoire, mais ignorer le premier travail de M. JESSEN.

(1) Cette remarque, que je crois avoir faite dans ma conférence de 1924, a été omise dans la rédaction trop concise qui en a été publiée en 1925; elle se trouve, je crois, explicitement imprimée pour la première fois, dans le cas particulier du cube à une infinité de dimensions, dans une communication de M. B. JESSEN au Congrès des mathématiciens scandinaves qui s'est réuni à Oslo en 1929.

## THÉORIE DES PROBABILITÉS DÉNOMBRABLES

moins un élément de  $E$  qui ait une probabilité positive, et qui ne soit pas distingué des autres après un nombre fini d'opérations, la longueur du plus grand des  $I_n$  ne tendra pas vers zéro pour  $n$  infini.

Précisons enfin un point important. On ne sait pas actuellement s'il existe des ensembles non dénombrables ayant une puissance inférieure à celle du continu. Mais nous pouvons montrer que s'il en existe, il est impossible d'y définir une répartition de la probabilité, à l'aide d'une partition parfaite (ou presque parfaite), autrement qu'en attribuant toute la probabilité au plus à une infinité dénombrable d'éléments  $e_h$  ( $h = 1, 2, \dots$ ) ayant chacun une probabilité positive, le reste de l'ensemble  $E$  considéré ayant une probabilité nulle. Si en effet cette probabilité  $\alpha$  est positive, nous pouvons remplacer par zéro les probabilités positives des éléments  $e_h$  et multiplier les autres par  $\frac{1}{\alpha}$ . Il n'y a plus après cette opération d'éléments à probabilité positive, mais il peut y avoir une infinité dénombrable d'éléments  $e'$  distingués des autres comme étant l'aboutissement de chaînes finies. Cela apparaît même comme nécessaire, car nous savons déjà que s'il n'en était pas ainsi  $E$  devrait avoir la puissance du continu ; mais nous allons montrer que, même s'il y a de telles chaînes, on n'échappe pas à cette conclusion en contradiction avec notre hypothèse.

Considérons en effet la partition obtenue, dont nous savons maintenant qu'aucune chaîne, finie ou infinie, n'aboutit à une probabilité positive, et la représentation linéaire, conservant la mesure, de cette partition, sur l'intervalle  $I$ , complété s'il le faut par des points  $t'$  qui correspondront aux éléments  $e'$ . Les longueurs de tous les intervalles  $I_n$  distingués dans cette représentation tendront vers zéro (autrement il y aurait des chaînes aboutissant à un intervalle  $i$ , qui devrait correspondre à un élément  $e$  à probabilité positive). Chaque point de  $I$  est donc l'aboutissement d'au moins une chaîne, non vide, et correspond donc à au moins un élément  $e$  de l'ensemble  $E$ , qui contrairement à l'hypothèse faite, doit avoir au moins la puissance du continu ; le résultat énoncé est donc démontré ; on montre aisément qu'il subsiste dans le cas d'une partition presque parfaite.

*On ne peut donc définir de véritables lois de probabilité (comme je l'ai indiqué dans mon mémoire déjà cité de 1925) que dans les ensembles finis, ou dénombrables, ou ayant la puissance du continu.*



4. — APPLICATION AU CUBE  $Q_\omega$

Revenons au cube  $Q_\omega$ , défini par les formules

$$0 \leq x_n < 1, \quad (n = 1, 2, \dots).$$

L'ensemble de ses points  $x$  ayant la puissance du continu, une partition parfaite est possible. Mais il est préférable de considérer une partition presque parfaite.

Utilisons, pour fixer les idées, la numération décimale. Désignons par

$$(3) \quad a_{n,1}, a_{n,2}, \dots, a_{n,p}, \dots$$

la suite des décimales de  $a_n$ , et rangeons tous les  $a_{n,p}$  en une suite unique

$$(4) \quad b_1, b_2, \dots, b_j, \dots$$

On obtient ainsi une partition, la donnée successive des  $b_j$  définissant une chaîne, et il suffit de les considérer comme les décimales d'un nombre unique  $t$  pour avoir la représentation linéaire de cette partition. Il faudrait considérer les deux représentations possibles pour chaque valeur décimale de chacun des  $x_n$  comme deux nombres distincts pour avoir une partition parfaite. En choisissant pour fixer les idées celle qui contient une infinité de zéros (de sorte que la donnée des  $p$  premières décimales définira toujours un intervalle semi-ouvert), et en lui attribuant toujours le poids qu'il s'agit d'attribuer à la valeur considérée de  $x_n$ , on est sûr que les chaînes vides ainsi écartées ont un poids total nul <sup>(1)</sup>. La partition est donc presque parfaite. Au point de vue de la représentation linéaire, les chaînes vides correspondent à un ensemble de valeurs de  $t$  qui a la puissance du continu, et une mesure nulle ; les autres valeurs de  $t$  sont obtenues une fois et une seule. Toute loi de probabilité définie dans  $Q_\omega$  à l'aide de la partition considérée a une représentation linéaire qui la définit complètement, mais qui n'est pas quelconque, puisqu'elle donne un poids nul à l'en-

(1) Elles apparaissent en effet comme groupées en une infinité dénombrable de groupes (non disjoints), correspondant chacun à un entier  $n$  et une valeur décimale de  $x_n$ , ayant chacun un poids nul. D'ailleurs la remarque générale de la note 5 suffit pour justifier le résultat énoncé.

Remarquons aussi que la région définie par une infinité d'inégalités de la forme (5) a pour mesure le produit infini  $\prod (b_j a_j)$ . Cette mesure peut être positive, si les  $a_j$  tendent vers zéro et les  $b_j$  vers un ; mais en général elle sera nulle ; elle l'est notamment lorsque les inégalités considérées définissent un cube de côté  $a < 1$ . Contrairement à ce qui se produit dans le cas d'un espace à un nombre fini de dimensions, un ensemble de mesure nulle peut contenir des points intérieurs ; son complément ne sera pas partout dense.

semble des valeurs exclues ; cet ensemble étant de mesure nulle, toutes les lois absolument continues sont effectivement obtenues.

La définition ainsi obtenue pour les lois de probabilité définies dans  $Q_n$  coïncide d'ailleurs exactement avec celle donnée au § 2. Si en effet on connaît la probabilité de tous les systèmes d'inégalités de la forme (2), et par suite aussi de ceux de la forme

$$(5) \quad a_\nu \leq x_\nu < b_\nu, \quad (0 \leq a_\nu < b_\nu \leq 1, \nu = 1, 2, \dots, n).$$

on connaît par la même le poids de toutes les régions  $E_\nu$  définies par la donnée d'un certain nombre de décimales  $a_{n,p}$ . La réciproque est aussi facile à vérifier, l'ensemble (5) étant la réunion d'une infinité dénombrable de régions  $E_\nu$ , extérieures les unes aux autres.

Ce qui précède s'applique en particulier au cas de la répartition homogène de la probabilité dans  $Q_n$ . On peut le définir, soit en disant que l'ensemble (5) a la probabilité

$$\prod_1^n (b_\nu - a_\nu),$$

soit en disant que la région  $E_n$  définie par la donnée de  $b_1, b_2, \dots, b_n$  a la probabilité  $10^{-n}$ . Alors la probabilité de n'importe quel sous-ensemble  $E'$  de  $E$ , qui est la mesure géométrique de  $E'$ , coïncide exactement avec la mesure linéaire de l'ensemble  $I'$ , lieu du point  $t$  quand  $x$  décrit  $E'$  ;  $E'$  et  $I'$  sont mesurables en même temps et ont même mesure. On remarque que dans ce cas la probabilité que l'un quelconque des  $x_n$  soit un nombre décimal étant nulle, les précautions qu'il faut prendre en général pour distinguer les deux représentations possibles de chaque nombre décimal deviennent inutiles.

##### 5. — L'INTÉGRATION AU SENS DE RIEMANN

La notion de partition, que l'on utilise ou non la possibilité d'une représentation linéaire, donne une base excellente à l'extension de la théorie de la mesure, tant au sens de JORDAN qu'au sens de M. LEBESGUE, et à celle de l'intégration, tant au sens de RIEMANN qu'à celui de M. LEBESGUE. Qu'il s'agisse d'un ensemble abstrait quelconque ou du cube  $Q_n$ , les régions distinguées dans la partition jouent le même rôle que dans les régions très petites distinguées dans la défini-

tion des intégrales de type classique (simples ou multiples), et il ne me paraît pas nécessaire d'exposer en détail l'extension évidente de notions bien connues ; de même que dans mon article déjà cité en 1925, je me contenterai d'insister sur les différences entre le cas classique et celui du cube à une infinité de dimensions (1).

Dans l'espace ordinaire, la théorie de l'intégration repose sur le fait qu'un volume fini peut être divisé en éléments très petits dans tous les sens ; une fonction continue étant d'ailleurs uniformément continue, son oscillation dans n'importe laquelle des régions ainsi distinguées est très petite ; il en résulte qu'elle est intégrable.

Il n'en est plus de même ici. Pour la partition que nous venons de définir, chacune des régions distinguées est définie par des inégalités limitant un nombre fini de coordonnées  $x_n$ , tandis que pour les autres l'intervalle de variation reste égal à l'unité. Ce sont des *parallélépipèdes* ayant toujours un certain nombre fini de côtés de longueurs inférieures à l'unité ou même très petites, tous les autres côtés restant égaux à un.

Il n'est d'ailleurs pas possible, quel que soit le procédé employé, d'obtenir des régions *très petites dans tous les sens*. Le sens même des mots soulignés, très clair dans le cas de l'espace ordinaire, a ici besoin d'être précisé. La distance de deux points  $x$  et  $y$  peut être par exemple l'un ou l'autre des nombres  $r$  et  $r'$  définis par

$$r = \text{Max } |y_n - x_n|, \quad (n = 1, 2, \dots, \infty),$$

$$r'^2 = \sum_1^{\infty} (y_n - x_n)^2,$$

et il peut arriver que  $r$  soit très petit et  $r'$  infini. Lorsque nous parlerons plus loin de la sphère, la distance considérée sera  $r'$  ; mais pour l'étude du cube  $Q_\omega$ , la première définition s'impose, de sorte qu'un volume très petit sera un volume intérieur à un cube de côté très petit. Or si, pour commencer, nous décomposons  $Q_\omega$  en cubes de côtés égaux à  $\frac{1}{10}$ , pour désigner un de ces cubes, il faudra se donner la suite infinie des

(1) Pour une étude plus complète de la mesure dans ce cube, considérée indépendamment de la théorie générale des ensembles abstraits, nous ne saurions trop recommander au lecteur de se reporter à un remarquable mémoire de M. B. JESSEN (*Acta mathematica*, vol. 63, 1934, p. 249-323), qui, à côté d'une excellente mise au point de cette théorie de la mesure, contient des résultats importants et nouveaux dont nous indiquerons quelques-uns dans la suite (nouveaux ou du moins dus à ce savant ; je ne puis dire au juste lesquels de ces résultats se trouvaient déjà dans sa thèse, soutenue en 1930, et rédigée en danois).

décimales  $a_{n,1}$  ( $n = 1, 2, \dots$ ), ce qui montre que l'ensemble de ces cubes a la puissance du continu. De même qu'il est impossible de définir une répartition continue de la probabilité dans un intervalle en le considérant comme une réunion de points dont chacun a une probabilité nulle, il est impossible ici de prendre une telle décomposition comme le point de départ d'une partition (on peut à la rigueur distinguer à chaque opération une infinité dénombrable de régions, mais pas davantage).

Cette circonstance est liée au fait qu'il n'y a pas de lemme de WEIERSTRASS-BOLZANO ; si par exemple nous appelons  $x^{(n)}$  le point pour lequel  $x_n = \frac{1}{2}$ , et  $x_\nu = 0$  pour  $\nu \neq n$ , la distance  $r$  de deux quelconques des points  $x^{(n)}$  est  $\frac{1}{2}$ , et il n'y a pas de point d'accumulation. Il résulte de ces circonstances qu'une fonction  $f(x)$  continue n'est pas toujours uniformément continue, et qu'une fonction constante dans chacun des cubes de côté  $\frac{1}{10}$  dont nous venons de parler n'est pas pour cela intégrable.

J'ai montré dans mon mémoire de 1925 déjà cité qu'on n'échappe à ces difficultés, pour l'étude des volumes situés dans un espace à une infinité de dimensions, que pour des volumes dont les dimensions successives sont de plus en plus petites (1), ce qui revient à dire, si l'on adopte toujours la même définition de la distance, qu'ils sont intérieurs à un parallélépipède  $P$  défini par des inégalités de la forme

$$a_n \leq x_n < b_n, \quad (n = 1, 2, \dots, \infty; \quad b_n - a_n \rightarrow 0).$$

Alors une fonction continue  $u$  peut, avec une approximation aussi grande que l'on voudra, être représentée par une fonction d'un nombre fini de coordonnées de  $x$ , et son intégrale dans  $P$  sera facile à définir, à condition de prendre comme unité le volume de  $P$  (et non celui de  $Q_\omega$  qui apparaît alors comme infini) ; on peut aussi, sans parler de l'unité de volume, dire qu'on a défini la valeur moyenne de  $u$ , et non son intégrale.

M. JESSEN, dans son mémoire de 1934 cité plus haut (*loc. cit.* 9), se place à un point de vue un peu différent, en se contentant de considérer

(1) J'avais déjà, dans mes Leçons d'analyse fonctionnelle publiées en 1922, attiré l'attention sur le rôle de ces volumes dans la théorie de l'intégration.

ce qu'on peut appeler la *convergence faible*. Il y a convergence faible de  $x$  vers une limite  $y$  quand toutes les différences  $y_n - x_n$  tendent vers zéro, tandis que la distance  $r$  considérée tout à l'heure ne tend vers zéro que si ces différences tendent *uniformément* vers zéro. Cette notion est naturellement liée à la partition du cube  $Q_\omega$ , car une suite de points  $x^{(\nu)}$  ( $\nu = 1, 2, \dots$ ) tend nécessairement vers  $x$  (au sens de M. JESSEN) si, après les  $\nu$  premières opérations de la partition (c'est-à-dire lorsqu'on connaît  $b_1, b_2, \dots, b_\nu$ ),  $x^{(\nu)}$  est dans la même région  $E_\nu$  que  $x$ ; en effet, quels que soient  $n$  et  $p$  fixes, si  $\nu$  est assez grand, parmi les nombres  $b_1, b_2, \dots, b_\nu$  supposés connus se trouvent les  $p$  premières décimales de  $x_n$ , qui déterminent ce nombre avec une précision d'autant plus grande que  $p$  est plus grand. Il y a aussi un lemme de WEIERSTRASS-BOLZANO, car si un ensemble contient une infinité de points, il en sera de même pour au moins un des  $10^\nu$  régions définies par la donnée de  $b_1, b_2, \dots, b_\nu$ , ce qui permet de former au moins une chaîne aboutissant à un point d'accumulation.

Cette convergence faible peut d'ailleurs se définir par l'intermédiaire d'une distance; il n'y a qu'à prendre l'un des nombres  $\rho$  et  $\rho'$  définis par

$$\begin{aligned} \rho &= \text{Max } a_n | y_n - x_n |, & (n = 1, 2, \dots, \infty) \\ \rho'^2 &= \sum b_n^2 (y_n - x_n)^2, \end{aligned}$$

les  $a_n$  étant des nombres positifs tendant vers zéro, et les  $b_n$  étant tels que la série  $\sum b_n^2$  soit convergente. La condition nécessaire et suffisante pour qu'il y ait convergence faible de  $x$  vers une limite  $y$  est que  $\rho$  tende vers zéro; de même pour  $\rho'$ . On remarque que, si chaque définition précise de  $\rho$  ou  $\rho'$  dépend de l'ordre des coordonnées (puisqu'il faut que  $a_n$  et  $b_n$  varient avec  $n$ ; les coordonnées ne jouent donc pas toutes le même rôle) la notion de convergence faible (comme pour les modes de convergence liés à  $r$  ou à  $r'$ ) est indépendante de cet ordre. On remarque aussi que cette définition de la distance donne au cube  $Q_\omega$  le caractère, qui lui manquait d'abord, d'un volume ayant ses dimensions successives de plus en plus petites <sup>(1)</sup>. Les résultats de

(1) Toutefois il ne faut pas oublier qu'elle prend ce caractère en vertu d'une définition en quelque sorte trompeuse de la distance. Il est plus naturel, comme nous l'avions fait d'abord, de considérer que les régions  $E_\nu$  distinguées successivement ne sont jamais petites dans tous les sens; que si un point  $x^{(\nu)}$  est, pour chaque valeur de  $\nu$ , choisi au hasard dans la même région  $E^\nu$  que  $x$ , la suite des  $x^{(\nu)}$  converge faiblement vers  $x$ , tandis qu'il y a une probabilité unité pour qu'au sens de la convergence forte  $x$  ne soit même pas pour cette suite un point d'accumulation.

M. JESSEN rentrent donc dans ceux que j'ai indiqués autrefois au sujet de l'intégration dans les volumes de ce genre. Une fonction  $f(x)$  continue, c'est-à-dire ici telle que  $f(x)$  tend vers  $f(y)$  quand  $x$  tend vers  $y$  au sens de la convergence faible, est toujours la limite uniforme, pour  $n$  infini, d'une fonction de  $x_1, x_2, \dots, x_n$ . Son oscillation maxima, dans n'importe laquelle des régions  $E_n, E_n', \dots$ , définies par la donnée de  $b_1, b_2, \dots, b_n$ , tend vers zéro pour  $n$  infini ; elle est donc intégrable.

Pour une fonction bornée, la condition nécessaire et suffisante pour qu'elle soit intégrable au sens de RIEMANN est que,  $\varepsilon$  et  $\varepsilon'$  étant arbitrairement petits, et  $n$  assez grand, l'oscillation dans n'importe laquelle de ces régions  $E_n, E_n', \dots$ , soit inférieure à  $\varepsilon$ , sauf pour des régions de poids total inférieure à  $\varepsilon'$ . Pour les fonctions non bornées, il y a naturellement une condition de convergence sur laquelle je n'insisterai pas.

On voit aisément enfin que l'intégrale, pas plus que dans la théorie classique, ne dépend de la définition précise des régions  $E_n$  ; deux partitions différentes, si elles donnent le même poids à chacune des régions définies par des formules du type (2), conduisent au même résultat. Toutes ces remarques s'appliquent d'ailleurs aussi bien au cas d'une répartition quelconque de la probabilité dans  $Q_\omega$  qu'au cas d'un ensemble abstrait où une partition parfaite soit possible, à cela près quelconque.

#### 6. — L'INTÉGRATION AU SENS DE M. LEBESGUE

A ce point de vue, plus encore qu'à celui de RIEMANN, la théorie de l'intégration dans  $Q_\omega$  est identique à ce qu'elle est dans le cas des fonctions d'une variable, et si l'on utilise la représentation linéaire, elle se ramène au cas de ces fonctions. Les quelques remarques que nous allons présenter ne sont pas spéciales au cas du cube  $Q_\omega$ , mais ont pour but de vous rappeler quelques points importants de la théorie de M. LEBESGUE.

On sait que cette théorie repose essentiellement sur la notion d'ensemble *mesurable* ; occupons-nous d'abord des ensembles linéaires situés dans l'intervalle unitaire I. Il était évident *a priori* que l'application du principe d'addition que l'on appelle principe des probabilités totales quand on s'occupe de calcul des probabilités, devait per-

mettre de mesurer certains ensembles, que l'on appellerait mesurables, les autres devant échapper à ces procédés. C'est le caractère de ces ensembles qu'il s'agissait de préciser.

Observons d'abord qu'un ensemble  $E$  a sûrement une mesure inférieure à  $M$  s'il peut être enfermé dans une infinité dénombrable d'ensembles de mesure inférieure à  $M$ . Sa *mesure extérieure*  $\overline{m}$  est la borne inférieure de ces limites supérieures  $M$  ; sa *mesure intérieure*  $\underline{m}$  se déduit de la considération de son complément dans  $I$ . Si  $\overline{m} = \underline{m}$ ,  $E$  est mesurable, et sa mesure  $m$  est la valeur commune de  $\overline{m}$  et de  $\underline{m}$ .

On en déduit aisément que, quelque petit que soit  $\varepsilon$ ,  $E$  peut, à  $\varepsilon$  près, être confondu avec un ensemble constitué par un nombre fini  $n$  d'intervalles  $L_1, L_2, \dots, L_n$ ,  $n$  étant d'autant plus grand que  $\varepsilon$  est plus petit ; « à  $\varepsilon$  près » signifie que les points de  $E$  extérieurs à  $E'$  et les points  $E'$  extérieurs à  $E$  forment un ensemble de mesure extérieure inférieure à  $\varepsilon$ . De cette remarque résulte ensuite aisément que la réunion de deux ensembles mesurables conserve ce caractère ; il en est de même de leur partie commune, et cette remarque s'étend au cas d'une infinité dénombrable d'ensembles mesurables. Aucune de ces opérations ne permet donc d'étendre la notion d'ensemble mesurable, et l'on a pu montrer que cela n'est possible d'aucune manière.

On peut en somme dire, bien que ces ensembles mesurables aient une généralité bien plus grande que les ensembles mesurables  $B$  (au sens de M. BOREL), qu'on les obtient assez simplement en partant des intervalles. On se rend compte qu'il doit exister des ensembles non mesurables, définis par des procédés qui n'aient pas pour point de départ le groupement de points en intervalles complets ; il doit exister d'autres groupements de points. Je ne veux pas insister sur les procédés par lesquels on a pu établir effectivement l'existence d'ensembles non mesurables. Ces ensembles n'en restent pas moins assez mystérieux, et l'on n'a pas pu, on ne pourra sans doute jamais, en donner d'exemple précis ; cela tient à ce que le caractère de groupements en intervalles se retrouve à la base des définitions de toutes les propriétés qui caractérisent certains ensembles, même lorsque l'on utilise des propriétés arithmétiques, comme celle qui caractérise les nombres rationnels, ou les nombres algébriques.

Remarquons que tout intervalle peut, avec une erreur aussi petite que l'on veut, être confondu avec un intervalle dont les extrémités

## THÉORIE DES PROBABILITÉS DÉNOMBRABLES

sont des nombres décimaux. En appliquant cette remarque aux intervalles  $I_1, I_2, \dots, I_m$ , considérés il y a un instant, on voit que :  $\varepsilon$  étant arbitrairement petit, tout ensemble mesurable peut, à  $\varepsilon$  près, être confondu avec un ensemble constitué par un nombre fini d'intervalles, dont chacun est un des intervalles  $I_n$  distingués dans la partition de  $I$  basée sur la numération décimale.

Le principe de la représentation linéaire nous montre alors que : l'énoncé précédent s'applique au cas des ensembles mesurables dans  $Q_\omega$  (comme dans tout ensemble abstrait), sans autre changement que de remplacer chaque intervalle  $I_n$  par l'ensemble  $E_n$  dont il est l'image.

Cette remarque importante nous servira au chapitre II.

### 7. — REMARQUES SUR L'HISTOIRE DE LA NOTION DE MESURE ET SON APPLICATION AUX PROBABILITÉS DÉNOMBRABLES

La théorie de la mesure a fait récemment des progrès importants, et quelques-uns des résultats dont nous parlerons au chapitre II datent de 1934. Mais la lecture de quelques mémoires où ces résultats sont précédés par un exposé des principes que nous venons de rappeler peut faire perdre de vue qu'il s'agit simplement d'une mise au point de principes connus. J'ai déjà parlé de mon mémoire de 1925 ; j'avais dès 1919 appliqué la notion de partition à la sphère à une infinité de dimensions ; mais il serait injuste, non seulement de ne pas citer les travaux antérieurs de M. M. FRÉCHET (1915) et P. J. DANIELL (1918-1919) sur l'intégration dans les espaces abstraits, mais même de ne pas remonter bien plus loin en arrière.

Toute la théorie de la mesure repose sur un principe d'addition. Il importe peu que l'on emploie le mot de probabilité, ou qu'on exprime ce principe en disant que la mesure est une fonctionnelle additive. Il a pris toute sa portée le jour où l'on a remarqué qu'il s'appliquait à l'addition d'une infinité dénombrable d'ensembles disjoints, mais non à l'addition d'une infinité non dénombrable d'ensembles ou d'éléments. Sur ce dernier point, la notion de répartition continue a dû enlever tout doute depuis bien longtemps. Pour le premier, il n'est pas sans intérêt de remarquer que GAUSS avait déjà parlé de la probabilité pour que, dans le développement en fraction continue d'un nombre choisi au hasard entre 0 et 1, le quotient complet de rang  $n$  dépasse



une valeur donnée ; c'était, dans un cas particulier, parler de la mesure d'un ensemble formé par la réunion d'une infinité dénombrable d'intervalles. La difficulté, pour GAUSS, était de calculer cette mesure. J'imagine qu'il eût été très étonné si on lui avait dit qu'il y ait du mérite à la définir ; il l'eût d'ailleurs été aussi, je n'en doute pas, si on lui avait dit qu'il faudrait démontrer le théorème de BOREL-LEBESGUE.

En somme, l'outil existait ; mais GAUSS ne pouvait avoir l'idée qu'il y ait intérêt à étudier les ensembles de points les plus généraux, soit sur une droite, soit dans un espace à  $n$  dimensions ou à une infinité de dimensions, soit même dans un espace abstrait quelconque. L'outil existait, mais son champ d'application maximum n'avait pas été défini (1). Le grand mérite de JORDAN, puis de MM. BOREL et LEBESGUE, a été de comprendre l'importance de cette étude générale des ensembles. Il a fallu alors énoncer avec précision certaines règles, évidentes dans des cas simples et qu'à cause de cette évidence même on n'avait cru utile de formuler, pour savoir si elles s'étendaient aux cas les plus généraux. Dans certains cas l'extension a été facile ; d'autres problèmes plus difficiles ont été posés et souvent résolus ; je ne peux essayer d'esquisser ici un exposé d'ensemble de ces travaux.

Mais lorsqu'il apparut qu'il y avait intérêt à introduire l'espace à une infinité de dimensions, et même des espaces abstraits plus généraux, la théorie de la mesure était assez avancée pour s'adapter sans effort à cette extension. Lorsqu'en 1909 M. BOREL étudia pour la première fois des probabilités dépendant d'une infinité dénombrable de paramètres, on ne peut pas dire qu'il ne savait pas de quoi il parlait. On peut d'ailleurs constater qu'il n'a fait que calculer des probabilités en appliquant d'une façon correcte les deux principes fondamentaux, le principe des probabilités composées qui donne la mesure de chaque élément du cube  $Q_n$ , et le principe des probabilités totales qui en déduit la mesure d'ensembles plus complexes ; on n'a rien fait depuis, au point de vue des principes, que de mieux définir le champ d'application de cette méthode.

La notion de partition est à la base de l'intégration classique, et surtout de l'extension de la notion d'intégrale de STIELTJES au cas

(1) On peut faire la même remarque à propos de l'apparition du calcul infinitésimal. L'outil existait chez ARCHIMÈDE mais il ne pouvait avoir l'idée qu'il fût intéressant d'étudier des courbes ou surfaces quelconques. Seul le progrès de la notion de fonction a rendu possible les travaux de LEIBNIZ et NEWTON.

des intégrales multiples. Lorsqu'en 1919 j'en ai exposé au Collège de France l'application à la sphère à une infinité de dimensions, on conçoit que, même sans connaître encore les travaux de P. J. DANIELL, je n'aie pas cru devoir attacher une grande importance à cette extension simple de notions classiques. Dans mon livre de 1922 je n'ai consacré que quelques pages au rôle des volumes à une infinité de dimensions de plus en plus petites, et ce n'est qu'en 1924, connaissant alors les travaux de P. J. DANIELL, et N. WIENER, que je me suis décidé à publier un exposé d'ensemble, encore un peu trop concis, de mes idées ; l'essentiel m'avait paru être de montrer à l'aide de la représentation linéaire, qu'il n'existe pas d'autres lois de probabilité que dans les ensembles finis, dénombrables, ou ayant la puissance du continu, et, pour ces derniers, s'ils ont une infinité de dimensions, qu'une théorie de l'intégration basée sur une véritable loi de probabilité n'est possible que dans le cas des fonctions qui sont presque partout limites uniformes de fonctions d'un nombre fini de variables.

Je ne veux pas dire d'ailleurs que les exposés plus récents de MM. JESSEN (*loc. cit.*, 9) et STEINHAUS <sup>(1)</sup> aient été inutiles ; pour prouver qu'ils étaient au contraire indispensables, il suffit de rappeler l'histoire du problème qui a conduit M. STEINHAUS à sa théorie générale <sup>(2)</sup>. Il s'agissait de montrer qu'une série entière dont les coefficients ont des modules donnés et des arguments choisis au hasard ont *en général* leur cercle de convergence comme coupure. En me reportant aux travaux cités par M. STEINHAUS, et au livre de MM. HADAMARD et MANDELBJROJDT sur la série de TAYLOR, j'ai pu constater, non sans surprise, que les savants qui s'étaient occupés de la question de 1917 à 1930, et qui avaient obtenu de très beaux résultats au point de vue topologique, ou bien n'avaient pas songé à donner à l'expression « en général » le sens précis que lui donne le calcul des probabilités, ou bien même y avaient songé et avaient reculé devant l'obscurité des principes. Le problème ne fut posé avec netteté et résolu que par M. STEINHAUS, en 1930.

La même année, sans penser que mes lecteurs risquaient de ne pas savoir de quoi je parlais, je m'étais placé au même point de vue que

(1) H. STEINHAUS. *Sur la probabilité de la convergence de séries*, Studia mathematica, t. II (1930), p. 21-39.

(2) H. STEINHAUS. *Über die Wahrscheinlichkeit dafür, dass der konvergenzkreis einer Potenzreihe ihre natürliche Grenze ist*, Math. Zeitschrift, t. XXXI (1930), p. 408-416.

PAUL LÉVY

M. STEINHAUS pour étudier la croissance des fonctions entières. De 1917 à 1930, de nombreux auteurs, sans penser plus que je ne l'avais fait que des explications sur les principes fussent nécessaires, avaient étudié des problèmes de probabilités dénombrables ; je citerai simplement les travaux de M. CANTELLI sur la loi forte des grands nombres, et ceux de MM. KHINTCHINE et KOLMOGOROFF sur la loi du logarithme itéré et sur la convergence des séries à termes aléatoires.

## CHAPITRE II

### Probabilités égales à zéro ou un

#### 8. — EXEMPLES SIMPLES. LE CUBE ET LA SPHÈRE (A UNE INFINITÉ DE DIMENSIONS)

Nous avons à nous occuper de cas dans lesquels on est assuré *a priori* qu'une probabilité est nécessairement égale à 0 ou 1. Des exemples simples de cette circonstance sont fournis par la remarque que si dans l'espace à une infinité de dimensions, deux volumes sont homothétiques, avec un rapport d'homothétie autre que 1, un point pris au hasard dans l'ensemble des deux (qu'il y ait ou non une partie commune) a une probabilité nulle d'être dans le plus petit ; le rapport des volumes, qui est  $k^n$  dans l'espace à  $n$  dimensions, devient en effet nul ou infini.

Appliquons cette remarque au cube  $Q_\omega$ . Désignons par  $V_1, V_2, \dots, V_n, \dots$ , des cubes homothétiques et concentriques à  $Q_\omega$ , de côtés croissants et qui tendent vers 1. Chacun d'eux ayant pour mesure zéro, il en est de même de leur réunion, et presque tout le volume  $Q_\omega$  apparaît comme constitué par l'ensemble  $S'$  des points extérieurs à tous ces cubes, et par suite situés à une distance nulle de la surface  $S$  de  $Q_\omega$ .

Précisons bien que nous appelons surface de  $Q$  le lieu des points pour lesquels l'un au moins des  $x_n$  a une des valeurs limites 0 ou 1. La probabilité qu'un point choisi au hasard dans  $Q_\omega$  soit sur  $S$  est nulle. Au contraire  $S'$  est le lieu des points pour lesquels la suite des  $x_n$  admet l'une au moins des valeurs 0 et 1 comme valeur d'accumulation ; or il y a une probabilité égale à un, non seulement pour que cette circonstance se produise, mais même pour que la suite des  $x_n$  soit

partout dense. La distance  $\delta$  à  $S$  étant définie comme étant la borne inférieure de la distance du point considéré aux différentes faces du cube, c'est-à-dire la borne inférieure de la suite contenant tous les nombres  $x_n$  et  $1-x_n$ , la condition  $\delta = 0$  caractérise les points de  $S'$  et a une probabilité égale à un.

On peut dire aussi que  $S'$  est l'ensemble dérivé de  $S$ . Étant donné un point  $x$  de  $S'$  n'appartenant pas à  $S$ , on peut trouver sur  $S$ , ou même à l'extérieur de  $Q_\omega$ , une suite de points tendant vers  $x$  ; mais il n'existe pas de ligne continue tendant vers  $x$ .

C'est ainsi que si l'on sépare le plan en deux régions par la cardioïde  $r = 1 - \cos \vartheta$  (en coordonnées polaires), et l'ensemble des segments situés sur les droites  $\vartheta = 0$  et  $\vartheta = \pm \frac{\pi}{n}$  ( $n = 1, 2, \dots, \infty$ ) et définis sur chacune de ces droites par  $1 - \cos \vartheta < r < 1$ , l'origine est un point frontière, mais qu'un point qui se déplace d'une manière continue ne peut pas tendre vers l'origine.

Revenons à l'espace à une infinité de dimensions. Les remarques qui précèdent s'appliquent à des volumes finis de formes absolument quelconques. Si nous choisissons une définition de la distance de deux points  $x, y$  qui, outre les conditions bien connues qui caractérisent la distance au sens de M. FRÉCHET, vérifie celle d'être une fonction linéaire et homogène des  $x_n - y_n$ , tout volume fini a cette propriété que, si l'on y choisit un point au hasard (avec probabilité également répartie), *il y a une probabilité unité que sa distance à la surface soit inférieure à n'importe quel nombre positif donné*. Mais l'exemple de la sphère va nous montrer que pour certains volumes se présentent des difficultés qui n'existent pas dans le cas du cube.

Désignons par  $V_\omega$  l'intérieur de la sphère de centre  $O$  et de rayon  $1$ , c'est-à-dire le lieu des points pour lesquels la somme  $\sum x_n^2$  est finie et inférieure à  $1$ . La distance sera ici la distance  $r'$  du § 5 ; si un point de  $V_\omega$  est à une distance  $r'$  du centre il est à une distance  $1 - r'$  de la surface  $S$ . Si nous considérons des sphères de centre  $O$  et de rayons croissant vers l'unité, leur réunion constitue tout  $V_\omega$  et l'ensemble extérieur à ces sphères n'existe pas. Le raisonnement fait tout à l'heure dans le cas du cube conduit ici à cette conclusion absurde qu'un point étant choisi au hasard dans  $V_\omega$ , il y a une probabilité unité qu'il appartienne à un ensemble qui ne contient aucun point. La véritable conclusion est donc que : *il n'existe dans la sphère aucune définition*

*de la mesure compatible à la fois avec le principe d'homothétie indiqué tout à l'heure et le principe des probabilités totales.*

La difficulté est la même que lorsqu'on parle d'un nombre entier choisi au hasard ; chaque entier a une probabilité nulle, et il n'en est pas de même de l'ensemble de tous les entiers. Le principe des probabilités totales est en défaut. Du moins il ne s'applique pas *sous sa forme complète*, mais s'applique *sous sa forme restreinte*, c'est-à-dire lorsqu'on n'ajoute qu'un nombre fini de cas. De même pour la sphère  $V_\omega$ . Dans les cas de ce genre, si nous parlons de probabilité, il ne faut pas oublier qu'il ne s'agit pas d'une vraie loi de probabilité, vérifiant le principe d'additivité complète, et que l'on puisse définir à l'aide d'une partition.

Dans ces cas, la définition de la probabilité est asymptotique. Si par exemple on dit que,  $n$  et  $n'$  étant deux entiers choisis au hasard, la probabilité qu'ils soient premiers entre eux est  $\frac{6}{\pi^2}$ , il est entendu qu'il s'agit de la valeur limite, pour  $N$  infini, d'une probabilité calculée d'abord en supposant  $n$  et  $n'$  au plus égaux à  $N$ . A la limite, le principe d'additivité totale disparaît. De même la mesure et l'intégration dans la sphère peuvent être définie par une formule asymptotique, qui ne conserve que le principe restreint. Je ne puis que mentionner ici cette théorie, que l'on trouvera exposée dans mes Leçons d'analyse fonctionnelle ; j'indique seulement qu'il résulte de ce caractère asymptotique que les différentes dimensions interviennent successivement, et que leur ordre, qui dans le cas du cube n'intervient pas, joue dans le cas de la sphère un rôle essentiel.

Ces réserves étant faites sur l'application à  $V_\omega$  du langage de la théorie de la mesure, on voit aisément que, si l'on partage la sphère en régions séparées par des plans  $x_1 = \text{const.}$ , celle qui contient le centre, ayant un rayon plus grand, l'emporte sur les autres : *Le voisinage du plan  $x_1 = 0$  constitue presque toute la sphère ; de même pour  $x_n = 0$  ; on peut donc négliger les parties de la sphère qui ne sont pas à la fois voisines de la surface et d'un nombre fini de ces plans.*

On remarque à nouveau l'absurdité du résultat que donnerait le principe d'additivité complète : un point choisi au hasard devrait avec une probabilité unité être à la fois au centre et sur la surface ; en effet en appliquant ce principe on aurait

$$\mathfrak{P}\{x_n = 0\} = \lim_{\varepsilon \rightarrow 0} \mathfrak{P}\{|x_n| < \varepsilon\} = \lim \mathbf{1} = \mathbf{1}.$$

et par suite

$$\mathfrak{F}\{x_1 = x_2 = \dots = x_n = \dots\} = 0,$$

et de même

$$\mathfrak{F}\{r' = 1\} = \lim_{\varepsilon \rightarrow 0} \mathfrak{F}\{r' > 1 - \varepsilon\} = \lim 1 = 1.$$

Ce rôle des plans  $x_n = 0$  a été indiqué depuis longtemps par M. Émile BOREL. Les mêmes conclusions s'appliquent si, au lieu de  $x_n$ , on considère une fonction linéaire de la forme  $\sum a_n x_n$ , avec  $\sum a_n^2 < \infty$ , condition nécessaire et suffisante pour que cette fonction soit défini dans toute la sphère ; à une facteur constant près, elle représente la distance à un plan qui contient le centre, et est presque partout nulle.

J'ai indiqué autrefois un résultat plus général : *pour une fonction  $u$  continue dans la sphère, il existe toujours une valeur  $m$  telle que le voisinage de la surface  $u = m$  constitue presque toute la sphère*. Si elle est uniformément continue, elle est donc presque partout presque égale à  $m$  ; pour une telle fonction, une seule valeur est prépondérante, et il n'y a pas de vraie répartition de la probabilité (1).

Pourtant il s'agit d'un ensemble ayant la puissance du continu ; une partition  $\gamma$  est possible. Mais il n'est pas certain (Cf. Note 5) qu'on puisse définir les poids attribués aux régions distinguées de manière que la probabilité obtenue soit une *mesure géométrique* (vérifiant notamment le principe de l'égalité des mesures de deux volumes égaux, et le principe d'homothétie déjà mentionné ; on peut les réunir en un principe de similitude). Les remarques qui précèdent prouvent que c'est en effet impossible. Il peut être utile de montrer par un exemple à quoi tient cette impossibilité.

Imaginons qu'on cherche à définir une partition d'une sphère de rayon  $\frac{1}{2}$  inscrite dans le cube  $Q_\omega$  en utilisant la partition du cube indiquée au § 4 ; il n'y aura naturellement qu'à tenir compte des portions intérieures à la sphère des régions  $E_\nu$  distinguées dans l'étude du cube. Alors les régions dont la frontière contient le centre seront prépondérantes, les autres étant négligeables (le nombre 10 étant

(1) Paul LÉVY. *Leçons d'analyse fonctionnelle*, p. 274-280. La démonstration utilise des résultats établis p. 270-273 ; il faut noter, ce que j'avais oublié de préciser, que ces résultats subsistent lorsque la variété désignée à cet endroit par C n'est pas d'un seul tenant. Sans cette remarque on pourrait avoir des doutes sur l'exactitude du théorème rappelé ici.

pair, les plans  $x_n = 0$  seront pris les uns après les autres comme plans de séparation, et le centre sera sur la frontière commune d'un nombre de régions indéfiniment croissant).

Si dans ces conditions une fonction  $u$  est continue au centre, et si son oscillation dans chacune des régions distinguées devient à partir d'un certain moment dans presque toutes ces régions inférieure à  $\varepsilon$ , c'est qu'elle est presque partout presque égale à sa valeur au centre. La partition indiquée, si une fonction est continue au centre, ne permet donc de l'intégrer  $u$  au sens de RIEMANN que si elle est presque partout presque égale à sa valeur au centre.

D'autres modes de partition permettent d'éviter ce rôle du centre ; il n'y a qu'à l'isoler d'abord par une petite sphère. Mais il y aura toujours un point qui jouera le rôle joué par le centre dans l'exemple précédent, et l'on n'arrivera à intégrer une fonction continue que si elle est presque partout presque égale à une constante déterminée.

Si on renonce à la continuité, on trouvera aisément des fonctions intégrables dont la moyenne sera une véritable moyenne (avec poids non nuls) entre différentes valeurs. Ainsi la fonction

$$u = \Sigma \pm \frac{1}{2^n},$$

où le signe précédant  $\frac{1}{2^n}$  est celui de  $x_n$ , est intégrable, si, comme dans le mode de partition que nous avons utilisé, chacun des plans  $x_n = 0$  est un plan de division ; ses valeurs possibles varient de  $-1$  à  $+1$ , avec une répartition uniforme de la probabilité dans cet intervalle.

#### 9. — LES LEMMES DE MM. BOREL ET CANTELLI

Revenons au cas du cube  $Q_\omega$ , et considérons les coordonnées  $x_n$  comme déterminées au cours d'expériences successives et indépendantes les unes des autres, de manière que la notion de probabilité coïncide avec celle de mesure. Soit  $E_n$  un événement dont la réalisation dépend de  $x_n$ , mais peut aussi dépendre de  $x_1, x_2, \dots, x_{n-1}$ . Si elle ne dépend que de  $x_n$ , les différents  $E_n$  sont indépendants les uns des autres ; nous désignerons la probabilité de chaque  $E_n$  par  $\alpha_n$ . Dans le cas général, il faut distinguer la probabilité *a priori*  $\alpha_n$  (évaluée avant le début des expériences), et la probabilité  $\alpha_n'$ , évaluée après

THÉORIE DES PROBABILITÉS DÉNOMBRABLES

les  $n - 1$  premières expériences ;  $x_n$  est la valeur probable (*a priori*) de  $x_n'$ .

Le lemme de M. BOREL, relatif au cas d'événements indépendants, est le suivant :

*Dans le cas où les  $E_n$  sont indépendants les uns des autres la probabilité que  $E_n$  soit réalisé pour une infinité de valeurs de  $n$  est zéro si la série  $\sum x_n$  est convergente et un si cette série est divergente ; elle n'est donc jamais comprise entre zéro et un.*

La première partie, relative au cas de convergence, est évidente. Quelque petit que soit  $\varepsilon$  positif, on peut en effet dans ce cas choisir  $n$  assez grand pour que

$$x_{n+1} + x_{n+2} + \dots < \varepsilon.$$

Or la probabilité totale de la réalisation de  $E_\nu$ , pour au moins une valeur de  $\nu$  supérieure à  $n$  est au plus égale à cette somme, donc inférieure à  $\varepsilon$ . Il en est de même, *a fortiori*, de la probabilité étudiée, qui ne peut être que nulle.

On peut aussi observer, si  $N$  est le nombre de réalisations de  $E_\nu$  pour des valeurs au plus égales à  $n$ , que

$$E \{ N \} = x_1 + x_2 + \dots + x_n < S,$$

$S$  étant la somme de la série  $\sum x_n$ , et le symbole  $E$  désignant la valeur probable

$$E \left\{ N > \frac{S}{\varepsilon} \right\} < \varepsilon,$$

ce qui donne une nouvelle démonstration du résultat énoncé.

Supposons maintenant la série  $\sum x_n$  divergente, ce qui entraîne la divergence du produit infini  $\prod (1 - x_n)$ . D'après le théorème des probabilités composées, la probabilité que  $E_\nu$  soit réalisé pour  $\nu = n$  et ne le soit ensuite pour aucune valeur de  $\nu$  supérieure à  $n$  est

$$\beta_n = x_n(1 - x_{n+1})(1 - x_{n+2}) \dots = 0.$$

La probabilité qu'il en soit ainsi pour au moins une valeur de  $n$  est donc  $\sum \beta_n = 0$ . Or, dire qu'il n'y a aucune valeur de  $\nu$  pour laquelle  $E_\nu$  soit réalisé pour la dernière fois est dire que cet événement est réalisé une infinité de fois. Il en est donc ainsi, sauf dans des cas de probabilité  $\sum \beta_n = 0$ , ce qui achève la démonstration du lemme de M. BOREL.



La première partie de ce lemme subsiste sans changement si les  $E_n$  ne sont pas indépendants les uns des autres (aucune des deux démonstrations indiquées n'utilise l'hypothèse qu'il y a indépendance) ; cette remarque est due à M. CANTELLI.

Au contraire, dans la deuxième partie, la valeur indiquée pour  $\beta_n$  implique cette indépendance. On peut évidemment, s'il n'y a pas indépendance, s'arranger pour que la probabilité étudiée soit comprise entre 0 et 1. Ainsi, supposons les événements  $E_n$  d'indices autres que 1 indépendants les uns des autres, mais dépendants de  $x_1$  (c'est plus général que de supposer qu'ils dépendent de la réalisation de  $E_1$ , ce qui impliquerait qu'on ne distingue que deux cas) ; la probabilité  $\alpha_n$  sera donc connue après la première expérience, et à ce moment on sera dans les conditions requises pour l'application du lemme de M. BOREL. Or nous pouvons prendre pour  $\alpha_n$  n'importe quelle fonction, mesurable et comprise entre 0 et 1, de  $x_1$ . On peut s'arranger (même si l'on ne distingue que deux cas), pour que la série  $\Sigma \alpha_n$  soit tantôt convergente et tantôt divergente, la probabilité  $\beta$  de sa divergence ayant n'importe quelle valeur  $\beta$  donnée entre 0 et 1 ; la probabilité que  $E_n$  soit réalisé une infinité de fois a cette même valeur  $\beta$ . On peut aussi s'arranger (à condition cette fois que les cas distingués soient en nombre infini) pour que la somme  $\Sigma \alpha_n$  soit finie, mais fonction de  $x_1$  et à valeur probable infinie ; par exemple  $\frac{1}{x_1}$ . Alors  $\Sigma \alpha_n$ , qui est la valeur probable de  $\Sigma \alpha_n$ , est infini, et malgré cela il y a une probabilité unité que  $E_n$  ne soit réalisé qu'un nombre fini de fois.

Cette remarque nous met sur la voie d'une circonstance très générale, dans la théorie des probabilités en chaîne : ce qu'il importe de connaître, c'est  $\alpha_n$ , et non  $\alpha_n$  ; c'est ce nombre qu'il faut indiquer à la personne chargée d'organiser l'expérience qui donne à  $E_n$  la probabilité  $\alpha_n$  ; c'est sa connaissance qui permet des prévisions, et si l'on connaît la suite des  $\alpha_n$ , on n'a pas besoin de savoir si ces nombres étaient donnés *a priori*, ou si chacun d'eux a été déduit du résultat des expériences précédentes ; les conséquences seront en tout cas les mêmes. Le lemme de M. BOREL se généralise donc par l'énoncé suivant : *la somme  $\Sigma \alpha_n$  et le nombre de réalisations de  $E_n$ , sauf dans des cas dont la probabilité est nulle, sont tous deux finis ou tous deux infinis* (1).

(1) Paul LÉVY. Propriétés asymptotiques des sommes de variables aléatoires enchaînées. *Bull. des Sciences Math.*, mars-avril 1935.

## THÉORIE DES PROBABILITÉS DÉNOMBRABLES

On peut, en s'inspirant de ce principe, étendre aux suites de variables aléatoires enchaînées la plupart des théorèmes de la théorie des variables aléatoires indépendantes, notamment le théorème de BERNOULLI, la loi forte des grands nombres, et la loi du logarithme itéré. Je renvoie, pour ces questions, à un mémoire qui doit paraître prochainement <sup>(1)</sup>.

### 10. — SUR LES ÉVALUATIONS SUCCESSIVES DE LA PROBABILITÉ DE E

Nous désignerons indifféremment par E un événement dont la réalisation peut dépendre de tous les  $x_n$ , et l'ensemble des points  $x$  de  $Q_m$  pour lesquels cet événement est réalisé. *Nous supposons essentiellement que l'événement E a une probabilité  $\alpha$  bien déterminée*, c'est-à-dire que l'ensemble E est mesurable et a pour mesure  $\alpha$ . La probabilité  $\mathfrak{P}_n \{E\}$ , évaluée lorsqu'on connaît  $x_1, x_2, \dots, x_n$ , est alors bien déterminée, sauf dans des cas de probabilité nulle, et est une fonction mesurable de ces variables. Cela se déduit aisément de ce que la partie commune à deux ensembles mesurables étant mesurable, la probabilité que E soit réalisé, lorsque  $x$  est choisi au hasard dans le volume défini par les formules (2), est bien déterminée; cette probabilité est la moyenne de  $\mathfrak{P}_n \{E\}$  dans ce volume.

**THÉORÈME.** — *Sauf dans un ensemble de mesure nulle,  $\mathfrak{P}_n \{E\}$  tend vers un dans l'ensemble E et vers zéro en dehors de E.*

En termes moins mathématiques, cela veut dire que l'on peut, presque sûrement, faire au bout d'un nombre d'expériences assez grand, mais fini, de prévisions presque sûres.

La démonstration est très simple. Le théorème est en effet évident si E se réduit à un nombre fini de régions  $R_n$  distinguées dans la partition de  $Q_\omega$  (c'est-à-dire, pour le mode de partition défini § 4, que la réalisation de E dépend d'un nombre fini de décimales  $b_n$ ). L'extension au cas général va résulter de ce qu'en tout cas E peut être assimilé à la réunion E' d'un nombre fini de régions  $R_n$ , l'erreur portant sur les points d'un certain ensemble F de mesure inférieure à un nom-

(1) *Journal de Mathématiques*. Volume dédié à M. GOURSAT (1936). C'est en raison de cette prochaine publication que je ne développe pas ici les résultats indiqués oralement à ma conférence du 20 décembre 1935.

bre arbitrairement petit, qu'il sera commode de désigner par  $\varepsilon\varepsilon'$ ,  $\varepsilon$  et  $\varepsilon'$  étant eux-mêmes arbitrairement petits.

Soit  $G_n$  l'ensemble constitué par les points de  $Q_\omega$ , extérieurs aux ensembles  $G_1, G_2, \dots, G_{n-1}$  supposés définis avant  $G_n$ , et pour lesquels la probabilité que  $x$  appartienne à  $F$ , évaluée lorsqu'on connaît  $x_1, x_2, \dots, x_n$ , dépasse  $\varepsilon$ . Soit  $G$  la réunion de tous les  $G_n$ ; c'est un ensemble mesurable, de mesure  $\eta$ . Dans chaque  $G_n$ , donc dans  $G$ , la densité de  $F$  dépasse  $\varepsilon$  (d'après le théorème de M. FUBINI, qui s'applique d'après l'hypothèse que  $F$  est mesurable, cette densité est en effet la valeur probable de  $\mathfrak{P}_n \{ F \}$  dans  $G_n$ ; cf. JESSEN, *loc. cit.*, 9, § 12); la mesure de  $F$  dépasse donc  $\eta\varepsilon$ . Comme elle est, par définition de  $F$ , inférieure à  $\varepsilon\varepsilon'$ , on a  $\eta < \varepsilon'$ .

Pour tout point extérieur à  $G$ , tous les  $\mathfrak{P}_n \{ F \}$  sont  $\leq \varepsilon$ . A partir du moment où  $n$  est assez grand, si on connaît  $x_1, x_2, \dots, x_n$ , on sait avec certitude si  $x$  appartient ou non à  $E'$ , et, si  $x$  est extérieur à  $G$ , la probabilité pour qu'il appartienne à  $E$  et non à  $E'$ , ou inversement, qui n'est autre que  $\mathfrak{P}_n \{ F \}$ , est  $\leq \varepsilon$  et le reste quand  $n$  augmente. Donc, sauf dans un ensemble  $G$  de mesure  $< \varepsilon'$ , la probabilité  $\mathfrak{P}_n \{ E \}$  est, pour  $n$  assez grand (c'est-à-dire pour  $n > N$ ,  $N$  ne dépendant que de  $\varepsilon$  et  $\varepsilon'$ ),  $\geq 1 - \varepsilon$  dans  $E$  et  $\leq \varepsilon$  en dehors de  $E$ . En faisant tendre vers zéro  $\varepsilon'$ , puis  $\varepsilon$ , on obtient le résultat annoncé.

Nous l'avons obtenu sans utiliser le principe de représentation linéaire exposé § 4 et 5. Si on utilise ce principe, il se ramène à peu près au théorème connu de M. LEBESGUE d'après lequel, si un ensemble linéaire est mesurable, sa densité au voisinage de chacun de ses points est égale à l'unité, sauf pour les points d'un ensemble de mesure nulle. Il y a toutefois entre le théorème de M. LEBESGUE, et la traduction linéaire du nôtre, une différence qui fait que la démonstration directe nous a paru préférable; elle donne à notre théorème un caractère plus élémentaire.

Pour se rendre compte de la portée de ce théorème, il faut se rappeler que, d'après le § 2, tous les problèmes de probabilités dénombrables se ramènent à la détermination successive des coordonnées d'un point de  $Q_\omega$ , dans les conditions où nous venons de nous placer.

## II. — LE LEMME DE MM. KOLMOGOROFF ET JESSEN

Le théorème précédent comprend comme cas particulier un théorème important dû à M. KOLMOGOROFF <sup>(1)</sup> et que M. JESSEN <sup>(2)</sup> à l'inverse de ce que nous venons de faire, prend comme point de départ pour établir plusieurs théorèmes sur l'intégration dans  $Q_\omega$  qui comprennent et dépassent celui que nous venons d'exposer. Je me contente de mentionner ces théorèmes, et ne m'occuperai que de celui qui lui sert de lemme.

Il s'agit du cas où l'ensemble  $E$  n'est pas modifié par le changement d'un nombre fini de coordonnées  $x_n$  ; il est donc caractérisé par une propriété uniquement asymptotique de la suite des  $x_n$ , telle que la convergence d'une série de la forme  $\sum \varphi_n(x_n)$ . Dans ce cas, la probabilité  $\mathfrak{P}_n \{ E \}$ , étant indépendante de  $x_1, x_2, \dots, x_n$ , n'est pas distincte de la probabilité *a priori*  $\mathfrak{P} \{ E \}$ . Comme d'autre part elle doit tendre presque partout vers l'une des limites 0 ou 1, il faut bien que  $\mathfrak{P} \{ E \}$  ait l'une des valeurs 0 ou 1.

Nous venons de nous placer dans le cas même considéré par M. JESSEN ; mais nous n'avons pas pleinement utilisé l'hypothèse faite. L'essentiel est que  $\mathfrak{P}_n \{ E \}$  soit indépendant de  $x_1, x_2, \dots, x_n$ . Donc :

**THÉORÈME.** — *Si un événement  $E$  dont la réalisation dépend seulement de la suite des  $x_n$  a une probabilité déterminée  $\alpha$ , et si l'évaluation de cette probabilité n'est pas modifiée par la donnée d'un nombre fini de variables  $x_n$  (sauf peut-être dans des cas dont la probabilité est nulle), on a nécessairement  $\alpha = 0$  ou  $\alpha = 1$ .*

Ce théorème et le précédent précisent bien la structure des ensembles mesurables, qui apparaît comme assez particulière. Il existe sûrement des ensembles pour lesquels on ne soit pas renseigné par la donnée de  $x_1, x_2, \dots, x_n$ , avec des chances d'erreur tendant presque sûrement vers zéro quand  $n$  augmente indéfiniment, sur la question de savoir

(1) A. KOLMOGOROFF. — *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Berlin, Julius Springer, 1933).

(2) *Loc. cit.* 9 (p. 270). Il y a lieu de noter que, compte tenu du principe de représentation linéaire indiqué § 4, ce lemme est à rapprocher d'un théorème sur les ensembles linéaires établi dès 1916 par M. BURSTIN (*Monatshefte für Math. und Physik*, t. XXVII, 1916, p. 163-165), qui est un corollaire du théorème de M. LEBESGUE, comme le lemme de M. JESSEN est un corollaire de notre théorème du § 10.

si  $x$  est dans l'ensemble. Mais, comme nous l'avons dit, on n'a pu jusqu'ici en *nommer* aucun ; à cause du principe de représentation linéaire, la difficulté est la même sur une droite et dans  $\mathbb{Q}_0$ , et le fait d'introduire une infinité de variables indépendantes ne donne aucun moyen nouveau pour la résoudre.

Pratiquement ces théorèmes ont une très grande portée, comme on va le voir par les applications qui suivent. Pour ne pas trop allonger l'exposé, nous omettons la vérification du fait qu'il s'agit d'ensembles mesurables ; elle nécessite parfois un peu d'attention, mais se fait sans difficulté sérieuse dans tous les cas dont nous allons parler.

## 12. — APPLICATION AUX SÉRIES NUMÉRIQUES

Il résulte évidemment du lemme de M. KOLMOGOROFF que la convergence d'une série aléatoire  $\sum u_n$  à termes indépendants les uns des autres a toujours pour probabilité 0 ou 1 ; il en est de même de sa sommabilité, quel que soit le procédé de sommation employé.

Dans le cas de convergence sûre ou presque sûre, si tous les termes sont rationnels, la probabilité pour que la somme soit rationnelle, ou appartienne à un groupe qui ne change pas par l'addition d'un nombre rationnel, est 0 ou 1 (comme exemple de tels groupes citons : les nombres algébriques ; les nombres dont le développement en fraction continue a ses quotients bornés). Dans cet énoncé, on peut remplacer le mot « rationnel » par algébrique, ou décimal, ou généraliser de bien des manières. Il s'applique en particulier au cas d'un nombre défini par le tirage au sort de ses décimales successives ; la probabilité qu'un nombre choisi au hasard entre 0 et 1 avec une répartition uniforme de la probabilité soit algébrique, celle qu'il soit rationnel, ou que chacune de ses décimales se trouve dans son développement avec une fréquence tendant vers une limite donnée, sont toujours 0 ou 1. Tous ces résultats connus se rattachent au lemme de M. KOLMOGOROFF.

Considérons maintenant le cas de divergence presque sûre ; soit  $S_n$  la somme des  $n$  premiers termes de la série, et  $f(n)$  une fonction indéfiniment croissante de  $n$ . L'expression

$$C = \overline{\lim}_{n \rightarrow \infty} \frac{S_n}{f(n)},$$

n'est pas modifiée par l'addition d'une constante ; la probabilité de

$C < c$  n'est donc pas modifiée par la connaissance d'un nombre fini de termes de la série. Il en résulte que  $C a$ , sauf dans des cas de probabilité nulle, une valeur bien déterminée (pouvant être nulle ou infinie).

J'ai établi en 1931 un résultat plus précis : *Dans le cas de divergence presque sûre, quelle que soit la fonction  $f(n)$ , la probabilité  $\mathfrak{P}\{E\}$  qu'il existe un  $N$  tel que l'on ait  $S_n < f(n)$  pour tout  $n > N$ , est toujours 0 ou 1 (dans le cas de divergence presque sûre). Pour le montrer, il suffit encore de montrer que cette probabilité n'est pas modifiée par l'addition d'une constante  $a$  aux premiers termes. Or on a*

$$\mathfrak{P}\{E\} = \int_{-\infty}^{+\infty} \varphi_n(s) d\mathfrak{P}\{S_n < s\}.$$

$\varphi_n(S_n)$  désignant la probabilité de la propriété considérée  $E$  lorsqu'on connaît  $S_n$  ; c'est évidemment une fonction monotone. L'augmentation de la probabilité due à l'addition d'une constante  $a$ , si par exemple  $a$  est positif, est alors

$$-\int_{-\infty}^{+\infty} \varphi_n(s) d\mathfrak{P}\{s - a \leq S_n < s\} = \int_{-\infty}^{+\infty} \mathfrak{P}\{s - a \leq S_n < s\} d\varphi_n(s).$$

Elle est donc au plus égale en valeur absolue au maximum de  $\mathfrak{P}\{s - a \leq S_n < s\}$  ; or, dans le cas où la divergence est presque sûre, la dispersion de  $S_n$  augmente indéfiniment avec  $\frac{1}{n}$ , c'est à-dire que dans un intervalle, variable ou non avec  $n$ , mais de longueur fixe  $a$ , on ne peut avoir qu'une probabilité tendant vers zéro. La variation de  $\mathfrak{P}\{E\}$ , ainsi bornée par une expression tendant vers zéro, est nulle, et l'on est dans les conditions voulues pour l'application du lemme.

### 13. — APPLICATION AUX SÉRIES ENTIÈRES

Considérons une série entière

$$f(z) = \sum_{\mathbf{1}}^{\infty} a_n z^n,$$

les  $a_n$ , réels ou complexes, étant choisis indépendamment les uns des autres suivant des lois au sujet desquelles nous supposerons d'abord

seulement qu'on ne puisse pas trouver une suite de constantes  $c_n$  telle que la série  $\sum a_n \neq c_n$  soit convergente. D'après le lemme de JESSEN, le rayon de convergence, qui n'est pas modifié par le changement d'un nombre fini de coefficients, a avec une probabilité unité, une valeur déterminée  $R$  (pouvant être nulle ou infinie). Il en est de même,  $M(r)$  désignant le module maximum de  $f(z)$  pour  $|z| = r$ ,  $\varphi(r)$  une fonction qui augmente indéfiniment quand  $r$  tend vers  $R$ , et aussi lentement qu'on veut si  $R$  est fini et plus rapidement que n'importe quelle puissance de  $r$  si  $R$  est infini, de

$$\overline{\lim}_{r \rightarrow R} \frac{M(r)}{\varphi(r)};$$

cette limite a avec une probabilité unité une valeur bien déterminée, qui peut être nulle ou infinie.

On peut préciser cet énoncé d'une manière analogue à celle indiquée tout à l'heure pour les séries numériques convergentes. Le raisonnement est un peu plus délicat ; quoique j'aie encore quelques difficultés dans les détails, je crois pouvoir énoncer le résultat suivant : *la probabilité pour qu'à partir d'un certain moment  $M(r)$  soit borné supérieurement par une fonction  $\varphi(r)$  indéfiniment croissante est toujours 0 ou 1 ;* (pour une fonction restant bornée le théorème est faux lorsque  $\sum |a_n|^2 R^{2n}$  est une série à convergence presque sûre, et trivial dans le cas contraire, que  $R$  soit fini ou infini).

Plaçons-nous maintenant au point de vue considéré en 1930 par M. STEINHAUS et par moi-même, dans nos mémoires cités § 7. Chaque  $a_n$  est de la forme  $c_n e^{2\pi i x_n}$ ,  $c_n$  étant donné, et  $x_n$  choisi au hasard entre 0 et 1 (avec répartition uniforme de la probabilité). Supposons qu'il s'agisse d'étudier l'ensemble  $F$  des directions ayant une propriété donnée, non modifiée par le changement d'un nombre fini de coefficients (par exemple celles des droites de JULIA, si  $R$  est infini, et celles des rayons aboutissant aux points singuliers situés sur la circonférence  $|z| = R$ , si  $R$  est fini). Je vais indiquer un raisonnement, qui m'a été communiqué récemment par M. L. SCHWARTZ, élève à l'École Normale Supérieure, et qui simplifie et généralise celui fait par M. STEINHAUS dans son mémoire cité.

Soit  $\alpha$  la probabilité pour qu'un arc donné du cercle trigonométrique contienne au moins un point définissant une direction qui ait la propriété considérée. Ou bien il existe au moins un arc pour lequel

$\alpha < 1$ , donc  $\alpha = 0$  ; dans ce cas, à cause de l'isotropie des données, si la définition de l'ensemble  $F$  conserve cette isotropie, on a  $\alpha = 0$  pour tous les arcs de même ouverture, et, comme on peut couvrir le cercle avec un nombre fini de tels arcs, l'ensemble est vide, sauf dans des cas de probabilité nulle. Dans le cas contraire,  $\alpha = 1$  pour tous les arcs possibles ; si l'on considère une infinité dénombrable d'arcs, de longueurs tendant vers zéro et couvrant le cercle une infinité de fois, il doit y avoir un point de  $F$  dans chacun de ces arcs ;  $F$  est donc partout dense. Donc : *en supposant que la probabilité  $\alpha$  ne change pas par une rotation d'ensemble de  $F$ , et toujours sauf dans des cas de probabilité nulle, l'ensemble  $F$  est, ou bien vide, ou bien partout dense. Si l'on sait qu'il contient au moins un point, et qu'il est fermé, il contient tous les points du cercle.*

Ainsi il y a une probabilité unité pour que le cercle de convergence soit une coupure (c'est le théorème de M. STEINHAUS) ; pour que, si  $R$  est infini, toutes les directions soient directions de JULIA, et même directions de BOREL, d'espèce maximum (résultats dus à M. L. SCHWARTZ ; ils étaient déjà connus dans des cas particuliers, mais n'avaient pas été démontrés d'une manière tout à fait générale).

De même, pour  $R$  infini, si  $\varphi(r)$  est une fonction telle qu'il y ait une probabilité unité que  $M(r)$  dépasse  $\varphi(r)$  pour des valeurs arbitrairement grandes de  $r$ , il y a aussi une probabilité unité pour que le résultat subsiste si l'on ne considère que les valeurs de  $z$  d'une part, de  $f(z)$  d'autre part, dont les arguments soient sur des arcs donnés, non réduits à un point. Il peut tomber en défaut si l'un de ces arcs se réduit à un point. Dans le même ordre d'idées, considérons le point de la circonférence  $|z| = r$  pour lequel on a exactement  $|f(z)| = M(r)$  ; au voisinage d'une valeur donnée de  $r$ , il est en général unique et varie d'une manière continue avec  $r$  ; mais il existe en général une infinité de valeurs de  $r$  pour lesquelles l'argument de ce point varie brusquement avec  $r$ , et en raison de cette circonstance, les arguments de ces directions peuvent avoir pour points d'accumulation tous les points du cercle trigonométrique sans le recouvrir tout entier une infinité de fois. La première de ces circonstances a toujours une probabilité égale à l'unité (d'après le théorème de M. L. SCHWARTZ) ; on montre aisément que la seconde a pour probabilité 0 ou 1 suivant que la variation de l'argument  $\theta$  entre deux sauts consécutifs est le terme général d'une série presque sûrement convergente ou presque sûrement divergente.



PAUL LÉVY

Si  $R$  est fini, mais  $\sum c_n^2 R^{2n}$  infini, on obtient des résultats tout à fait analogues aux précédents pour la croissance de  $M(r)$  lorsque  $r$  tend vers  $R$ . Mais lorsque  $\sum c_n^2 R^{2n}$  est fini,  $M(r)$  reste presque sûrement fini, et tous ces résultats tombent en défaut [mais il peut arriver qu'ils s'appliquent dans ces cas aux dérivées de  $f(z)$  d'ordres suffisamment élevés].

#### 14. — REMARQUE

Dans ce qui précède, nous nous sommes bornés à des théorèmes très généraux montrant que certaines probabilités n'ont jamais de valeurs comprises entre 0 et 1. Il reste à chercher dans quels cas elles ont la valeur 0 et dans quels cas elles ont la valeur 1. C'est un problème souvent plus difficile, et dont la solution nécessite l'emploi de méthodes pouvant varier d'un cas à l'autre.

Le cas le plus simple, pour ce genre de problèmes, nous est fourni par le lemme de M. BOREL. L'application au problème traité par M. BOREL du lemme de M. JESSEN montre simplement que la probabilité étudiée est toujours 0 ou 1 ; les raisonnements que j'ai rappelés au § 9, afin de vous donner un exemple de ces résultats plus précis, montrent dans quels cas elle a chacune de ces valeurs.

La difficulté dont il s'agit est actuellement résolue dans un grand nombre de cas. Ainsi l'on sait dans quels cas la probabilité de la convergence d'une série à termes aléatoires indépendants les uns des autres est 0 et dans quels cas elle est 1 ; ce résultat est dû à MM. KHINTCHINE et KOLMOGOROFF. Mais il reste dans cet ordre d'idées beaucoup de problèmes non résolus. Ainsi M. FRÉCHET m'a signalé récemment le problème de la sommabilité d'une suite de variables aléatoires indépendantes par la méthode des moyennes arithmétiques ; il s'agit par exemple de savoir si l'on a

$$\lim_{n \rightarrow \infty} \frac{u_1 + u_2 + \dots + u_n}{n} = 0.$$

La probabilité de cette circonstance ne peut être que 0 ou 1, et il s'agirait, d'après les caractéristiques des lois de probabilité dont dépendent les  $u^n$ , de donner une règle aussi simple que possible permettant de savoir si elle est 0 ou si elle est 1. Malgré la simplicité apparente de ce problème, il n'est pas résolu, du moins à ma connaissance.

Manuscrit reçu le 20 décembre 1935.

---

Le Gérant : E. SCHNEIDER.

---

IMP. DES PRESSES UNIVERSITAIRES DE FRANCE, Paris-St-Amand. — 9-12-1936.