

# Les probabilités dénombrables et leur rapport à la théorie de la mesure.

Par

Hugo Steinhaus (Léopol).

M. E. Borel<sup>1)</sup> a montré le premier l'intérêt qui s'attache à l'étude de probabilités dénombrables et il a donné des applications arithmétiques trouvées sur cette voie, parmi lesquelles le théorème suivant connu comme „le paradoxe de Borel“ a attiré l'attention des analystes :

La probabilité pour que la fréquence du chiffre 0 dans le développement dyadique d'un nombre pris au hasard soit égale à  $\frac{1}{2}$  a pour valeur l'unité; on appelle fréquence du chiffre 0 la valeur limite du quotient par  $n$  du nombre de fois que ce chiffre figure parmi les  $n$  premiers chiffres du développement<sup>2)</sup>.

On trouve chez divers auteurs<sup>3)</sup> l'énoncé suivant du même théorème:

*Presque tous* les nombres  $\alpha$  ont la propriété que la fréquence des zéros dans leurs développements dyadiques est égale à  $\frac{1}{2}$ ; *presque tous* signifie que la mesure de Lebesgue de l'ensemble des  $\alpha$  privés de la propriété en question est nulle. Pour obtenir cet énoncé il suffit de modifier verbalement la démonstration primitive de M. Borel sans toucher à son idée.

Le but de cette Note est d'établir un système des postulats pour les probabilités dénombrables qui permettra une fois pour toutes de passer d'une interprétation à l'autre dans les recherches de ce genre. A vrai dire nous nous bornons aux problèmes, où le nombre

<sup>1)</sup> Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. mat. Palermo*, tome XXVII, 1<sup>o</sup> sem. 1909, pp. 247—271.

<sup>2)</sup> loc. cit., chapitre II, n<sup>o</sup> 10, pp. 259—260.

<sup>3)</sup> P. e. F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914, pp. 419—422.

des cas possibles est le même dans chaque épreuve, mais les épreuves sont en nombre infini dénombrable; ces problèmes constituent la première catégorie de M. Borel. Le problème le plus simple de deux cas seuls possibles et également probables conduit au théorème de M. Borel.

Au § 1 nous aurons donc à formuler l'axiomatique d'un jeu de „pile-face“ en généralisant le calcul de probabilités ordinaire qui suffit pour traiter ce jeu quand on considère seulement les parties finies; nous allons montrer ensuite comment une interprétation convenable de notions utilisées dans cette axiomatique fournit une définition axiomatique des ensembles mesurables ( $L$ ); ce sera précisément la définition donnée par M. Sierpiński<sup>1)</sup>. Le § 2 donnera des applications immédiates aux problèmes de la fréquence de chiffres et de la probabilité de convergence de certaines séries numériques, dont le terme général a un signe indéterminé. Au § 3 nous revenons au paradoxe de M. Borel pour lui donner une forme un peu plus précise que celle de M. G. H. Hardy<sup>2)</sup>; nous remarquons que ce n'est pas la forme définitive; le problème de la meilleure formule reste donc à résoudre. Au § 4 nous énonçons une condition nécessaire et suffisante pour qu'un nombre soit „normal“ au sens de M. Borel et nous indiquons quelques généralisations de notre théorie.

### § 1.

Considérons une suite contenant une infinité dénombrable des épreuves successives indépendantes, les seuls cas possibles et s'excluant mutuellement à chaque épreuve étant „rouge“ ( $\rho$ ) et „noir“ ( $\nu$ ) et supposons que les deux cas soient également probables. Une partie ( $a$ ) soit — par définition — une succession déterminée

$$\nu\rho\dots\nu\nu\rho\rho\dots\nu\dots$$

des  $\nu$  et  $\rho$ .

Introduisons  $A$  = l'ensemble de tous les  $a$  possibles,  $E, E' \dots$  = les sous-ensembles de  $A$  et  $\mathfrak{M}$  = la classe de tous les  $E$ .

<sup>1)</sup> Sur une définition axiomatique des ensembles mesurables ( $L$ ). *Bull. Acad. des Sc. de Cracovie*, 1919; pp. 173—178.

<sup>2)</sup> Some problems of Diophantine Approximation, *Acta Mathematica* tome 37 1914, part II.

Supposons qu'il soit possible de donner  $\mathfrak{R} =$  une certaine classe des  $E$  (donc partie de  $\mathfrak{M}$ ) et une fonction d'ensemble  $\mu(E)$  définie pour tous les  $E$  de  $\mathfrak{R}$  de manière que l'on ait:

1°  $\mu(E) \geq 0$  pour tout  $E$  appartenant à  $\mathfrak{R}$

2°  $\overset{n}{E}$  appartient à  $\mathfrak{R}$ , si  $\overset{n}{E}$  désigne un ensemble composé de tous les  $a$  à  $n$  premiers coups donnés (communs);  $n = 0, 1, 2 \dots$  ou  $\infty$ .

2° Si  $\overset{n}{E}$  et  $\overset{n}{E'}$  ne diffèrent qu'au  $i$ -ième coup ( $i \geq n$ ) on a  $\mu(\overset{n}{E}) = \mu(\overset{n}{E'})$ .

2°  $\mu(A) = 1$ .

3° Si  $E_i$  à appartient à  $\mathfrak{R}$  pour  $i = \overline{1}, 2 \dots$  et  $E_i E_k = 0$  pour  $i \neq k$ ,

alors  $\sum_{i=1}^n E_i, \sum_{i=1}^{\infty} E_i$  appartiennent à  $\mathfrak{R}$  et

$$\mu\left(\sum_{i=1}^n E_i\right) = \sum_{i=1}^n \mu(E_i); \quad \mu\left(\sum_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

4° Si  $E_1 \supset E_2$  et  $E_1, E_2$  appartiennent à  $\mathfrak{R}$ ,  $E_1 - E_2$  appartient à  $\mathfrak{R}$ .

5° Si  $E$  appartient à  $\mathfrak{R}$  et  $\mu(E) = 0$ , toute partie  $E'$  de  $E$  appartient à  $\mathfrak{R}$ .

L'énoncé des postulats étant très succinct, quelques remarques explicatives ne seront pas superflues:

Le système des postulats conduit à une classe  $\mathfrak{R}$  composée des ensembles de parties pour lesquels le problème de la détermination d'une probabilité générale est possible. La fonction  $\mu(E)$  est la probabilité cherchée; elle ne prend que des valeurs finies; 1°, 2°, 3° et 4° montrent même que l'on a  $0 \leq \mu(E) \leq 1$ .

Le postulat 2° exige que les parties finies considérées dans la théorie classique des probabilités aient une probabilité définie.

2° exprime le principe de l'égalité des chances pour rouge et noir, qui est une hypothèse; nous remarquons que  $\overset{n}{E}, \overset{n}{E'}$  appartiennent à  $\mathfrak{R}$  en vertu de 2°.

2° détermine la constante  $\mu(A)$  conformément à l'usage du calcul classique des probabilités; sans ce postulat on pourrait satisfaire au système en prenant  $\mathfrak{M}$  pour  $\mathfrak{R}$  et zéro pour  $\mu$ .

3° est le principe connu d'après lequel la probabilité d'une somme logique des événements incompatibles est égale à la somme des probabilités correspondantes aux termes

4° attribue une probabilité (déterminée ensuite par 3°) à un ensemble composant si l'ensemble total et l'autre ensemble composant ont des probabilités déterminées.

5° implique — en vertu de 1°, 3° et 4° — que  $\mu(E') = 0$  si  $\mu(E) = 0$  et  $E'$  est partie de  $E$ . Si un événement a une probabilité nulle tout événement plus spécial a aussi une probabilité nulle.

Nos postulats sont d'accord avec les principes du calcul classique des probabilités. Seuls les postulats 3°, qui généralise pour  $n = \text{infini}$ , un principe classique valable seulement pour  $n = \text{fini}$  et 2° pour  $n = \infty$ , qui attribue une probabilité déterminée à l'ensemble composée d'une seule partie infinie, font exception. Quand on s'occupe seulement des ensembles  $\overset{n}{E}$  pour  $n$  fini (qui ne sont rien d'autre que les parties finies), et de ce qui en résulte par les opérations d'addition et de soustraction appliquées un nombre fini de fois on obtient à l'aide de nos postulats les valeurs  $\mu$  qui correspondent d'après le calcul ordinaire des probabilités à ces combinaisons et on peut faire le calcul sans se servir de deux postulats partiels exceptionnels.

Considérons maintenant le produit  $\mathfrak{K}$  de toutes les classes  $\mathfrak{K}$  qui remplissent les postulats 1°—5°. On vérifie immédiatement que  $\mathfrak{K}$  contient les  $\overset{n}{E}$  et tous les ensembles provenant des  $\overset{n}{E}$  par les opérations d'addition et soustraction en nombre fini. On voit aussi que si il est possible de trouver une fonction  $\mu(E)$  pour tous les  $E$  de  $\mathfrak{K}$  telle que  $\mu$  et  $\mathfrak{K}$  satisfassent à 1°—5°,  $\mu(E)$  sera égale à la probabilité ordinaire pour les parties finies et leurs combinaisons finies. Il faut remarquer que l'introduction de  $\mathfrak{K}$  comme une sorte de „ $\mathfrak{K}$  minimum“ lève l'indétermination qui restait encore  $\overline{\mathfrak{K}}$  n'étant pas unique; il reste à démontrer que  $\mathfrak{K}$  est un  $\overline{\mathfrak{K}}$ . —

Faisons correspondre à toute partie infinie

$$a \text{ (p. e. } \nu q \nu \nu q q \dots \nu q \dots)$$

un nombre réel  $\alpha$

$$(0,010011\dots 01)$$

à savoir ce nombre dont le développement dyadique commence par 0, et dont le  $n$ -ième chiffre après la virgule est 0 ou 1 suivant que le  $n$ -ième coup de  $\alpha$  était noir ou rouge.

Cette convention fait correspondre à tout ensemble  $E$  des parties infinies des ensembles ponctuels sur  $\langle 01 \rangle$ , aux ensembles  $\overset{n}{E}$  certains

intervalles de longueur  $\frac{1}{2^n}$  et spécialement à  $A$  l'intervalle  $\langle 01 \rangle$ .

Nous désignerons les ensembles-images par les mêmes lettres que les ensembles des parties et nous lirons les postulats 1°—5° comme portant sur des ensembles ponctuels linéaires.

Dans cette interprétation les postulats 1°—5° ne sont qu'une répétition presque verbale des postulats dont s'est servi M. Sierpiński <sup>1)</sup> pour définir les ensembles mesurables ( $L$ ) et leur mesure. La seule différence est que le postulat II de M. Sierpiński attribue à tout intervalle  $J$  fini et fermé sa longueur comme mesure  $\mu(J)$  tandis que notre postulat 2° est divisé en trois propositions  $2_1^0, 2_2^0, 2_3^0$  (ce qu'il fallait faire pour mettre à jour le lien entre la théorie des jeux finis et celle des jeux infinis). La définition de  $\mathfrak{L}$  comme produit est la même.

La classe  $\mathfrak{L}$  est précisément la classe des ensembles mesurables ( $L$ ). Ce théorème a été démontré par M. Sierpiński à l'aide de postulats 1°, II, 3°, 4°, 5°, qui constituent son système. Or, II implique évidemment  $2_{1,2,3}^0$ . Si nous démontrons encore que 1°, 2°, 3°, 4°, 5° impliquent II nous établirons l'équivalence de deux systèmes et par là le théorème cité.

Démonstration.  $2_1^0$  signifie: l'intervalle fermé  $\langle \frac{q}{2^n}, \frac{q+1}{2^n} \rangle$  appartient à  $\mathfrak{L}$ , si  $n=0, 1, 2, \dots$  et  $q < 2^n, q=0, 1, 2, \dots$  fini. En effet, les parties ayant  $n$  premiers coups communs fournissent comme images tous les points de  $\langle \frac{q}{2^n}, \frac{q+1}{2^n} \rangle$  les extrémités y comprises. Le cas  $n=\infty$  doit être traité à part: il signifie que tout point de  $\langle 01 \rangle$  appartient à  $\mathfrak{L}$ .

$2_2^0$  signifie  $\mu \langle \frac{q}{2^n}, \frac{q+1}{2^n} \rangle = \mu \langle \frac{q'}{2^n}, \frac{q'+1}{2^n} \rangle$  ( $q < 2^n, q' < 2^n$ ) (et pour  $n=\infty$  que  $\mu(E) = \mu(E')$  si  $E, E'$  sont composés de points uniques).

$2_3^0$  signifie  $\mu \langle 01 \rangle = 1$ .

$2_1^0$  et  $3^0$  impliquent que tous les intervalles dyadiques fermés  $\langle \frac{a}{2^n}, \frac{b}{2^n} \rangle$  de  $\langle 01 \rangle$  appartiennent à  $\mathfrak{L}$ ; en effet un tel intervalle peut être considéré comme une somme des intervalles:

<sup>1)</sup> Sur une définition axiomatique des ensembles mesurables ( $L$ ). *Bull. de l'Acad. des sc. de Cracovie*, 1919; pp. 173—178.



$$\left\langle \frac{q}{2^n}, \frac{q+1}{2^n} \right\rangle + \left\langle \frac{q+1}{2^n}, \frac{q+2}{2^n} \right\rangle + \dots$$

car deux intervalles contigus (qui ont géométriquement un point commun p. e.  $\frac{q+1}{2^n}$ ) peuvent être considérés comme images de deux ensembles des parties  $\overset{n}{E}, \overset{n}{E'}$  sans éléments communs, comme il y a deux parties différentes qui donnent la même image (p. e.  $\frac{q+1}{2}$ ), une partie appartenant à  $\overset{n}{E}$  l'autre à  $\overset{n}{E'}$ .

3° et 2° donnent  $\mu \left\langle \frac{q}{2^n}, \frac{q+1}{2^n} \right\rangle = \frac{1}{2^n}$ , donc — d'après ce qui précède et 3°: —  $\mu \left\langle \frac{a}{2^n}, \frac{b}{2^n} \right\rangle = \frac{b-a}{2^n}$  ( $0 < a < b < 2^n$ ;  $a, b$  entiers,  $n=0, 1 \dots$  fini). A tout intervalle dyadique fermé  $J$  correspond donc sa longueur comme  $\mu(J)$ . Nous avons vu que tout ensemble  $E$  composé d'un point unique quelconque  $a$  de  $\langle 01 \rangle$  appartient à  $\bar{\mathfrak{R}}$ . Il s'ensuit d'après 4°, que les intervalles dyadiques ouverts — (d'une ou de deux côtés) appartiennent à  $\bar{\mathfrak{R}}$ . Or, si  $E$  consiste d'un seul points il fait partie des intervalles dyadiques si courts que l'on veut et 1° 3° et 4° donnent  $\mu(E) = 0$ . Il s'ensuit que  $\mu(J) =$  longueur de  $J$  pour tous les intervalles dyadiques, ouverts ou fermés, de  $\langle 01 \rangle$ . Tout intervalle de  $\langle 01 \rangle$  étant une somme d'un nombre fini ou d'une infinité dénombrable de tels intervalles. 3° achève la démonstration du postulat II, d'après lequel à tout intervalle  $J$  de  $\langle 01 \rangle$  correspond sa longueur comme  $\mu(J)$ . —

Maintenant nous pouvons affirmer sans démonstration que la classe  $\mathfrak{R}$  est la classe des ensembles mesurables ( $L$ ) de  $\langle 01 \rangle$  et que  $\mu(E)$  est égal pour  $E$  appartenant à  $\mathfrak{R}$  à la mesure lebesgienne de  $E$ .

L'indépendance mutuelle de postulats est une question à part. Sans vouloir l'aborder nous remarquons, ce qui suit: Si on appelle (pour le moment) 3°<sub>1</sub> et 3°<sub>2</sub> les deux parties du postulat 3°, dont la première concerne la sommation  $\overset{n}{\Sigma}$  d'un nombre fini de termes, la seconde  $\overset{\infty}{\Sigma}$  celle d'un nombre infini de termes, on peut énoncer suivamment le problème que se pose (entre bien d'autres) M. Ugo Broggi dans sa dissertation sur les axiomes du calcul des proba-

bilités<sup>1)</sup>:  $3_2^0$  est-il une conséquence de  $1^0$ ,  $2_3^0$  et  $3_1^0$ ? M. Broggi donne au § 4<sup>2)</sup> de sa thèse une démonstration pour l'affirmative. Pour réfuter son résultat il suffit de citer la généralisation de la mesure lebesguienne pour tous les ensembles  $E$  de l'intervalle  $<01>$  donnée par M. Banach<sup>3)</sup>, qui démontre l'existence d'une fonction  $\mu(E)$  satisfaisant à  $1^0$ ,  $2_3^0$ ,  $3_1^0$  sans satisfaire à  $3_2^0$ ; là  $\mu(E)$  est égale à la mesure lebesguienne de  $E$ , si  $E$  est mesurable ( $L$ ) et  $\mu(E) = \mu(F)$  si  $E$  est congru à  $F$ ; le domaine  $\mathfrak{E}$  contient tous les ensembles ponctuels de  $<01>$ . M. Broggi considère deux fonctions différentes satisfaisant à  $1^0$ ,  $2_3^0$ ,  $3_1^0$ ,  $\mu(E)$  et  $\mu'(E)$  et il pose immédiatement

$$\mu' = f(\mu)$$

$f$  étant le symbole d'une fonction univoque. Je ne crois pas que cette manière de raisonner soit admissible.

## § 2.

Les applications du principe démontré au § 1 sont immédiates:

1. La probabilité d'une partie déterminée comme p. e.

*ovovov...*

est égale à la mesure d'un ensemble composé d'un seul point, donc à zéro. Zéro peut donc représenter la probabilité d'un événement possible. Nous appellerons un événement dont la probabilité est zéro *presque impossible* et un événement dont la probabilité est unité *presque certain*. Il est aisé de voir que la *presque-certitude* d'un événement n'implique pas sa certitude.

2. Soit  $\varphi(n)$  une fonction positive, indéfiniment croissante et  $C$  une constante (= indépendante de  $n$ ) qui peut avoir aussi les valeurs  $+\infty$ ,  $-\infty$ . Quelle est la probabilité que pour  $\varphi$  et  $C$  donné on ait

$$(1) \quad \limsup_{n \rightarrow \infty} \frac{\nu(n)}{\varphi(n)} = C$$

si  $\nu(n)$  désigne le nombre de coups noirs qui se présentent dans les  $n$  premières épreuves?

<sup>1)</sup> Die Axiome der Wahrscheinlichkeitsrechnung, Göttingen, Dieterich'sche Universitätsdruckerei, 1907.

<sup>2)</sup> loc. cit., pp. 16—19.

<sup>3)</sup> Sur le problème de la mesure, Fundamenta mathematicae IV, Varsovie 1923, pp. 7—33.

Je dis que cette probabilité ne peut être que 1 ou 0. En effet, en se servant de nôtre principe, nous avons seulement à chercher la mesure d'un certain ensemble ponctuel (dont la mesurabilité est facile à démontrer<sup>1)</sup>). Cet ensemble  $E$  est évidemment l'ensemble des points de  $\langle 01 \rangle$  dont les développements dyadiques satisfont à (1) en interprétant  $\nu(n)$  comme le nombre de zéros parmi les  $n$  premiers chiffres. Soient  $\alpha, \beta$  deux nombres de  $\langle 01 \rangle$  différant de  $\frac{1}{2^q}$ ,  $q$  étant un nombre naturel. En excluant les nombres  $\alpha$  dont le développement n'est pas unique et qui forment un ensemble dénombrable les mesures ne changent pas; nous pouvons donc dire que les développements dyadiques de  $\alpha$  et  $\beta$  ne diffèrent que par les  $q$  premiers chiffres tout au plus. La limite (1) aura donc la même valeur pour  $\alpha$  et pour  $\beta$ . Il s'ensuit que  $\alpha$  et  $\beta$  appartiennent tous les deux à  $E$  ou bien tous les deux au complémentaire de  $E$ . On voit que la mesure de  $E$  dans  $\langle 0, \frac{1}{2^q} \rangle$  est égale à la mesure de  $E$  dans  $\langle \frac{1}{2^q}, \frac{2}{2^q} \rangle \dots$  égale à la mesure de  $E$  dans  $\langle \frac{2^q - 1}{2^q}, 1 \rangle$ . Tout intervalle pouvant être approché par des intervalles dyadiques, la mesure de  $E$  est toujours égale dans deux intervalles de longueurs égales. Donc la „densité“ de  $E$  est constante et  $E$  ne peut avoir que la mesure 0 ou 1<sup>2)</sup>. c. q. f. d.

<sup>1)</sup>  $\frac{\nu(n)}{\varphi(n)}$  est une fonction du nombre développé  $\alpha$ ,  $f_n(\alpha)$ ;  $f_n(\alpha)$  est (comme  $\nu(n)$ ) évidemment mesurable, car son image géométrique consiste d'un nombre fini des segments rectilignes parallèles à l'axe des  $\alpha$ . La limite supérieure d'une suite des fonctions mesurables est elle même mesurable, donc  $\limsup \frac{\nu(n)}{\varphi(n)}$  est une fonction mesurable  $g(\alpha)$ ; les solutions de l'équation  $g(\alpha) = C$  forment par là un ensemble mesurable.

<sup>2)</sup> Supposons, par l'impossible,  $0 < k < 1$  et soit  $m(x)$  la mesure de la portion de  $E$  contenue dans l'intervalle  $\langle 0, x \rangle$ . Deux cas sont à distinguer ou bien  $m(1) = 0$ , alors  $k = 0$ ; ou bien  $m(1) > 0$  alors quel que soit  $\varepsilon > 0$  l'ensemble  $E$  contient un sous ensemble parfait  $P$  de mesure plus grande que  $m(1) - \varepsilon$ . L'ensemble parfait  $P$  peut être couvert par un nombre fini d'intervalles  $(\alpha_1, \beta_1), (\alpha_2, \beta_2) \dots (\alpha_q, \beta_q)$  de longueur totale comprise entre  $m(1) - \varepsilon$  et  $m(1) + \varepsilon$ . Puisque, par hypothèse  $m'(x) = k$  partout, on a:  $m(\beta_i) - m(\alpha_i) = k(\beta_i - \alpha_i)$ ; par suite  $\sum_{i=1}^q [m(\beta_i) - m(\alpha_i)] = k \sum_{i=1}^q (\beta_i - \alpha_i) < k[m(1) + \varepsilon]$ .  $P$  étant un sous-ensemble de  $E$  la mesure de la portion de  $P$  contenue dans  $\langle \alpha_i, \beta_i \rangle$  n'est pas supérieure



En interprétant  $\nu(n)$  arithmétiquement la limite supérieure (1) devient une fonction de  $\alpha$ ,  $\alpha$  étant le nombre dont on considère le développement. Je dis que cette fonction de  $\alpha$  est égale presque partout à une constante. En effet, le raisonnement de tout à l'heure prouve que cette fonction — qui est évidemment mesurable comme limite supérieure d'une suite des fonctions mesurables — possède des périodes de longueurs  $\frac{1}{2^q}$ ,  $q$  étant un nombre naturel arbitraire, quand on néglige un certain ensemble des  $\alpha$  de mesure nulle. Un théorème connu de M. Burstin<sup>1)</sup> nous assure que cette fonction est une constante presque partout.

3. La détermination de  $\varphi(n)$  et de  $C$  de manière que la probabilité de (1) soit 1 est facile, quand on suit la voie inverse, à savoir quand on part de théorèmes du calcul élémentaire des probabilités.

En effet, le théorème de Bernoulli affirme que,  $\varepsilon > 0$ ,  $1 > \eta > 0$  étant donnés, on a pour  $n > N(\varepsilon, \eta)$

$$(2) \quad \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| < \varepsilon$$

avec une probabilité qui surpasse  $1 - \eta$ .

Passons à nos images; l'ensemble  $E$  de points dont les développements ont la propriété de remplir l'inégalité (2) à partir d'un certain  $n$  a dans  $\langle 01 \rangle$  une mesure plus grande que  $1 - \eta$ . Par un raisonnement identique à celui des N<sup>os</sup> précédents on conclut que cette mesure — qui est  $> 0$  — est 1. Donc presque tous les points ont la propriété (2). Désignons par  $D(\varepsilon)$  l'ensemble à mesure

nulle de points privés de cette propriété;  $\sum_{m=1}^{\infty} D\left(\frac{1}{m}\right)$  aura aussi une

mesure nulle. Les points  $\alpha$  n'appartenant pas à  $\sum_{m=1}^{\infty} D\left(\frac{1}{m}\right)$  remplissent tous l'inégalité

à  $m(\beta_i) - m(\alpha_i)$ ; donc on a:  $m(1) - \varepsilon \leq \text{mes}(I') \leq \sum [m(\beta_i) - m(\alpha_i)] \leq k[m(1) - \varepsilon]$ . C'est-à-dire:  $m(1) - \varepsilon \leq km(1) - \varepsilon k$ . En passant à la limite pour  $\varepsilon \rightarrow 0$   $m(1) \leq km(1)$ . d'où tenant compte de ce que  $m(1) > 0$  dans le cas considéré et  $k \geq 1$  par sa nature même, on déduit  $k = 1$  e. f. d.

<sup>1)</sup> Über eine spezielle Klasse reeller periodischer Funktionen, Monatshefte für Mathematik und Physik, XXVI; § 2, Satz 3, pp. 234–236.

$$\left| \frac{\nu(n)}{n} - \frac{1}{2} \right| < \frac{1}{m}$$

pour  $n > Q(\alpha, m)$ ;  $m$  étant si grand que l'on veut ces points remplissent la relation

$$(3) \quad \lim_{n \rightarrow \infty} \frac{\nu(n)}{n} = \frac{1}{2}$$

et la mesure de leur ensemble est unité dans  $\langle 01 \rangle$ . C'est précisément le théorème de M. Borel sous sa forme arithmétique; or, nos postulats et le principe d'équivalence permettent de passer immédiatement à l'interprétation originale.

En réalité M. Borel a démontré son théorème pour une base quelconque  ${}_nq^u$  au lieu de  ${}_n2^u$ . Cela revient à calculer la probabilité pour des parties où il y a  ${}_nq^u$  couleurs également probables. On voit que presque tous les nombres  $\alpha$  ont la propriété

$$\lim_{n \rightarrow \infty} \frac{\nu_p(n)}{n} = \frac{1}{q} \quad (p=0, 1, 2 \dots q-1)$$

quand  $\nu_p(n)$  signifie le nombre des fois que l'on rencontre  $p$  parmi les  $n$  premiers chiffres du développement  $q$ -adique de  $\alpha$ . Si  $E(q)$  désigne l'ensemble (de mesure nulle) de nombres privés de la pro-

priété en question pour une base donnée  $q$ , on voit que  $\sum_{q=2}^{\infty} E(q)$

a une mesure nulle. Donc presque tous les nombres ont cette propriété par rapport à toutes les bases  $q$ , que la fréquence relative de tous les chiffres est égale. Ce sont les nombres que M. Borel appelle absolument normaux. La détermination effective d'un tel nombre était — d'après l'opinion exprimée en 1909<sup>1)</sup> par ce géomètre ingénieux — un problème très difficile; M. Sierpiński a résolu ce problème en 1917<sup>2)</sup>.

4. Soit  $\{c_n\}$  une suite numérique. Quelle est la probabilité qu'en tirant les signes  $+$  — d'une urne qui contient toujours autant des boules  $\oplus$  que de boules  $\ominus$ , on rendra convergente la série

$$(4) \quad \sum_{n=1}^{\infty} \pm c_n?$$

<sup>1)</sup> loc. cit. p. 261.

<sup>2)</sup> Une démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et une détermination effective d'un tel nombre; Bulletin de la Société Mathématique de France, t. 45, 1917.

Je dis que cette probabilité est égale à la mesure de l'ensemble  $E$  des points  $\alpha$  ( $0 \leq \alpha \leq 1$ ) pour lesquels la série

$$\sum_{n=1}^{\infty} c_n f_n(\alpha)$$

est convergente, en désignant par  $\{f_n\}$  la suite des fonctions définies par les égalités:

$$f_n(t) = (-1)^k \quad \text{pour} \quad \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \quad \left( \begin{array}{l} k=0, 1 \dots 2^n - 1 \\ n=1, 2 \dots \end{array} \right)$$

$$f_n(1) = -1.$$

Démonstration. Il s'agit de la probabilité pour que la suite:  $+ \dots \pm \dots$  appartienne à un certain ensemble de telles suites à savoir à l'ensemble  $S$  qui est caractérisé par la propriété de rendre (4) convergente. Faisons correspondre à  $+$  le chiffre 0, à  $-$  le chiffre 1; ainsi à toute suite  $\{\pm\}$  nous ferons correspondre un nombre

$$\alpha = 0, \alpha_1 \alpha_2 \dots \alpha_n \dots \quad (\alpha_i = 0 \text{ ou } 1 \text{ pour tout } i).$$

A l'ensemble  $S$  correspond un ensemble ponctuel  $P$  et nous savons — par notre principe d'équivalence — que la probabilité cherchée est égale à la mesure de  $P$  pourvu que  $P$  soit mesurable. Or  $P$  est identique à  $E$  quand on néglige les nombres à développement dyadique ambigu. En effet, d'après la définition de  $f_n$

$$f_n(\alpha) = \pm 1 \quad \text{suivant que} \quad \alpha_n = \begin{array}{l} 0 \\ 1 \end{array}$$

quand  $0, \alpha_1 \alpha_2 \dots \alpha_n$  est le développement unique de  $\alpha$ ; il s'ensuit que

$$\sum_{n=1}^{\infty} c_n f_n(\alpha) \equiv \sum_{n=1}^{\infty} \pm c_n$$

terme à terme. Donc pour que la suite  $+ \dots \pm$  appartienne à  $S$  il faut et il suffit d'une part que l' $\alpha$  correspondant appartienne à  $P$  et il faut et il suffit d'autre part que  $\alpha$  appartienne à  $E$ . L'identité de  $E$  et de  $P$  est donc démontrée en négligeant un ensemble de mesure nulle. Or,  $E$  est évidemment mesurable, comme ensemble des points de convergence d'une série des fonctions mesurables. Il s'en-

suît (que  $P$  est mesurable et) que la probabilité de la convergence de (4) est égale à la mesure de  $E$ .

D'après un théorème récent de M. H. Rademacher<sup>1)</sup> sur la convergence des séries  $\sum_{n=1}^{\infty} c_n f_n(t)$  ou  $\{f_n(t)\}$  est la suite définie à la page précédente, la convergence de

$$\sum_{n=1}^{\infty} c_n^2$$

suffit pour rendre convergente presque partout la série

$$\sum_{n=1}^{\infty} c_n f_n(t).$$

Si donc p. e.:

$$c_n = \frac{1}{n} \text{ où } c_n = \frac{1}{n^{1/2+\varepsilon}}, \text{ où plus généralement } c_n = O\left(\frac{1}{\sqrt{n}(\log n)^{1/2+\varepsilon}}\right),$$

la probabilité de la convergence de  $\sum_{n=1}^{\infty} \pm c_n$  est égale à 1.

Il est aisé de voir que les suites  $\{c_n\}$  for. ent deux catégories: celles pour lesquelles la convergence de  $\sum_{n=1}^{\infty} \pm c_n$  est presque certaine et celles pour lesquelles elle est presque impossible.

En effet, toutes les fonctions  $f_n$  à partir de la  $q+1$ -ième admettent  $\frac{1}{2^q}$  comme période dans  $\langle 01 \rangle$ . L'ensemble  $E$  des points de convergence de  $\sum c_n f_n$  ne dépendant pas de  $q$  premiers termes, la densité de  $E$  dans  $\langle 0, \frac{1}{2^q} \rangle$ ,  $\langle \frac{1}{2^q}, \frac{2}{2^q} \rangle \dots \langle \frac{2^q-1}{2^q}, 1 \rangle$  est la même; or  $q$  est arbitraire, donc la densité est constante et  $E$  est de mesure 0 où 1 c. q. f. d.

<sup>1)</sup> Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen, *Mathematische Annalen*, 87 (1922), pp. 112—138; spécialement le chapitre VI.

## § 3.

5. On a démontré plusieurs fois <sup>1)</sup> rigoureusement le théorème suivant dû à Laplace:

„En désignant par  $\nu(n)$  le nombre des coups noirs à  $n$  épreuves et par  $f(t, n)$  la probabilité que l'on ait

$$(5) \quad \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq \frac{t}{\sqrt{2n}} \quad (t > 0, n = n. nat.)$$

on aura

$$\lim_{n \rightarrow \infty} f(t, n) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-\tau^2} d\tau$$

$t$  étant constant.

En désignant par  $F(t, n)$  l'ensemble des points pour lesquels le nombre  $\nu(n)$  des zéros du développement dyadique satisfait à (5) et par  $X$  la mesure de  $X$  quelque soit l'ensemble mesurable  $X$ , nous aurons pour l'ensemble-limite restreint  $Z_r$  de la suite  $\{F(t, n)\}$  l'inégalité:

$$|Z_r| \leq \lim_{n \rightarrow \infty} |F(t, n)| = \lim_{n \rightarrow \infty} f(t, n) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-\tau^2} d\tau < 1. \text{ Or, l'ensemble}$$

$S(t')$  des points pour lesquels on a

$$\limsup_{n \rightarrow \infty} \sqrt{2n} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq t' < t$$

est par définition de l'ensemble-limite restreint <sup>2)</sup> et du  $\limsup$  une partie de  $Z_r$ , nous aurons donc

$$|S(t')| \leq |Z_r| < 1,$$

ce qui entraîne d'après un raisonnement analogue au § 2, N° 2

$$|S(t')| = 0$$

<sup>1)</sup> p. e. C. Jordan. Cours d'Analyse de l'Ecole Polytechnique, première édition, Paris, Gauthier-Villars, 1883, tome II, pp. 187—190. M. Bernays a donné dans un cours professé à Goettingue tous les calculs nécessaires pour cette démonstration.

<sup>2)</sup> L'ensemble limite restreint de  $\{E_n\}$  est par définition  $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n$ ; l'ensemble limite complet est le complémentaire à l'ensemble limite restreint de  $\{E_n\}$  si  $\bar{E}_n$  désigne le complément de  $E_n$ .



On a donc  $\left| \sum_{k=1}^{\infty} S(k) \right| = 0$ . En dehors de  $\sum_{k=1}^{\infty} S(k)$  — donc *presque partout* — on aura d'après la définition de  $S(k)$

$$(6) \quad \limsup_{n \rightarrow \infty} \sqrt{2n} \left| \frac{v(n)}{n} - \frac{1}{2} \right| = +\infty.$$

Considérons maintenant l'ensemble-limite complet  $Z_c$  de  $\{F(t, n)\}$ . Nous aurons

$$|Z_c| \geq \lim_{n \rightarrow \infty} |F(t, n)| = \lim_{n \rightarrow \infty} f(t, n) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-\tau^2} d\tau > 0.$$

Or, l'ensemble  $I(t')$  pour lequel on a

$$\liminf_{n \rightarrow \infty} \sqrt{2n} \left| \frac{v(n)}{n} - \frac{1}{2} \right| \leq t'$$

contient  $Z_c$  si  $t' > t$  d'après les définitions du *lim inf* et de l'*ensemble-limite complet*; on a donc

$$|I(t')| \geq |Z_c| > 0, \quad |I(t')| = 1.$$

On a donc presque partout

$$\liminf_{n \rightarrow \infty} \sqrt{2n} \left| \frac{v(n)}{n} - \frac{1}{2} \right| \leq t';$$

$t'$  étant un nombre positif arbitraire cela entraîne

$$(7) \quad \liminf_{n \rightarrow \infty} \sqrt{2n} \left| \frac{v(n)}{n} - \frac{1}{2} \right| = 0$$

*presque partout*.

6. Démontrons d'abord une inégalité élémentaire:

$$\sum_{p=0}^n \binom{n}{p} (2p - n)^2 \leq 2^n (2k - 1)!! n^k$$

$k, n$ , étant des nombres naturels et  $(2k - 1)!!$  désignant le produit  $1 \cdot 3 \cdot 5 \dots (2k - 1)$ .

Si  $n = 1$  l'inégalité devient évidente pour tous les  $k$ :

$$\binom{1}{1} (0 - 1)^{2k} + \binom{1}{0} (2 - 1)^{2k} \leq 2 \cdot (2k - 1)!!$$

En procédant par induction nous écrirons

$$\sum_{p=0}^{n+1} \binom{n+1}{p} (2p-n-1)^{2k} = \sum_{p=0}^n \binom{n}{p} [(2p-n+1)^{2k} + (2p-n-1)^{2k}] =$$

$$= \sum_{p=0}^n \binom{n}{p} 2 \left[ (2p-n) + \binom{2k}{2} (2p-n)^{2k-2} + \dots + 1 \right] \leq (\text{en supposant (8)})$$

$$\leq 2^{n+1} \left[ (2k-1)!! n^k + \binom{2k}{2} (2k-3)!! n^{k-1} + \dots + 1 \right];$$

or

$$(\text{lemme}) \quad \binom{2k}{2s} (2k-2s-1)!! \leq \binom{k}{s} (2k-1)!! \quad \text{pour } s \leq k,$$

car

$$\frac{2k \cdot 2k-1 \dots 2k-2s+1}{1 \cdot 2 \dots 2s} \cdot 2k-2s-1 \dots 3 \cdot 1 \leq \frac{k \cdot k-1 \dots k-s+1}{1 \cdot 2 \dots s} \cdot 2k-1 \dots 3 \cdot 1$$

si il est vrai que

$$\frac{2k \cdot 2k-2 \dots 2k-2s+2}{s+1 \cdot s+2 \dots 2s} \leq k \cdot k-1 \dots k-s+1,$$

et cette dernière inégalité est une conséquence de

$$2^s \leq (s+1)(s+2) \dots (s+s).$$

qui vaut pour tout  $s$  naturel. Le lemme donne:

$$\sum_{p=0}^{n+1} \binom{n+1}{p} (2p-n-1)^{2k} \leq 2^{n+1} (2k-1)!! \sum_{s=0}^k \binom{k}{s} n^{k-s} =$$

$$= 2^{n+1} (2k-1)!! (n+1)^k$$

ce qui achève la démonstration <sup>1)</sup>.

La condition nécessaire et suffisante pour que parmi les  $n$  premiers chiffres du développement dyadique d'un nombre  $\alpha$  il ait  $p$  zéros et  $n-p$  unités est que  $\alpha$  soit situé dans un intervalle appartenant à une certaine collection composée de  $\binom{n}{p}$  intervalles n'empiétant pas l'un sur l'autre dont les longueurs sont  $\frac{1}{2^n}$ . La somme

<sup>1)</sup> J'emprunte à un cours lithographié de M. Sierpiński (Léopol 1918, Teoria mnogości II, p. 147) l'indication que M. Madzia a employé l'induction pour  $k=2$ . M. Hausdorff (loc. cit.) mentionne l'existence d'une inégalité pour tout  $k$  fixe.

des longueurs est  $\binom{n}{p} \frac{1}{2^n}$ .  $n$  étant fixe, considérons tous les intervalles  $I_\varepsilon$  tels que les  $p$  correspondants remplissent la condition

$$(9) \quad \left| \frac{p}{n} - \frac{1}{2} \right| > \varepsilon \quad (\varepsilon > 0).$$

La somme des longueurs pour  $p$  fixe étant  $\binom{n}{p} \frac{1}{2^n}$  et les collections correspondantes aux différents  $p$  n'empiétant pas l'une sur l'autre, la somme des longueurs de tous les  $I_\varepsilon$  sera

$$\sum_p^* \binom{n}{p} \frac{1}{2^n}$$

le symbole  $\sum_p^*$  signifiant que l'indice  $p$  ne prend que ces valeurs  $0, 1, 2, \dots, n$  qui obéissent à (9). On aura (pour  $k$  naturel)

$$\sum_{p=0}^n \binom{n}{p} \left( \frac{p}{n} - \frac{1}{2} \right)^{2k} \frac{1}{2^n} \geq \sum_p^* \binom{n}{p} \left( \frac{p}{n} - \frac{1}{2} \right)^{2k} \frac{1}{2^n} > \sum_p^* \binom{n}{p} \frac{\varepsilon^{2k}}{2^n}$$

et par conséquent

$$\begin{aligned} \sum_p^* \binom{n}{p} \frac{1}{2^n} &< \frac{1}{2^n (2n\varepsilon^2)^k} \cdot \sum_{p=0}^n \binom{n}{p} (2p - n)^{2k} < \quad (\text{vu (8)}) \\ &< \frac{(2k-1)!!}{2^k} \cdot \frac{1}{(2n\varepsilon^2)^k} < \frac{k!}{(2n\varepsilon^2)^k} < \frac{k^k \cdot e^{-k} \sqrt{2\pi k} \cdot e^2}{(2n\varepsilon^2)^k}. \end{aligned}$$

Le premier terme est la somme des longueurs des  $I_\varepsilon$  (qui s'excluent mutuellement), dans le dernier terme de l'inégalité posons

$$2n\varepsilon^2 = (1+\eta) \log n \quad (n \geq 3)$$

$$k = [(1+\eta) \log n]$$

—  $\eta$  étant positif et moindre que 1 et  $[x]$  désignant la partie entière de  $x$  — et bornons nous aux  $n \geq 2$ . On aura: somme des longueurs des

$$\begin{aligned} I_\varepsilon &< e^2 \frac{[(1+\eta) \log n]^{[(1+\eta) \log n]} e^{-[(1+\eta) \log n]} \sqrt{2\pi [(1+\eta) \log n]}}{((1+\eta) \log n)^{[(1+\eta) \log n]}} < \\ &< e^2 \cdot e^{1-(1+\eta) \log n} \sqrt{2\pi (1+\eta) \log n} < 2e^3 \sqrt{\pi} \frac{\sqrt{\log n}}{n^{1+\eta}}. \end{aligned}$$

En désignant par  $Z_n$  l'ensemble composé de tous les  $I_\varepsilon$  ( $\eta$  et  $n$ , et par conséquent  $\varepsilon$  étant fixe) on aura donc

$$|Z_n| < 2e^3 \sqrt{\pi} \frac{\sqrt{\log n}}{n^{1+\eta}} \quad (n \geq 3).$$

L'ensemble-limite complet  $L_\varepsilon$  de la suite  $\{Z_n\}$  a une mesure nulle, car la série

$$\sum_{n=3}^{\infty} \frac{\sqrt{\log n}}{n^{1+\eta}}$$

et a fortiori la série  $\sum_{n=1}^{\infty} |Z_n|$  est convergente. En dehors de  $L_\varepsilon$  on a

$$(10) \quad \limsup_{n \rightarrow \infty} \left| \sqrt{\frac{2n}{\log n}} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \right| \leq \sqrt{1+\eta};$$

En effet, si  $\alpha$  n'appartient pas à  $L_\varepsilon$  il n'appartient pas aux  $Z_n$  à partir de  $n = N(\alpha)$ ; si  $\nu(n)$  désigne le nombre de zéros parmi les  $n$  premiers chiffres de  $\alpha$  le fait que  $\alpha$  n'appartient pas à  $Z_n$  se traduit par l'inégalité contradictoire à (9):

$$\left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq \varepsilon$$

qui équivaut à

$$\left| \sqrt{\frac{2n}{\log n}} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \right| \leq \sqrt{1+\eta}$$

en vertu de la valeur donnée à  $\varepsilon$ . Cette dernière inégalité vaut pour  $n \geq N(\alpha)$  et implique par là (10).  $L_\varepsilon$  ayant une mesure nulle (10) s'applique à presque tous les  $\alpha$ ; or  $\eta$  est arbitrairement petit, (10) entraîne donc

$$(11) \quad \limsup_{n \rightarrow \infty} \left| \sqrt{\frac{2n}{\log n}} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \right| \leq 1$$

*presque partout.*

(11) peut aussi être interprétée comme il suit: il est *presque certain* que l'écart absolu entre  $\frac{1}{2}$  et le quotient du nombre des coups noirs par le nombre total  $n$  des coups ne surpasse pas  $\sqrt{\frac{\log n}{2n}}$  pour les parties très longues. L'expérience confirme aussi ce résultat.

## § 4.

7. Du point de vue du calcul des probabilités il n'y a rien de changé si l'on considère les coups dans un ordre différent de leur ordre naturel ou bien si l'on considère seulement les coups appartenant à une suite partielle déterminée (p. e. les coups aux indices pairs) en ignorant tous les autres. au lieu d'envisager tous les coups comme ils se suivent dans leur ordre naturel. Cette remarque presque banale à ce point de vue conduit à certaines conséquences arithmétique en vertu du principe d'équivalence. On voit ainsi p. e. la validité de la formule (11) pour presque tous les nombres, quand  $\nu(n)$  signifie le nombre des zéros figurant à la 2<sup>ième</sup>, 4<sup>ième</sup>, 2<sup>n</sup><sup>ième</sup> place du développement dyadique. Un autre exemple fournit le problème de déterminer la mesure de l'ensemble de *tous* les nombres du  $\langle 01 \rangle$  qui montrent 0 comme  $q_1$ -ième ...  $q_n$ -ième ... chiffre du développement  $q_n$  étant une suite d'indices donnée (p. e.  $q_n = 10^{10^n}$ ). Le calcul des probabilités donne immédiatement *zéro* comme réponse, car telle est la probabilité de tirer toujours „noirs“ à la  $q_1$ -ième ...  $q_n$ -ième ... épreuve; en effet on écarte les coups d'indices  $\neq q_n$  et on obtient ainsi *une seule* partie, dont la probabilité est *zéro*.

Ce raisonnement nous permet d'énoncer les théorèmes suivants:

I. Soit  $A$  un ensemble ponctuel mesurable situé sur l'intervalle  $\langle 01 \rangle$  et  $\{q_n\}$  une suite à termes naturels et différents; soit  $B$  l'ensemble de *tous* les nombres de  $\langle 01 \rangle$  dont les chiffres d'indices  $q_1, q_2 \dots q_n \dots$  que nous appellons  $\alpha_1, \alpha_2 \dots \alpha_n \dots$  donnent pour valeur de la fraction dyadique infinie

$$0, \alpha_1 \alpha_2 \dots \alpha_n \dots$$

un nombre  $\alpha$  de  $A$ .  $B$  est mesurable et sa mesure est égale à celle de  $A$ ,

Démonstration:  $A$  est mesurable, l'ensemble des parties correspondantes a donc une probabilité déterminée (d'après § 1). D'après la remarque du début de ce § cette probabilité est égale à celle de l'ensemble correspondant à  $B$ .  $B$  est donc mesurable et  $|B| = |A|$  c. q. f. d.

II. Soit  $A$  un ensemble ponctuel situé sur  $\langle 01 \rangle$ , soit  $q_n$  la suite définie au théorème I et  $B$  un ensemble composé *seulement*



des nombres ayant la propriété supposée dans l'énoncé du théorème I; alors la mesure extérieure de  $B$  ne dépasse pas celle de  $A$ .

Démonstration: Remplaçons  $A$  par un ensemble mesurable  $\bar{A}$  qui contient  $A$  et dont la mesure est égale à la mesure extérieure de  $A$ <sup>1)</sup>, remplaçons  $B$  par un ensemble  $\bar{B}$  qui contient seulement et tous les nombres ayant la propriété signalée par rapport à  $\bar{A}$ ; les ensembles  $A$  et  $B$  vérifient les conditions du théorème I ce qui implique la mesurabilité de  $\bar{B}$  et  $|\bar{B}| = |\bar{A}|$ . Or  $B \subset \bar{B}$ ,  $|\bar{A}| = \text{mesure extérieure de } A$ , donc: mesure extérieure de  $B \leq \text{mesure extérieure de } A$ , c. q. f. d.

III. Si  $A$  et  $B$  sont deux ensembles ponctuels de  $\langle 01 \rangle$ ,  $\{\varrho_n\}$  une suite à termes naturels différents et  $\{\sigma_n\}$  une autre suite du même genre et si les chiffres d'indices  $\varrho_1, \varrho_2 \dots \varrho_n$  d'un nombre de  $B$  donnent toujours un nombre de  $A$  et les chiffres d'indices  $\sigma_1, \sigma_2 \dots \sigma_n$  d'un nombre de  $A$  donnent toujours un nombre de  $B$ , alors les mesures extérieures de  $B$  et  $A$  sont égales.

Démonstration: Théorème II.

On pourrait rendre les démonstrations ci-dessus tout à fait exactes en prouvant rigoureusement (à l'aide des postulats du § 1) que la probabilité qu'une partie appartienne à un ensemble des parties caractérisées par les propriétés du  $\varrho_1$ -ième...  $\varrho_n$ -ième... coup est définie et égale à la probabilité qu'une partie appartienne à un ensemble des parties dans lesquelles les mêmes propriétés conviennent au 1<sup>er</sup>, 2<sup>ième</sup>...  $n$ -ième coup, si cette dernière probabilité est définie. Cela reviendrait au fond à démontrer directement les théorèmes I, II, III à l'aide de la théorie de la mesure. Or, nous croyons que les démonstrations du texte qui s'appuient sur le principe d'équivalence d'une part et sur la remarque générale du début de ce §, qui est évidente du point de vue du calcul des probabilités sont à la fois persuasives et naturelles. Néanmoins nous indiquons à la fin de l'ouvrage le chemin exacte. (N° 10).

8. Appellons avec M. Borel „normal“ un nombre  $\alpha$  dont le développement satisfait à la relation (3). Appellons  $\alpha$  normal par rap-

<sup>1)</sup> Si  $A_n$  est un ensemble composé d'une infinité dénombrable d'intervalles tels que  $A_n \supset A$  et que  $|A_n| - \text{mes. ext. } A < \frac{1}{n}$  alors  $\bigcap_{n=1}^{\infty} A_n$  est l'ensemble cherché  $A$ . D'après la définition de mesure extérieure les  $A_n$  existent.

port à  $\beta$  ( $\alpha R \beta$ ) si les chiffres du développement de  $\alpha$  qui occupent les mêmes rangs que les chiffres 0 du développement de  $\beta$ , considérés à eux-mêmes satisfont à (3). D'après la remarque du N° 7,  $\beta$  étant fixe et  $\neq \frac{p}{2^n}$  ( $p=0, 1, 2 \dots n=0, 1, 2$ ), presque tous les  $\alpha$  de  $< 01 >$  sont normaux par rapport à  $\beta$ . Nous allons démontrer, que l'on a même le théorème:

„Tout  $\alpha$  normal est normal par rapport à presque tous les  $\beta$ .

Démonstration: Soit  $\alpha$  un nombre normal déterminé et soit  $k_1(\nu)$  l'indice du  $\nu$ -ième zéro dans le développement de  $\alpha$  (Si p. e.  $\alpha = 0, 10110 \dots$ ,  $k_1(1) = 2$ ,  $k_1(2) = 5$ ). Considérons maintenant les chiffres de  $\beta$  qui occupent le  $k_1(1)$ -ième...  $k_1(\nu)$ -ième rang. En désignant par  $N_1(\nu)$  le nombre des zéros parmi ces chiffres nous aurons — pour presque tous les  $\beta$  —

$$(12) \quad \lim_{\nu \rightarrow \infty} \left| \frac{N_1(\nu)}{\nu} - \frac{1}{2} \right| = 0$$

comme nous avons démontré par la remarque générale du début du § n° 7. Soit  $k_2(\varrho)$  l'indice de la  $\varrho$ -ième unité dans le développement de  $\alpha$ . Si  $N_2(\varrho)$  signifie le nombre des zéros parmi les chiffres de  $\beta$  occupant le  $k_2(1) \dots k_2(\varrho)$ -ième rang, on aura pour presque tous les  $\beta$

$$(13) \quad \lim_{\varrho \rightarrow \infty} \left| \frac{N_2(\varrho)}{\varrho} - \frac{1}{2} \right| = 0.$$

Considérons seulement ces chiffres de  $\alpha$  qui font face aux zéros de  $\beta$  et effaçons tous les autres. De  $\nu(n)$  premiers zéros (que l'on trouve parmi les  $n$  premiers chiffres de  $\alpha$ ) il restera seulement  $N_1(\nu)$ , de  $\varrho(n)$  premières unités il restera seulement  $N_2(\varrho)$ . Je dis, que (3), (12) et (13) impliquent

$$(14) \quad \lim_{n \rightarrow \infty} \frac{N_1(\nu(n))}{N_1(\nu(n)) + N_2(\varrho(n))} = \frac{1}{2}.$$

$$\text{En effet: } N_1(\nu) = \frac{\nu}{2} + o(\nu), \quad \nu(n) = \frac{n}{2} + o(n)$$

$$N_2(\varrho) = \frac{\varrho}{2} + o(\varrho), \quad \varrho(n) = \frac{n}{2} + o(n).$$

ce qui implique

$$N_1(\nu) = \frac{n}{4} + o\left(\frac{n}{2} + o(n)\right) = \frac{n}{4} + o(n)$$

$$N_2(\varrho) = \frac{n}{4} + o(n).$$

D'autre part (14) conduit au théorème proposé;  $N_1(\nu)$  est le nombre des zéros non-éffacés,  $N_1(\nu) + N_2(\varrho)$  le nombre total des chiffres non-éffacés qui restent de  $n$  premiers chiffres de  $\alpha$ , (14) donne ainsi a fortiori:

$$(15) \quad \lim_{N \rightarrow \infty} \frac{N_1}{N} = \frac{1}{2}$$

pour presque tous les  $\beta$ , où  $N_1$  désigne le nombre des zéros non-éffacés parmi les  $N$  premiers chiffres non-éffacés de  $\alpha$ , c. q. f. d.

Pour démontrer le théorème inverse à savoir:

„Si  $\alpha$  est normal par rapport à presque tous les  $\beta$ ,  $\alpha$  est normal“, supposons le contraire; cette supposition équivaut à l'existence d'une suite croissante des nombres naturels  $\{n_k\}$  telle que

$$(16) \quad \nu(n_k) = \beta n_k + o(n_k)$$

$$\varrho(n_k) = (1 - \beta)n_k + o(n_k)$$

avec  $\beta \neq \frac{1}{2}$ , les notations étant les mêmes, que nous avons employé tout à l'heure. En répétant le procédé expliqué pendant la démonstration du théorème directe, nous trouverons — pour presque tous les  $\beta$  — parmi ces chiffres de  $\alpha$  qui restent de  $n_k$  premiers après avoir conservé seulement ceux qui font face aux zéros de  $\beta$ .  $N_1(\nu(n_k))$  zéros et  $N_2(\varrho(n_k))$  unités. Les formules:

$$N_1(\nu) = \frac{\nu}{2} + o(\nu), \quad N_2(\varrho) = \frac{\varrho}{2} + o(\varrho)$$

avec (16) donnent

$$N_1(\nu(n_k)) = \frac{\beta}{2} n_k + o(\beta n_k + o(n_k)) + \frac{1}{2} o(n_k)$$

$$N_2(\varrho(n_k)) = \frac{1 - \beta}{2} n_k + o((1 - \beta)n_k + o(n_k)) + \frac{1}{2} o(n_k);$$

on en déduit

$$N_1(\nu(n_k)) = \frac{\beta}{2} n_k + o(n_k), \quad N_2(\varrho(n_k)) = \frac{1 - \beta}{2} n_k + o(n_k)$$

et

$$\lim_{k \rightarrow \infty} \frac{N_1(\varrho(n_k))}{N_1(\nu(n_k)) + N_2(\varrho(n_k))} = \beta \neq \frac{1}{2}$$

pour presque tous les  $\beta$ , contrairement à l'hypothèse du théorème. Nous voyons aussi qu'il suffirait de supposer que  $\alpha$  est normal par rapport à un ensemble de  $\beta$  de mesure extérieure positive, pour conclure que  $\alpha$  est normal. La condition nécessaire et suffisante pour qu'un nombre soit normal est qu'il soit normal par rapport à presque tous les nombres.

9 Au lieu des probabilités égales pour „rouge“ et „noir“ introduisons le cas plus général, où ces probabilités sont

$$p, q \ (p > 0, q > 0, p + q = 1).$$

Nos postulats du § 1 s'adaptent encore à ce cas si l'on change convenablement  $2_2^0$ :

„Si  $\vec{E}$  et  $\vec{E}'$  ne diffèrent que pour le  $i$ -ième coup“ ( $i \leq n$ ), ce coup étant rouge pour  $\vec{E}$  et noir pour  $\vec{E}'$ , on a  $\mu(\vec{E}) : \mu(\vec{E}') = p : q$ “.

A toute partie infinie  $\alpha$  on fera correspondre un développement dyadique comme au § 1, mais on lira ce développement comme  $\alpha$ ,  $\alpha$  étant défini par l'intermédiaire de formules suivantes:

$$(\text{déf.}) \quad \alpha = [0, \alpha_1, \alpha_2 \dots \alpha_n \dots] \quad \alpha_i = \begin{cases} 0 \\ 1 \end{cases} \text{ pour } \begin{matrix} \text{„noir“} \\ \text{„rouge“} \end{matrix}$$

au  $i$ -ième coup;

$$(\text{déf.}) \quad [0, 0] = 0, [0, 1] = q \quad [0, \alpha_1 \alpha_2 \dots \alpha_n \alpha_{n+1}] = [0, \alpha_1 \alpha_2 \dots \alpha_n] \text{ si } \alpha_{n+1} = 0;$$

$$(\text{déf.}) \quad [0, \alpha_1 \alpha_2 \dots \alpha_n \alpha_{n+1}] = [0, \alpha_1 \alpha_2 \dots \alpha_n] + q \cdot q^{\sum_{i=1}^n \alpha_i} \cdot p^{\sum_{i=1}^n \alpha_i} \text{ si } \alpha_{n+1} = 1;$$

$$(\text{déf.}) \quad [0, \alpha_1 \alpha_2 \dots \alpha_n \dots] = \lim_{n \rightarrow \infty} [0, \alpha_1 \alpha_2 \dots \alpha_n].$$

Géométriquement: on divise  $\langle 01 \rangle$  à raison de  $q : p$  et on pose 0, 0 ou 0, 1 suivant que le nombre que nous voulons développer tombe à gauche ou à droite du point de division; après, on divise les deux intervalles  $q, p$  chacun d'une manière semblable (à raison  $q : p$ ) ce qui donne le 2<sup>ième</sup> chiffre du développement et on procède ainsi indéfiniment, ce qui fournit une suite des chiffres  $\alpha_1, \alpha_2 \dots$  et il n'y a ambiguïté qu'au cas où  $\alpha$  serait lui-même un des points de division; ces nombres exceptionnels (dont l'ensemble est dénombrable) donnent deux développements différents. Nos formules apprennent comment il faut calculer le nombre développé  $\alpha$  si les chiffres  $\{\alpha_i\}$  sont données.

Maintenant il faut répéter le raisonnement du § 1 pour obtenir les postulats de M. Sierpiński. Les probabilités des ensembles

de parties  $\alpha$  seront encore égales aux mesures lebesguiennes des ensembles des  $\alpha$  correspondants.

Le théorème de Bernoulli, que

$$\left| \frac{\nu(n)}{n} - q \right| < \varepsilon \quad \text{pour } n > N(\varepsilon, \eta)$$

avec une probabilité qui surpasse  $1 - \eta$ , fournit

$$\lim_{n \rightarrow \infty} \frac{\nu(n)}{n} = q, \quad \lim_{n \rightarrow \infty} \frac{\varrho(n)}{n} = p$$

pour presque tous les  $\alpha$ ,  $\nu(n)$  ( $\varrho(n)$ ) désignant le nombre de zéros (unités) parmi les  $n$  premiers chiffres du développement dyadique généralisé conformément aux indications de ce N°

Le théorème de Laplace affirme que la probabilité  $f(t, n)$  de l'inégalité

$$\left| \frac{\nu(n)}{n} - q \right| \leq \frac{t\sqrt{pq}}{\sqrt{n}} \quad (t \geq 0)$$

tend vers

$$\frac{1}{\sqrt{\pi}} \int_0^t e^{-\tau^2} d\tau$$

avec  $n \rightarrow \infty$ , si  $t$  reste fixe; ceci fournit

$$\lim_{n \rightarrow \infty} \sup_{\inf} \left| \frac{\nu(n)}{n} - q \right| = +\infty$$

pour presque tout les  $\alpha$ . Nous n'insistons pas d'avantage sur ces généralisations et nous remarquons seulement qu'il faut sans doute des nouveaux artifices de calcul pour obtenir une formule analogue à (11).

Quand on remplace la base 2 par une autre base quelconque (p. e. 10) on peut étudier les probabilités des parties à plusieurs cas possibles (p. e.) à chaque épreuve (p. e. 10 cas).

10. Pour démontrer rigoureusement les résultats du N° 7, il faut établir le théorème suivant:

Si  $E$  est un ensemble des parties ayant une probabilité définie  $\mu(E)$  et si l'on considère une suite des indices différents  $i_1, i_2 \dots i_n$  et l'ensemble  $\bar{E}$  des parties que l'on obtient en faisant correspondre à toute partie  $\alpha$  de  $E$  toutes les parties  $\bar{\alpha}$  dont les  $i_1, i_2 \dots i_n$ -ièmes ...



coups sont identiques au 1<sup>er</sup>, 2<sup>ième</sup>...  $n$ -ième... coup de  $a$ , alors la probabilité  $\mu(\overline{E})$  de  $\overline{E}$  est définie et égale à  $\mu(E)$ .

Démonstration. Changeons la définition de la probabilité comme il suit: conservons les postulats du § 1 avec la seule modification que  $\overline{E}$  reçoive une nouvelle signification; à savoir celle de l'ensemble de toutes les parties à  $i_1$ -<sup>er</sup>,  $i_2$ -ième, ...  $i_n$ -ième coups données. En employant pour  $E$  la vieille et pour  $\overline{E}$  la nouvelle définition de la probabilité on obtient évidemment la même valeur de  $\mu$  pour  $E$  et pour  $\overline{E}$ . Or, pour rendre ce raisonnement valide il faut démontrer encore l'équivalence de deux définitions. En faisant correspondre à  $E$  et à  $\overline{E}$  les ensembles ponctuels comme au § 1 nous retombons précisément sur le théorème I du N° 7 concernant les ensembles ponctuels  $A$  et  $B$ , qu'il faut maintenant établir directement, sans l'aide du calcul des probabilités.

A cet effet remarquons d'abord que la correspondance existant par hypothèse du théorème I du N° 7 entre  $A$  et  $B$ , donne pour  $A'$ , partie de  $A$ , comme image  $B'$ , partie de  $B$ , pour l'image d'une somme la somme des images, et pour l'image du complémentaire le complémentaire des images. Si  $A$  est  $\langle \frac{q}{2^n}, \frac{q+1}{2^n} \rangle$  ( $n=0, 1, \dots, q=0, \dots, 2^n-1$ ) le  $B$  correspondant consiste d'un nombre fini des segments non-empiétants dont la longueur totale est  $\frac{1}{2^n}$  ( $=|A|$ ). Il s'ensuit que l'image  $B$  d'un ensemble  $A$  d'un seul élément est de mesure nulle. Maintenant on voit aisément qu'à tout ensemble  $A$  identique à un segment (ouvert ou fermé) dont les extrémités sont de forme  $\frac{p}{2^n}, \frac{q}{2^n}$  ( $n=0, 1, \dots, p=0, 1, \dots, 2^n, q=0, 1, \dots, 2^n, p \leq q$ ) correspond un  $B$  de même mesure, consistant d'un nombre fini des segments et d'un certain ensemble  $O$  de mesure nulle. Ce résultat s'étend sur tous les segments à extrémités dyadiques

$$\left( \frac{p}{2^m}, \frac{q}{2^m}; n=0, 1, \dots, m=0, 1, \dots, p=0, 1, \dots, 2^m, q=0, 1, \dots, 2^m, \frac{p}{2^m} \leq \frac{q}{2^m} \right).$$

Soit  $A$  un ensemble mesurable; on peut enfermer les points de  $A$  dans un système des segments dyadiques non-empiétants de longueur totale proche de  $|A|$  et on peut enfermer l'ensemble complémentaire à  $A$  (par rapport à  $\langle 01 \rangle$ ) dans un système analogue de

longueur totale proche de  $1 - |A|$ . Si  $B$  est l'ensemble correspondant à  $A$ , les systèmes correspondants aux deux systèmes définis tout à l'heure auront comme longueurs totales les mêmes nombres; un de ces système enferme  $B$  et sa longueur totale est proche de  $|A|$ , l'autre enferme le complémentaire à  $B$  et sa longueur totale est proche de  $1 - |A|$ , il faut donc que  $B$  soit mesurable et  $|B| = |A|$  c. q. f. d.<sup>1)</sup>.

<sup>1)</sup> M. F. Bernstein a publié en 1912 dans les *Mathematische Annalen* (vol. 71, pp. 417—439) un travail sur l'application de la théorie des ensembles à un problème de la théorie des perturbations séculaires. Il démontre entre d'autres — comme M. Borel dans le travail cité (Chap. III, 264—269) — des résultats intéressants sur les fractions continues. La même question est aussi traitée plus tard chez M. Burstin (loc. cit.). Or, dans tous ces travaux „probabilité“ n'est qu'une expression pour „mesure“ ( $B$  ou  $L$ ) ou bien elle n'intervient pas du tout. Dans les deux cas le lien qui attache le vrai calcul de la probabilité des jeux à la mesure reste caché. La dualité qui fait correspondre à tout théorème d'un domaine un théorème analogue de l'autre domaine n'y apparaît pas encore d'une manière nette.

Göttingue, 18 juin 1922.

---