

Nouveaux fondements du calcul des probabilités.

(Définition de la probabilité fondée sur la théorie des ensembles).

Par

Antoine Łomnicki (Léopol).

Introduction.

Le calcul des probabilités occupe parmi les sciences mathématiques une place tout à fait infime. Le manque de définitions nettes et, ce qui en est la conséquence, quantité des paradoxes, ne peuvent en effet, malgré l'immense champ d'application de cette branche des mathématiques, lui valoir la faveur de la génération moderne des mathématiciens, habituée à des raisonnements subtils et rigoureux.

C'est ainsi, par exemple, que Poincaré commence son „*Calcul des probabilités*“¹⁾ par cet aveu désespéré:

„On ne peut guère donner une définition satisfaisante de la *probabilité*“.

Dans l'article „*La théorie des probabilités*“ de M. S. Mazurkiewicz²⁾ on trouve cette remarque fort peu encourageante:

„La théorie des probabilités n'est pas un élément indispensable de l'instruction mathématique; il est toutefois bien désirable pour un mathématicien d'en connaître les principes généraux... Les notions fondamentales y sont incomplètement fixées: elles contiennent beaucoup de difficultés inexpliquées“.

Dans une vaste étude récemment publiée M. R. v. Mises³⁾ déclare simplement:

¹⁾ II éd. Paris 1912, p. 24.

²⁾ Poradnik dla samouków t. I, 1915, p. 438.

³⁾ *Grundlagen der Wahrscheinlichkeitsrechnung*, Math. Zeitschrift, Berlin 1919, V Band, p. 52–99.

„... in der Tat kann man den gegenwärtigen Zustand kaum anders als dahin kennzeichnen, dass *die Wahrscheinlichkeitsrechnung heute eine mathematische Disziplin nicht ist*“.

Il construit cependant une théorie des probabilités qu'il conçoit comme une science naturelle („Die vorliegende Darstellung geht von der Auffassung aus, dass *die Wahrscheinlichkeitsrechnung eine Naturwissenschaft gleicher Art wie die Geometrie oder die theoretische Mechanik ist*“) et qui est tout à fait insuffisante, puisqu'elle exclue d'avance „*une besondere Klasse mathematischer Fragen, die in neuerer Zeit unter Verwendung der Ausdruckweise der Wahrscheinlichkeitsrechnung zur Sprache gebracht wurden...*“, c'est à dire, justement les problèmes qui intéressent le plus les mathématiciens contemporains.

Les tentatives de donner un système d'axiomes suffisant pour construire le calcul des probabilités soit se bornèrent aux probabilités d'un nombre fini de cas¹⁾, soit se contentèrent de donner les lois principales de la théorie des probabilités, sans en examiner la suffisance et l'indépendance réciproque et sans tenir compte d'une manière suffisante des probabilités pour les cas „inégalement possibles“, c'est à dire — selon notre terminologie — pour les ensembles à distribution hétérogène des poids (voir p. 83), où la loi de la multiplication des probabilités ne s'applique pas toujours²⁾. Dans toutes ces tentatives on trouve de plus de nombreux termes étrangers aux mathématiques et non définis (par exemple: événement, possibilité, indépendance des événements, cause etc.), ce qui rend la question encore plus obscure.

En m'efforçant de m'expliquer les notions fondamentales du calcul des probabilités et de remonter jusqu'aux sources, où prend son origine le flot trouble des paradoxes (p. ex. ceux de Bertrand), j'ai acquis la conviction que *le calcul des probabilités tout entier est un problème bien déterminé de la théorie des ensembles, étroitement lié à la théorie de la mesure des ensembles et absolument indépendant de tout facteur métaphysique ou empirique.*

Les paradoxes naissent, en particulier, lorsqu'on parle de la probabilité par rapport aux ensembles, pour lesquels la notion de la

¹⁾ P. ex. J. Łukasiewicz. Die logischen Grundlagen der Wahrscheinlichkeitsrechnung, Cracovie 1913.

²⁾ Cf. G. Bohlmann, Encyklop. der math. Wissensch. I, Band 2, p. 859.

mesure n'a pas été préalablement définie (tel est le cas d'ensemble de toutes les droites d'un plan) et, en général, lorsque les données du problème ne sont pas précisées d'une façon assez nette.

Ainsi, la difficulté qui concerne la notion-même, n'est plus dans la théorie des probabilités, mais bien dans la théorie de la mesure, donc dans celle des ensembles. Ainsi conçue, la théorie des probabilités devient une science tout au moins aussi „mathématique“ que la théorie des ensembles et, quant à son applicabilité à telle ou telle branche des sciences expérimentales ou même mathématiques, cela n'intéresse point la théorie des probabilités, mais bien chacune de ces branches. C'est pourquoi je ne me propose pas de m'occuper ici de ces questions.

Le présent Mémoire a pour but de présenter une définition „constructive“ de la probabilité mathématique, embrassant tous les cas particuliers qu'il est possible de traiter à l'état actuel du développement de la théorie des ensembles, à savoir: la définition classique des probabilités finies, celle des probabilités continues (uniquement pour les ensembles de points), ensuite celle des probabilités dénombrables et enfin celle des probabilités non-dénombrables mesurables. Je développe tous les raisonnements, en m'appuyant conséquemment sur la théorie des ensembles.

Il est vrai que dans la littérature mathématique on voit ça et là rattacher certains problèmes de la théorie des probabilités à la théorie des ensembles¹⁾; cependant je n'ai jamais rencontré une application systématique de ce principe²⁾. Déjà la notation traditionnelle „ p “, sans indices ni variables, prouve combien on se rendait peu compte de ce que la probabilité est une *fonction d'ensemble*.

En me basant sur la définition construite de cette façon, j'envisage sous ce jour nouveau les notions fondamentales de la théorie des probabilités. Je discute ensuite les principales lois formelles du calcul, surtout celle des probabilités composées, et j'applique les résultats

¹⁾ Cf. p. ex. F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig 1914, p. 416 et E. Borel, *Remarques sur certaines questions de probabilité*, Bull. de la Soc. math. 1905, p. 123—128.

²⁾ Je n'ai pas eu occasion de consulter la dissertation de R. Lämmel (1904), citée par Czuber, où l'auteur distingue 144 types de problèmes de probabilité, en combinant entre eux les ensembles qui se comportent différemment au point de vue d'énumération, de densité et de mesure.

obtenus à la démonstration de la loi des erreurs de Gauss, en mettant en même temps en évidence les hypothèses mathématiques, sur lesquelles repose cette loi. Enfin je montre où faut-il chercher les sources de différents paradoxes dans les probabilités dites „géométriques“; je discute, en particulier, le paradoxe de Bertrand et le problème de Buffon.

Je désire que cet essai d'élucider les principes du calcul des probabilités contribue, ne fût-ce que partiellement, à dissiper la méfiance — d'ailleurs parfaitement justifiée par l'état actuel de cette science — qu'elle inspirait à ceux qui presque à chaque pas de leur travail ont fait usage de ce calcul. Je me permettrai de citer à ce sujet les paroles significatives de mon professeur feu M. Smoluchowski¹⁾:

„Bien que le champ d'applications du calcul des probabilités se soit énormément élargi, l'analyse rigoureuse des notions fondamentales de ce calcul a fait des progrès insignifiants; aujourd'hui encore il est vrai qu'aucune autre branche de mathématiques ne repose sur des fondements aussi vagues et précaires“.

Cette constatation était pour moi le plus fort encouragement pendant le travail.

§ 1. Les éléments constitutifs de la notion de la probabilité.

Lorsqu'on parle de probabilité, on a toujours en vue deux *ensembles* m et \mathcal{M} , dont le premier fait partie du second (c'est à dire, en est un *sous-ensemble*, vrai ou non; dans le dernier cas les ensembles m et \mathcal{M} coïncident). Les *éléments* de ces ensembles s'appellent „cas“ dans la théorie classique, à savoir: les éléments de m „cas favorables pour un événement“ et les éléments de \mathcal{M} „cas possibles“. A chaque élément de \mathcal{M} et, à plus forte raison, de m , qui en fait partie, on fait correspondre un nombre positif fini, inférieur à un nombre constant L arbitrairement donné; appelons-le: *poids* de l'élément. Si à tous les éléments on fait correspondre le même poids, ou — comme on dit dans la théorie classique — si tous les „cas“ ou „événements“ sont „également possibles“, la distribution des poids sera dite *homogène*; dans le cas contraire nous

¹⁾ Ueber den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik, Die Naturwissenschaften, Berlin 1918, Heft 17.

l'appellerons *hétérogène*. Nous avons défini ainsi pour tout ensemble \mathcal{M} une fonction de ses éléments $\varphi_{\mathcal{M}}(e)$ qui, dans le cas spécial de la distribution homogène, est constante pour chaque élément e :

$$\varphi_{\mathcal{M}}(e) = \delta$$

et qui, en tout cas, est une fonction bornée. Nous admettons, de plus, que c'est une fonction mesurable, donc sommable.

Quelle distribution des poids doit-on adopter dans chaque problème donné ou — ce qui revient au même — quelle forme et quelles propriétés doit avoir la fonction $\varphi(e)$ — cela n'appartient pas au calcul des probabilités, mais doit être donné du dehors par la science qui veut appliquer notre théorie à la résolution de ses problèmes. Ici on peut tout au plus discuter quelle influence exercent sur les règles de notre théorie divers genres déterminés des fonctions $\varphi(e)$. Nous allons le faire dans la suite pour quelques cas les plus importants au point de vue des applications. Quelques auteurs ¹⁾ appellent la fonction qui représente la distribution des poids „densité“ de la probabilité; ils ne l'appliquent cependant qu'aux ensembles continus.

Enfin, le dernier élément constitutif, indispensable à la construction d'une définition de la probabilité est la notion de la mesure d'un ensemble ou — comme on dit improprement dans la théorie classique — du „nombre des cas“. Suivant la mesure attribuée à l'ensemble des poids correspondant à un ensemble donné \mathcal{M} , on peut obtenir des probabilités différentes pour les mêmes „événements“ ou — d'une manière plus précise — pour le même ensemble m dans le même ensemble \mathcal{M} et pour la même distribution des poids.

Nous allons donc restreindre notre définition de la probabilité aux ensembles pour lesquels la notion de la mesure a été établie, notamment aux ensembles de points d'un espace à un nombre fini quelconque de dimensions et aux ensembles de nombres, isomorphes des précédents, comme l'ensemble des nombres réels, des paires de nombres (p. ex. celui des nombres complexes ordinaires), l'ensemble des systèmes de trois nombres (p. ex. celui des vecteurs) et ainsi de suite, en nous bornant naturellement aux ensembles me-

¹⁾ P. ex. Markoff, *Wahrscheinlichkeitsrechnung*, 1911.

surables. Nous emploierons la mesure de Lebesgue étendue également sur les ensembles non-bornés et, lorsque cette mesure est nulle, celle de Carathéodory¹⁾ n'étant pas nulle dans un espace à un nombre inférieur de dimensions, nous emploierons la mesure de Carathéodory.

L'emploi de la notion de la mesure dans une forme si générale dépasse considérablement les cadres de la théorie classique des probabilités, où la mesure habituelle de volume, de surface et de longueur était suffisante. Or, nous introduisons dès le premier abord les notions les plus générales de la théorie moderne des ensembles pour faire voir d'une part le rapport intime, qui existe entre les notions de la théorie des ensembles et celles du calcul des probabilités, et d'autre part l'étendue de ses applications. Par contre, nous n'allons pas développer ici la théorie des probabilités pour les ensembles, pour lesquels la notion de la mesure n'est pas suffisamment fixée (p. ex. l'ensemble des droites d'un plan, l'ensemble de toutes les cordes d'un cercle), bien que les problèmes concernant de tels ensembles occupent une place assez notable dans la théorie classique, où ils sont traités, à la vérité, sans aucun esprit de critique.

Pour déterminer une probabilité il faut donc spécifier:

1° L'ensemble \mathcal{M} , dans lequel on examine la probabilité (ensemble des „cas possibles“).

2° L'ensemble m , dont on examine la probabilité dans \mathcal{M} (ensemble des „cas favorables pour l'événement considéré“).

3° La distribution des poids, c'est à dire, la fonction assignant à tout élément e de \mathcal{M} un nombre $\varphi(e)$ (degré de „possibilité“ de chaque „cas“).

4° Le mode de mensuration de l'ensemble de tous les poids assignés aux éléments de \mathcal{M} et le mode de mensuration de chacun de ses sous-ensembles (mesurables).

Le symbole désignant une probabilité doit par conséquent renfermer ceux de tous les quatre éléments de cette notion, du moins si l'on veut éviter qu'il ne soit *plurivoque*. Or, nous nous sommes décidés une fois pour toutes à n'employer ici que la mesure de Lebesgue

¹⁾ Ueber das lineare Mass von Punktmengen — eine Verallgemeinerung des Längenbegriffs, Götting. Nachr. 1914.

Ce mémoire renferme une définition de la mesure p-dimensionnelle dans un espace à q dimensions.

ou celle de Carathéodory; nous pouvons donc renoncer à la représenter dans nos symboles, à moins de nécessité absolue.

Ceci posé, la probabilité ne dépendra que des ensembles \mathcal{M} et m et de la fonction $\varphi(e)$. Plus exactement — en employant le langage de la théorie des ensembles — la probabilité est une *fonction d'ensemble* définie dans l'ensemble \mathcal{M} pour chacun de ses sous-ensembles m mesurables (sauf, peut-être, certains sous-ensembles „singuliers“, qui correspondent aux points singuliers de la définition commune de la fonction à une ou plusieurs variables) et dépendant, en outre, d'une fonction arbitrairement choisie $\varphi_{\mathcal{M}}(e)$ définie pour chaque élément de \mathcal{M} .

Le symbole de la probabilité, qui spécifie ces trois éléments constitutifs, peut donc être, par exemple, de la forme suivante:

$$(1) \quad p_{\mathcal{M}}(m; \varphi_{\mathcal{M}}(e)),$$

ce qu'on lira: *la probabilité de m dans \mathcal{M} où la distribution des poids est déterminée par $\varphi_{\mathcal{M}}(e)$* . Lorsque la distribution des poids est homogène, on peut supprimer le signe de la fonction $\varphi_{\mathcal{M}}(e)$, mais il ne faut jamais négliger d'indiquer explicitement ceux des ensembles m et \mathcal{M} . L'emploi notoire du signe p dépourvu d'arguments, en usage constant dans la théorie classique, donne toujours lieu à des explications verbales supplémentaires et peut donner lieu à des facheux malentendus.

§ 2. Probabilité dans les ensembles finis.

Pour plus de clarté nous allons construire la définition de la probabilité successivement pour les ensembles de plus en plus compliqués: d'abord pour les ensembles finis, ensuite pour les ensembles dénombrables et enfin pour les ensembles non-dénombrables.

Nous appelons *probabilité d'un sous-ensemble m de l'ensemble \mathcal{M} dans l'ensemble \mathcal{M} composé d'un nombre fini b d'éléments à poids $\delta_1, \delta_2, \dots, \delta_b$, le quotient, dont le dividende est la somme des poids des éléments de m et dont le diviseur est celle des poids des éléments de \mathcal{M} .*

Par conséquent, \mathcal{M} étant composé d'éléments

$$e_1, e_2, \dots, e_b$$

et m étant composé de a éléments

$$e_{i_1}, e_{i_2}, \dots, e_{i_a}$$

de \mathcal{N} , on a :

$$(2) \quad p_{\mathcal{N}}(m) = \frac{\delta_{i_1} + \delta_{i_2} + \dots + \delta_{i_a}}{\delta_1 + \delta_2 + \dots + \delta_b}$$

En termes de géométrie, on peut se représenter \mathcal{N} comme un ensemble fini de points d'un espace à un nombre fini quelconque de dimensions et s'imaginer une „ordonnée“ dressée dans chacun de ces points et qui soit de longueur égale au poids de l'élément correspondant. La probabilité de m dans \mathcal{N} s'obtient comme une fraction qui a pour numérateur la mesure de l'ensemble-somme des poids des éléments de m et pour dénominateur la mesure de l'ensemble-somme des poids des éléments de \mathcal{N} . Dans le cas considéré la mesure est simplement la somme des longueurs des segments qui représentent les poids; c'est donc la mesure linéaire.

Si l'espace dans lequel on a placé l'ensemble \mathcal{N} est à q dimensions, celui où sont repartis les poids assignés aux éléments de \mathcal{N} est évidemment à $q + 1$ dimensions. La mesure lebesguienne de l'ensemble-somme des poids, comme mesure $(q + 1)$ dimensionnelle, est nulle, le nombre des éléments de \mathcal{N} étant fini. On doit donc employer une mesure à un nombre plus petit de dimensions. Dans le cas considéré c'est la mesure 1-dimensionnelle ou linéaire qu'il faut prendre, ce qui veut dire qu'on emploie la mesure de Carathéodory et non plus celle de Lebesgue. Dans le cas contraire on obtiendrait comme expression de la probabilité un quotient non-défini de deux zéros, à moins de renoncer à définir la probabilité comme quotient de deux mesures ou comme limite, vers laquelle tend ce quotient.

Or, pour les ensembles \mathcal{N} continus, c'est précisément la définition universellement employée et, puisque nous nous proposons de parvenir à une définition de la probabilité commune à toutes les espèces d'ensemble \mathcal{N} et embrassant comme cas particulier celle, qui vient d'être donnée pour les ensembles finis, nous sommes contraints d'employer dès le début la notion de la mesure. Je crois, en outre, que cette manière de concevoir la probabilité est la plus recommandable, ne fût ce que pour les raisons didactiques, car elle nous offre aussitôt une image, facile à saisir, des „cas“ et de divers degrés de leur „possibilité“.

Lorsque la distribution des poids est homogène, c'est à dire, lorsque chaque élément de \mathcal{N} a le même poids δ , on a:

$$(3) \quad p_{\mathcal{N}}(m) = \frac{a \cdot \delta}{b \cdot \delta} = \frac{a}{b}$$

conformément à la définition classique, où la probabilité signifie le rapport du nombre des cas „favorables“ à celui de tous les cas „possibles“, dans l'hypothèse que tous les cas sont „également possibles“.

On sait que la probabilité définie de la manière classique satisfait à certaines lois formelles très simples, sur lesquelles le calcul entier est fondé et que l'on s'efforçait à utiliser pour construire une définition axiomatique de la probabilité. Nous allons établir cependant l'existence des probabilités qui satisfont aussi bien à notre définition qu'aux définitions classiques concernant les ensembles à distribution hétérogène des poids, mais pour lesquelles le théorème sur la *multiplication* des probabilités n'est pas vrai. L'applicabilité ou non-applicabilité de cette loi „fondamentale“ est intimement liée au mode de distribution des poids et nous allons le bien mettre en évidence. Nous établirons notamment une condition nécessaire et suffisante, que doit remplir un mode de distribution des poids, pour que la probabilité soit *multiplicative*, c'est à dire, pour que le théorème sur la multiplication des probabilités soit vrai. Nous donnerons en même temps un exemple d'un mode de distribution où la probabilité n'est pas multiplicative.

Dans un système d'axiomes de la théorie générale des probabilités il faut donc en tout cas supprimer l'axiome concernant la multiplication des probabilités. Les cas échappant à cet axiome se présentaient également dans la théorie classique. On disait alors que les „cas“ étaient dépendants. Or, si l'on se donne la peine de chercher dans la littérature ce que veut dire l'expression „cas dépendants“, on n'en trouve que chez A. Markoff¹⁾ une explication assez nette et d'ailleurs très proche de la nôtre. C'est pourquoi nous préférons de nous servir désormais du terme bien défini *probabilité multiplicative* au lieu du terme confus — au moins jusqu'à présent — de „cas dépendants“. La mise en lumière du sens de ce dernier terme ou plutôt l'emploi à sa place de la notion parfaitement précise de

¹⁾ l. c., p. 158.

la „multiplicativité“ est une des qualités peut-être les plus importantes de notre définition nouvelle de la probabilité.

Les lois fondamentales du calcul des probabilités s'expriment dans notre système de notation, comme il suit:

I. Lorsque m est vide, on a

$$p_{\mathcal{N}}(m; \varphi_{\mathcal{N}}(e)) = 0.$$

II. Lorsque m est identique à \mathcal{N} , on a

$$p_{\mathcal{N}}(m; \varphi_{\mathcal{N}}(e)) = 1.$$

III. m_1 et m_2 étant deux sous-ensembles disjoints (sans éléments communs) de \mathcal{N} et $m_1 + m_2$ désignant leur somme (l'ensemble de tous les éléments de ces deux ensembles), on a toujours:

$$(4) \quad p_{\mathcal{N}}(m_1 + m_2; \varphi_{\mathcal{N}}(e)) = p_{\mathcal{N}}(m_1; \varphi_{\mathcal{N}}(e)) + p_{\mathcal{N}}(m_2; \varphi_{\mathcal{N}}(e)).$$

Remarque: Si les ensembles m_1 et m_2 avaient des éléments communs, on aurait (en symboles abrégés):

$$p(m_1 + m_2) = p(m_1) + p(m_2) - p(D(m_1, m_2)),$$

où $D(m_1, m_2)$ désigne l'ensemble des éléments communs à m_1 et m_2 ¹⁾.

On exprime la loi III tout court, en disant que la probabilité est une fonction *additive*. Dans la théorie classique cette loi porte le nom du théorème sur la probabilité totale.

IV. m_1 et m_2 étant des sous-ensembles quelconques des ensembles \mathcal{N}_1 et \mathcal{N}_2 (où m_1 et m_2 ainsi que \mathcal{N}_1 et \mathcal{N}_2 peuvent avoir des éléments communs et même être identiques), on a — en désignant d'une façon générale par AB l'ensemble de toutes les paires que l'on peut former des éléments de A et de B —

$$(5) \quad \begin{aligned} p_{\mathcal{N}_1 \mathcal{N}_2}(m_1 m_2; \overline{\varphi}_{\mathcal{N}_1 \mathcal{N}_2}(e)) &= \\ &= p_{\mathcal{N}_1}(m_1; \varphi_{\mathcal{N}_1}(e)) \cdot p_{\mathcal{N}_2}(m_2; \overline{\varphi}_{\mathcal{N}_2}(e)) \end{aligned}$$

pour tous les sous-ensembles m_1 , m_2 et $m_1 \cdot m_2$. Les signes $\varphi_{\mathcal{N}_1}(e)$, $\varphi_{\mathcal{N}_2}(e)$, $\overline{\varphi}_{\mathcal{N}_1 \mathcal{N}_2}(e)$ représentent les fonctions d'élément e qui définissent la distribution des poids dans les ensembles \mathcal{N}_1 , \mathcal{N}_2 et $\mathcal{N}_1 \mathcal{N}_2$.

Or, c'est cette loi qui ne s'applique qu'aux probabilités à une

¹⁾ Nous le désignons ici par $D(m_1, m_2)$ et non par $m_1 \cdot m_2$ ou $m_1 m_2$, parce-que le symbole $m_1 m_2$ a dans notre théorie une signification différente (voir IV).

distribution convenable des poids (en particulier homogène) et nous allons démontrer que cette loi n'est pas vraie pour le cas général. Elle porte le nom du théorème sur la probabilité composée, mais il serait bien plus naturel de l'appeler *loi multiplicative*. Nous dirons donc d'une fonction d'ensemble qui satisfait à la loi IV qu'elle est une fonction *multiplicative*.

Il est aisé de vérifier que dans les ensembles finis à distribution homogène des poids la probabilité remplit toutes les lois I...IV, pourvu que l'ensemble des paires soit également à distribution homogène des poids. Pour vérifier, par exemple, la loi multiplicative, on remarquera que m_1 contenant a_1 éléments et m_2 en contenant a_2 , on peut former $a_1 \times a_2$ paires des éléments de ces ensembles. D'une façon analogue on forme $b_1 \times b_2$ paires d'éléments, en associant ceux de \mathcal{M}_1 à ceux de \mathcal{M}_2 . Soit δ_1 le poids commun aux éléments de \mathcal{M}_1 et δ_2 le poids commun à ceux de \mathcal{M}_2 . Toutes les paires des éléments ayant les poids égaux, par exemple Δ , nous avons par définition:

$$(5a) \quad p_{\mathcal{M}_1 \mathcal{M}_2}(m_1 m_2) = \frac{a_1 \cdot a_2 \cdot \Delta}{b_1 \cdot b_2 \cdot \Delta} = \frac{a_1 \cdot a_2}{b_1 \cdot b_2} = \frac{a_1 \cdot \delta_1 \cdot a_2 \cdot \delta_2}{b_1 \cdot \delta_1 \cdot b_2 \cdot \delta_2} = \\ = p_{\mathcal{M}_1}(m_1) \cdot p_{\mathcal{M}_2}(m_2).$$

Cette formule fait apparaître clairement l'utilité du mode adopté de l'écriture. Pour obtenir la probabilité composée il faut calculer la probabilité de chaque événement par rapport à l'ensemble correspondant, donc $p(m_1)$ par rapport à \mathcal{M}_1 , et $p(m_2)$ par rapport à l'ensemble \mathcal{M}_2 . La difficulté fondamentale dans la théorie des probabilités consiste justement dans l'évaluation correcte de ces probabilités par rapports aux ensembles \mathcal{M}_1 et \mathcal{M}_2 . Il arrive que l'ensemble \mathcal{M}_2 est une modification de l'ensemble \mathcal{M}_1 et que cette modification dépend d'un événement, auquel appartient la probabilité $p_{\mathcal{M}_1}(m_1)$. Dans ce cas on dit dans la théorie classique que les probabilités $p(m_1)$ et $p(m_2)$ appartiennent aux événements dépendants et que la loi de la probabilité composée ne s'applique pas sans restriction. Cette loi n'était vraiment pas applicable, si l'on voulait écrire d'une manière inadmissible $p_{\mathcal{M}_1}(m_1) \cdot p_{\mathcal{M}_1}(m_2)$ à la place de $p_{\mathcal{M}_1}(m_1) \cdot p_{\mathcal{M}_2}(m_2)$. (Bohmann, l. c., emploie dans ce cas l'écriture $p(a, b) = p(a) \cdot p_a(b)$, où a, b désignent les „événements“). Il faut remarquer pourtant, que l'ensemble \mathcal{M}_2 peut différer de l'ensemble

\mathcal{N}_1 non seulement à cause d'un événement dont la probabilité est $p_{\mathcal{N}_1}(m_1)$, mais aussi à cause des circonstances qui ont modifié l'ensemble \mathcal{N}_1 , tout en étant indépendantes du premier événement. Donc il n'est pas suffisant de dire: la loi IV s'applique, si les probabilités sont indépendantes; par contre il suffit de dire: il faut évaluer correctement $p(m_1)$ et $p(m_2)$ par rapport aux ensembles correspondants \mathcal{N}_1 et \mathcal{N}_2 .

Nous dirons donc en résumant: la loi (5 a) de multiplication des probabilités homogènes s'applique sans aucune restriction.

Remarque. Avant d'aller plus loin nous signalerons encore un exemple¹⁾ dans lequel on verra la notion de la probabilité composée confondue avec la notion de la probabilité d'une autre nature. On cherche la probabilité pour qu'un nombre entier $1 \leq n \leq 90$ soit divisible à la fois par 10 et par 6; or, les probabilités partielles étant $\frac{1}{10}$ et $\frac{1}{6}$, la probabilité „composée“ n'est pas $\frac{1}{10} \cdot \frac{1}{6}$ mais $\frac{1}{10} \cdot \frac{1}{3}$; on regarde donc les événements comme dépendants et par cette voie on cherche à élucider ce paradoxe apparent. Or, ce que l'on veut ici évaluer, ce n'est point la probabilité composée dans notre sens: $p_{\mathcal{N}_1 \mathcal{N}_2}(m_1 m_2)$, mais la probabilité simple $p_{\mathcal{N}_1}(D(m_1, m_2))$ (voir III, remarque). On peut — il est vrai — transformer cette probabilité en probabilité composée: $p_{\mathcal{N}_1 m_1}(m_1 D(m_1, m_2))$, comme il est facile de le voir, et appliquer la formule (5 a), qui donne en effet $p_{\mathcal{N}_1}(m_1) \cdot p_{m_1}(D(m_1, m_2))$, mais la voie la plus simple est sans doute celle d'utiliser la formule III, remarque:

$$p_{\mathcal{N}_1}(D(m_1, m_2)) = p_{\mathcal{N}_1}(m_1) + p_{\mathcal{N}_1}(m_2) - p_{\mathcal{N}_1}(m_1 + m_2).$$

Passons au cas général de distribution hétérogène des poids.

Les lois I, II et III sont faciles à vérifier pour une distribution hétérogène des poids dans les ensembles finis. Lorsqu'on en veut vérifier, par exemple, la loi additive, il suffit d'écrire les formules:

$$\begin{aligned} p_{\mathcal{N}}(m_1; \varphi_{\mathcal{N}}(e)) &= \frac{\delta_{i_1} + \delta_{i_2} + \dots + \delta_{i_{a_1}}}{\delta_1 + \delta_2 + \dots + \delta_b}, & p_{\mathcal{N}}(m_2; \varphi_{\mathcal{N}}(e)) &= \\ &= \frac{\delta_{k_1} + \delta_{k_2} + \dots + \delta_{k_{a_2}}}{\delta_1 + \delta_2 + \dots + \delta_b}, \end{aligned}$$

¹⁾ Voir p. ex. E. Borel, *Éléments de la théorie des probabilités*, Paris 1909, p. 29 et 30.

de les ajouter ensuite membre à membre et de comparer le résultat à la formule de $p_{\mathcal{N}}(m_1 + m_2; \varphi_{\mathcal{N}}(e))$ donnée par cette loi.

Or, quant à la loi IV, il est évident à priori, que la distribution des poids étant supposée arbitraire, cette loi ne s'applique pas toujours. En effet, on ne sait pas d'avance quels poids doit-on attribuer aux paires des éléments, lorsque les poids des éléments de \mathcal{N}_1 sont $\delta_1, \delta_2, \dots, \delta_{n_1}$ et ceux des éléments de \mathcal{N}_2 sont $\delta'_1, \delta'_2, \dots, \delta'_{n_2}$. On peut donc attribuer à leurs paires des poids tout à fait arbitraires; par exemples, les sommes des poids des éléments formant la paire, les quotients ou les produits de ces poids, ou même des poids qui en sont absolument indépendants. Dans le dernier cas la fonction $\bar{\varphi}_{\mathcal{N}_1 \mathcal{N}_2}(e)$ — et par conséquent l'expression $p_{\mathcal{N}_1 \mathcal{N}_2}(m_1 m_2; \bar{\varphi}_{\mathcal{N}_1 \mathcal{N}_2}(e))$, qui en dépend — prend des valeurs tout à fait arbitraires et, d'une façon générale, différentes pour les $\bar{\varphi}_{\mathcal{N}_1 \mathcal{N}_2}(e)$ différentes. Donc elle n'est pas constamment égale au produit

$$p_{\mathcal{N}_1}(m_1; \varphi_{\mathcal{N}_1}(e)) \cdot p_{\mathcal{N}_2}(m_2; \bar{\varphi}_{\mathcal{N}_2}(e)),$$

comme la loi IV l'exige.

Supposons, par exemple, que \mathcal{N} soit composé de deux éléments e_1 et e_2 et assignons le poids δ au premier, le poids 2δ au second. La probabilité de l'ensemble m_1 composé de l'élément e_1 seul est

$$p_{\mathcal{N}}(m_1) = \frac{\delta}{\delta + 2\delta} = \frac{1}{3},$$

tandis que celle de m_2 , que nous supposerons composé de l'élément e_2 , est

$$p_{\mathcal{N}}(m_2) = \frac{2\delta}{\delta + 2\delta} = \frac{2}{3}.$$

Le produit est $\frac{2}{9}$. Or, $m_1 m_2$ ne se compose que d'une seule paire (e_1, e_2) et $\mathcal{N}\mathcal{N}$ de quatre paires

$$(e_1, e_1), (e_1, e_2), (e_2, e_1) \text{ et } (e_2, e_2).$$

Assignons à chacune de ces paires le poids égal à la somme des poids de ses éléments constitutifs:

$$2\delta, 3\delta, 3\delta \text{ et } 4\delta$$

Nous aurons:

$$p_{\mathcal{M}\mathcal{N}}(m_1 m_2; s(e)) = \frac{3\delta}{12\delta} = \frac{1}{4} \neq \frac{2}{9},$$

ce qui prouve que la loi IV ne s'applique pas ici.

En prenant les quotients des poids au lieu de leurs sommes, on obtient tout pareillement:

$$p_{\mathcal{M}\mathcal{N}}(m_1 m_2; q(e)) = \frac{\frac{1}{2}}{1 + \frac{1}{2} + 2 + 1} = \frac{1}{9} \neq \frac{2}{9}.$$

Par contre, lorsqu'on attribue à chacune des paires le poids égal au produit des poids de ses éléments;

$$\delta^2, 2\delta^2, 2\delta^2 \text{ et } 4\delta^2,$$

on obtient:

$$p_{\mathcal{M}\mathcal{N}}(m_1 m_2; p(e)) = \frac{2\delta^2}{9\delta^2} = \frac{2}{9}$$

de sorte que la loi multiplicative se trouve dans ce cas vérifiée.

Nous allons démontrer le théorème suivant, dont l'importance est fondamentale:

Théorème. *Pour que la probabilité soit multiplicative pour tous deux sous-ensembles m_1 et m_2 de \mathcal{M}_1 et \mathcal{M}_2 , il faut et il suffit que le poids de chaque paire composée d'un élément de \mathcal{M}_1 et d'un élément de \mathcal{M}_2 soit égal au produit des poids de ces éléments multiplié par un facteur arbitraire constant (c'est-à-dire, le même pour toutes les paires)*

Ainsi toutes les fois que pour une raison quelconque on est forcé de distribuer aux paires les poids suivant une loi différente, la probabilité n'est pas multiplicative ou — comme on l'énonce dans la théorie classique — „les événements sont dépendants“.

Démonstration. Les probabilités de m_1 dans \mathcal{M}_1 et de m_2 dans \mathcal{M}_2 s'expriment par les quotients:

$$(6) \quad p_{\mathcal{M}_1}(m_1; \varphi_{\mathcal{M}_1}(e)) = \frac{\delta_{i_1} + \delta_{i_2} + \dots + \delta_{i_{a_1}}}{\delta_1 + \delta_2 + \dots + \delta_{b_1}}$$

$$p_{\mathcal{M}_2}(m_2; \bar{\varphi}_{\mathcal{M}_2}(e)) = \frac{\delta'_{k_1} + \delta'_{k_2} + \dots + \delta'_{k_{a_2}}}{\delta'_1 + \delta'_2 + \dots + \delta'_{b_2}}$$

et la probabilité de l'ensemble des paires $m_1 m_2$ dans l'ensemble de toutes les paires $\mathcal{M}_1 \mathcal{M}_2$ s'exprime par la formule:

$$(7) \quad p_{\mathcal{M}_1 \mathcal{M}_2}(m_1 m_2; \bar{\bar{\varphi}}_{\mathcal{M}_1 \mathcal{M}_2}(e)) = \frac{\Delta_{i_1 k_1} + \Delta_{i_1 k_2} + \dots + \Delta_{i_1 k_{a_2}} + \Delta_{i_2 k_1} + \Delta_{i_2 k_2} + \dots + \Delta_{i_{a_1} k_{a_2}}}{\Delta_{11} + \Delta_{12} + \dots + \Delta_{1b_2} + \Delta_{21} + \Delta_{22} + \dots + \Delta_{b_1 b_2}}$$

En multipliant les formules (6) membre à membre, on obtient la formule:

$$(8) \quad \frac{p_{\mathcal{N}_1}(m_1; \varphi_{\mathcal{N}_1}(e)) \cdot p_{\mathcal{N}_2}(m_2; \bar{\varphi}_{\mathcal{N}_2}(e))}{\delta_1 \delta'_1 + \delta_1 \delta'_2 + \dots + \delta_1 \delta'_{b_1} + \delta_2 \delta'_1 + \delta_2 \delta'_2 + \dots + \delta_{b_1} \delta'_{b_2}}$$

où le numérateur et le dénominateur contiennent — conformément à la loi de la multiplication des polynômes — autant de sommandes qu'il y en a dans ceux de la fraction (7).

Nous affirmons que pour que les membres droits des formules (7) et (8) soient égaux, quels que soient m_1 et m_2 , il faut et il suffit qu'on ait

$$\Delta_{rs} = c \cdot \delta_r \cdot \delta'_s,$$

où Δ_{rs} désigne le poids de la paire dont les éléments ont les poids δ_r et δ'_s .

1° La condition est suffisante. En effet, si l'on substitue aux Δ_{rs} dans la formule (7) les produits des poids correspondants multipliés par c , on obtient — toutes réductions faites — la fraction (8).

2° La condition est nécessaire. En effet, les membres droits des formules (7) et (8) étant par hypothèse égaux pour tous les sous-ensembles possibles de \mathcal{N}_1 et \mathcal{N}_2 , on a, en désignant tout court par c_1 et c_2 les dénominateurs qui restent invariables quand on passe d'un sous-ensemble à un autre:

$$\frac{1}{c_1} (\Delta_{i_1 k_1} + \Delta_{i_1 k_2} + \dots + \Delta_{i_1 k_{a_1}} + \Delta_{i_2 k_1} + \Delta_{i_2 k_2} + \dots + \Delta_{i_{a_1} k_{a_2}}) = \frac{1}{c_2} (\delta_{i_1} \delta'_{k_1} + \delta_{i_1} \delta'_{k_2} + \dots + \delta_{i_{a_1}} \delta'_{k_{a_2}}),$$

c'est-à-dire:

$$\Delta_{i_1 k_1} + \Delta_{i_1 k_2} + \dots + \Delta_{i_1 k_{a_1}} + \Delta_{i_2 k_1} + \Delta_{i_2 k_2} + \dots + \Delta_{i_{a_1} k_{a_2}} = c (\delta_{i_1} \delta'_{k_1} + \delta_{i_1} \delta'_{k_2} + \dots + \delta_{i_{a_1}} \delta'_{k_{a_2}}).$$

Cette égalité étant vraie, quels que soient m_1 et m_2 , elle subsiste en particulier dans le cas, où chacun de ces sous-ensembles n'est composé que d'un seul élément. On aura donc par exemple $\Delta_{i_1 k_1} = c \cdot \delta_{i_1} \delta'_{k_1}$ et d'une façon générale, en formant successivement toutes les paires possibles:

$$\Delta_{11} = c \cdot \delta_1 \delta'_1, \Delta_{12} = c \cdot \delta_1 \delta'_2, \dots, \Delta_{rs} = c \cdot \delta_r \delta'_s, \dots$$

c'est-à-dire, précisément la condition proposée.

J'insiste sur le fait que la multiplicativité n'est pas une propriété des ensembles \mathcal{N}_1 et \mathcal{N}_2 pris isolément, mais bien une propriété de l'ensemble des paires de leurs éléments. Cette distinction, peu importante pour le cas de la distribution homogène des poids (que l'on peut d'ordinaire étendre sans scrupules à l'ensemble des paires), est cependant d'une importance capitale pour les distributions hétérogènes, dont on ne sait de quelle façon faut-il communiquer cette hétérogénéité aux paires formées d'éléments des deux ensembles considérés.

Envisageons maintenant de plus près l'influence qu'exerce la valeur des poids sur la probabilité. En premier lieu, il est évident que la probabilité, comme quotient, ne change pas de valeur lorsqu'on fait varier simultanément celle de tous les poids dans une même proportion. On peut donc, pour plus de commodité, „normer“ les poids de façon que leur somme soit égale à 1. Il suffit à ce but de diviser chaque poids par la somme des poids de tous les éléments de \mathcal{N} . La distribution des poids modifiée de la sorte sera dite *normée*. Ce procédé conduit à une simplification notable des formules, notamment à la disparition des dénominateurs compliqués dans la plupart de démonstrations. Les poids étant des nombres positifs finis, leur somme est distincte de 0 et de ∞ . Les poids normés sont donc également les nombres positifs finis.

C'est à présent que se révèle en toute netteté la signification mathématique des poids: on constate que *chaque poids normé représente la probabilité de l'élément considéré dans tout l'ensemble donné* (ou plus exactement, la probabilité de l'ensemble composé de cet élément unique). On peut donc appeler les poids normés: *probabilités élémentaires*, bien entendu sans attacher à ce terme un sens métaphysique et sans avoir à craindre un cercle vicieux consistant à définir „la probabilité par la probabilité“. La probabilité d'un élément e_i dans \mathcal{N} est, en effet,

$$p_{\mathcal{N}}(e_i) = \frac{\delta_i}{\delta_1 + \delta_2 + \dots + \delta_i + \dots + \delta_b}$$

et, son poids normé étant aussi

$$r_i = \frac{\delta_i}{\delta_1 + \delta_2 + \dots + \delta_i + \dots + \delta_b},$$

on a:

$$p_{\mathcal{N}}(e_i) = r_i.$$

Ensuite, il est intéressant de savoir comment se comporte la probabilité, lorsqu'on fait varier les poids autrement que dans une même proportion. Supposons, par exemple, que l'on modifie leur valeur de façon que les nouveaux poids soient:

$$\delta'_1 = a_1 \delta_1, \delta'_2 = a_2 \delta_2, \dots, \delta'_b = a_b \delta_b.$$

La probabilité p_i d'un élément e_i devient évidemment

$$(9) \quad p'_i = \frac{a_i \delta_i}{a_1 \delta_1 + a_2 \delta_2 + \dots + a_i \delta_i + \dots + a_b \delta_b}$$

ou, si on a eu soin de normer les poids primitifs,

$$(10) \quad p'_i = \frac{a_i p_i}{a_1 p_1 + a_2 p_2 + \dots + a_i p_i + \dots + a_b p_b}.$$

La formule de cette transformation est très importante au point de vue de la théorie classique des probabilités. D'une part, elle projette un jour nouveau sur la loi de Bayes concernant la probabilité de la cause et, d'autre part, elle est en relation directe avec la notion de „l'espoir mathématique“.

La loi de Bayes s'exprime en effet justement par la formule (10), pourvu qu'on assigne un sens convenable aux poids a_1, a_2, \dots, a_b et aux probabilités primitives p_1, p_2, \dots, p_b . Ce sens est le suivant:

Certain effet ne peut être produit que par les causes A_1, A_2, \dots, A_n , mais pas en même mesure par chacune d'elles. La cause A_1 est par rapport à l'effet considéré de poids p_1 (poids normé ou probabilité; d'ailleurs on peut se passer sans cette hypothèse, en normant les poids plus tard, dans la formule toute faite), la cause A_2 est de poids p_2 et ainsi de suite. En outre, le fait que c'est la cause A_i qui se présente et non pas une autre n'est pas également „possible“ pour chaque cause, mais il a des poids a_1, a_2, \dots, a_b différents pour chacune. Dans ce cas, la probabilité pour que ce soit la cause A_i qui surviendra et produira précisément l'effet considéré (et non pas un autre) s'exprime par notre formule (10) ou bien par la formule (9), si les poids primitifs n'ont pas été normés auparavant.

Cette formule montre aussi que les probabilités des éléments ne subissent pas la même variation que les poids (multiplication par le facteur a_i), mais une variation proportionnelle à celle des poids. En désignant par N le dénominateur de notre formule, nous

aurons:

$$p'_1 = \frac{a_1 p_1}{N}, p'_2 = \frac{a_2 p_2}{N}, \dots, p'_b = \frac{a_b p_b}{N}$$

et

$$(11) \quad N = a_1 p_1 + a_2 p_2 + \dots + a_b p_b = a_1 \cdot \frac{p_1}{p'_1} = a_2 \cdot \frac{p_2}{p'_2} = \dots = a_b \cdot \frac{p_b}{p'_b}$$

C'est ce coefficient constant qui s'appelle *espoir mathématique* de la grandeur $x = a_1, a_2, \dots, a_b$. La formule (11) peut être aussi interprétée comme une transformation de la formule évidente

$$p_1 + p_2 + \dots + p_b = 1$$

par l'introduction des nouveaux poids.

Ces deux exemples suffisent pour faire voir comment diverses notions de la théorie classique, même très éloignées entre elles, viennent — grâce à notre définition — se lier l'une à l'autre, pour ainsi dire, organiquement, tandis que dans la théorie classique où elles ne servent que pour exprimer le même fait, à savoir, que les „événements“ ont souvent des poids différents, elles sont introduites d'une façon tout à fait artificielle et comme „deus ex machina“. Les notions que l'on cherchait à dévoiler dans des traités classiques, en leur consacrant des chapitres, ici se trouvent en quelque sorte comprimées dans une seule formule qui est au point de vue logique une prémisses indispensable pour la considération des lois fondamentales de ce calcul.

L'analyse des suites des „essais“ devient également plus accessible grâce à l'introduction des poids. On peut, en effet, considérer chaque „essai“ comme un élément, dont le poids est égal à la probabilité de l'„événement“ examiné. L'ensemble des essais devient donc en même temps un ensemble des éléments et on peut y appliquer directement les lois du calcul des probabilités. Si une suite des essais s'effectue dans des conditions invariables, la distribution des poids dans l'ensemble des essais est homogène. Elle est hétérogène dans le cas contraire. En admettant une suite infinie des essais, on entre dans la seconde partie du calcul des probabilités, consacrée aux ensembles *dénombrables*.

Nous nous sommes arrêtés un peu plus longtemps sur les probabilités finies. Nous l'avons fait pour deux raisons: d'abord, pour présenter un système complet des lois et des notions sur les-

quelles repose cette partie — close et libre de fautes logiques — du calcul des probabilités, partie ayant pour elle seule un très vaste **champ** d'applications. D'autre part, parce que les considérations qui vont suivre seront modelées précisément sur ces *faits simples*, ce qui se manifestera d'une manière particulièrement apparente au moment de l'introduction des intégrales à la place des sommes finies. La notation un peu compliquée, qui a été employée jusqu'ici, peut être considérablement simplifiée, si l'on renonce à des généralisations nouvelles pour n'envisager que les probabilités finies (comme dans les mathématiques des assurances, par exemple).

§ 3. Probabilité dans les ensembles dénombrables.

Êtant donné un ensemble dénombrable \mathcal{M} des éléments $e_1, e_2, \dots, e_n, \dots$ à poids $\delta_1, \delta_2, \dots, \delta_n, \dots$, si on veut calculer la probabilité d'un sous-ensemble quelconque m composé des éléments $e_{i_1}, e_{i_2}, \dots, e_{i_n}, \dots$, on ne peut faire usage de la définition admise pour les ensembles finis, car aussi bien le „nombre“ des éléments que la mesure de la somme de leurs poids est infinie (du moins dans le cas général; on peut, en effet, donner des exemples des distributions des poids telles que leur somme soit convergente).

Toutefois, en considérant les sous-ensembles finis $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_n, \dots$ de \mathcal{M} , composés respectivement d'éléments $e_1; e_1$ et $e_2; e_1, e_2$ et $e_3; \dots$, on peut calculer pour chaque \mathcal{M}_n la probabilité de l'ensemble des éléments communs de m et \mathcal{M}_n dans \mathcal{M}_n . Posons $m_n = D(m, \mathcal{M}_n)$.

Si les probabilités finies

$$p_{\mathcal{M}_1}(m_1), p_{\mathcal{M}_2}(m_2), \dots, p_{\mathcal{M}_n}(m_n), \dots$$

tendent vers une limite, cette limite prendra le nom de „probabilité dans un ensemble dénombrable“ ou bien (d'après M. Borel²⁾) tout court „probabilité dénombrable“.

On l'écrira:

$$(12) \quad p_{\mathcal{M}}(m) = \lim_{n \rightarrow \infty} p_{\mathcal{M}_n}(m_n).$$

On peut l'exprimer aussi, en disant qu'on examine la limite du quotient des mesures (linéaires) des sommes des poids assignés

¹⁾ *Les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Math. di Palermo, XXVII, 1 sem. 1909, p. 247—271.

aux sous-ensembles finis suivants:

$$m_1 \text{ et } \mathcal{M}_1, m_2 \text{ et } \mathcal{M}_2, \dots, m_n \text{ et } \mathcal{M}_n, \dots$$

A l'aide des théorèmes connus concernant la somme et le produit des limites on démontre aisément que la probabilité dénombrable ainsi définie satisfait aux lois I, II et III. La loi IV n'est satisfaite que lorsque la distribution des poids est multiplicative.

Notre définition a été construite, bien entendu, dans l'hypothèse que la suite considérée des probabilités partielles admet une limite. Or, il n'en est pas toujours ainsi, malgré qu'une telle suite soit toujours *bornée* (puisque en tout cas $0 \leq p_{\mathcal{M}_n}(m_n) \leq 1$). Même la distribution des poids étant homogène, la limite peut ne pas exister, de sorte que l'„événement“ considéré est dans ce cas — conformément à notre définition — *dépourvu de toute probabilité déterminée*.

Dans le cas d'une distribution homogène des poids notre formule de la probabilité dénombrable peut être mise sous une forme plus claire. Désignons par c_i l'unité ou le zéro suivant que l'élément à indice i appartient à m ou non. Nous aurons:

$$(13) \quad p_{\mathcal{M}}(m) = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n c_i \cdot \delta}{n \cdot \delta} = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n c_i}{n}.$$

Cette expression, connue dans l'Analyse sous le nom de *moyenne de Cesaro*, a été objet de plusieurs études et on a constaté que le quotient

$$\frac{c_1 + c_2 + \dots + c_n}{n}$$

tend vers la limite $\frac{1}{2}$ — il est vrai — *presque partout* (c'est-à-dire, pour presque toutes les suites $c_1, c_2, \dots, c_n, \dots$), mais, quand même, *pas pour toutes* les suites. Notamment, si à chacune de telles suites divergentes on fait correspondre un nombre dont le développement diadique se compose précisément de termes de cette suite, l'ensemble des nombres ainsi obtenu est — il est vrai — non dénombrable, mais sa mesure (lebesguienne) est nulle.

Il est facile, en effet, de construire une suite $\{c_n\}$ telle que la probabilité oscille toujours entre $\frac{1}{3}$ et $\frac{2}{3}$. En voici un exemple:

100 1111 00000000 1111 1111 1111 1111 000000000000...

Nous n'insisterons pas davantage sur cette définition, aujourd'hui universellement admise, à l'aide de laquelle on a résolu déjà beaucoup de beaux problèmes (par exemple, celui de Tchebycheff sur la probabilité pour qu'une fraction soit irréductible)¹⁾; il est toutefois à remarquer que dans le cas d'une distribution homogène des poids la probabilité d'un ensemble fini dans un ensemble dénombrable est toujours *nulle*, ce qui a donné lieu à des attaques „du point de vue expérimental“ contre cette définition, malgré que dans une expérience on n'ait jamais à comparer un nombre fini de cas à un nombre infini.

Cependant, si on admet, par contre, une suite infinie d'expériences (ou essais), la question devient moins „paradoxale“. En général, dans les vérifications „expérimentales“ de toute sorte il serait désirable de se baser sur la théorie des suites d'essais finies et tout au plus dénombrables. Les principaux points de cette théorie sont traités dans le mémoire déjà cité de M. Borel. L'attention y est attirée surtout sur la convergence et la divergence des processus de ce genre. Nous observerons en terminant que les poids ne peuvent être normés que lorsque leur somme est convergente. Dans ce cas, même le sous-ensemble m étant fini, la probabilité de m dans \mathcal{M} cesse d'être égale à 0.

§ 4. Probabilité dans les ensembles non-dénombrables.

Nous allons maintenant étendre la notion de la probabilité sur les ensembles non-dénombrables de points de l'espace à r dimensions ou, ce qui revient au même, sur les systèmes (x_1, x_2, \dots, x_r) de r nombres réels. Pour plus de clarté nous allons nous servir du langage géométrique. Nous ferons correspondre à chaque point, tout comme auparavant, un certain poids que nous nous figurerons sous forme d'un segment, à savoir, sous forme de l'ordonnée dressée en ce point. Désignons par \mathcal{M}' l'ensemble-somme des ordonnées qui correspondent à un ensemble non-dénombrable \mathcal{M} et par $\varphi(x_1, x_2, \dots, x_r)$ ou par $\varphi_{\mathcal{M}}(e)$ la fonction mesurable bornée, partout positive, qui détermine cette correspondance.

Nous appellerons *probabilité de m dans \mathcal{M}* le quotient ou la limite des quotients des mesures lebesguiennes $(r+1)$ -dimensionnelles des sommes de leurs ordonnées, si la mesure de \mathcal{M}' diffère de 0.

¹⁾ Cf. A. Markoff, l. c., p. 148.

La limite des quotients est à prendre dans le cas où \mathcal{M} est non borné: on calcule le quotient des mesures pour la partie de \mathcal{M}' comprise dans un domaine borné à $(r+1)$ dimensions et on fait augmenter ce domaine de façon qu'il contienne, à la limite, l'ensemble \mathcal{M}' tout entier. Si le quotient des mesures tend en même temps vers une limite, c'est cette limite que l'on appellera „probabilité de l'ensemble m dans l'ensemble non-borné \mathcal{M} “. Si, pour un m_1 quelconque, cette limite n'existe pas, la probabilité $p_{\mathcal{M}}(m_1)$ est indéterminée et m_1 sera dit „sous-ensemble singulier“ de la fonction $p_{\mathcal{M}}(m)$ ¹⁾.

Les mesures à employer sont $(r+1)$ -dimensionnelles. Si, par exemple, \mathcal{M} est situé dans un plan, les ordonnées occupent, d'une façon générale, un certain volume et c'est précisément le quotient des volumes (ou sa limite, lorsque \mathcal{M} est non-borné) qui représente la probabilité de m dans \mathcal{M} .

Nous ne définirons pas, pour le moment, la probabilité de m dans \mathcal{M}' pour le cas où les volumes (mesures 3-dimensionnelles) de m' et \mathcal{M}' sont nuls malgré que les ensembles m et \mathcal{M} eux-mêmes soient non-dénombrables.

La définition qui vient d'être présentée embrasse les probabilités continues, traitées déjà dans la théorie classique. Elles concernent la probabilité d'un domaine m dans un domaine plus grand \mathcal{M} à nombre égal de dimensions. Telles sont, par exemple, la probabilité d'un volume dans un autre volume ou même dans tout l'espace à 3 dimensions, celle d'un domaine plan dans un autre domaine plan ou dans le plan tout entier, celle des ensembles de segments d'une droite dans des segments plus longs ou dans la droite tout entière. Il s'y rattachent également les probabilités souvent envisagées aujourd'hui dans la théorie des ensembles et concernant un ensemble arbitraire (pas nécessairement un „domaine“) dans un domaine ou dans un autre ensemble arbitraire à nombre égal de dimensions.

Nous avons exclu le cas où les mesures des deux ensembles-sommes d'ordonnées sont nulles malgré que les ensembles m et \mathcal{M} soient non-dénombrables. Ce cas n'a pas été étudié jusqu'à présent

¹⁾ Cette remarque concerne également les ensembles dénombrables, lorsque la limite d'une suite de probabilités partielles n'existe pas.

au point de vue de la théorie des ensembles, quoique la théorie classique se soit occupée même des cas pareils. Ainsi, par exemple, on peut étudier la probabilité pour qu'un point soit situé sur un arc donné d'une courbe, en admettant qu'il peut occuper une position quelconque sur cette courbe tout entière¹⁾; il en est de même des surfaces courbes etc. Si l'on employait dans ces cas la mesure de Lebesgue, on aboutirait toujours à un quotient de deux zéros, c'est-à-dire à une probabilité indéterminée, de sorte qu'à l'aide de la définition précédente, qui est aujourd'hui universellement admise dans la théorie des ensembles, il nous serait impossible de résoudre des problèmes si intéressants. Il est donc nécessaire d'élargir de nouveau notre définition, en introduisant la notion des *mesures d'ordre inférieur* pour les ensembles placés dans un espace à un nombre supérieur de dimensions. Ainsi, il faut introduire la mesure linéaire pour les ensembles situés dans un plan, la mesure plane et linéaire pour les ensembles à 3 dimensions et, d'une façon générale, la mesure q -dimensionnelle pour l'espace à r -dimensions (où $q < r$).

Or, on trouve ces notions toutes prêtes dans les travaux de Carathéodory, comme nous l'avons déjà signalé dans § 1. Avant d'énoncer la définition générale, nous ferons encore observer une différence essentielle qui existe entre la mesure de Lebesgue et celle de Carathéodory.

La mesure (L) est finie pour tout ensemble borné, tandis que la mesure (C) peut devenir infiniment grande même pour les ensembles contenus dans les domaines finis, comme le prouvent, par exemple, les courbes ayant un arc infiniment grand dans un domaine fini plan et ne remplissant aucun domaine. Il suffit à titre d'exemple de construire un „volute“ composé de segments de longueur: $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots$

Ces préparatifs terminés, nous pouvons reprendre le fil conducteur de notre définition.

Lorsque la mesure (L) $(r+1)$ -dimensionnelle de \mathcal{M}' (ensemble-somme des ordonnées) est nulle et la mesure (C) r -dimensionnelle ne l'est pas, on envisagera le quotient de ces mesures d'ordre inférieur. Si ce quotient est fini, c'est lui que l'on appellera „probabilité de m dans \mathcal{M} “

¹⁾ Cf. E. Czuber, *Geometrische Wahrscheinlichkeiten und Mittelwerte*, Leipzig 1884.

Dans le cas où la mesure (C) r -dimensionnelle est également nulle on passe à la mesure $(r - 1)$ -dimensionnelle et ainsi de suite. Or, ce procédé ne peut-être répété qu'un nombre fini de fois. En effet, r est un nombre fini, car nous ne nous occupons ici que des espaces à un nombre fini de dimensions, et les mesures 1-dimensionnelles ne sont sûrement pas nulles, puisque les poids, et par conséquent les ordonnées qui les figurent, sont, par hypothèse, positives, donc de mesure linéaire distincte de 0.

Ceci dit, il ne nous reste à discuter que le cas des *probabilités limites*; c'est-à-dire, telles qu'en procédant comme plus haut on arrive aux mesures de m' et \mathcal{M}' qui sont infiniment grandes toutes les deux. Deux cas sont à distinguer:

1° Les parties des ensembles m et \mathcal{M}' contenues dans les domaines finis de l'espace à $(r + 1)$ -dimensions ont les mesures de Carathéodory finies. On procède alors selon la définition primitive, c'est-à-dire, on considère $\lim p_{\mathcal{M}'_n}(m_n)$.

2° Les parties des ensembles m' et \mathcal{M}' contenues dans les domaines finis ont une mesure infinie. Ce n'est que pour les cas spéciaux que l'on sait dire dans ces conditions si le quotient des mesures admet une limite ou non. Tel est, notamment, le cas où l'ensemble des ordonnées est dénombrable (à ce cas nous appliquerons les définitions du § 3), ensuite le cas où m ne diffère de \mathcal{M} que par un ensemble de mesure finie (il est évident que dans ce cas nous attribuerons au quotient des mesures la valeur 1).

Dans d'autres cas la probabilité de m dans \mathcal{M} n'est pas définie. Un tel sous-ensemble m sera dit aussi „sous-ensemble singulier de la fonction $p_{\mathcal{M}}(m)$ “.

Nous sommes ainsi arrivés à la fin de notre définition. Pour la résumer, nous dirons brièvement:

La probabilité d'un ensemble dans un autre s'exprime par le quotient des mesures (à nombre de dimensions le plus grand de ceux pour lesquels elles sont toutes les deux distinctes de 0) des ensembles-sommes des poids correspondant à ces ensembles ou par la limite de ce quotient, quand ces mesures sont infiniment grandes.

Cette définition comprend toutes les précédentes et en particulier celle de la probabilité dans les ensembles finis (donnée au § 1), puisque, les ensembles considérés étant finis, toutes les mesures

successives sont nulles sauf la mesure linéaire, qui est toujours finie comme somme des longueurs d'un nombre fini de segments.

Elle comprend également toutes les définitions classiques des probabilités continues, dénombrables etc., excepté les cas où on employait le mot „probabilité“ sans avoir établi au préalable la notion de la mesure. On trouvera dans le § 6 quelques remarques concernant cette application, incorrecte à mon avis, de la notion de la probabilité.

La possibilité d'enfermer dans une seule définition claire tous les cas traités est entièrement due à ce que nous nous sommes décidés d'introduire dès le premier abord la notion du poids, et partant, d'envisager aussitôt les „événements à divers degré de possibilité“. En effet, si nous nous étions bornés au quotient des mesures des ensembles m et \mathcal{M} , nous aurions obtenu un quotient de deux zéros aussi bien pour les probabilités finies que pour les probabilités dénombrables. Dans ce cas il nous serait impossible de les comprendre dans une définition commune avec les probabilités continues.

La question se pose, si la définition que nous venons de proposer satisfait aux lois formelles du calcul des probabilités, c'est-à-dire, aux lois I—IV. Or, pour les lois I et II cela se voit immédiatement et la loi additive III résulte du fait que la mesure de Lebesgue ainsi que celle de Carathéodory est une fonction additive. Quant à la loi IV, elle reclame une analyse un peu plus profonde, qui nous conduira d'ailleurs — comme nous allons le voir — à un résultat bien simple n'étant qu'une généralisation du théorème démontré pour les ensembles finis (page 47).

Pour commencer, cherchons à représenter la probabilité par une formule analytique. Envisageons en premier lieu le cas, où les mesures lebesguiennes des ensembles m' et \mathcal{M}' ne sont simultanément ni nulles, ni infinies. Elles sont dans ce cas — comme on le sait — les intégrales de Lebesgue de la fonction $\varphi(x_1, x_2, \dots, x_r)$, qui représente la distribution des poids (des ordonnées), prises, bien entendu, dans les ensembles m et \mathcal{M} .

Si on veut employer les intégrales ordinaires, prises dans les domaines contenant m et \mathcal{M} , on n'a qu'à remplacer $\varphi(x_1, x_2, \dots, x_r)$ par la fonction $f(x_1, x_2, \dots, x_r)$, qui lui soit égale pour les points de ces ensembles et qui soit égale à 0 pour tous les autres points. On a donc, dans l'espace à $r + 1$ dimensions, le suivant quotient des

intégrales r^{tes} de Lebesgue (se réduisant, comme on sait, aux intégrales multiples ordinaires dans les cas réguliers, qui seuls entrent en jeu dans les applications):

$$(14) p_{\mathcal{M}}(m; \varphi_{\mathcal{M}}(e)) = \frac{\int \int_{(m)} \dots \int \varphi(x_1, x_2, \dots, x_r) dx_1 dx_2 \dots dx_r}{\int \int_{(\mathcal{M})} \dots \int \varphi(x_1, x_2, \dots, x_r) dx_1 dx_2 \dots dx_r}.$$

Si les mesures en question étaient infinies, on représenterait la probabilité par la limite vers laquelle tend ce quotient, lorsqu'on fait augmenter indéfiniment le domaine correspondant.

Si la mesure lebesgienne de \mathcal{M}' est nulle, il faut passer à la mesure r -dimensionnelle de Carathéodory. Elle peut être représentée par une expression que nous appellerons *intégrale de Carathéodory*. Pour expliquer ce que nous entendons par ce nom, transportons nos considérations dans l'espace à 3 dimensions. Admettons que m et \mathcal{M} sont plans et, par suite, que m' et \mathcal{M}' (les ensembles-sommes des ordonnées ou des poids) sont à 3 dimensions. Leurs mesures 3-dimensionnelles (les intégrales de Lebesgue) sont donc nulles.

Or, les mesures de surface de ces ensembles 3-dimensionnels peuvent être finies; elles peuvent, par exemple, remplir les surfaces latérales de certains cylindres. Nous appellerons, pour un cas régulier, „intégrale de Carathéodory“ l'intégrale ordinaire qui sert à calculer la surface latérale de ce cylindre, c'est-à-dire l'intégrale:

$$\int_{(\mathcal{M})} f(x, y) ds.$$

En étendant cette définition aux ensembles irréguliers quelconques, nous allons appeler „intégrale de Carathéodory prise dans un ensemble \mathcal{M} “ simplement la mesure carathéodory'enne de l'ensemble \mathcal{M}' .

Or, si la mesure (C) q -dimensionnelle de \mathcal{M}' (où $q > r$) est distincte de 0, la probabilité de m dans \mathcal{M} , s'exprime par le quotient de deux intégrales $(q-1)^{\text{tes}}$ de cette nouvelle espèce. Le symbole

$\int f(x, y) ds$ désignant un intégrale simple, on obtient pour la probabilité cherchée la formule

$$(15) \quad p_{\mathcal{M}}(m; \varphi_{\mathcal{M}}(e)) = \frac{\int \int_{(m)} \dots \int \varphi(x_1, x_2, \dots, x_r) d\sigma}{\int \int_{(\mathcal{M})} \dots \int \varphi(x_1, x_2, \dots, x_r) d\sigma}$$

Les simples propriétés des intégrales de cette espèce se laissent déduire sans peine de celles de la mesure (C) signalées par Carathéodory. Nous croyons que la généralisation proposée peut se montrer utile aussi dans d'autres parties de l'Analyse: elle permettra, en effet, d'exprimer par une seule formule bien concise l'aire de certaines surfaces dans l'espace à 3 dimensions, les aires et les volumes dans celui à 4 dimensions et, d'une façon générale, les grandeurs à un nombre quelconque de dimensions contenues dans les espaces à un nombre de dimensions plus grand.

Sans insister ici sur les considérations évidentes concernant les *probabilités limites*, nous allons continuer la discussion, en nous bornant aux ensembles à nombre de dimensions le plus petit, c'est-à-dire, aux ensembles \mathcal{M} linéaires.

La probabilité s'exprime dans ce cas par un quotient de deux intégrales (lebesguiennes pour le cas le plus général et ordinaires pour les cas réguliers):

$$(16) \quad p_{\mathcal{M}}(m; \varphi_{\mathcal{M}}(e)) = \frac{\int_{(m)} \varphi(x) dx}{\int_{(\mathcal{M})} \varphi(x) dx} = \frac{\int_{x_1}^{x_2} f(x) dx}{\int_b^a f(x) dx}$$

Si tous les poids sont égaux, on a $\varphi(x) = c$, par conséquent:

$$p_{\mathcal{M}}(m) = \frac{c \times m}{c \times M} = \frac{m}{M}$$

où m et M désignent les mesures des ensembles m et \mathcal{M} .

On a donc le théorème suivant:

Théorème. *La distribution des poids étant homogène, la probabilité de m dans \mathcal{M} est égale au quotient des mesures des ensembles m et \mathcal{M} eux-mêmes (donc non seulement au quotient des mesures des ensembles-sommes des ordonnées m' et \mathcal{M}') — pourvu, bien entendu, que ce quotient ne soit pas indéterminé.*

Sans entrer en discussion minutieuse des cas exceptionnels, voyons encore si le théorème concernant la multiplication des probabilités est vrai pour la distribution homogène des poids.

Or, cela résulte déjà du théorème qui précède. Désignons respectivement par m_1, m_2, M_1 et M_2 les mesures des ensembles $\mathcal{M}_1, \mathcal{M}_2, \mathcal{N}_1$ et \mathcal{N}_2 .

L'ensemble $\mathcal{N}_1 \mathcal{N}_2$ composé de toutes les paires possibles d'éléments de \mathcal{N}_1 et de \mathcal{N}_2 , se présentera comme un ensemble plan de mesure $M_1 \times M_2$. D'une façon analogue l'ensemble $m_1 m_2$ sera de mesure $m_1 \times m_2$, de sorte que, en conservant la distribution homogène des poids aussi pour ces ensembles à 2 dimensions, on aura:

$$p_{\mathcal{N}_1 \mathcal{N}_2}(m_1, m_2) = \frac{m_1 \times m_2}{M_1 \times M_2} = p_{\mathcal{N}_1}(m_1) \times p_{\mathcal{N}_2}(m_2)$$

c. q. f. d.

Il existe toutefois — comme nous allons le démontrer — des distributions hétérogènes des poids pour lesquelles le théorème sur la multiplication des probabilités subsiste.

Soit $\varphi_1(x)$ la fonction (bornée et mesurable, donc sommable) qui caractérise la distribution des poids dans \mathcal{N}_1 et $\varphi_2(y)$ une fonction analogue pour \mathcal{N}_2 . Soit enfin $f(x, y)$ une fonction — naturellement à deux variables — qui représente la distribution des poids dans l'ensemble des paires $\mathcal{N}_1 \mathcal{N}_2$.

On demande que le théorème sur la multiplication des probabilités s'applique, c'est-à-dire, que l'on ait:

$$\frac{\iint_{(m_1 m_2)} \widehat{f}(x, y) dx dy}{\iint_{(\mathcal{N}_1 \mathcal{N}_2)} f(x, y) dx dy} = \frac{\int_{(m_1)} \varphi_1(x) dx}{\int_{(\mathcal{N}_1)} \varphi_1(x) dx} \times \frac{\int_{(m_2)} \varphi_2(y) dy}{\int_{(\mathcal{N}_2)} \varphi_2(y) dy}$$

Les intégrales du membre droit de cette égalité peuvent être également mises sous la forme d'intégrales doubles¹⁾. On aura donc:

$$\frac{\iint_{(m_1 m_2)} f(x, y) dx dy}{\iint_{(\mathcal{N}_1 \mathcal{N}_2)} f(x, y) dx dy} = \frac{\iint_{(m_1 m_2)} \varphi_1(x) \times \varphi_2(y) dy dx}{\iint_{(\mathcal{N}_1 \mathcal{N}_2)} \varphi_1(x) \times \varphi_2(y) dx dy}$$

¹⁾ Ce théorème, connu pour les cas réguliers (pour les domaines), a été généralisé pour les ensembles quelconques par Lebesgue et Fubini (Cf. De la Vallée-Poussin, 1912, t. II, p. 117 et suivantes).

Cette égalité subsiste, par hypothèse, pour tous les sous-ensembles (mesurables) des ensembles donnés \mathcal{N}_1 et \mathcal{N}_2 . Désignons par c_1 et c_2 les dénominateurs constants et par c leur quotient. Il reste:

$$c \times \int_{(m_1 m_2)} \varphi_1(x) \varphi_2(y) dx dy = \int \int_{(m_1 m_2)} f(x, y) dx dy.$$

Or, puisque cette égalité subsiste, quel que soit l'ensemble dans lequel on effectue l'intégration, les fonctions sous le signe d'intégrale sont toujours égales:

$$(17) \quad f(x, y) = c \times \varphi_1(x) \times \varphi_2(y).$$

Il est donc démontré que l'égalité (17) est une condition nécessaire pour que la loi de multiplication des probabilités soit applicable. En substituant cette formule à des expressions correspondantes dans la formule de la probabilité, il est aisé de prouver qu'inversement, l'égalité (17) est en même temps une condition suffisante pour que la loi considérée s'applique.

Nous pouvons donc énoncer le théorème suivant, qui est d'une importance capitale pour tout le calcul des probabilités:

Théorème. *Pour qu'on puisse appliquer la loi des probabilités „composées“, il faut et il suffit que la distribution des poids remplisse la condition suivante: le poids de chaque paire des éléments est égal au produit des poids de ces éléments multiplié par un coefficient constant pour toutes les paires.*

Nous avons ainsi obtenu — on s'en aperçoit aussitôt — pour ce cas très général la même loi, qui a été établie pour les probabilités finies (voir p. 47). Cette loi nous explique d'où vient cette relation étroite qui existe entre le théorème sur les probabilités „composées“ d'une part et la distribution des poids ou „densité“ de la probabilité de l'autre. En même temps, elle donne une signification précise et bien définie aux termes assez confus des „événements dépendants“ et „indépendants“. Il serait plus juste de dire: une distribution des poids *multiplicative* ou *non-multiplicative* dans un ensemble des paires.

Pour les ensembles à un nombre plus grand de dimensions les raisonnements sont tout à fait analogues. Par contre, nous n'avons pas examiné au point de vue de la loi multiplicative les probabilités qui s'expriment par un quotient des intégrales de Carathéodory. Aussi est-il facile d'étendre aux ensembles non-dénom-

brables (et, en particulier, *continus*) la formule (10) et, par suite, la loi de Bayes et la notion de l'espoir mathématique.

En terminant par ceci les considérations générales sur les principes du calcul des probabilités, nous allons montrer sur un exemple particulier, avec quelle facilité se laissent résoudre certains problèmes complexes de la théorie classique, lorsqu'on les aborde de notre point de vue.

§ 5. Une démonstration nouvelle de la loi de distribution des erreurs d'observation.

Particulièrement importante au point de vue des applications est la distribution des poids *symétrique*, c'est-à-dire, telle que les éléments équidistants de l'origine des axes des ordonnées aient des poids égaux.

Ainsi par exemple, pour les ensembles plans, la distribution des poids est dite *symétrique*, lorsque les extrémités des segments qui représentent ces poids sont situées sur une surface quelconque de révolution ayant Z pour axe de rotation.

Nous appellerons une distribution: *parfaitement symétrique*, lorsque la symétrie se conserve également pour l'ensemble des paires d'éléments, qui est à considérer, bien entendu, dans l'espace à un nombre double de dimensions.

Nous montrerons à présent, comment la loi connue de la distribution des erreurs (loi de Gauss) se déduit d'une manière très simple de notre théorie, sans qu'il faille faire appel à des hypothèses artificielles ou opérer avec des formules asymptotiques.

On demande qu'une distribution des poids dans l'ensemble de tous les points d'une droite soit *multiplicative* (pour l'ensemble des paires de ces points), qu'elle soit en même temps *parfaitement symétrique* et, enfin, qu'elle s'exprime par une fonction *continue positive*. En termes connus de théorie des erreurs cela revient à dire: les erreurs positives et négatives de même valeur absolue sont regardées comme „également probables“, de sorte que la fonction représentant la répartition des „probabilités“ de ces erreurs est *symétrique*. Cette fonction doit être, en outre, continue et partout positive. La question posée est la suivante: quelle doit être la distribution de ces „probabilités“ pour qu'on puisse y appliquer les lois fondamentales du calcul des probabilités (en particulier, la loi IV

ou loi multiplicative) et pour que la symétrie subsiste aussi pour les paires des erreurs, représentées par les points du plan? La „probabilité“ considérée doit, en outre, tendre vers 0, lorsque les erreurs croissent indéfiniment.

Dans l'énoncé qui précède il faut évidemment remplacer le mot „probabilité“ (mis entre „“) par le mot „poids“.

Ceci établi, désignons par $\varphi(x)$ la fonction inconnue, continue et partout positive, qui représente la distribution cherchée des poids. Désignons encore par $f(x, y)$ la fonction qui représente la distribution des poids dans l'ensemble composé de toutes les paires possibles des erreurs (s'étendant, par conséquent, sur le plan tout entier).

La distribution des poids pour les points situés sur l'axe X est donc:

$$f(x, 0) = \varphi_1(x)$$

et pour ceux de l'axe Y :

$$f(0, y) = \varphi_2(y).$$

La distribution demandée étant multiplicative, on a en vertu du théorème, page 62:

$$(18) \quad f(x, y) = c \times \varphi_1(x) \times \varphi_2(y).$$

De plus, la distribution représentée par $f(x, y)$ étant symétrique, $z = f(x, y)$ présente une surface de révolution, et par suite doit être de la forme:

$$(19) \quad z = f(x, y) = F(\sqrt{x^2 + y^2}).$$

En rapprochant les formules (18) et (19), on obtient l'équation fonctionnelle suivante à trois fonctions inconnues:

$$(20) \quad c \times \varphi_1(x) \times \varphi_2(y) = F(\sqrt{x^2 + y^2}).$$

Au moyen de différentiation logarithmique suivant les variables x et y ¹⁾, on en déduit ce résultat surprenant que toutes les trois fonctions inconnues de cette équation doivent avoir la même forme bien déterminée:

$$(21) \quad \varphi(u) = a e^{mu^2}.$$

La distribution cherchée des poids sur l'axe X (la fonction $\varphi_1(x)$) doit, par conséquent, avoir la forme suivante:

$$(22) \quad \varphi(x) = a e^{mx^2}.$$

¹⁾ Cf. H. Poincaré, *Calcul des probabilités*, 1912, p. 42.

Cette démonstration concise de la loi de Gauss suppose toutefois que la fonction cherchée est dérivable, de sorte que l'on se croit forcé d'ajouter cette nouvelle restriction à nos anciennes hypothèses.

Or, la restriction en question se montre parfaitement superflue et, partant, étrangère à la nature de notre problème. C'est notamment, grâce à une remarque que MM. Banach et Ruziewicz ont bien voulu me communiquer, que j'ai trouvé une démonstration tout à fait élémentaire et n'exigeant pas l'existence des dérivées des fonctions inconnues.

Cette démonstration, basée uniquement sur les hypothèses antérieurement admises, à savoir, que la fonction cherchée est continue et positive (de sorte que le poids d'aucun élément n'est égal à 0), est la suivante:

Posons $y = 0$ dans la formule (20). On en conclut que pour chaque x :

$$\varphi_1(x) = \frac{F(x)}{c \times \varphi_2(0)} = \frac{F(x)}{c_1},$$

puisque $\varphi_2(0)$ est par hypothèse un nombre fixe et différent de 0. D'une façon analogue, en posant $x = 0$, on déduit de cette formule que pour chaque y :

$$\varphi_2(y) = \frac{F(y)}{c \times \varphi_1(0)} = \frac{F(y)}{c_2}.$$

Soit, pour plus de brièveté: $\frac{c}{c_1 \times c_2} = k$.

La formule (20) prend donc la forme suivante:

$$(23) \quad k \times F(x) \times F(y) = F(\sqrt{x^2 + y^2}).$$

Nous avons ainsi obtenu une équation fonctionnelle ne contenant qu'une seule fonction inconnue, ce qui est une conséquence directe de la symétrie. Or, non seulement $f(x, y)$, mais aussi $F(x)$ est une fonction symétrique. La formule (23) implique en effet que

$$F(x) = \frac{F(\sqrt{x^2 + y^2})}{k \times F(y)},$$

ce qui prouve que $F(x)$ ne change pas de valeur, lorsqu'on y substitue $-\sqrt{x^2 + y^2}$ à x .

On peut donc mettre $F(x)$ sous la forme

$$F(x) = \psi(x^2),$$

d'où

$$k \times \psi(x^2) \times \psi(y^2) = \psi(x^2 + y^2).$$

Déjà cette forme d'une équation fonctionnelle suffit pour faire naître la supposition que ψ est une fonction *exponentielle*. Logarithmons donc notre équation — ce qui est parfaitement légitime, la fonction cherchée étant par hypothèse continue et positive. Nous aurons:

$$\log k + \log \psi(x^2) + \log \psi(y^2) = \log \psi(x^2 + y^2).$$

Posons

$$\log \psi(x^2) = g(x^2).$$

Donc

$$(24) \quad \log k + g(x^2) + g(y^2) = g(x^2 + y^2).$$

Posons

$$\log k + g(u) = h(u)$$

et ajoutons $\log k$ aux deux membres de l'équation (24). Nous aurons:

$$h(x^2) + h(y^2) = h(x^2 + y^2).$$

En prenant maintenant pour x^2 et y^2 deux valeurs positives arbitraires v_1 et v_2 , nous arrivons finalement à une équation fonctionnelle bien simple, qui est la suivante:

$$(25) \quad h(v_1) + h(v_2) = h(v_1 + v_2)$$

et qui subsiste pour tous v_1 et v_2 positifs.

Or, la seule solution continue (et même l'unique solution mesurable¹⁾) de cette équation — étudiée déjà par Cauchy — est représentée, comme on sait, par la fonction linéaire

$$h(v) = mv,$$

où m désigne un constant.

En revenant à nos notations antérieures, nous aurons successivement:

$$g(v) = h(v) - \log k = mv - \log k;$$

$$\log \psi(x^2) = mx^2 - \log k;$$

$$\psi(x^2) = F(x) = e^{mx^2 - \log k} = \frac{1}{k} e^{mx^2}.$$

¹⁾ Les démonstrations en ont été données (dans les hypothèses les plus générales) par M. Fréchet et par MM. Sierpiński et Banach, Fund. Math. t. I, p. 116 et 124.

Or, étant donné que $\varphi_1(x) = \frac{F(x)}{c_1}$, posons $\frac{1}{k \times c_1} = a$ et nous aurons finalement la loi suivante de distribution des poids sur l'axe X :

$$(26) \quad \varphi(x) = \varphi_1(x) = a e^{mx^2}$$

conformément au résultat précédent que nous avons obtenu en supposant l'existence de la dérivée.

Si, en outre, on tient compte de la condition accessoire d'après laquelle $\varphi(x)$ tend vers 0, lorsque x tend vers ∞ , on a

$$m = -k^2$$

et la formule (26) prend la forme connue de la théorie des erreurs

$$(27) \quad \varphi(x) = a e^{-k^2 x^2}.$$

Ce résultat peut s'énoncer sous la forme du théorème suivant:

Il n'existe qu'une seule distribution des poids à la fois continue, multiplicative, parfaitement symétrique et telle que $\lim_{x \rightarrow \infty} \varphi(x) = 0$, à savoir, la distribution suivant la loi des erreurs de Gauss:

$$\varphi(x) = a e^{-k^2 x^2}$$

Un raisonnement analogue, mais appliqué au cas de l'espace à 3-dimensions, permet d'établir une relation de même genre entre notre théorie et la loi de distribution des vitesses, donnée par Maxwell:

$$\varphi(x, y, z) = a e^{-k^2(x^2 + y^2 + z^2)}.$$

Il n'existe dans les ensembles à 3 dimensions qu'une seule distribution des poids à la fois continue, multiplicative et parfaitement symétrique, notamment, la distribution suivant la loi des vitesses de Maxwell.

Le théorème précédent nous explique bien suffisamment le rôle assez mystérieux de la fonction e^{mx^2} dans le calcul des probabilités; on la rattachait jusqu'à présent à la formule asymptotique de Stirling ou bien on employait des hypothèses assez obscures. Or, nous venons de voir que la loi des erreurs résulte au contraire des hypothèses fort simples ne faisant usage que des notions du calcul des probabilités et n'invoquant aucune hypothèse supplémentaire, créée ad hoc. Il en est de même de la loi de Maxwell.

Ainsi les objections soulevées par Bertrand, répétées par Poincaré et acceptées par les physiciens, paraissent tout à fait

mal fondées. Cette confusion est issue, comme je crois, de l'emploi imprécis du terme „phénomènes indépendants“ qui n'a pourtant d'autre signification dans le calcul des probabilités que celle du postulat de „multiplicativité“.

La voie qui nous a conduit à la démonstration de la loi des erreurs me paraît liée si naturellement à la théorie des probabilités et fondée sur des prémisses si évidentes et générales, qu'elle est particulièrement propre à entrer dans un exposé rigoureux de la théorie des erreurs.

La distribution homogène des poids est comprise dans la formule (22) comme un cas spécial, car pour $m = 0$ on a:

$$\varphi(x) = a.$$

Enfin, les poids exprimés par la formule (22) peuvent être aussi normés, c'est-à-dire, divisés par un facteur constant et tel qu'on ait:

$$\int_{-\infty}^{+\infty} a e^{-k^2 x^2} dx = 1.$$

On en déduit que $a = \frac{k}{\sqrt{\pi}}$, ce qui donne une distribution normée des poids:

$$(28) \quad \varphi(x) = \frac{k}{\sqrt{\pi}} e^{-k^2 x^2}.$$

Dans ce cas, la probabilité ne sera plus représentée par un quotient de deux intégrales, mais par une seule intégrale. Ce procédé peut être fréquemment appliqué à des cas plus généraux.

On pourrait pousser encore plus loin l'analogie avec les ensembles finis et se demander quelle est la signification des poids normés. Dans le cas des ensembles finis les poids normés désignaient les „probabilités élémentaires“ ou probabilités d'un élément. Ici une „probabilité élémentaire“ doit être entendue d'une façon un peu différente.

Or, $\varphi(x)$ étant évidemment égal à la *dérivée* de la probabilité évaluée pour un segment quelconque qui se termine par le point x , on pourrait appeler ce poids *vitesse* de variation de la probabilité dans le point donné. On appelle aussi $\varphi(x)$ la „densité“ de la probabilité.

§ 6. Remarques sur l'application de la notion de la probabilité aux ensembles autres que ceux des points. Le paradoxe de Bertrand et le problème de Buffon.

Au cours du développement de calcul classique des probabilités on se posait quelquefois des problèmes concernant la probabilité d'un sous-ensemble dans un ensemble dont la mesure n'a pas été préalablement fixée d'une façon assez nette et générale. On ne se demandait pas quelle est la *mesure* d'un sous-ensemble par rapport à celle de l'ensemble entier, mais on se demandait quel est le *nombre* des éléments d'un ensemble par rapport à celui des éléments d'un autre. Même chez C z u b e r ¹⁾ on rencontre une certaine indécision en ce qui concerne le choix entre les *nombres cardinaux* et la *mesure* comme notion constitutive fondamentale de celle de la probabilité. La probabilité dans les ensembles transfinis, C z u b e r la fait dépendre d'une „structure“ des ensembles pas plus précisée.

Cependant, dans tous les problèmes de ce genre il est sans doute question du rapport des *mesures* (convenablement définies) des ensembles considérés. Lorsqu'on a voulu, par exemple, assigner une mesure à l'ensemble des droites d'un plan, même en admettant que la distribution de leurs poids est homogène, on faisait correspondre à chaque droite un point du plan et on considérait implicitement la mesure de surface de cet ensemble des points comme mesure de celui des droites. Or, aucun procédé fixé d'établir cette correspondance n'existant pas à ce temps là (et même aujourd'hui), on la variait même pour un problème donné. On obtenait de cette façon plusieurs ensembles différents de points d'un plan et, par suite, des aires différentes pour exprimer la mesure de l'ensemble des droites.

A défaut de travaux systématiques sur ce sujet, nous ne pouvons pas rejeter à priori aucune notion, même la plus fantastique, de la *mesure*, créée pour des ensembles tels que celui de toutes les droites d'un plan, de tous les segments d'un plan, de toutes les cordes d'un cercle, de toutes les positions possibles d'une courbe sur une surface sphérique (comme c'est le cas, par exemple, des problèmes astronomiques étudiés par Poincaré) etc.

En modifiant pour un problème donné le mode de mensuration de l'ensemble considéré, on peut obtenir plusieurs résultats différents

¹⁾ l. c., I B., II édition 1908, p. 82—86.

et il n'y a là rien de paradoxal. On peut tout au plus reprocher aux auteurs de ces „paradoxes“ de ne pas avoir mis en évidence d'une façon assez claire la relation entre la probabilité et l'espèce de la mesure employée, comme c'est le cas, par exemple, du paradoxe de Bertrand.

La chose devient plus grave, lorsque dans les problèmes de ce genre on ne tient pas compte explicitement de la possibilité de plusieurs solutions différentes et lorsqu'on regarde implicitement le résultat acquis au moyen d'une mesure spécialement choisie (qui n'est pas nécessairement la plus convenable au point de vue mathématique — si on peut toutefois parler dans les mathématiques des conventions convenables et pas convenables) comme le seul résultat possible. Ce qui est pire, on „vérifie“ ensuite un résultat pareil par „expérience“, comme on l'a fait pour le problème connu de Buffon concernant la projection d'une aiguille sur un plan couvert des droites parallèles¹⁾, et on fait naître ainsi l'idée que ce résultat soit le seul justifié.

Or, en étudiant le problème de Buffon à notre point de vue, j'ai trouvé naturellement, outre de la vieille solution, une solution nouvelle, aussi bien justifiée mathématiquement que la première et même plus suggestive.

Je renonce à développer ici plus amplement ces faits très simples, d'autant plus que — comme je me suis convaincu — M. Borel était déjà parvenu au même résultat²⁾.

Il serait très intéressant de voir si les expériences faites au sujet du problème de Buffon, mais exécutées dans les conditions différentes de celles des essais „classiques“ ne puissent fournir aussi des résultats différents, mais prévus théoriquement; en d'autres termes,

¹⁾ On trouve une description de ces expériences chez Czuber, *Geometrische Wahrscheinlichkeiten*, p. 85.

²⁾ Notamment, dans le livre de E. Borel, *Le hasard*, Paris 1914, p. 85 on trouve cette remarque:

„On peut aisément traiter un problème qui est la généralisation du précédent (probabilité pour qu'une corde d'un cercle ait une longueur supérieure à une longueur donnée) et déduire de la solution de ce problème celle du problème de l'aiguille. On obtiendra ainsi pour le problème de l'aiguille trois solutions différentes, suivant que l'on aura choisi la première, la seconde ou la troisième des solutions proposées par Bertrand. Comme la solution expérimentale du problème de l'aiguille n'est pas contestée, on est conduit, dans ce cas *pratique*, à constater que la première des solutions de Bertrand est la seule bonne“.

si les sciences expérimentales n'exigent que l'on choisisse parmi tous les procédés possibles de mesurer un ensemble (d'éléments autres que les points) un procédé spécial se réalisant dans les expériences pratiques. Cette question, si importante qu'elle ne soit au point de vue des *applications* du calcul des probabilités, ne peut exercer, bien entendu, aucune influence sur la liberté créatrice d'une théorie mathématique.

Poincaré conçoit la question à un point de vue différent, mais (au moins théoriquement) aussi bien justifié que celui qui fut exposé ici et, peut-être, même équivalent à lui, bien que moins „évident“. Nous avons supposé implicitement, pendant toute la discussion de ces problèmes, que la distribution des poids est homogène, c'est-à-dire — pour employer l'ancienne expression — que tous les éléments des ensembles considérés sont „également possibles“. Les différentes solutions d'un même problème s'obtenaient par les différents choix des *mesures*. Or, Poincaré parvient au même groupe des résultats différents, en faisant varier non pas le système des *mesures*, mais celui des *poids*, c'est-à-dire — en termes classiques — en regardant les éléments des ensembles considérés comme présentant divers „degrés de possibilité“. Cette hypothèse, bien qu'elle soit équivalente à la nôtre, me paraît moins naturelle à cause de l'absence d'une „raison suffisante“ qui nous fasse distinguer les uns des autres les éléments tout à fait semblables d'un même ensemble, par exemple, les droites situées dans un plan. Cependant, le fait que l'on peut parvenir à des résultats identiques par ces deux voies différentes est très instructif par lui-même. Ce fait devient plus clair dans la théorie des transformations des intégrales par l'introduction des nouvelles variables d'intégration.

Léopol, le 19 Novembre 1920.
