ANNALES DE L'I. H. P.

HANS REICHENBACH

Les fondements logiques du calcul des probabilités

Annales de l'I. H. P., tome 7, nº 5 (1937), p. 267-348

<http://www.numdam.org/item?id=AIHP_1937__7_5_267_0>

Les fondements logiques du calcul des probabilités

par

M. Hans REICHENBACH,
(Istanbul).

CHAPITRE I.

L'ÉTAT DES PROBLÈMES.

Les fondements logiques du calcul des probabilités offrent autant d'intérêt au philosophe qu'au mathématicien. Ce furent, il est vrai, les mathématiciens qui développèrent ce calcul et qui en firent l'instrument ingénieux et indispensable à toute science qui veut aboutir à un traitement quantitatif des phénomènes naturels. Les philosophes reconnurent bientôt que cette découverte avait une portée beaucoup plus vaste, que ce calcul mathématique renfermait des problèmes touchant les fondements mêmes de la connaissance; depuis lors ils ont toujours essayé de poursuivre des études dans une autre direction qui ne développe pas les conséquences ramifiées du calcul, mais qui s'oriente au contraire vers une reconstruction des fondements sur lesquels est basé tout l'édifice de la théorie.

Si je tire, avec cette remarque, une ligne de démarcation entre l'intérêt des mathématiciens et celui des philosophes, je ne veux pas dire cependant que cette ligne ait jamais séparé les hommes autant que les problèmes. Au contraire, on voit les mêmes hommes travailler dans les deux domaines; je n'ai qu'à citer les noms d'un Leibniz et d'un Laplace pour montrer que les esprits productifs dans les deux domaines sont pour la plupart ou bien des mathématiciens philosophes ou des philosophes mathématiciens. Le nom de l'Institut, dont la très aimable invitation m'a donné la possibilité d'exposer ici mes idées, rappelle un des représentants les plus éminents de cette union entre les mathématiques et la philosophie, de l'héritier de la mission historique d'un Leibniz et d'un Laplace. Je suis heureux d'avoir cette occasion d'exprimer mon admiration profonde envers Henri Poincaré dont les ouvrages sur le calcul des probabilités ont fait sur moi dès le temps de mes études universitaires la plus grande impression; et je me permets d'interpréter ma présence ici comme le signe que la tradition de Henri Poincaré, celle de l'entente cordiale entre les mathématiques et la philosophie, continue à caractériser l'esprit de cet Institut.

Les problèmes philosophiques relatifs aux probabilités se divisent en deux groupes principaux. Le premier concerne les *relations internes* du système mathématique des probabilités, la connexion entre les hypothèses et les théorèmes qu'on en déduit, la formation de la structure logique, la définition des concepts, etc. Le second s'occupe des *relations externes* du concept de probabilité, à savoir des *relations entre probabilité et réalité;* il s'agit ici du problème de l'applicabilité du calcul à des objets réels, de la validité des résultats du calcul pour le monde physique. Cette distinction correspond à la distinction entre le *système mathématique* et *le système physique* du calcul des probabilités; d'une façon analogue, la discussion philosophique se ramifie en une *analyse logique* et une *analyse épistémologique* du problème de la probabilité.

La séparation exprimée dans cette distinction ne doit pas être conçue comme rigide; les deux groupes de problèmes ne sont pas indépendants l'un de l'autre. Le système mathématique des probabilités a été construit de façon à être adapté aux besoins de la pratique; d'autre part, la question de l'applicabilité dépend de la forme conceptuelle du système mathématique, et ne peut pas être résolue avant l'analyse complète de ce système. Malgré ces relations, il convient de respecter dans l'analyse philosophique la distinction indiquée, car ces deux groupes de problèmes répondent à des questions logiquement séparées.

Le problème des relations internes ou problème logique de la probabilité a trouvé sa première solution dans les systèmes d'un Pascal, d'un Fermat, d'un Laplace. On posait au début un certain nombre d'axiomes, ou de définitions, par exemple la définition de la probabilité comme rapport entre le nombre des cas favorables et celui des cas possibles, la notion des cas dits « également possibles », etc. A partir de cet ensemble d'hypothèses et de définitions on développait le calcul des probabilités qui constituait déjà à cette époque une discipline assez compliquée. On s'aperçut bientôt que cette discipline ne différait en aucune manière des autres chapitres des mathématiques quant à la rigueur de ses relations internes, que l'incertitude liée à la notion de « probable » ne s'infiltrait pas dans le système, que les relations entre prémisses et conséquences relevaient du type strict de la déduction logique. La probabilité ne commence, pour ainsi dire, qu'avec le passage à la réalité, tandis que le système mathématique est tout aussi rigoureux que toute autre discipline mathématique déductive, comme par exemple la géométrie.

Avec la découverte de cette différence entre le système des relations internes et la signification du concept de probabilité, on était entré dans la voie qui mène à la séparation stricte de ces deux questions. La définition de la probabilité comme rapport entre les cas favorables et les cas possibles offrait des difficultés; d'abord pour le problème épistémologique dont nous allons parler, et ensuite pour certains cas d'application qui ne présentent pas la structure logique des cas « également possibles », comme par exemple les statistiques sociales. On remplaça alors cette définition par une autre suivant laquelle la probabilité est la limite d'une fréquence, définition que nous trouvons déjà chez Poisson (¹); et l'on s'aperçut que ce changement de définition n'entraînait aucun changement des relations internes du calcul. A partir de cette découverte il n'y avait qu'un pas jusqu'à l'idée de renoncer à toute définition de la probabilité, de se contenter de construire le calcul des probabilités comme système relationnel, et de considérer l'interprétation du concept fondamental de ce système comme un problème concernant plutôt les applications du calcul que ce calcul lui-même. Dans l'élaboration de cette idée on pouvait invoquer comme modèle la conception analogue de la géométrie, selon laquelle le système relationnel de la géométrie peut être entièrement séparé de son interprétation physique.

Ce développement n'a pu être achevé que grâce à l'extrême rigueur

(¹) S. D. POISSON, *Recherches sur la probabilité des jugements*, Paris 1837. La même définition se trouve, sous une forme très précise, dans BOOLE, *Laws of Thought*, London, 1854, p. 295.

qui distingue les mathématiques actuelles des mathématiques anciennes. C'est avant tout la méthode axiomatique, introduite par les travaux classiques de Hilbert sur la géométrie, qui constitue l'instrument de ces constructions mathématiques. Les deux conceptions, celle de l'interprétation de la fréquence et celle du système purement relationnel, n'ont donc pu trouver leur forme définitive que dans une série de recherches récentes parmi lesquelles je citerai celles de Bohlmann, Borel, von Mises, Tornier, Dörge, Copeland, Kolmogoroff.

A côté de la méthode axiomatique, on rencontre aussi des recherches d'un degré élevé de complication mathématique, concernant des théorèmes assez éloignés de la base axiomatique, mais portant sur la construction des concepts fondamentaux. Déjà le théorème de Bernouilli avait révélé une connexion fort intéressante entre un théorème déduit des axiomes fondamentaux et l'interprétation de la fréquence; cette connexion a été approfondie dans des recherches modernes que M. Darmois (1) nous a exposées dans une très intéressante étude. D'autre part les recherches de MM. Borel et Copeland sur l'existence de séries d'une certaine irrégularité supposée dans le théorème de Bernoulli montrent que la discussion des fondements du système probabilitaire est susceptible de conduire à des considérations mathématiques très compliquées.

Le problème des relations externes, ou problème épistémologique de la probabilité, a parcouru, lui aussi, diverses étapes de développement progressif. La correspondance entre le système mathématique des probabilités et la réalité physique a fait depuis le début l'objet de nombreuses discussions. Elle a trouvé sa première expression dans le concept des cas dits « également possibles », terme qui visait à exprimer le lien logique entre la symétrie des cas élémentaires des jeux de hasard, et l'égalité des fréquences observées. Cette correspondance, si naturelle qu'elle puisse paraître à première vue, ne semble pas moins étrange dès qu'on la soumet à un examen un peu plus approfondi; le besoin de la justifier s'est fait sentir et dès le début on trouve un certain nombre de travaux consacrés à cette justification. C'est dans le cadre de ces réflexions qu'on a introduit *le principe de la raison manquante*, considéré comme complément négatif *du principe de la raison suffi-*

(1) *Recherches philosophiques*, IV, 1934-1935, p. 348.

sante. Ce dernier justifie l'arrivée de *l'effet* si la *cause* est réalisée; son complément négatif doit justifier *l'égalité des fréquences d'apparition* si l'on *ne connaît pas de cause* pour qu'un des cas élémentaires soit privilégié.

Je n'ai pas besoin d'entreprendre ici la critique de cette idée dont le rôle a été restreint à la fonction historique d'ouvrir la discussion épistémologique du problème des probabilités. On voit bien, et la critique moderne l'a exposé plus d'une fois, que ce principe de la raison manquante, qui attribue à l'ignorance humaine la fonction de déterminer la fréquence dans des séries d'évènements objectifs, n'est que l'expression linguistique du fait qu'on n'avait pas trouvé d'explication à cette régularité objective. Permettez-moi d'ajouter qu'aujourd'hui ce problème, qui est d'ailleurs restreint aux jeux de hasard et ne joue aucun rôle dans les autres applications du calcul des probabilités, a trouvé une solution simple que Poincaré a indiquée. Nous ne nous y attarderons pas; seule la fonction historique qui se rattache à cette idée nous intéresse ici.

De cette idée dérive la conception subjective de la probabilité d'après laquelle la probabilité n'est qu'un expédient de l'ignorance humaine, un moyen provisoire inévitable parce que l'esprit humain ne peut pas arriver à une connaissance parfaite des causes qui provoquent les phénomènes physiques. La forme célèbre dont Laplace a revêtu cette idée, en imaginant une intelligence surhumaine aux yeux de laquelle « l'avenir comme le passé serait présent » (1) est devenue le prototype de la conception subjective de la probabilité. Cette conception, cependant, ne convient pas au problème; le fait que les phénomènes montrent, objectivement, des régularités statistiques s'oppose à une conception subjective de la probabilité. Cournot qui a soutenu la thèse du caractère objectif de la probabilité a répondu à Laplace que son intelligence surhumaine ne renoncerait point à l'usage des lois de probabilité, qu'elle ne différerait de l'homme qu'en ce qu'elle serait un meilleur mathématicien des probabilités. « Une intelligence supérieure à l'homme ne différerait de l'homme qu'en ce qu'elle se tromperait moins souvent que lui, ou même ne se tromperait jamais dans l'appli-

(1) P. S. LAPLACE, *Essai philosophique sur les probabilités*, édition nouvelle dans la collection « *Les maîtres de la pensée scientifique* », Paris, 1921, p. 3.

cation de cette donnée de la raison... elle pousserait plus loin que nous et appliquerait mieux la science de ces rapports mathématiques, tous liés à la notion du hasard ([1]) ».

Cependant la conception objective à laquelle les idées de Cournot ont frayé le chemin est déterminée dans ses débuts, elle aussi, par le principe de la raison manquante. Les prétentions d'objectivité sont basées, pour les anciens représentants de cette conception, sur un pouvoir législatif de la raison, dont le fameux principe n'est qu'une expression particulière. Cette idée est l'émanation d'une conception aprioriste de la logique selon laquelle la logique, quoique fondée sur la seule raison, prescrirait des lois à la nature. La situation des lois de probabilité ne diffère donc pas, selon cette conception, de celle des lois générales de la logique; les lois de la probabilité sont des « données de la raison » d'après l'expression même de Cournot. La correspondance entre le calcul des probabilités et les phénomènes naturels n'est qu'un cas particulier de la correspondance générale entre raison et réalité.

Cette conception rationaliste de la probabilité, produit ultérieur du rationalisme des XVII^e^ et XVIII^e^ siècles, a conduit la théorie des probabilités dans une nouvelle direction qui a joué un rôle dominant jusqu'à nos jours bien qu'elle se soit séparée de l'interprétation aprioriste : nous voulons parler de l'idée que la théorie des probabilités est à concevoir comme une logique généralisée. Leibniz, ce visionnaire de la logique, a anticipé ce développement en demandant « une nouvelle espèce de logique, qui traiterait des degrés de probabilité »; ce ne fut cependant qu'un siècle et demi plus tard que ce projet fut réalisé. C'est Boole, dans son *Investigation of the laws of thought*, qui nous a présenté le premier système d'une logique des probabilités. Quoique la construction de ce système relève encore de la conception rationaliste de la logique [Boole conçoit les lois de la logique comme « standards not derived from without, but deeply founded in the constitution of human faculties » ([2])], elle présente déjà une perfection surprenante, due à la forme mathématique de cette logique; cette logique probabilitaire fait partie de l'œuvre même dont l'apparition doit être considérée comme la naissance de la logistique moderne. Depuis Boole, cette conception

([1]) M. A. COURNOT, *Théorie des chances et des probabilités*, Paris, 1843, p. 104.
([2]) BOOLE, *loc. cit.*, p. 2.

logique de la probabilité s'est maintenue et développée jusqu'à nos jours; je ne mentionnerai que les noms de Venn, Peirce et Keynes.

Malgré le grand progrès marqué par l'avènement de la conception logique des probabilités, on ne peut pas considérer le problème épistémologique comme résolu tant que la notion de probabilité est rattachée à la conception aprioriste de la logique. Elle bute, dans ce cadre, contre plus d'obstacles que l'interprétation de la logique classique elle-même; la « croyance rationnelle » d'un Boole et d'un Keynes serait une espèce d'*a priori synthétique*, et c'est là un concept qu'une philosophie scientifique ne peut pas admettre. La logique classique, l'*a priori analytique,* admet une interprétation d'après laquelle la logique n'est qu'un système de règles linguistiques, n'empiétant pas sur le contenu des énoncés; la logique probabilitaire conçue comme une « croyance rationnelle » par contre échapperait à cette interprétation.

La raison en est que la logique de probabilités renferme le problème de l'inférence inductive. Cette inférence doit nous conduire à des conclusions dépassant le contenu des prémisses; elle ne se justifie donc pas comme transformation équivalente des propositions. Cette inférence n'a pas lieu, il est vrai, dans le système mathématique des probabilités, qui ne contient que des déductions strictes; mais elle est nécessaire pour les applications de ce système à la réalité. Elle intervient dans le problème de la correspondance entre les formules du calcul et les phénomènes de la nature; et la « croyance rationnelle » qui doit justifier l'inférence inductive n'est que le palliatif inventé pour voiler l'abîme qui sépare la raison et la réalité. C'est pourquoi l'inférence inductive constitue le nœud de toutes les difficultés du problème épistémologique de la probabilité.

Certains représentants de l'interprétation par la fréquence ont cru pouvoir résoudre d'une façon simple le problème de la validité du calcul des probabilités. Si la probabilité est définie comme une fréquence, disent-ils, et si les lois du calcul sont déduites de cette définition, il s'ensuit que ces lois seront valables pour des phénomènes statistiques. C'est là, cependant, une pseudo-solution. La définition supposée dans le calcul concerne une série infinie, tandis que la série donnée physiquement est finie; l'inférence inductive apparaît donc ici dans la transition de la partie observée de la série à sa prolongation. On ne peut pas se soustraire à ce problème en disant que les mathématiques s'occupent seulement de phénomènes idéalisés et que les relations entre le système conceptuel et la réalité ne touchent plus le mathématicien. Cela reviendrait seulement à faire passer le problème entre les mains des philosophes; mais comme les philosophes, en général, refusent de s'occuper si intimement des questions concernant les mathématiques, la responsabilité retomberait sur les philosophes mathématiciens qui préféreraient pouvoir bénéficier de la collaboration des mathématiciens philosophes.

Quelle que soit la manière dont on veuille formuler le problème de la coordination entre le système relationnel des probabilités et la réalité, la solution du problème épistémologique des probabilités devra comprendre une solution du problème de l'induction. C'est là l'origine des difficultés du problème épistémologique de la probabilité, mais aussi la raison de son importance. L'inférence inductive est le nerf de la méthode scientifique; toute solution correcte du problème de la probabilité renfermera donc une réponse au problème fondamental du progrès de la connaissance qui part d'observations pour aboutir à des prédictions.

J'ai essayé de donner dans ce qui précède un bref aperçu de l'état des problèmes constituant le point de départ des recherches que je vais exposer ci-après. Le problème épistémologique est à la base de ces recherches, et la solution que j'indiquerai fera usage des conceptions d'une logique probabilitaire, en évitant cependant une interprétation aprioriste. D'autre part, il m'a semblé nécessaire d'entrer d'abord dans une analyse des relations internes du système mathématique des probabilités. Les constructions axiomatiques présentées par d'autres auteurs ne m'ont pas satisfait parce qu'elles se sont trop éloignées des applications. Il me semble que parfois, dans la construction des systèmes probabilitaires par les mathématiciens, le souci d'élégance et de totalité ait refoulé les exigences des applications physiques et de l'analyse épistémologique. J'ai donc cherché à construire un système du calcul des probabilités qui soit assez large pour englober toutes les applications de ce concept en physique, et qui soit en même temps basé sur des hypothèses assez simples pour rendre possible la résolution du problème épistémologique. L'exposé de mes idées montrera si j'ai réussi à atteindre ce but (1).

(1) L'analyse de la théorie des probabilités exposée dans ce qui suit a été publiée pour la première fois dans mon article : *Axiomatik der Wahrscheinlichkeitsrechnung*.

CHAPITRE II.

CALCUL ÉLÉMENTAIRE DES PROBABILITÉS.

1. Considérations logiques préliminaires. — La construction du système des probabilités que je vais présenter ici a pour but d'englober, en un seul système, tous les concepts mathématiques et logiques qui jouent un rôle dans ce calcul. En poursuivant ce but, notre première tâche sera de construire la forme logique de l'énoncé probabilitaire.

Le premier caractère logique qu'il faut envisager ici est la constatation que l'énoncé probabilitaire concerne une *relation*. Nous ne parlons pas, par exemple, de la probabilité en général de trouver sur cette table un dé avec le « six » en haut; nous ne considérons cette probabilité que si nous *jetons* d'abord le dé; en d'autres termes la probabilité concerne une relation entre les deux événements : « lancement du dé » et « apparition du six ».

Le second caractère logique concerne la description des événements. Nous donnons cette description en précisant seulement la classe à laquelle l'événement appartient; et nous omettons tous les caractères individuels qui distinguent l'événement considéré des autres appartenant à la même classe. Dans notre exemple nous avons deux classes, A et B; A est la classe indiquée par la description « lancement du dé », B est la classe indiquée par la description « apparition du six ».

Le troisième caractère logique consiste en ce que l'énoncé ne concerne pas seulement les événements individuels en question, mais tous les événements des deux classes, sous la seule condition que ces événements soient donnés dans un certain ordre, qu'ils constituent donc une *suite*.

(*Math. Zeitschr.*, 34, 1932, p. 568). La conception logique de la théorie des probabilités a été exposée dans *Wahrscheinlichkeitslogik* (*Berichte d. Berliner Akademie, math. Kl.*, 1932); la première publication concernant mes idées sur l'induction se trouve dans *Die logischen Grundlagen des Wahrscheinlichkeitsbegriffs*, Erkenntnis 3, 1933, p. 401. Toutes ces idées ont été résumées et exposées en détail dans mon livre *Wahrscheinlichkeitslehre*, Leiden 1935, Sijthoff, auquel je renvoie pour beaucoup de détails et de considérations supplémentaires que j'ai dû supprimer dans l'exposition abrégée présentée ici.

Nous écrivons donc l'énoncé probabilitaire sous la forme (¹)

(1) $$(i)(x_i \varepsilon A \underset{p}{\supset} y_i \varepsilon B).$$

En langage ordinaire : « pour tous les i, si x_i appartient à la classe A, cela implique avec la probabilité p que y_i appartient à la classe B ». La relation dénotée par « $\underset{p}{\supset}$ » est appelée « implication de probabilité ».

Exemple : x_i état d'un homme pendant un examen médical, A tuberculose, y_i état de l'homme une année plus tard, B mort.

Les deux événements x_i et y_i sont coordonnés; la règle de cette coordination doit être donnée, ainsi que l'ordre dans lequel la variable x_i parcourt toutes ses valeurs. La relation entre x_i et y_i peut être l'identité, c'est-à-dire x_i peut être identique à y_i; par exemple on peut considérer la probabilité pour qu'un malade atteint de tuberculose présente simultanément certains ulcères, etc.

Au lieu de (1), nous introduirons une notation abrégée

(2) $$(A \underset{p}{\supset} B),$$

dont la signification est donnée par (1).

Les deux notations (1) et (2) seront appelées *notations implicatives;* il est commode d'introduire une autre notation que nous appellerons *notation fonctionnelle*. Nous écrirons

(3) $$P(A, B) = p$$

et nous lirons : « la probabilité de A à B est égale à p », ou « la probabilité de B, par rapport à A, est égale à p ». La définition de cette expression est, elle aussi, donnée par (1). L'avantage de la forme fonctionnelle consiste en ce qu'elle est mieux adaptée aux besoins des mathématiques.

Dans certains cas il convient d'employer des fonctions propositionnelles au lieu des termes concernant des classes; nous remplacerons

(¹) J'emploie les signes logistiques de Russell, sauf pour la négation pour laquelle j'utilise un signe différent. Nous aurons donc :

$a \vee b$ (a ou b)	$a \supset b$ (a implique b)	(i) (pour tous les i).
$a.b$ (a et b)	$a \equiv b$ (a équivalent à b)	$x \varepsilon A$ (x est membre
$\bar{a}$ (non a)	$(\exists x)$ (il existe un x)	de la classe A).

donc

$$x_i \varepsilon \text{A} \quad \text{par} \quad \varphi x_i,$$
$$y_i \varepsilon \text{B} \quad \text{par} \quad \psi y_i,$$

et nous écrirons

$$(i)(\varphi x_i \underset{p}{\supset} \psi y_i), \tag{4}$$

$$\text{P}(\varphi x_i, \psi y_i) = p. \tag{5}$$

Introduisons enfin une dernière abréviation. Dans (1)-(5), nous considérons tous les événements y_i pour lesquels φx_i est vrai; cette fonction propositionnelle fait donc un triage parmi les éléments y_i. On peut biffer tous les y_i pour lesquels φx_i n'est pas vrai; en dénotant par z_k les éléments restants, qui forment une suite partielle de y_i nous pouvons écrire

$$\text{P}(\psi z_k) = p. \tag{6}$$

Nous avons dans ce cas une *notation à un terme;* les autres formes sont des *notations à deux termes*.

Dans (6), z_k est une variable restreinte; elle parcourt les éléments d'une série coordonnée à la fonction propositionnelle ψ. Appelons cette série le *fondement* coordonné à la fonction propositionnelle ψ.

Dans (5), x_i et y_i sont des variables libres, liées seulement à un certain ordre; le choix nécessaire se fait ici par la condition que φx_i doit être vrai. φ définit donc le fondement de ψ.

La notation à deux termes est préférable à l'autre, pour les raisons suivantes :

1° Elle est plus claire, parce que le fondement y est indiqué explicitement;

2° Il arrive que nous ayons à introduire une restriction supplémentaire pour y_i, en faisant une sélection de la série des y_i. Dans ce cas, le premier terme est inévitable. Les probabilités dites *relatives* dont il s'agit ici ne diffèrent pas, en principe, des autres; chaque probabilité est relative.

Nous adopterons donc dorénavant la notation à deux termes; ce n'est que pour l'analyse de la conception logique de la probabilité qu'il est commode de choisir la notation à un terme. Nous y reviendrons.

Enfin, une dernière remarque. L'implication de probabilité peut être considérée comme une généralisation de l'implication générale, ou implication formelle, de Russell :

(7) $$(i)(\varphi x_i \supset \psi y_i).$$

Cette implication dit que chaque fois que φx_i est vrai, ψy_i est vrai aussi; (4) dit que dans ce cas ψy_i n'est vrai que quelquefois, mais dans un rapport déterminé. La dernière interprétation présuppose cependant déjà l'interprétation de la fréquence.

Il y a, pourtant, une différence essentielle. A (7) correspond l'implication individuelle

(8) $$\varphi x_1 \supset \psi y_1,$$

pour une paire x_1, y_1 arbitrairement choisie dans les séries. Une expression analogue pour l'implication probabilitaire n'existe pas, ou bien, nous n'en avons pas besoin. Nous aurons à examiner cette question plus loin.

2. **Les axiomes.** — J'ai développé la forme logique de l'énoncé probabilitaire, mais je me suis abstenu de dire ce que je comprends par *probabilité*. Je ne vais pas le préciser davantage maintenant. L'analyse logique de la géométrie nous a montré qu'on peut parfaitement traiter une discipline mathématique sans être obligé de donner une interprétation de ses concepts fondamentaux. Appliquons cette idée aux probabilités, et développons le calcul des probabilités sans donner de définition du terme *probabilité*.

Au lieu de donner une définition de ce terme, je vais indiquer un système d'axiomes qui déterminent l'usage qu'on peut faire du terme en question. En possession de ce système d'axiomes, nous pouvons répondre à toute question concernant la vérité de combinaisons logiques dans lesquelles on rencontre le terme « *probabilité* ». Cela nous suffira pour l'instant; nous remettons à plus tard le problème de l'interprétation.

Conformément aux deux notations développées plus haut, la notation implicative et la notation fonctionnelle, nous aurons deux formes du système des axiomes. Ces deux formes sont équivalentes, bien qu'elles ne se correspondent pas directement dans les formules individuelles.

AXIOMES DU CALCUL DES PROBABILITÉS.

Forme implicative.

I. *Univocité :*

$$(p \neq q) \supset [(A \underset{p}{\supset} B).(A \underset{q}{\supset} B) \equiv (\overline{A})]$$

II. *Normalisation :*

1. $$(A \supset B) \supset (A \underset{p}{\supset} B).(p = 1)$$

2. $$(A \underset{p}{\supset} B) \supset (p \geqq 0)$$

III. *Addition :*

$$(A \underset{p}{\supset} B).(A \underset{q}{\supset} C).(A.B \supset \overline{C}) \supset (A \underset{r}{\supset} B \vee C).(r = p + q)$$

IV. *Multiplication :*

$$(A \underset{p}{\supset} B).(A.B \underset{u}{\supset} C) \supset (A \underset{r}{\supset} B.C).(r = p.u)$$

Forme fonctionnelle.

α. *Normalisation :*

1. $$P(A,\ A \vee B) = 1$$

2. $$P(A,\ B.\overline{B}) = 0$$

3. $$0 \leqq P(A,\ B)$$

β. *Addition :*

$$P(A,\ B \vee C) = P(A,\ B) + P(A,\ C) - P(A,\ B.C)$$

γ. *Multiplication :*

$$P(A,\ B.C) = P(A,\ B) \cdot P(A.B,\ C)$$

Le logicien préférera la forme implicative. Elle permet d'exprimer les relations d'implication logique valables entre les formules composées; de plus, elle donne la possibilité d'exprimer le principe de l'univocité, ce qui n'est pas réalisable avec la forme fonctionnelle. Cet axiome d'univocité I s'exprime de la façon suivante : si entre deux classes d'événements A et B, deux degrés différents de probabilité sont valables, la classe A ne possède aucun élément. Cette forme est nécessaire à cause de l'adaptation générale du système aux règles de la logique.

L'axiome II, 1 montre la connexion entre la certitude et la probabilité 1 : la certitude implique la probabilité 1, mais la réciproque n'est pas vraie. La probabilité 1 est donc un concept plus général.

Or, ce système montre la nécessité d'une règle s'ajoutant au système des axiomes et que j'appelle la règle d'existence. Elle s'exprime ainsi : *si une probabilité inconnue est déterminée numériquement par les probabilités données, cette probabilité existe* [1]. Cette règle nous permet par exemple d'inférer de l'axiome III que si les deux probabilités p et r existent, la probabilité q existe aussi et est égale à la différence $r - p$, inférence qui ne serait pas valable sans l'admission préalable de la règle précédente. Cette règle s'applique aussi au système fonctionnel; ici elle s'exprime par l'autorisation d'appliquer les règles des équations, aux termes entiers.

Pour le mathématicien, la forme fonctionnelle est la plus intéressante. Elle offre la possibilité de traiter des expressions concernant une probabilité sous forme d'*équations;* elle est donc mieux adaptée à l'usage du calcul mathématique. D'autre part, elle ne renonce pas à formuler les opérations logiques. Celles-ci apparaissent à l'intérieur des parenthèses; ici, les opérations logistiques trouvent leur domaine d'application. Les parenthèses divisent donc les termes en deux domaines : à l'extérieur des parenthèses s'étend le domaine des mathématiques, à l'intérieur celui où règne la logistique. Le développement du calcul mettra en évidence plus loin leur collaboration idéale; les parenthèses constituent les barrières qui empêchent les termes d'enfreindre les limites de leur domaine.

Examinons le mécanisme du calcul sur des exemples simples.

Substituons en α, 1 le terme $B \vee \overline{B}$ au lieu du terme B :

$$P(A, A \vee B \vee \overline{B}) = 1.$$

La logistique nous enseigne

$$A \vee B \vee \overline{B} \equiv B \vee \overline{B}.$$

[1] Par *déterminé* nous entendons le fait que les axiomes de notre système déterminent la probabilité inconnue comme fonction continue (pour les valeurs numériques données) des probabilités données.

En vertu de cette équivalence, nous pouvons substituer l'un à l'autre les deux membres de cette relation, dans l'équation précédente, et nous obtenons

$$P(A, B \vee \overline{B}) = 1.$$

Appliquons maintenant l'axiome β, en substituant $\overline{B}$ à C; avec α, 2 nous obtenons :

$$(9) \qquad P(A, B \vee \overline{B}) = P(A, B) + P(A, \overline{B}) = 1.$$

Cette proposition est connue sous le nom de *proposition du complément*. On peut facilement en déduire, en la combinant avec α, 3, qu'une probabilité ne peut jamais être plus grande que l'unité.

Le théorème particulier de l'addition, exprimé en III, est contenu en β comme cas particulier. Pour le montrer, il faut prouver que $(A.B \supset \overline{C})$ implique $P(A, B.C) = 0$. Notre calcul nous donne ce résultat, évident lorsqu'on considère le contenu des énoncés, de la façon suivante :

La logistique nous fournit la formule

$$(10) \qquad [A \supset D] \supset [D \equiv A \vee D]$$

qui signifie que si A est sous-classe de D, la classe de réunion entre A et D sera identique à la classe D. Substituons $\overline{B.C}$ à D :

$$(11) \qquad [A \supset \overline{B.C}] \supset [\overline{B.C} \equiv A \vee \overline{B.C}].$$

Or, nous avons

$$[A.B \supset \overline{C}] \equiv [A \supset \overline{B.C}].$$

Il s'ensuit que notre condition $A.B \supset \overline{C}$ nous permet de substituer $\overline{B.C}$ à $A \vee \overline{B.C}$; en l'appliquant dans la formule

$$P(A, A \vee \overline{B.C}) = 1,$$

dérivée de α, 1 en substituant $\overline{B.C}$ pour B, nous obtenons

$$P(A, \overline{B.C}) = 1.$$

Avec la proposition du complément, nous avons donc

$$P(A, B.C) = 0$$

et pour ce cas, β prend la forme du théorème particulier de l'addition.

3. **La règle de l'élimination.** — Par des opérations semblables, on peut déduire toutes les formules du calcul des probabilités de notre système d'axiomes. Je me borne ici à quelques exemples :

L'équivalence logistique

$$[B \vee \bar{B}].C \equiv C$$

nous autorise à écrire

$$P(A, C) = P(A, [B \vee \bar{B}].C).$$

En appliquant la loi de la distribution, nous obtenons

$$P(A, C) = P(A, B.C \vee \bar{B}.C).$$

Cette formule se résoud selon le théorème particulier de l'addition déjà démontré :

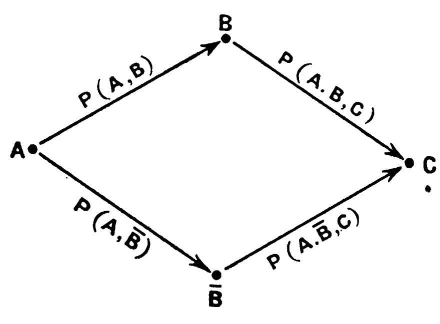
$$P(A, C) = P(A, B.C) + P(A, \bar{B}.C).$$

Avec le théorème de la multiplication, nous avons

$$P(A, C) = P(A, B) \cdot P(A.B, C) + P(A, \bar{B}) \cdot P(A.\bar{B}, C). \tag{12}$$

Nous appellerons ce théorème la *règle de l'élimination*, parce qu'il nous apprend à éliminer un événement B intercalé (*fig.* 1).

Fig. 1.



On peut employer notre théorème pour calculer la valeur de $P(A.\bar{B}, C)$:

$$P(A.\bar{B}, C) = \frac{P(A, C) - P(A, B) \cdot P(A.B, C)}{1 - P(A, B)}. \tag{13}$$

Il s'ensuit que seulement trois des quatre valeurs P(A, B), P(A, C), P(A.B, C), $P(A.\bar{B}, C)$ peuvent être choisies arbitrairement; la quatrième valeur est alors déterminée [1]. D'autre part, les valeurs numériques des trois variables indépendantes sont liées par deux inégalités, découlant du postulat qui maintient la valeur $P(A.\overline{B}, C)$ dans l'intervalle zéro à 1 :

$$(14) \qquad 1 - \frac{1 - P(A, C)}{P(A, B)} \leqq P(A.B, C) \leqq \frac{P(A, C)}{P(A, B)}.$$

Ces conditions sont comparables aux conditions restrictives concernant les valeurs numériques des trois paramètres qui en géométrie déterminent un triangle.

Les deux inégalités précédentes sont suffisantes pour notre calcul; on peut démontrer que d'autres semblables en découlent. Par exemple, on peut démontrer que, si les deux inégalités (14) sont remplies, il s'ensuit que

$$(15) \qquad P(A, B . C) \leqq P(A, B),$$

$$(16) \qquad P(A, B \vee C) \geqq P(A, B).$$

4. L'indépendance. — L'indépendance est définie par le cas particulier où

$$(17) \qquad P(A.B, C) = P(A, C).$$

Dans ce cas, le théorème particulier de la multiplication est valable :

$$(18) \qquad P(A, B.C) = P(A, B)\cdot P(A, C).$$

Notre notation montre en (17) que l'indépendance est une relation à trois termes : C est indépendant de B par rapport à A. On peut facilement montrer que cette relation est symétrique en B et C; cela se déduit de la relation (23).

L'indépendance, par contre, n'est pas transitive; c'est-à-dire : si B est indépendant de C, et C indépendant de D, il ne s'ensuit pas que B est indépendant de D. On peut illustrer cela par un exemple. Prenons deux dés reliés avec un fil court tel que les coups des dés ne puissent plus être indépendants; appelons B un coup de six du premier dé, D un

---

[1] Il est commode de considérer les trois valeurs $P(A, B)$, $P(A, C)$, $P(A.B, C)$ comme variables indépendantes; la valeur $P(A.\overline{B}, C)$ serait alors une variable dépendante.

coup de six du second. Pour C nous choisissons un troisième dé non couplé. Nous aurons alors : B est indépendant de C, C est indépendant de D, mais B est dépendant de D.

Une autre propriété de l'indépendance consiste en ce qu'elle n'est pas *associable* (1). Pour expliquer cette propriété, partons d'un cas de trois suites B, C, D telles que chacune d'elles soit indépendante de chacune des deux autres, par rapport à une quatrième suite A. Peut-on en conclure que chaque suite est indépendante des deux autres, prises ensemble?

La première condition est exprimée par les relations

$$(19)\qquad \left\{\begin{array}{l} P(A.B, C) = P(A.D, C) = P(A, C), \\ P(A.C, D) = P(A.B, D) = P(A, D), \\ P(A.D, B) = P(A.C, B) = P(A, B). \end{array}\right.$$

La seconde serait exprimée par les relations

$$(20)\qquad \left\{\begin{array}{l} P(A.B.C, D) = P(A, D), \\ P(A.C.D, B) = P(A, B), \\ P(A.D.B, C) = P(A, C). \end{array}\right.$$

Mais ces relations (20) ne découlent pas des relations (19). On peut le montrer en comptant les équations qui déterminent les termes à gauche dans (20) (2); nous nous bornons ici à donner la démonstration par la construction d'un exemple où les équations (19) sont vérifiées, tandis que les équations (20) ne le sont pas. Prenons les suites suivantes définies par répétition de la période écrite :

$$(21)\qquad \left\{\begin{array}{l} A\ A\ A\ A\ A\ A\ A\ A\ \ldots, \\ B\ \overline{B}\ B\ \overline{B}\ B\ \overline{B}\ B\ \overline{B}\ \ldots, \\ C\ C\ \overline{C}\ \overline{C}\ C\ C\ \overline{C}\ \overline{C}\ \ldots, \\ D\ \overline{D}\ \overline{D}\ D\ D\ \overline{D}\ \overline{D}\ D\ \ldots. \end{array}\right.$$

(1) Nous désignons par « associable » une propriété des relations qui peut être conçue comme l'inverse de la propriété appelée « distributive ». Une relation r sera associable si elle satisfait la condition

$$(B\,r\,D).(C\,r\,D) \supset (B.C)\,r\,D.$$

Cette condition correspond à la loi de distributivité de l'arithmétique si l'on remplace r par la multiplication et le point par l'addition. Cette loi est définie de telle sorte qu'elle demande l'équivalence des deux côtés; le terme « distributif » cependant dérive de la considération de l'opération de droite à gauche.

(2) Cf. *Wahrscheinlichkeitslehre*, p. 114.

Les probabilités (19) sont ici toutes égales à $\frac{1}{2}$, mais les probabilités à gauche dans (20) sont toutes égales à 1.

C'est pour cette raison que nous dirons que l'indépendance n'est pas associable. Les suites pour lesquelles non seulement les relations du type (19), mais aussi celles du type (20) sont vérifiées, s'appelleront des *suites complètement indépendantes.*

5. **La règle de Bayes.** — Abordons un autre exemple. Le théorème de la multiplication est symétrique du côté gauche, mais asymétrique du côté droit. Il fournit donc une deuxième sorte de décomposition lorsqu'on échange B et C :

$$(22)\qquad P(A, B.C) = P(A, B)\cdot P(A.B, C) = P(A, C)\cdot P(A.C, B);$$

Donc

$$(23)\qquad P(A.C, B) = \frac{P(A, B)\cdot P(A.B, C)}{P(A, C)}.$$

Introduisons au dénominateur la règle de l'élimination :

$$(24)\qquad P(A.C, B) = \frac{P(A, B)\cdot P(A.B, C)}{P(A, B)\cdot P(A.B, C) + P(A, \overline{B})\cdot P(A.\overline{B}, C)}.$$

Nous tombons sur la formule de Bayes, qui détermine la « probabilité rétrospective » comme fonction des « probabilités prospectives » (cf. *fig.* 1). P(A, B) est la probabilité dite *a priori* de la cause B; je préfère employer l'expression « probabilité initiale », proposée par M. v. Mises.

Ces exemples doivent suffire. On peut démontrer par des opérations semblables que tous les théorèmes de la probabilité se déduisent des axiomes donnés. L'essentiel, cependant, est que cette déduction se fait par des méthodes purement formelles, qui ne font aucunement appel au contenu physique des formules. Ce résultat est dû dans notre calcul au fait que le formalisme s'étend aussi bien aux opérations logiques qu'aux opérations mathématiques. Il fonctionne comme un mécanisme logico-mathématique et donne automatiquement des résultats justes.

6. **L'interprétation de la fréquence.** — Après avoir étudié les qualités

du système relationnel des probabilités, passons à la question de l'*interprétation* du concept fondamental. Nous allons définir la probabilité comme limite d'une fréquence. Cette définition occupe la position logique d'une définition de coordination : nous coordonnons au concept de probabilité, défini implicitement par le système axiomatique, une interprétation externe. Le procédé trouve son analogue dans les définitions de coordination de la géométrie, par lesquelles on passe du système formel de la géométrie à un système interprété. Si l'on coordonne la notion de « point » de la géométrie aux petites particules de matière, la notion « ligne droite » à un rayon de lumière, etc., on arrive à la géométrie physique.

Cette coordination ne peut cependant pas s'effectuer arbitrairement. Elle n'est admissible que si l'on peut démontrer que les objets coordonnés satisfont aux axiomes précédents. L'interprétation de la probabilité par la fréquence exige donc qu'on puisse démontrer que les axiomes posés au début sont satisfaits; or, il est effectivement possible de donner une pareille démonstration.

Introduisons le symbole de la fréquence relative :

$$F^n(A, B) = \frac{\overset{n}{\underset{1}{\mathbf{N}_i}}(x_i \,\varepsilon\, A).(y_i \,\varepsilon\, B)}{\overset{n}{\underset{1}{\mathbf{N}_i}}(x_i \,\varepsilon\, A)}, \tag{25}$$

où $\overset{n}{\underset{1}{\mathbf{N}_i}}(x_i \varepsilon A)$ signifie le nombre des entités x_i entre x_1 et x_n qui satisfont la condition $x_i \varepsilon A$. Ensuite définissons

$$P(A, B) = \lim_{n=\infty} F^n(A, B). \tag{26}$$

L'existence de cette limite est le seul postulat à exiger de nos séries. Nous n'avons pas besoin d'un principe d'irrégularité. Nous examinerons plus loin les problèmes qui se rattachent à cette notion.

Il est facile de montrer qu'avec cette définition tous nos axiomes sont satisfaits. Pour α, $1-\alpha$, 3 la démonstration est banale. Prenons ensuite le théorème de l'addition. Nous avons

$$F^n(A, B \vee C) = \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).[(y_i \varepsilon B) \vee (z_i \varepsilon C)]}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A)}$$

$$= \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(y_i \varepsilon B)}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A)} + \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(z_i \varepsilon C)}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A)}$$

$$- \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(y_i \varepsilon B).(z_i \varepsilon C)}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A)}.$$

Le troisième terme est nécessaire puisque les combinaisons $(y_i \varepsilon B)$ $(z_i \varepsilon C)$ ont été comptées deux fois, dans les termes précédents. En appliquant à plusieurs reprises la définition du symbole F^n, nous aurons

$$(27) \qquad F^n(A, B \vee C) = F^n(A, B) + F^n(A, C) - F^n(A, B.C).$$

Si les trois limites du second membre existent, il en sera de même pour le premier, ce qui constitue précisément l'axiome β.

Le théorème de la multiplication est également satisfait en vertu de la forme générale que nous lui avons donnée. Nous avons

$$(28) \qquad F^n(A, B.C) = \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(y_i \varepsilon B).(z_i \varepsilon C)}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A)}$$

$$= \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(y_i \varepsilon B)}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A)} \cdot \frac{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(y_i \varepsilon B).(z_i \varepsilon C)}{\displaystyle\mathop{\mathbf{N}}_{i\;1}^{n}(x_i \varepsilon A).(y_i \varepsilon B)}$$

$$= F^n(A, B) \cdot F^n(A.B, C).$$

Ici aussi, le premier membre aura une limite si le second en a une : c'est là précisément notre axiome γ.

Des remarques analogues s'appliquent à la règle d'existence, qui peut être démontrée, elle aussi, comme conséquence nécessaire de la définition de la probabilité au moyen de la fréquence. En effet, cette règle précise que si

$$\alpha_i = \varphi(\beta_i, \gamma_i, \ldots)$$

est une fonction régulière de $\beta_i, \gamma_i, \ldots$, et si $\beta_i, \gamma_i, \ldots$ convergent vers des limites $\beta, \gamma, \ldots, \alpha_i$ doit aussi converger vers une limite α.

L'importance de cette démonstration consiste en ce qu'elle montre que la présupposition de la limite suffit pour le calcul des probabilités. Nous n'avons pas besoin d'autres conditions. C'est pourquoi il nous est permis de choisir, comme objets de notre calcul des probabilités interprété, des suites mathématiques données par une prescription mathématique. Par exemple, les décimales du nombre π nous fournissent un modèle d'une série si nous y considérons les chiffres pairs comme définissant le cas B, les chiffres impairs comme définissant le cas $\overline{B}$.

CHAPITRE III.

THÉORIE DE L'ORDRE DES SUITES PROBABILITAIRES.

7. Les probabilités de phase. — Dans la théorie élémentaire des probabilités nous n'avons considéré que l'*association externe* des suites, sans entrer dans une analyse de leur *structure interne*. En appliquant l'interprétation de la fréquence, nous pouvons formuler cette différence en disant : nous n'avons encore exigé pour nos suites que l'existence d'une limite de la fréquence, et nous avons ensuite considéré tous les théorèmes qui découlent de ce postulat pour l'association des suites; mais nous n'avons pas encore abordé les problèmes qui concernent l'ordre interne des suites. C'est ce problème que nous allons étudier maintenant.

En abordant ce problème, nous ne serons pas obligés de restreindre notre théorie à des types de suites d'un ordre particulier, comme par exemple le type des suites qui apparaissent dans les jeux de hasard.

L'avantage de notre théorie est précisément d'embrasser toutes les sortes de types possibles et de traiter le problème de la probabilité sous sa forme la plus générale. Le problème de l'ordre se présente pour nous d'une autre façon : nous considérons chaque suite comme une donnée dont nous avons à étudier la structure interne, qu'elle soit du type irrégulier des suites des jeux de hasard, ou d'un type tout à fait régulier, ou enfin qu'elle appartienne à un type intermédiaire, à ranger entre les cas extrêmes de l'irrégularité complète et de la stricte régularité. Nous allons étudier les suites comme le botaniste étudie une plante : il en examine les feuilles, les fleurs, compte les anthères, etc., et détermine l'espèce particulière à laquelle sa plante appartient. Pour la classification des suites probabilitaires nous procéderons de même exactement comme le botaniste qui construit une classification en introduisant des définitions de types particuliers, dont il étudie les propriétés. Nous adapterons évidemment nos définitions aux besoins des applications physiques, qui renferment comme l'on sait beaucoup de types différents, parmi lesquels le type des jeux de hasard ne constitue qu'un cas d'importance secondaire.

En abordant cette tâche, nous allons analyser tout d'abord certains *moyens de caractérisation structurale*, c'est-à-dire des moyens qui nous permettent de caractériser la structure de l'ordre des suites. Nous ferons usage d'une méthode que M. v. Mises a introduite dans la théorie des probabilités, dans un but d'ailleurs plus restreint que celui que nous poursuivons; nous caractérisons la structure d'une suite probabilitaire en donnant des énoncés sur les *probabilités dans des suites partielles*.

Les types particuliers de notre théorie seront caractérisés par l'égalité des probabilités dans certaines suites partielles, ou par l'égalité de cette probabilité à celle de la suite principale. Nous n'exigerons pas, cependant, une condition si forte que celle du postulat de l'irrégularité de M. v. Mises; il nous suffira d'exiger l'invariance de la probabilité pour certaines suites partielles bien définies. En variant le domaine de ces suites partielles, nous arriverons aux définitions de divers types de suites probabilitaires.

Le premier moyen de caractérisation structurale est la sélection par le prédécesseur. On sait que, par exemple, une suite de pile et face est ordonnée de façon que la probabilité d'obtenir « pile » après un coup de pile est égale à celle de trouver « pile » en général; c'est-à-dire la suite partielle déterminée par « pile » comme prédécesseur a la même limite de fréquence que la suite principale. Dans le mouvement brownien, au contraire, il n'en est pas ainsi. Pour exprimer symboliquement de pareilles conditions il nous faut introduire de nouveaux signes dans notre calcul.

Reprenons d'abord la notation complète dont nous avons fait usage dans notre formule (1). Une probabilité relative au premier prédécesseur s'écrira

$$(i)([x_i \varepsilon A].[y_i \varepsilon B] \underset{p_1}{\supset} [y_{i+1} \varepsilon B]). \tag{29}$$

Introduisons comme équivalente à cette expression, l'abréviation

$$P(A.B, B^1) = p_1.$$

L'indice 1 indique la *phase*, et nous appellerons cette sorte de probabilité une *probabilité de phase*. On voit qu'il est facile d'étendre cette façon d'écrire à d'autres prédécesseurs; on écrira par exemple :

$$P(A.B.B^1, B^2) \qquad P(A.B, B^2) \qquad P(A.\overline{B}, B^3)$$

dont le sens est évident.

En employant ces probabilités de phase, nous n'avons pas besoin de renoncer à notre distinction entre le calcul formel des probabilités, et le calcul interprété; c'est-à-dire nous sommes libres de considérer les probabilités des suites partielles comme une propriété inconnue, ou de les considérer comme limites de la fréquence. Nous pouvons donc considérer nos probabilités comme termes purement formels auxquels nous appliquons le mécanisme de notre calcul; les lettres munies d'un indice de phase y seront traitées de la même façon que les autres lettres, et pourront êtres soumises à toutes les opérations logistiques. Cette extension de notre calcul ne signifie pas autre chose que l'adoption du postulat d'après lequel les suites partielles définies par les opérations de phase sont aussi des suites probabilitaires (ou, dans l'interprétation de la fréquence, ont une limite de la fréquence). Nos axiomes sont donc également valables pour les formules contenant les symboles de phases.

En outre, nous avons besoin de deux nouveaux axiomes concernant les probabilités de phase. Contrairement aux autres, ces nouveaux axiomes ne sont valables que pour des suites infinies, et pour le cas où A est compact (c'est-à-dire ne contient que des membres du type A) :

δ. *Axiomes de phase :*

1. $$P(A^{\alpha}.C, B^{\beta}) = P(A.C, B^{\beta}),$$

2. $$P(A.C^{\alpha}\ldots E^{\sigma}, B^{\tau}) = P(A.C^{\alpha-\rho}\ldots E^{\sigma-\rho}, B^{\tau-\rho}).$$

Ces axiomes découlent d'une façon simple de l'interprétation de la fréquence ; ils expriment la propriété d'une suite infinie d'après laquelle on peut éliminer au commencement de la suite une tranche finie sans que la limite de la suite soit altérée. Pour notre calcul formel, cependant, il faut les énoncer expressément. Ajoutons qu'on peut démontrer aussi pour ces axiomes la validité de la règle d'existence [1].

8. Les suites sans rémanence. — Le premier type que nous définissons sera caractérisé par la condition que la sélection par un groupe quelconque de prédécesseurs n'ait aucune influence. Pour l'écrire, il est commode de considérer une disjonction *complète* et *séparative* de r termes $B_1 \vee$, $B_2 \vee, \ldots, \vee B_r$. Le terme « séparatif » introduit ici indiquera que chaque B_i exclut les autres. La notation à employer pour une telle disjonction est plus simple que celle pour une alternative $B \vee \overline{B}$.

Nous définissons les suites sans rémanence par la condition

$$P(A.B^{1}_{i_1}\ldots B^{\nu-1}_{i_{\nu-1}}, B^{\nu}_{i_\nu}) = P(A, B_{i_\nu}) \qquad (\nu = 1, 2, 3, \ldots.) \tag{30}$$

De cette définition découle la validité du théorème particulier de la multiplication, pour ces suites. Nous avons

$$\begin{aligned} P(A, B^{1}_{i_1}\ldots B^{\nu}_{i_\nu}) &= P(A, B^{1}_{i_1}\ldots B^{\nu-1}_{i_{\nu-1}})\cdot P(A.B^{1}_{i_1}\ldots B^{\nu-1}_{i_{\nu-1}}, B^{\nu}_{i_\nu}) \\ &= P(A, B^{1}_{i_1}\ldots B^{\nu-1}_{i_{\nu-1}})\cdot P(A, B_{i_\nu}). \end{aligned}$$

En appliquant la même décomposition au premier terme à droite, et en répétant le raisonnement, nous arrivons à la formule

$$P(A, B^{1}_{i_1}\ldots B^{\nu}_{i_\nu}) = P(A, B_{i_1})\ldots P(A, B_{i_\nu}), \tag{31}$$

c'est-à-dire au théorème annoncé.

9. Deux modes de dénombrement. — Le théorème particulier de la

[1] Cette démonstration, cependant, diffère un peu de celles valables pour les autres axiomes; cf. *Wahrscheinlichkeitslehre*, p. 132.

multiplication n'est valable, pour les suites sans rémanence, que pour un certain mode de dénombrement que nous appellerons *dénombrement par empiètement*. Pour l'illustrer, prenons un exemple.

La suite

$$\text{(32)} \qquad B\ B\ \overline{B}\ \overline{B}\ B\ B\ \overline{B}\ \overline{B}\ \ldots$$

fournie par la répétition identique de la période écrite, n'est pas sans rémanence, mais elle remplit la condition (30) au moins pour le premier prédécesseur. Nous avons ici

$$P(A, B) = \frac{1}{2},$$

$$P(A.B, B^1) = \frac{1}{2}, \qquad P(A.\overline{B}, B^1) = \frac{1}{2},$$

donc

$$P(A, B.B^1) = P(A, B)\cdot P(A.B, B^1) = P(A, B)\cdot P(A, B) = \frac{1}{4}.$$

Nous pouvons vérifier, en effet, cette dernière relation si nous comptons les combinaisons BB avec empiètement, c'est-à-dire si nous les comptons sous la forme

$$B\ B\ \overline{B}\ \overline{B}\ B\ B\ \overline{B}\ \overline{B}\ \ldots$$

Il est clair cependant que nous aurons un résultat différent dans un *dénombrement par sections*, c'est-à-dire de la forme

$$\text{(33)} \qquad B\ B\,|\,\overline{B}\ \overline{B}\,|\,B\ B\,|\,\overline{B}\ \overline{B}\,|\ \ldots$$

Dans ce cas la probabilité d'une combinaison BB sera égale à $\frac{1}{2}$.

Pour caractériser ce mode de dénombrement, nous avons besoin d'un autre moyen de caractérisation structurale. Nous introduisons des sélections S données par des prescriptions, et nous envisagerons des probabilités de la forme

$$P(A.S, B), \quad P(A.S.B, B^1), \quad P(A.S.B.B^1, B^2).$$

Ce sont particulièrement les sélections $S_{\lambda\varkappa}$ fournies par des progressions arithmétiques qui nous intéressent parce qu'elles définissent le dénombrement par sections; nous les appelons *scindements réguliers*. Nous dirons que :

Un terme y_i de la suite principale appartient à la suite partielle déterminée par le *scindement régulier* $S_{\lambda\varkappa}$, si l'indice i remplit la condition

$$i = \varkappa + (m - 1)\cdot\lambda \qquad (m = 1, 2, 3, \ldots;\ \varkappa = 1, 2, \ldots, \lambda).$$

Le dénombrement indiqué (33) s'écrirait par exemple

$$(34) \qquad P(A.S_{21}, B.B^1) \quad \text{ou aussi} \quad P(A.S_{22}^1, B.B^1).$$

La longueur du scindement régulier serait ici $\lambda = 2$. La probabilité (34) se décompose en

$$(35) \qquad P(A.S_{21}, B.B^1] = P(A.S_{21}, B)\cdot P(A.S_{21}.B, B^1).$$

On voit que, dans notre exemple (32), nous avons

$$(36) \qquad P(A.S_{21}, B) = \frac{1}{2}, \qquad P(A.S_{21}.B, B^1) = 1.$$

La probabilité (35) ne satisfait pas le théorème particulier de la multiplication à cause précisément du second terme dans la relation précédente.

10. Le domaine d'invariance. — Pour chaque suite, il y a des sélections qui conduisent à une limite de fréquence différente, et d'autres qui la laissent invariante. Cette idée nous fournit une classification des sélections : nous comprenons l'ensemble des sélections du second type sous le nom de *domaine d'invariance*. Il convient, cependant, d'exiger pour cette définition des propriétés supplémentaires : nous exigerons que non seulement la probabilité principale, mais aussi que chaque probabilité de phase reste invariante pour ces sélections. Nous dirons donc que :

Une sélection S appartient au domaine d'invariance de la suite, si elle remplit les conditions

$$(37) \qquad \left\{\begin{array}{c} P(A.S, B) = P(A, B), \\ P\left(A.S^\alpha.B_{i_1}^1 \ldots B_{i_{\nu-1}}^{\nu-1}, B_{i_\nu}^\nu\right) = P\left(A.B_{i_1}^1 \ldots B_{i_{\nu-1}}^{\nu-1}, B_{i_\nu}^\nu\right), \\ (\nu = 1, 2, 3; \ldots;\ \alpha = 1, 2, 3; \ldots). \end{array}\right.$$

Pour la définition de types de suites particulières nous aurons besoin de caractériser le domaine d'invariance de la suite. Il serait cependant pratiquement impossible de nommer toutes les sélections du domaine d'invariance; nous nous contenterons de postuler un minimum pour l'effectif de ce domaine. C'est-à-dire la définition du type de suite sera construite de façon que la suite appartienne au type en question si au moins certaines sélections prescrites appartiennent au domaine d'invariance. Nous ne nous préoccuperons pas de savoir s'il existe d'autres sélections appartenant à ce domaine.

11. Les suites normales. — Les suites normales sont définies de la façon suivante :

Une suite est appelée normale, si elle est sans rémanence et si les scindements réguliers appartiennent à son domaine d'invariance; c'est-à-dire si les relations suivantes sont remplies à côté des relations (30)

$$(38)\quad \left\{\begin{aligned} &P(A.S_{\lambda\varkappa}, B_i) = P(A, B_i) \\ &P(A.S_{\lambda\varkappa}^{\alpha}.B_{i_1}^1 \ldots B_{i_{\nu-1}}^{\nu-1}, B_{i_\nu}^{\nu}) = P(A.B_{i_1}^1 \ldots B_{i_{\nu-1}}^{\nu-1}, B_{i_\nu}^{\nu}) = P(A, B_{i_\nu}) \\ &(\varkappa = 1, 2, \ldots, \lambda;\ \alpha = 1, 2, \ldots, \lambda;\ \nu = 1, 2, \ldots, \lambda;\ \lambda = 1, 2, 3, \ldots). \end{aligned}\right.$$

La signification de cette définition consiste en ce que le théorème particulier de la multiplication est satisfait non seulement pour le dénombrement par empiétement, mais aussi pour le dénombrement par sections. Nous avons

$$\begin{aligned} P(A.S_{\lambda 1}^1, B_{i_1}^1 \ldots B_{i_\lambda}^{\lambda}) &= P(A.S_{\lambda 1}^1, B_{i_1}^1 \ldots B_{i_{\lambda-1}}^{\lambda-1}) \cdot P(A.S_{\lambda 1}^1.B_{i_1}^1 \ldots B_{i_{\lambda-1}}^{\lambda-1}, B_{i_\lambda}^{\lambda}) \\ &= P(A.S_{\lambda 1}^1, B_{i_1}^1 \ldots B_{i_{\lambda-1}}^{\lambda-1}) \cdot P(A, B_{i_\lambda}). \end{aligned}$$

En continuant cette décomposition, nous obtiendrons

$$(39)\qquad P(A.S_{\lambda 1}^1, B_{i_1}^1 \ldots B_{i_\lambda}^{\lambda}) = P(A, B_{i_1}) \ldots P(A, B_{i_\lambda}),$$

c'est-à-dire le théorème particulier de la multiplication en dénombrement par sections. Il s'ensuit que le théorème de Bernouilli est valable pour nos suites normales, dans les deux modes de dénombrement.

Ajoutons encore que cette définition des suites normales est identique à celle donnée par M. Copeland qui appelle les suites normales « admissible numbers ». L'article de M. Copeland [1], antérieur à ma première

[1] A. H. COPELAND, *Admissible numbers in the theory of probability* (*American Journ. of Math.*, vol. L, n° 4, 1928, p. 535). M. Copeland définit les suites normales par les deux postulats de l'invariance de la limite dans les suites partielles. triées par les progressions arithmétiques, et l'indépendance complète de ces suites partielles.

publication sur ce sujet, m'était resté inconnu; je saisis volontiers l'occasion de reconnaître ici la priorité de M. Copeland quant aux suites normales. La théorie de M. Copeland est supérieure à la mienne sur un point important : M. Copeland a pu indiquer une méthode permettant de construire des suites normales. Il a donc montré que les postulats définissant les suites normales ne renferment aucune contradiction interne, et permettent au contraire de construire ces séries à l'aide de prescriptions mathématiques précises.

Le problème de l'irrégularité trouve, me semble-t-il, sa solution dans ces remarques. Il est possible de donner une définition de l'irrégularité telle que le théorème de Bernoulli soit applicable, sans enfreindre le principe d'après lequel une suite mathématique infinie ne doit pas exclure la possibilité d'une définition par prescription ([1]). Notre définition correspond aussi aux applications pratiques. Si l'on note, par exemple, tous les habitants d'une ville dans l'ordre alphabétique de leurs noms, et si l'on considère le cas d'une maladie tuberculose, comme définissant le cas B, on obtiendra une suite qu'on pourra appeler une suite normale. Si l'on fait cependant une sélection, en choisissant seulement les habitants d'un certain quartier de la ville d'hygiène insuffisante, on obtiendra une fréquence supérieure pour la tuberculose. L'existence de cette sélection à fréquence différente ne nous empêche pas, cependant, de classer la série originale comme série normale. Dans les jeux de hasard, nous trouvons des suites plus particulières encore pour lesquelles il nous est techniquement impossible de construire une sélection de fréquence différente, sans nous servir des caractères des épreuves; cette propriété s'exprime dans notre théorie par le fait que le domaine d'invariance est plus grand que celui de l'autre suite. Cela ne nous empêche pas, cependant, de traiter également les deux sortes de suites comme des suites normales parce que cette définition n'impose au domaine d'invariance qu'une extension minima.

J'avais choisi le même chemin dans ma première publication (*Math Zeitschr.*, 34, 1932, p. 603). On peut montrer l'équivalence de cette définition à celle donnée ci-dessus, qui remplace le second postulat par un autre exigeant que les suites soient sans rémanence et qui étend le premier postulat sur les probabilités de phase.

([1]) Ce principe ne doit pas être confondu avec le principe plus restreint d'après lequel une suite infinie n'est admise que si l'on *connaît* effectivement une définition par prescription.

12. Transfert de probabilité. — Nous obtenons un type de suite plus général en admettant une rémanence de la probabilité. Si cette généralisation ne consistait, cependant, qu'en la renonciation à la condition (30), le traitement mathématique des suites renfermerait des difficultés insurmontables. Dans le cas général, il n'y a aucune relation entre les probabilités de phase de la forme

$$P(A.B, B^1), \quad P(A.B.B^1, B^2), \quad P(A.B.B^1.B^2, B^3), \quad \ldots.$$

Nous n'envisagerons donc pas d'emblée le cas le plus général, mais nous construirons la généralisation par étapes, en commençant avec le cas le plus simple.

Une première étape s'obtient lorsque la rémanence dépend uniquement du premier prédécesseur. Ce cas sera appelé le cas du *transfert de probabilité* et nous le définirons par la condition suivante :

$$(40) \qquad P(A.B^1_{i_1}\ldots B^{\nu-1}_{i_{\nu-1}}, B^{\nu}_{i_\nu}) = P(A.B^{\nu-1}_{i_{\nu-1}}, B^{\nu}_{i_\nu}).$$

Cette condition a été étudiée la première fois par Markoff, et ces suites portent aussi le nom de chaînes de Markoff. Pour les réaliser physiquement, on peut se servir de la méthode suivante : on prend deux mécanismes de probabilité (par exemple des roulettes à secteurs de largeur différente) fournissant les deux probabilités p_1 et q_1; si l'on a obtenu le cas B, on joue avec la probabilité p_1, si l'on a obtenu $\overline{B}$, on joue avec la probabilité q_1 pour B. Le premier terme de la suite est arbitraire.

Les probabilités données seront donc

$$(41) \qquad P(A.B, B^1) = p_1, \qquad P(A.\overline{B}, B^1) = q_1.$$

Notre calcul permet un traitement très simple de ces suites. La première question concerne l'existence d'une probabilité moyenne. En appliquant la règle de l'élimination, nous avons

$$\begin{aligned} P(A, B) = P(A, B^1) &= P(A, [B \vee \overline{B}].B^1) \\ &= P(A, B)\cdot P(A.B, B^1) + P(A, \overline{B})\cdot P(A.\overline{B}, B^1). \end{aligned}$$

En introduisant la notation

$$(42) \qquad P(A, B) = p,$$

nous pouvons écrire cette équation, en vertu de (41)

$$(43) \qquad p = p.p_1 + (1-p).q_1.$$

Elle constitue une relation entre les trois probabilités p, p_1, q_1. Comme p_1 et q_1 sont données, la règle de l'existence nous permet d'inférer que p existe. Sa valeur s'obtient par la solution de (43)

$$p = \frac{q_1}{1 - p_1 + q_1}. \tag{44}$$

Pour discuter cette relation, introduisons les notations

$$p_1 = p + \varepsilon, \qquad q_1 = p - \eta \tag{45}$$

(44) devient alors

$$\frac{p}{1-p} = \frac{\eta}{\varepsilon}. \tag{46}$$

Il s'ensuit que le signe de ε et η est toujours le même, mais il faut distinguer le cas des signes positifs de celui des signes négatifs.

Dans le premier cas, il y a une *tendance à persévèrer* dans la suite, une persistance de B si B s'est produit, et une persistance de $\overline{B}$ si $\overline{B}$ s'est produit; nous parlerons alors *d'entraînement des probabilités*. Dans le second cas, il y a *tendance au changement*, passage accéléré de B à $\overline{B}$ et de $\overline{B}$ à B; nous parlerons alors de *compensation des probabilités*.

Malgré la condition (40), il existe ici aussi une influence du second prédécesseur; elle ne se produit pas directement cependant, mais à l'aide d'un transport exécuté par le premier prédécesseur intermédiaire. Si par exemple B s'est produit, on trouve de préférence un B comme successeur, et par suite aussi un B comme second successeur. C'est pour cette raison que nous employons pour ces suites le nom de transfert de probabilité.

Nous avons donc

$$P(A.B, B^2) \neq P(A, B).$$

D'autre part, l'idée que le second prédécesseur n'a aucune influence directe s'exprime par la relation

$$P(A.B.B^1, B^2) = P(A.\overline{B}.B^1, B^2) = P(A.B^1, B^2),$$

qui découle de (40).

La détermination de

$$P(A.B, B^2) = p_2 \tag{47}$$

se fait à l'aide de la règle de l'élimination. Nous avons

$$\begin{aligned} P(A.B, B^2) &= P(A.B, [B^1 \vee \overline{B^1}].B^2) \\ &= P(A.B, B^1)\cdot P(A.B.B^1, B^2) + P(A.B, \overline{B}^1)\cdot P(A.B.\overline{B}^1, B^2). \end{aligned}$$

Avec les notations (41) et (45), nous avons également

$$p_2 = (p + \varepsilon)^2 + (1 - p - \varepsilon)\cdot(p - \eta).$$

En vertu de (46) cette équation se transforme en

$$p_2 = p + \varepsilon_1, \qquad \varepsilon_1 = \varepsilon \frac{\varepsilon}{1-p}. \tag{48}$$

Posons

$$p + \varepsilon < 1; \tag{49}$$

on peut facilement démontrer que $\varepsilon_1 < \varepsilon$, c'est-à-dire que p_2 est plus proche de p, que p_1.

Le calcul est tout à fait analogue pour un prédécesseur plus éloigné. Nous aurons

$$\left\{\begin{aligned} p_{\nu+1} = P(A.B, B^{\nu+1}) = P(A.B, [B^\nu \vee \overline{B^\nu}].B^{\nu+1}) \\ = \quad P(A.B, B^\nu)\cdot P(A.B.B^\nu, B^{\nu+1}) \\ + P(A.B, \overline{B}^\nu)\cdot P(A.B.\overline{B}^\nu, B^{\nu+1}) \\ p_{\nu+1} = p_\nu . p_1 + (1 - p_\nu).q_1. \end{aligned}\right. \tag{50}$$

Pour appliquer le principe de récurrence, posons

$$p_\nu = p + \varepsilon_{\nu-1}. \tag{51}$$

Avec (50), on en déduit que

$$p_{\nu+1} = p + \varepsilon_\nu, \qquad \varepsilon_\nu = \varepsilon_{\nu-1}\frac{\varepsilon}{1-p}, \tag{52}$$

p_2 ayant la forme (51), il s'ensuit que (52) est valable pour chaque ν. En substituant successivement la valeur de ε_ν, nous obtenons

$$p_\nu = p + \varepsilon\left(\frac{\varepsilon}{1-p}\right)^{\nu-1}. \tag{53}$$

On en déduit, compte tenu de (49), la convergence de p_ν vers p. On voit que le cas $\varepsilon > 0$ implique une convergence unilatérale vers p, et le cas $\varepsilon < 0$ une convergence alternée. Un théorème analogue se démontre pour q_ν.

ε sera appelé le *degré de transfert*. Le cas $\varepsilon = 1 - p$ signifie une dégénérescence telle que la convergence des p_ν vers p n'a plus lieu. Ce cas correspond à des suites d'une régularité stricte. Par exemple, la suite alternée

$$\mathrm{B}\ \overline{\mathrm{B}}\ \mathrm{B}\ \overline{\mathrm{B}}\ \mathrm{B}\ \overline{\mathrm{B}}\ \ldots \tag{54}$$

peut être conçue comme le cas $p = \frac{1}{2}$, $\varepsilon = -\frac{1}{2}$. On voit qu'il y a transition continue entre les suites normales et les suites de régularité stricte comme (54); cette transition est représentée pour $p = \frac{1}{2}$ par la transition de ε de la valeur 0 à la valeur $-\frac{1}{2}$. Ceci représente un nouvel argument en faveur de notre idée suivant laquelle les suites normales ne constituent qu'un cas particulier dans une théorie plus générale des probabilités.

Remarquons encore qu'on peut joindre la condition (37) avec $S = S_{\lambda\varkappa}$, à la condition (40), c'est-à-dire qu'on peut exiger que le domaine d'invariance des suites avec transfert de probabilité renferme les scindements réguliers. Il est facile de montrer que dans ce cas une suite partielle déterminée par la sélection $S_{\lambda\varkappa}$ et dans laquelle on introduirait un nouveau dénombrement de terme à terme, correspondra à une suite avec transfert telle que la probabilité p_1 du transfert soit égale à notre probabilité p_λ d'après (53), pour $\lambda = \nu$.

13. La grille. — Nous avons défini la probabilité par une fréquence dans une suite, ou dans une suite partielle. Il y a, cependant, un trait caractéristique que nous n'avons pas encore exprimé dans notre définition: c'est l'idée que la probabilité *reste constante* dans la suite, ou dans la suite partielle. Imaginons une roulette telle qu'on puisse y réaliser chaque valeur de la probabilité en changeant la largeur relative des secteurs; supposons qu'on joue avec la probabilité p_1 dans le premier coup, p_2 dans le second coup, etc., de façon que les $p_1, p_2, \ldots$ varient arbitrairement de coup à coup, mais convergent vers une limite p. La suite obtenue montrerait toutes les propriétés d'une suite normale; d'ailleurs même si l'on rétrécissait cette définition en agrandissant le domaine d'invariance, on n'arriverait jamais à distinguer la suite obtenue avec la roulette réglable d'une suite normale.

Pour exprimer l'idée de la constance de la probabilité, nous avons besoin d'un nouveau *moyen de caractérisation structurale*. Si nous réalisons la suite de nombreuses fois en réglant toujours la roulette de la même façon, nous obtiendrons une *grille de probabilité*

$$(55)\qquad \left\{\begin{array}{ccccc} y_{11} & y_{12} & \ldots & y_{1n} & \ldots \\ y_{21} & y_{22} & \ldots & y_{2n} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ y_{k1} & y_{k2} & \ldots & y_{kn} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{array}\right.$$

Les limites horizontales seront toutes égales à p; mais nous retrouverons les $p_1, p_2, \ldots$ dans les limites verticales de la grille. L'idée de la constance de la probabilité s'exprimerait, par contre, dans la condition que toutes les limites verticales soient aussi égales à p. Nous trouvons donc dans la grille probabilitaire un nouveau moyen pour exprimer les propriétés des suites considérées.

Pour pouvoir écrire les relations dans la grille, introduisons deux indices supérieurs qui indiquent la place dans la grille, et non plus la phase. L'indice courant sera indiqué par répétition hors des parenthèses, donc

$$P(A, B^{ki})^i$$

désignera la probabilité dans une suite horizontale et

$$P(A, B^{ki})^k,$$

la probabilité dans une suite verticale.

Le fait remarquable est qu'il n'existe aucun moyen de déduire des limites horizontales une propriété des limites verticales. Un simple exemple le démontrera. Plaçons partout à gauche de la diagonale des termes égaux à B, mais continuons chaque suite horizontale de façon qu'elle acquière les propriétés d'une suite normale de probabilité p. Toutes les limites verticales resteront néanmoins égales à 1.

Les énoncés suivants doivent donc être considérés comme des *définitions* qui établissent arbitrairement certaines propriétés de la grille. D'autre part, quant aux applications, il faudra envisager séparément chaque problème donné pour voir si l'on doit le rattacher à une de ces définitions et préciser ce choix.

La grille homogène est définie par les conditions

$$P(A, B^{ki})^i = P(A, B^{ki})^k. \tag{56}$$

La grille convergente est définie par

$$\left\{\begin{aligned} &P(A, B^{ki})^i = p \qquad (k = 1, 2, 3, \ldots), \\ &\lim_{i=\infty} P(A, B^{ki})^k = p. \end{aligned}\right. \tag{57}$$

Cette grille correspond aux suites obtenues par la roulette réglable.

Nous définissons encore *l'indépendance par combinaisons* :

$$P(A . B_{l_1}^{k+\rho_1, i} \ldots B_{l_{\nu-1}}^{k+\rho_{\nu-1}, i}, B_l^{ki})^i = P(A, B_l^{ki})^i \qquad (k = 1, 2, 3, \ldots). \tag{58}$$

$$P(A . B_{l_1}^{k, i+\rho_1} \ldots B_{l_{\nu-1}}^{k, i+\rho_{\nu-1}}, B_{l_\nu}^{ki})^k = P(A, B_\nu^{ki})^k \qquad (i = 1, 2, 3, \ldots). \tag{59}$$

$$(\nu = 1, 2, 3, \ldots).$$

(58) définit cette propriété pour les suites horizontales, (59) pour les suites verticales. Cette définition est nécessaire parce que l'indépendance de chaque paire de suites ne suffit pas pour fonder l'indépendance par combinaisons; nous avons démontré cela plus haut. Avec le terme introduit au paragraphe 4, nous pouvons donner à notre définition la forme suivante : les suites d'une grille sont indépendantes par combinaisons, si pour chaque ν un groupe de ν suites arbitrairement choisi est complètement indépendant.

Nous dirons qu'une grille est composée de *suites normales au sens restreint*, si les suites horizontales et verticales sont normales au sens plus large, défini plus haut, si elles sont indépendantes par combinaisons selon (58) et (59), et si la grille est homogène selon (56).

Une autre propriété qui peut être réalisée aussi si les suites ne sont pas normales sera appelée *invariance de grille*. Nous la définissons par

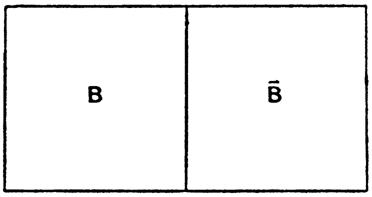
$$P(A . B_{l_0}^{ki} \ldots B_{l_{\nu-1}}^{k, i+\nu-1}, B_{l_\nu}^{k, i+\nu})^i = P(A . B_{l_0}^{ki} \ldots B_{l_{\nu-1}}^{k, i+\nu-1}, B_{l_\nu}^{k, i+\nu})^k, \tag{60}$$

$$(\nu = 1, 2, 3, \ldots).$$

Cette définition signifie que les probabilités de phases des suites horizontales se retrouvent dans les probabilités des combinaisons des suites verticales. Elle présuppose que le premier membre de (60) possède la même valeur pour tous les k, et le second membre pour tous les i.

Illustrons la signification de ce qui précède par un exemple emprunté à la théorie cinétique des gaz. Considérons un récipient divisé par une paroi en deux capacités égales B et $\overline{B}$ (*fig.* 2), contenant l'une, B, de l'oxygène, l'autre, $\overline{B}$, de l'azote.

Fig. 2.



Si nous enlevons la paroi, il y aura mélange des gaz; ce phénomène s'exprimera à l'aide d'une grille probabilitaire. Omettons les suites parcourues par les molécules de l'azote; celle des molécules de l'oxgyène s'arrangeront alors sous la forme d'une grille, si nous représentons le cours d'une molécule par une suite horizontale dont les éléments correspondent aux états de la molécule notés régulièrement par de courts intervalles de temps. Ces suites horizontales présentent une rémanence de la probabilité, parce qu'une molécule de B se trouvera en général, un instant plus tard, encore dans B (et de même pour $\overline{B}$); elles convergent cependant vers la probabilité moyenne $p = \frac{1}{2}$. La première suite verticale est composée par des seules B; la seconde présente déjà quelques $\overline{B}$, et les suivantes se rapprochent de plus en plus de la probabilité $p = \frac{1}{2}$. La grille est donc convergente. De plus, elle possède la propriété que nous avons appellée *invariance de grille* : les probabilités de phase, valables pour la rémanence de chaque molécule, se retrouvent dans les limites verticales. Si p_ν est la probabilité pour que la molécule, trouvée dans B, se trouve aussi dans B après ν termes de la suite horizontale, ce p_ν sera aussi la probabilité réalisée dans la $\nu^{\text{ième}}$ colonne verticale.

Dans la théorie cinétique des gaz, on a analysé cette propriété en raisonnant *de l'ensemble temporel à l'ensemble spatial*, et l'on a cru pouvoir fonder ces raisonnements sur le théorème de Bernoulli. On peut, il est vrai, déduire certains théorèmes concernant les suites verticales à la condition que la grille soit *finie* dans cette direction. Quand il s'agit de suites normales horizontales, ou de suites horizontales avec transfert de probabilité, on peut démontrer, à l'aide du théorème de Bernoulli, que les suites verticales d'une fréquence très différente de p sont très rares, si l'on compte, dans la direction horizontale, l'ensemble des verticales. C'est ce résultat qu'on a considéré comme justification des inférences de l'ensemble temporel à l'ensemble spatial.

Ce théorème supposant une longueur finie des suites verticales, on s'est appuyé sur le fait que les systèmes physiques sont en effet, finis dans cette direction; le nombre des molécules d'un système isolé, par exemple, est fini. Mais il ne faut pas oublier que ces systèmes sont aussi bien finis dans la direction du temps, au moins tant qu'il est question d'observations par des êtres humains. L'infini constitue, toujours pour la physique, une idéalisation; on en a besoin pour exprimer d'une façon simple certaines déductions que nous faisons déjà pour des suites finies et qui vont des éléments observés vers une prolongation de la série. Or, nous introduisons des déductions de ce type aussi bien pour la direction verticale que pour la direction horizontale; il faut donc introduire l'idéalisation d'une série infinie aussi bien pour l'une des directions que pour l'autre. Sans cela le schéma mathématique ne s'adapterait plus aux besoins de la physique.

Il est facile de donner des exemples de raisonnements employés par des physiciens, et qui attribuent aux suites verticales des qualités particulières, non déductibles des propriétés des suites horizontales. Imaginons que le phénomène du mélange se poursuive progressivement dans l'espace. En d'autres termes, imaginons qu'au début, seules les molécules de la partie supérieure du récipient quittent le domaine B, suivies ensuite par celles qui sont posées immédiatement en dessous, etc., de façon que le phénomène corresponde à un arrangement de la grille dans lequel toutes les places à gauche de la diagonale soient occupées par B. Un tel processus ne peut pas être exclu par le théorème de Bernoulli appliqué sur les suites horizontales. Il ne suffit pas de répondre que la longueur verticale de la grille étant finie, nous n'aurons qu'à attendre jusqu'à ce que la dernière suite horizontale ait traversé la diagonale; le physicien veut dire davantage, il veut dire qu'un tel mélange asymétrique n'arrive pas, ou au moins est très improbable. Une telle conception admet nécessairement que la grille présente des propriétés semblables dans la direction d'espace et dans celle du temps, hypothèse que nous avons introduite sous le nom *d'invariance de* *grille*. Une théorie mathématique embrassant toutes les conséquences dont le physicien a besoin, ne doit pas négliger cette hypothèse.

Nous n'avons donné dans ce qui précède qu'un court résumé de la théorie de l'ordre des suites probabilitaires. Le domaine qu'elle embrasse est, comme on le voit, très vaste et offre encore de nombreuses possibilités au mathématicien qui étudiera les autres types de suites et leurs propriétés.

Cependant, notre résumé nous permet déjà de formuler le trait caractéristique de cette théorie de l'ordre. La structure intérieure des suites se caractérise toujours à l'aide de théorèmes posant l'égalité de certaines probabilités concernant soit des suites partielles de la série originale, soit des suites verticales et horizontales d'une grille. L'analyse de la structure des suites probabilitaires ne constitue donc pas un problème logique ou épistémologique différant de celui de la détermination de la probabilité d'une suite; car si nous pouvons déterminer la probabilité de nos suites, nous pouvons aussi dire si deux de ces probabilités sont égales. Notre construction du calcul des probabilités montre donc que le problème logique et épistémologique de la probabilité ne concerne que le concept général de la probabilité défini par nos axiomes, ou par l'interprétation de la fréquence; les questions de la structure des suites probabilitaires ne comportent aucune question supplémentaire de nature philosophique.

CHAPITRE IV.

ÉLARGISSEMENTS DU CONCEPT DE SUITE PROBABILITAIRE.

14. **Suites primitives de probabilité.** — Nous avons développé le calcul des probabilités comme discipline formelle, séparée de l'interprétation de la fréquence qui joue le rôle d'une interprétation physique, matérielle, de ce système. Comme dans tous les cas semblables, il existe cependant non seulement *une* interprétation matérielle, unique, mais plusieurs; nous allons en construire une qu'on appelle l'interprétation géométrique.

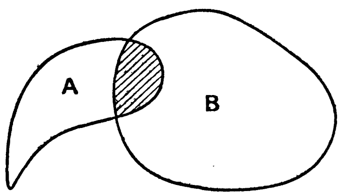
Imaginons qu'on ait défini, sur un plan, une fonction $\varphi(x, y)$ qui détermine la mesure M(A) d'un domaine A par

$$\text{(61)} \qquad M(A) = \iint_A \varphi(x, y)\, dx\, dy.$$

Considérons deux domaines A et B du plan; nous établissons l'interprétation géométrique (*fig.* 3) par la définition

$$\text{(62)} \qquad P(A, B) = \frac{M(A.B)}{M(A)}.$$

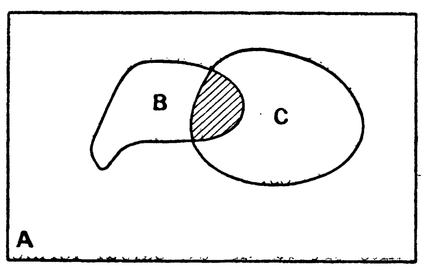
Fig. 3.



Ici, A.B signifie le domaine commun aux deux domaines A et B.

Il convient de choisir $M(A) = 1$, et les autres domaines B et C à

Fig. 4.



l'intérieur du domaine A (*fig.* 4). Avec cette condition supplémentaire, nous aurons

$$\text{(63)} \qquad \left\{ \begin{array}{l} P(A, B) = M(B), \qquad P(A, C) = M(C), \\ \qquad P(A.B, C) = \dfrac{M(B.C)}{M(B)}. \end{array} \right.$$

Il est facile de voir qu'avec ces définitions tous les axiomes du système formel des probabilités sont remplis, si nous considérons B.C comme classe commune, et $B \vee C$ comme classe de réunion de B et C.

Il s'ensuit que le mathématicien qui établit des formules du système formel des probabilités peut les considérer comme des énoncés concernant soit des fréquences, soit des mesures de domaines géométriques. Ce sont seulement les applications du concept de probabilité en physique qui font attribuer une position privilégiée à l'interprétation par la fréquence; cependant, l'interprétation géométrique trouve, elle aussi, une fonction logique dans ces applications, fonction que nous allons étudier.

Jusqu'à présent, les éléments x_i, y_i, ... de nos séries étaient classés selon des classes A, B, ..., fixes. On peut généraliser le type de suite probabilitaire en introduisant une *classification variable*. Chaque élément x_i, y_i, ... sera muni d'une *marque* indiquée par deux coordonnées ξ, η, et interprétée comme point sur notre plan; nous pourrons alors introduire une classification quelconque et la traiter comme les classifications A, B, ... employées auparavant. Nous appelons une suite de cette forme *suite primitive de probabilité*, pour indiquer qu'elle n'est pas directement une suite de probabilité, mais qu'un nombre infini de suites de probabilité peut être *dérivé* de cette suite. Nous avons affaire à la théorie des probabilités dites *géométriques*, ou probabilités à *marque continue*. Il est bien clair que le nombre des dimensions de notre *espace de marques*, ou *espace caractéristique*, est arbitraire, entre 1 et l'infini; nous nous bornerons à écrire nos formules pour l'espace à deux dimensions, simplification qui n'influence pas en principe nos résultats.

On connaît le rôle important que jouent les probabilités de ce type, dans les applications physiques. Mais on a oublié de dire que la justification logique de cette généralisation se trouve dans le fait de la possibilité de ces deux interprétations du calcul formel des probabilités. Cette justification revient à un isomorphisme entre l'interprétation géométrique et celle de la fréquence.

Il est facile de comprendre la signification de cet isomorphisme. Si l'on a déjà introduit une classification des éléments de la série, par exemple la classification B et $\overline{\text{B}}$ pour les y_i, on peut naturellement définir une fonction $\varphi(\xi, \eta)$ telle qu'on ait, pour la probabilité P(A, B) *définie par la fréquence*, la relation

$$\text{P}(\text{A}, \text{B}) = \int_{\text{B}}\!\!\int \varphi(\xi, \eta)\, d\xi\, d\eta. \tag{64}$$

Il n'est cependant pas possible d'en conclure, sans plus, que la *même* fonction $\varphi(\xi, \eta)$ rend ce même service pour toute autre classification C, D, ..., qu'elle nous permet de mesurer la probabilité de la disjonction $P(A, B \vee C)$ par une intégrale analogue sur le domaine de réunion $B \vee C$, etc. Seul l'isomorphisme indiqué nous autorise à faire usage de cette forme simple de calcul, constamment employée dans les applications. Notre construction axiomatique du système des probabilités, la distinction entre le système formel et ses interprétations, justifie donc également la généralisation du concept de suite probabilitaire réalisée dans les probabilités à caractère continu.

Ajoutons une remarque sur la fonction $\varphi(\xi, \eta)$. Dans la théorie des probabilités géométriques, on ne considère pas seulement des additions à un nombre fini de termes, mais on y introduit des additions d'un nombre infini de termes telles que la somme converge vers une limite. C'est pourquoi on a besoin d'un axiome de la forme : si l'on a des domaines D_1, D_2, ... convergeant vers une limite D, et si leurs probabilités coordonnées p_1, p_2, ... convergent vers la limite p, ce p est la probabilité de D. Cet axiome de continuité ne peut pas être déduit de nos axiomes donnés plus haut; la théorie des suites primitives des probabilités a donc besoin d'un axiome supplémentaire. Ce fait, cependant, ne peut pas nous surprendre; dans la théorie de l'ordre nous avons rencontré beaucoup de cas qui exigent l'introduction d'axiomes supplémentaires. On doit les considérer comme des conditions définissant le cas particulier envisagé; de la même manière nous avons à considérer l'axiome de la continuité comme une des conditions définissant le concept de suite primitive de probabilité, qui ne s'épuise pas dans le postulat de l'existence de marques continues pour les éléments de la suite.

On peut remplacer l'axiome de continuité par le postulat de l'existence d'une fonction $\varphi(\xi, \eta)$ intégrable (dans le sens de Riemann) telle qu'elle fournisse une représentation des fréquences sur l'espace caractéristique. D'autre part, la possibilité logique d'une telle fonction représentative repose sur l'isomorphisme des deux interprétations; c'est-à-dire cet isomorphisme nous garantit que le postulat de l'existence d'une telle fonction ne renferme pas de contradictions.

Avec ces remarques, nous terminons notre discussion des probabilités à marque continue. Notre analyse en a établi les fondements logiques; toutes les autres questions sont purement mathématiques, et je n'ai pas besoin de parler de l'abondance et des richesses des résultats obtenus, dans ce domaine, par les travaux de générations de mathématiciens.

15. **Suites continues de probabilité.** — L'élargissement du concept de suite de probabilité que nous venons de discuter se fait à l'aide du passage de marques discrètes à des marques continues. Il y a encore une autre généralisation du concept de suite de probabilité telle que le passage à une variation continue ne concerne pas la marque, mais l'élément y_i de la suite. Au lieu d'une suite d'éléments discrets y_i nous avons là une suite continue d'une variable y_t; on peut imaginer l'indice t par exemple comme indiquant le temps, tel qu'à chaque instant t il y a un élément y_t correspondant de la suite.

Il est bien clair qu'une telle généralisation ne peut pas partir du concept des suites normales, mais doit être rattachée au concept des suites avec rémanence. Car si la marque ξ coordonnée à l'élément y_t est située à un lieu déterminé de l'espace caractéristique, la marque $\xi + d\xi$ d'un élément y_{t+dt} ne peut pas être très loin de ce lieu, à cause de la continuité; cela veut dire qu'il y a là une rémanence de probabilité se manifestant comme une *tendance à persévérer*. On voit de plus que la condition du transfert de probabilité, la restriction de l'influence au prédécesseur immédiat, s'exprime pour ces suites dans la condition que seulement *le lieu* du point caractéristique détermine la rémanence, tandis qu'une influence du second prédécesseur prendrait ici la forme d'une contribution de *la vitesse* de ce point à la rémanence. L'extension de cette idée aux dérivées plus élevées du mouvement du point caractéristique se conçoit de même.

Dans cette généralisation du concept de suite de probabilité, l'interprétation de la fréquence est remplacée par une *interprétation de durée temporelle*. Le signe

$$\underset{0}{\overset{\tau}{\mathfrak{L}_t}}(\xi \,\varepsilon\, B)$$

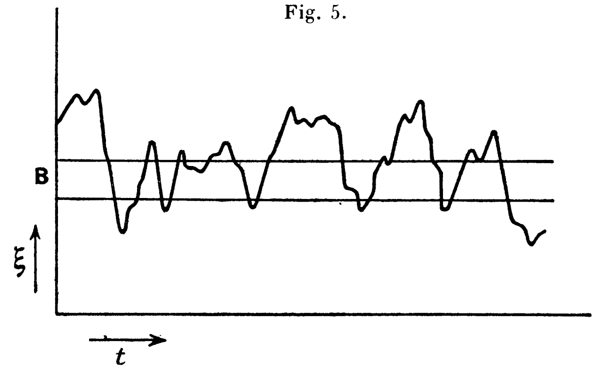
désignera la longueur du temps pendant lequel le point caractéristique se trouve dans B, à l'intérieur de l'intervalle temporaire de o à τ. Nous définissons alors la probabilité de trouver le point caractéristique dans B par

$$\mathrm{P}(\mathrm{A}, \mathrm{B}) = \lim_{\tau = \infty} \frac{1}{\tau} \int_{0}^{\tau}{}_{t} (\xi \varepsilon \mathrm{B}). \tag{65}$$

Il s'agit donc ici d'une interprétation géométrique de notre système d'axiomes.

Il existe en Physique beaucoup d'exemples de ce type de suite probabilitaire. Le mouvement brownien, par exemple, est représenté par une courbe oscillante dont l'ordonnée est une variable aléatoire continue

Fig. 5.



(*fig.* 5); c'est-à-dire, à chaque instant la prolongation de la courbe ne peut être prédite qu'avec une certaine probabilité. On peut parler d'un *état de probabilité*, existant pour chaque point de la courbe, et coordonnant à chaque direction une probabilité pour que la courbe la choisisse pour sa prolongation. On a souvent traité de pareils problèmes en coupant la courbe aléatoire en de petites sections, et en traitant la suite de ces sections comme une suite discrète de probabilité avec rémanence fixe. Mais on n'a pas besoin d'introduire une telle simplification qui n'est qu'approximativement valable; on peut traiter le problème d'une suite continue de probabilité avec toute la rigueur mathématique requise.

Pour atteindre ce but, nous avons à traiter la rémanence à l'aide d'un passage à la limite concernant des distances temporelles infiniment petites, passage accompagné d'un accroissement de la probabilité de rémanence vers l'unité. Prenons le cas du transfert de probabilité : la probabilité $p + \varepsilon$ que nous avons étudié là doit avoir ici la forme [1]

$$1 - \alpha\, dt,$$

telle qu'elle tende vers 1 si dt tend vers zéro. α est une constante. Appelons B un domaine de l'espace caractéristique, formé par la bande infiniment étendue entre deux lignes horizontales (*fig.* 5), la largeur (mesurée verticalement) restant arbitraire; nous avons alors les quatre probabilités suivantes [2] :

$$(66) \qquad \begin{cases} P(A.B, B^{dt}) = \quad p_1 \quad = 1 - \alpha\, dt, \\ P(A.B, \overline{B}^{dt}) = 1 - p_1 = \alpha\, dt, \\ P(A.\overline{B}, B^{dt}) = \quad q_1 \quad = \beta\, dt, \\ P(A.\overline{B}, \overline{B}^{dt}) = 1 - q_1 = 1 - \beta\, dt. \end{cases}$$

Nous appelons α et β les *coefficients de transfert;* leurs valeurs dépendent de la largeur du domaine B. En appliquant l'équation (44), nous obtenons

$$(67) \qquad P(A, B) = p = \frac{\beta\, dt}{\alpha\, dt + \beta\, dt} = \frac{\beta}{\alpha + \beta}.$$

Il y a donc une probabilité moyenne p pour obtenir le point caractéristique dans le domaine B.

Nous passons maintenant au calcul de la probabilité $f(t)$ pour que le point caractéristique reste dans le domaine B pendant tout le temps t.

---

(1) Cette façon d'écrire ne signifie pas autre chose que la définition que la fonction f définie par (68) admet un développement en série de Taylor.

(2) Les relations (66) renferment la supposition que les probabilités p_1 et q_1 ne dépendent pas explicitement du temps t. Or, la probabilité pour que le point caractéristique quitte le domaine B dépend de la localisation du point à l'intérieur de B; un point situé près de la marge supérieure ou inférieure du domaine par exemple aura plus de chances de quitter le domaine que ne l'aurait un point situé vers le milieu. La probabilité p_1 doit donc être considérée comme une valeur moyenne déterminée par la distribution des points caractéristiques à l'intérieur d'une section verticale de B (pour $t = \text{const.}$); des considérations analogues s'appliquent à q_1. La constance temporelle de α et β représente donc l'hypothèse que la distribution des points caractéristiques sur une section verticale de la figure 5 est stationnaire, c'est-à-dire ne change pas au cours du temps. Si l'on veut échappper à cette restriction, il faut introduire un traitement plus général de ces suites dans lequel la nature continue de ξ soit prise en considération; une méthode pareille a été indiquée par l'auteur dans *Stetige Wahrscheinlichkeitsfolgen* (*Zeitschr. f. Phys.*, 53, 1929, p. 274). Un résumé de cette méthode se trouve dans *Wahrscheinlichkeitslehre*, § 81.

Nous avons

$$f(t) = \lim_{\substack{dt=0 \\ n\,dt=t}} \mathrm{P}(\mathrm{A.B}, \mathrm{B}^{dt}.\mathrm{B}^{2dt}\ldots\mathrm{B}^{n\,dt}). \tag{68}$$

A cause de la condition du transfert de probabilité le terme à droite de la virgule se scinde en un produit de n facteurs de la forme $\mathrm{P}(\mathrm{A.B}, \mathrm{B}^{dt})$; nous avons donc

$$f(t) = \lim_{\substack{dt=0 \\ n\,dt=t}} (1-\alpha\,dt)^n.$$

Mettons

$$\alpha\,dt = \frac{\alpha t}{n} = \frac{1}{m} \qquad \text{ou} \qquad n = m\alpha t,$$

nous aurons donc

$$\begin{aligned} f(t) &= \lim_{m=\infty}\left(1-\frac{1}{m}\right)^{m\alpha t} \\ &= \left[\lim_{m=\infty}\left(1-\frac{1}{m}\right)^{m}\right]^{\alpha t} = \left[\lim_{m=\infty}\left(1+\frac{1}{m}\right)^{m}\right]^{-\alpha t} = e^{-\alpha t} \end{aligned} \tag{69}$$

avec la définition bien connue de l'exponentielle. C'est là la formule de l'amortissement exponentiel, employée dans tant d'applications du calcul des probabilités en physique.

On peut rattacher à ce résultat d'autres réflexions; par exemple on obtient, pour le nombre ν du changement de B à $\overline{\mathrm{B}}$, dans l'unité de temps

$$\nu = \alpha p = \frac{\alpha\beta}{\alpha+\beta}. \tag{70}$$

Quant à la démonstration de ce résultat, je dois me référer à mon livre Wahrscheinlichkeitslehre [1], où l'on trouve aussi d'autres considérations concernant ce type de suites probabilitaires.

On y trouve aussi un traitement plus général où l'espace caractéristique n'est pas simplement divisé en deux domaines B et $\overline{\mathrm{B}}$, mais est conçu comme l'espace de la variable continue ξ, conception qui réunit donc les deux passages à la continuité, concernant la marque et l'élément de la suite.

---

[1] § 45, § 81.

CHAPITRE V.

LE PROBLÈME DES APPLICATIONS.

16. L'interprétation de la fréquence. — Les idées que j'ai exposées jusqu'à présent concernent le calcul *mathématique* des probabilités. Pour employer la terminologie du début de cet article, j'ai exposé les relations *internes* sous forme d'un calcul embrassant, en même temps, les relations logiques et mathématiques. Je n'ai pas pu développer ce calcul en détail et présenter toutes ses diverses conséquences; je me suis borné à donner seulement une esquisse des idées principales. La construction des théorèmes découlant des axiomes est une tâche purement mathématique; les mathématiciens l'ont poussée, en des travaux remontant à presque trois siècles, à un degré de perfection montrant la productivité et la fécondité de cette admirable science, que constituent les mathématiques. Ici je n'ai voulu exposer que le squelette logique de ces opérations, le cadre relationnel dans lequel cette construction a été érigée.

Notre analyse nous a conduit à un résultat logique intéressant concernant une certaine dualité valable déjà dans le domaine du système mathématique. Le calcul des probabilités, comme système mathématique, admet deux formes de conception : un système formel, et un système interprété. Le premier est le système des relations rattachées au symbole P, système qui permet de formuler tous les théorèmes des probabilités, sans exception, mais qui s'abstient de toute interprétation de ce symbole, lequel ne figure dans les formules que sous forme d'une entité inconnue. Le second est le système dans lequel la probabilité est interprétée comme la limite de fréquence d'une série infinie. Cette série est un objet mathématique, elle aussi; c'est pourquoi nous restons dans le domaine des mathématiques même avec ce système interprété. Nous pouvons comparer ce système à celui de la géométrie si celle-ci est interprétée comme géométrie analytique, c'est-à-dire si les objets coordonnés aux concepts fondamentaux de la géométrie sont des triplets de coordonnées, des équations linéaires, etc.; cette interprétation fournit donc une représentation de la géométrie sur l'analyse. De la même façon, l'interprétation du calcul des probabilités à l'aide de la limite d'une fréquence réalise une représentation du système formel sur l'arithmétique.

La question se pose s'il y a une dépendance entre ces deux systèmes, si l'interprétation de la fréquence peut être déduite du système formel à l'aide de théorèmes de limite comme celui de Bernouilli. L'existence d'une autre interprétation du système formel, telle que l'interprétation géométrique, nous montre cependant que cette question doit être résolue négativement; une analyse nous explique ce fait en nous montrant que les théorèmes de limite ne nous fournissent que des énoncés *probabilitaires* sur la fréquence, jamais des énoncés *certains*. Ce qu'on peut démontrer c'est seulement que dans certains cas particuliers ([1]) l'existence d'une limite de la fréquence, égale à la probabilité formelle p de la suite, peut être affirmée avec la probabilité 1. Il s'en suit que si l'on ajoute une interprétation convenable du terme « probabilité 1 », par exemple dans le sens d'une certitude pratique, on peut en déduire un théorème sur l'existence d'une limite; une telle déduction ne posséderait pas, cependant, de grande importance épistémologique, et il serait préférable de considérer de pareilles réflexions seulement comme une nouvelle expression de la compatibilité de l'interprétation de la fréquence avec le système formel. Mais *compatibilité* ne signifie pas *dépendance;* le système interprété par la fréquence est logiquement indépendant du système formel.

L'existence de ce système mathématique interprété des probabilités est un résultat intéressant; elle nous montre que dans le domaine des mathématiques elles-mêmes il existe une théorie des probabilités qui ne diffère pas des autres branches des mathématiques. Ses concepts sont tout aussi « clairs et distincts », pour employer les termes d'un grand philosophe, que ceux des autres disciplines mathématiques; on n'y trouve rien qui corresponde au caractère indécis, imprécis, de la proba-

---

([1]) Ces cas concernent des grilles probabilitaires, formées par des suites normales, et pour lesquelles serait valable un axiome de continuité exigeant la commutativité des opérations *limite* et *probabilité*, telle que la limite d'une série de probabilités de certains termes soit égale à la probabilité de la limite de ces termes (*Cf.* n° 14, ci-dessus). Cependant un tel axiome ne doit pas être considéré comme axiome du système général des probabilités; sa validité est restreinte à des cas particuliers. Même dans le système formel des probabilités on peut donner aux théorèmes de limite, tels que le théorème de Bernouilli, une version qui ne suppose pas cet axiome.

bilité figurant dans les applications physiques. Même le concept de l'irrégularité qui, dans certains systèmes actuels des probabilités comporte des complications touchant à des paradoxes, se réduit à des notions simples, appartenant au cadre des concepts mathématiques usuels. Les séries infinies des probabilités sont donc des séries mathématiques ordinaires et les énoncés concernant leurs limites sont de même nature que les énoncés employés dans toutes les autres branches des mathématiques.

C'est là un résultat qui peut satisfaire le mathématicien; le physicien, cependant, nous demande une théorie des probabilités appliquées aux phénomènes physiques. Il s'agit donc maintenant de traiter ce problème avec des méthodes non moins précises, de montrer que ce problème de l'application physique admet une solution utilisant des notions aussi « claires et distinctes » que celles des mathématiques pures même si cette théorie nous conduit à des conséquences surprenantes et qui transforment toutes nos conceptions de la nature logique de la connaissance empirique.

Nous commencerons par exposer un ensemble d'idées qui servira de base à toute notre construction et que l'on peut intituler *la « toute-puissance » de l'interprétation de la fréquence.*

On trouve dans les applications du concept « probable » deux formes qui semblent, à première vue, tout à fait différentes. L'une est la forme statistique de la probabilité, se présentant dans les statistiques sociales, biologiques, physiques, etc.; sa ressemblance avec la notion mathématique de probabilité est évidente. L'autre se trouve dans l'usage du mot « probable » concernant des faits de la vie courante, des événements historiques, des hypothèses scientifiques, etc., usage qui n'est pas accompagné d'une détermination statistique, et qui confère au mot « probable » ce caractère indécis et vague dont les mathématiciens se sont si fortement effarouchés qu'ils ont préféré céder l'analyse de ce concept aux philosophes.

Je ne crois pas, cependant, que l'appréhension que soulève ce concept soit justifiée. On trouve dans le langage courant beaucoup de termes employés d'une façon inexacte et vague qui, pourtant, n'entraînent pas une imprécision de principe. Le charpentier qui mesure la longueur d'une poutre, interrogé quant à la signification de ce concept « longueur », fera, je suppose, une réponse assez confuse et ne trahissant que peu de relation avec le concept « longueur » des géomètres; si nous n'acceptons pas sa réponse comme preuve d'une différence entre ces deux concepts, c'est parce que nous pouvons substituer à son interprétation confuse une autre, celle des géomètres, sans que la correspondance entre ses mots et ses actions soit détruite. Le même procédé s'applique au terme « probable ». Si nous réussissons à substituer à l'usage vague du mot dans le langage courant une signification exacte, apte à justifier son usage, nous serons autorisés à dire que notre interprétation est correcte.

L'interprétation de la fréquence nous donne cette possibilité. Le terme « probable » admet toujours une interprétation dans le sens d'une fréquence même si l'individu qui emploie le mot nous assure qu'il n'a pas pensé à une fréquence. C'est ce que j'appelle la « toute-puissance » de l'interprétation de la fréquence.

Des exemples illustreront cette idée. L'historien qui parle de la probabilité que Jules César ait visité l'île de Bretagne, considère, il est vrai, un événement isolé; nous pouvons néanmoins interpréter son degré de probabilité comme une propriété statistique si nous examinons la méthode qui l'a conduit à cette supposition. Il l'appuie sur des rapports de chroniqueurs, et il juge de l'authenticité de ces rapports d'après des expériences concernant des rapports d'un type analogue. C'est donc *une statistique* qui le conduit à sa probabilité, et nous sommes à même de l'interpréter comme la prédiction que dans une série de cas semblables futurs la fréquence des suppositions confirmées sera de tel et tel degré.

On objectera que la classe dans laquelle il faut incorporer le cas individuel envisagé n'est pas bien définie. Cela est vrai, mais la pratique nous enseigne qu'on sait toujours bien choisir une classe apte à nous fournir une statistique utilisable. Dans ce choix nous suivons le principe de préférer la classe la plus restreinte. Si par exemple dans un cas de tuberculose le malade montre certains symptômes à la radiographie, nous choisissons pour calculer la probabilité de sa mort la statistique de la classe définie par « cas de tuberculose montrant ce symptôme sous les rayons Röntgen », et non pas la statistique générale des cas de tuberculose. Le principe de ce choix peut être justifié, d'une façon simple, par l'interprétation de la fréquence.

On objectera peut-être que la fréquence dans cette classe *sert de base* à la détermination de la probabilité, mais qu'elle ne peut pas nous fournir le *sens* de l'énoncé probabilitaire. L'historien veut dire quelque chose sur Jules César, et non pas sur une classe de chroniques; l'interprétation de la fréquence ne s'applique pas, dira-t-on, lorsqu'il s'agit d'un événement individuel.

Voici l'occasion d'introduire notre principe que l'opinion personnelle de celui qui parle ne nous touche point. Nous pouvons lui démontrer que s'il accepte notre interprétation, toutes ses actions rattachées à l'énoncé probabilitaire deviennent raisonnables, — voilà qui est suffisant.

La forme de notre démonstration est assez simple. L'interprétation de la fréquence nous fournit cette justification parce que, si nous préférons toujours la supposition la plus probable, cela implique selon cette interprétation que nous aurons le plus grand nombre de succès. Tout ce qui dépasse cette signification, toute tentative de voir dans l'assertion concernant l'événement individuel une qualité correspondant au degré de la probabilité, est superflue parce qu'elle ne peut jamais influencer notre comportement. Le principe pragmatiste disant qu'il n'y a de sens dans une proposition que dans la mesure où elle se traduit en des actions humaines, décide en faveur de l'interprétation de la fréquence; et une philosophie scientifique qui cherche des critères objectifs du concept de la signification ne peut pas se soustraire, me semble-t-il, à ce principe.

Prenons des exemples. Si quelqu'un me dit, que demain le temps sera beau avec une probabilité de 90 pour 100, cela ne dit rien, selon notre interprétation, sur le temps de demain, mais dit seulement quelque chose sur la fréquence des beaux jours dans une série de jours succédant à un jour de conditions météorologiques semblables à celles d'aujourd'hui. Néanmoins, il est raisonnable de faire une excursion demain, parce que si je suis ce principe de nombreuses fois, mes excursions se feront, dans un rapport de 90 pour 100, sous un ciel bleu. Ou si le médecin sait qu'une opération, dans le cas d'un cancer d'un certain type, promet un succès de 80 pour 100, il fera l'opération parce qu'au moins dans 80 pour 100 des cas il aura sauvé une vie humaine. Le cours des actions dans notre vie renferme une série assez longue d'actions pour justifier l'emploi de ce principe statistique dans nos décisions.

Avec cette interprétation statistique change, cependant, le sens que nous attribuons à une prédiction concernant un événement individuel. En nous décidant pour la prédiction la plus probable, nous ne voulons pas dire que cette prédiction soit *vraie*. Ce que nous faisons est comparable à la mise du joueur : nous *misons* sur cette proposition comme le joueur mise sur sa supposition. Une connaissance des probabilités nous permet de choisir *la mise la plus favorable*, — c'est là la seule qualité que nous puissions raisonnablement demander à nos prédictions. La probabilité détermine, disons-nous, le *poids* de la mise. Ce concept de poids prend la place de la probabilité du cas singulier; nous rattachons dans cette conception une mesure fictive à la mise singulière, qui ne se manifeste que pour des cas très nombreux, mais qui représente le poids avec lequel la prédiction du cas singulier entre dans nos calculs. L'interprétation de la fréquence nous conduit donc à une interprétation nouvelle des énoncés concernant l'avenir : les prédictions ne sont pas des *propositions vraies*, mais des *mises* possédant un certain *poids*.

L'importance de ces idées consiste en ce qu'elles nous permettent d'étendre l'interprétation de la fréquence à toutes les applications du concept « probable » dans les sciences. On sait bien que les énoncés scientifiques ne sont jamais certains; ils renferment toujours des prévisions, et dès les débuts de la philosophie on les a considérés comme *probables* et non pas comme *vrais*. Quelques philosophes croient cependant qu'il faut distinguer ce concept de probabilité de celui de la statistique, par conséquent reconnaître l'existence d'une probabilité philosophique différente de la probabilité mathématique. L'origine de cette *disparité* des deux concepts de la probabilité doit être cherchée dans le fait que ces philosophes ne possédaient pas de théorie qui justifie l'application de l'interprétation de la fréquence au cas singulier. Notre théorie cependant surmonte ces difficultés; c'est pourquoi nous sommes autorisés à soutenir *l'identité* des deux concepts de probabilité. Dans les deux concepts de mise et de poids nous possédons les outils qui nous permettent d'appliquer le concept statistique de la probabilité au cas singulier; c'est là la raison qui nous autorise à parler de la toute-puissance de l'interprétation de la fréquence. Nous appliquerons donc ces deux outils à la construction d'une conception nouvelle de la science; l'ensemble des énoncés scientifiques n'est pas constitué par des *propositions vraies*, mais par des *mises*, et le concept de *poids* se substitue à celui de la *vérité*.

Avec cette conception nous nous sommes engagés dans le chemin qui mène à la conception logique de la probabilité. L'idée de Leibniz de développer la théorie des probabilités en « une nouvelle espèce de logique », en une logique généralisée substituant l'échelle continue de la probabilité à l'alternative vrai-faux, paraît une conséquence nécessaire de la découverte que toutes les propositions de la science sont des énoncés probabilitaires. La théorie des probabilités ne signifie pas une discipline spéciale restreinte à des applications statistiques dans le sens usuel du mot; elle constitue plutôt la forme logique de la pensée empirique, la logique de la science. On ne l'avait pas reconnu aussi longtemps qu'on s'était tenu à la schématisation qui remplace la probabilité d'un degré élevé par la vérité, et la probabilité d'un degré inférieur par la fausseté; mais à la lumière d'une analyse plus profonde, cette logique de l'alternative se présente comme une approximation à laquelle on doit substituer pour des buts plus exacts une logique probabilitaire.

Ce sera notre tâche de construire cette logique probabilitaire; et notre calcul des probabilités, qui renferme déjà des opérations logiques dans ses formules, facilitera cette transcription du système sous forme d'une logique cohérente. Cependant, avant de commencer ce travail, nous avons à intercaler une remarque concernant l'applicabilité de notre système mathématique des probabilités à des phénomènes physiques.

L'application d'un système relationnel mathématique présuppose la démonstration que les relations du système sont valables pour les objets physiques auxquels elles sont appliquées. On a discuté cette question depuis longtemps pour le calcul qui nous occupe, et les difficultés épistémologiques de la théorie des probabilités prennent leur origine dans ce problème. Notre construction axiomatique du système des probabilités nous permet facilement de donner une réponse à cette question. Nous avons montré que toutes les lois de la probabilité se déduisent de l'interprétation de la fréquence. Nous n'avons pas eu besoin d'autres suppositions, parce que nous avions montré que les théorèmes particuliers du calcul en question se réduisent aux théorèmes généraux, si l'on ajoute des conditions restrictives qui prennent toujours la forme d'une égalité entre deux degrés de probabilité. Le problème épistémologique de la probabilité trouve donc une réponse simple : les lois des probabilités sont applicables si les séries physiques considérées possèdent des limites de fréquence.

On peut considérer cette réponse comme une nouvelle manifestation de la toute-puissance de l'interprétation de la fréquence; cette interprétation nous délivre du problème de la validité des lois probabilitaires.

Cependant, c'est là le dernier succès de l'interprétation statistique; en allant un pas plus loin, nous rencontrons des frontières à sa toute-puissance.

Les difficultés qui apparaissent sont contenues dans le concept de *limite*. La limite d'une série infinie est une chose bien claire pour une série mathématique, parce que les séries mathématiques sont données par l'énoncé d'une règle de formation, donc *données par compréhension*, comme l'on dit en logique. Chaque proposition sur la convergence de la série est vérifiable; nous pouvons, par exemple, calculer la place n à partir de laquelle l'écart de la limite reste plus petit qu'une grandeur ε choisie arbitrairement, etc. La série d'événements physiques, cependant, est *donnée par extension*, c'est-à-dire par énumération des termes individuels, l'un après l'autre; or, étant donné la *vita brevis* des êtres humains, il est difficile de comprendre comment on pourrait arriver jamais à une connaissance parfaite de la série physique, et comment nous pouvons répondre aux questions des mathématiciens concernant sa convergence. En effet, les séries des physiciens et statisticiens sont finies; et comme on peut toujours prolonger une série finie de façon qu'elle converge vers une limite quelconque, la section donnée de la série ne nous fait rien savoir sur sa prolongation future.

On pourrait essayer d'échapper à cette difficulté en construisant un calcul des probabilités concernant des séries finies. La possibilité mathématique existe; on peut facilement démontrer que nos axiomes du calcul élémentaire sont strictement valables même pour des séries finies, et la théorie de l'ordre peut être construite, elle aussi, pour de telles séries quoiqu'ici les développements prennent une forme assez compliquée et deviennent un cas « d'epsilontique » extrême. Mais une telle solution n'avancerait pas le problème épistémologique. Le physicien n'est pas satisfait par une information, si exacte soit-elle, sur les relations entre ses séries finies; il veut savoir quelque chose sur le prolongement de ses séries, et s'il est même assez modeste pour ne pas insister sur une connaissance des propriétés des séries à l'infini, dans le sens strict du mot, il demande au moins des informations sur une prolongation vers un *infini pratique*, c'est-à-dire sur une prolongation assez longue qui, en tout cas, renferme les mêmes difficultés épistémologiques que l'infini strict. Une « finitisation » du calcul des probabilités ne nous délivrerait donc point de cette difficulté particulière rattachée au besoin du physicien de *prévoir*, de *prédire* la prolongation de la série.

Il existe un seul chemin menant hors de cette difficulté : c'est la renonciation à la logique à deux valeurs. Le critique mathématicien peut objecter qu'il nous est impossible d'arriver à une proposition *vraie* concernant la limite; mais cela n'exclut pas la possibilité d'arriver à une supposition *probable* concernant la convergence. Ou plutôt : il peut être raisonnable de *faire une mise* sur une certaine valeur de la limite; et cette mise peut posséder un certain poids. Nous voyons donc que les concepts de la logique probabilitaire nous offrent la possibilité de trouver une solution à ce problème qui concerne les fondements de la statistique appliquée et qui, du fait de l'interprétation de la fréquence, se retrouve dans chaque application du mot « probable ».

D'autre part, nous reconnaissons qu'avec cette tournure du problème de la limite, l'idée d'une logique probabilitaire, qui s'était déjà présentée auparavant, est amplifiée et approfondie. Si nous considérons probable seulement un énoncé sur la limite d'une fréquence, nous introduisons une probabilité d'un niveau plus élevé, à savoir un *énoncé probabilitaire sur une probabilité*. La logique de la probabilité doit donc être construite de façon à admettre des probabilités de divers niveaux.

Les réflexions qui précèdent me semblent indiquer la nécessité inévitable de chercher la solution du problème des probabilités dans la construction d'une logique probabilitaire. C'est sur cette tâche que nous allons maintenant fixer notre attention.

17. La logique probabilitaire. — Au commencement de nos recherches nous avons analysé la structure des énoncés de probabilité. En faisant usage du concept logique de *fonction propositionnelle*, nous avons introduit en (5) et (6) les deux notations suivantes :

$$P(\varphi x_i, \psi y_i) = p, \tag{71}$$

$$P(\psi z_k) = p. \tag{72}$$

Les symboles φx_i, ψy_i, ψz_k, utilisés dans ces formules, désignent des suites d'événements; la probabilité figure donc ici comme propriété d'une suite d'événements. Cela correspond à la conception statistique de la probabilité qui la considère comme une fréquence d'événements.

On peut, cependant, adopter une autre conception qui, quoique ne différant pas matériellement de la première, s'en distingue du point de vue formel. On peut considérer la probabilité, non comme qualité des suites d'événements, mais comme qualité des suites *de propositions* *coordonnées;* ou, en passant à l'interprétation de la fréquence, on peut interpréter la probabilité non comme fréquence d'événements, mais comme fréquence de propositions coordonnées. La distinction de ces deux conceptions joue un certain rôle dans les discussions des logiciens. Boole ([1]) l'avait déjà envisagée et quoique les mathématiciens aient raison d'estimer exagérée l'importance qu'on attribue à cette question [voir, à ce sujet, le clair exposé de M. Borel ([2]), dans une critique du livre de Keynes], notre désir de précision logique même dans les détails nous oblige à la mentionner.

Si l'on veut introduire cette autre conception dans les formules, on peut se servir de guillemets pour indiquer le passage de la considération des objets à celle de leurs symboles; nous aurions donc à écrire

$$P(\text{«}\, \varphi . x_i . \psi . y_i \,\text{»}) = p, \tag{73}$$

$$P(\text{«}\, \psi z_k \,\text{»}) = p. \tag{74}$$

Nous appellerons cette conception la *conception logique* de la probabilité, en l'opposant à la *conception mathématique* exposée auparavant.

Il est hors de doute que la question de savoir laquelle de ces deux conceptions représente la « vraie » conception de la probabilité est dépourvue de sens. Toutes les deux sont possibles, et de plus, elles sont équivalentes, parce que le dénombrement des événements fournit les mêmes valeurs numériques que celui des propositions coordonnées. A cause de cette équivalence numérique, combinée avec le fait que tous les axiomes se déduisent de l'interprétation de la fréquence, il existe un isomorphisme strict entre ces deux conceptions; à chaque théorème appartenant à l'une correspond un théorème dans l'autre, et leurs applications à la réalité ont exactement la même signification en ce qui concerne notre connaissance du monde extérieur. Ceci nous permet de parler de l'identité de la conception logique et de la conception mathématique de la probabilité; et nous pouvons interpréter cette thèse, en revenant à un langage strict, comme une identité de structure combinée avec une identité de toutes les conséquences observables.

C'est seulement en vue de considérations logiques qu'il faut au moins

---

([1]) G. Boole, *An investigation of the laws of thought*, London, 1854, p. 247-248.

([2]) J. Borel, *A propos d'un traité de probabilités* (*Revue philosophique*, t. 98, 1924, p. 324).

mentionner cette distinction. Nous allons concevoir la probabilité comme une généralisation de la vérité; et comme la vérité est une propriété des propositions, non pas des objets, cette conception de la probabilité supposera le passage de la conception mathématique à la conception logique qui considère la probabilité comme propriété d'une suite de propositions. Dans le cadre de ces considérations, il convient de choisir la notation à un seul terme, pour obtenir une analogie avec la vérité qui est conçue d'ordinaire comme propriété d'un seul terme. Introduisons, de plus, le nom de *suite propositionnelle* pour la combinaison (ψz_k) de la fonction propositionnelle ψ avec son *fondement*, la série des événements z_k; nous arrivons alors à un parallélisme entre le terme

$$P(« \psi z_k ») = p,$$

que nous lisons : « la probabilité de la suite propositionnelle « (ψz_k) » est égale à p », et le terme

$$V(« \psi z_1 ») = 1, \tag{75}$$

que nous lisons : « la valeur de vérité de la proposition individuelle « ψz_1 » est égale à 1 », ou plus simplement : « la proposition « ψz_1 » est vraie ». L'élément de la logique probabilitaire est donc la suite propositionnelle, comme la proposition est l'élément de la logique à deux valeurs.

Ce qui précède constitue une série de considérations logiques préliminaires à l'exposé de la conception logique de la probabilité. Après les avoir indiquées, nous n'aurons plus besoin d'y revenir; la distinction entre la considération des objets et la considération des symboles n'a plus aucune importance pour tout ce qui va suivre. Nous allons donc omettre dans nos formules les guillemets trop embarrassants, en convenant que les parenthèses du symbole P( ) renferment dans leurs fonctions celle des guillemets.

Avant de développer plus avant la logique probabilitaire, intercalons une remarque concernant les éléments de la logique à deux valeurs. La base de cette logique est formée par la considération des *connexions de propositions* et des valeurs de vérités, coordonnées à ces connexions. Les connexions sont faites par les particules « ou », « et », « implique », « équivalent », et la valeur de vérité de la connexion est présentée en fonction de la valeur de vérité des propositions a, b élémentaires à l'aide de la *table de valeurs* suivante :

I. — *Table de valeurs de la logique à deux valeurs.*

a. Négation.

a	$\bar{a}$
I	O
O	I

b. Disjonction, produit, implication, équivalence.

a	b	$V(a \vee b)$	$V(a.b)$	$V(a \supset b)$	$V(a \equiv b)$
I	I	I	I	I	I
I	O	I	O	O	O
O	I	I	O	I	O
O	O	O	O	I	I

Les deux premières colonnes sont les colonnes des arguments, séparées par un gros trait des colonnes des fonctions. Par exemple, « a ou b » est vrai dans les trois premiers cas, mais faux dans le cas où a est faux et b est faux, etc.

En passant à la logique probabilitaire, pour laquelle les éléments sont des suites propositionnelles, il faut définir d'abord les opérations de connexion. En désignant la suite propositionnelle par (φx_i) nous introduirons les définitions suivantes :

$$(76) \quad \left\{ \begin{array}{l} (\varphi x_i) \vee (\psi y_i) =_{Df} (\varphi x_i \vee \psi y_i) \\ (\varphi x_i) \ . \ (\psi y_i) =_{Df} (\varphi x_i \ . \ \psi y_i) \\ (\varphi x_i) \supset (\psi y_i) =_{Df} (\varphi x_i \supset \psi y_i) \\ (\varphi x_i) \equiv (\psi y_i) =_{Df} (\varphi x_i \equiv \psi y_i) \end{array} \right.$$

Leur signification est la suivante : la connexion de deux suites propositionnelles est définie par la suite propositionnelle des connexions de ses éléments. Cette définition permet l'application de notre calcul des probabilités.

Dans la logique de la probabilité il s'ajoute, cependant, aux précédentes une nouvelle connexion qui n'a pas d'analogie dans la logique de l'alternative, à savoir *l'opération de la sélection*, exprimée dans nos formules par la virgule. Nous avons là aussi

$$(77) \qquad (\varphi x_i), (\psi y_i) =_{Df} (\varphi x_i, \psi y_i)$$

Le terme du second membre peut être interprété de la façon suivante : si φx_i est vrai, le terme est égal à ψy_i, si φx_i est faux, il n'existe aucun terme. Nous éliminons donc ici tous les termes où φx_i est faux, en sorte que nous obtenons une suite partielle (ψz_k), telle que z_k signifie les éléments dans le nouveau dénombrement.

Nous pouvons construire maintenant les tables de valeurs de la logique des probabilités, valeurs qui sont déterminées par les formules de notre système. Il existe cependant, une différence essentielle entre la logique probabilitaire et celle de l'alternative. Dans cette dernière, la valeur de vérité de la connexion est déterminée par les valeurs de vérité des éléments ; dans la logique probabilitaire, cependant, la valeur de vérité de la connexion dépend d'un troisième paramètre, pour lequel nous pouvons choisir la valeur de vérité de la connexion faite par l'opération de la sélection. Il s'agit là d'une considération que nous avons déjà rencontrée plus haut ; nous avons vu que, par exemple, la probabilité de la disjonction n'est pas déterminée par les probabilités des éléments singuliers, mais présuppose l'indication de la probabilité du produit logique, laquelle de son côté revient à l'indication de la probabilité relative entre les deux éléments. On peut appeler cette dernière le *degré de couplage* entre les deux suites propositionnelles ; donc la probabilité de la connexion de deux suites propositionnelles dépend des probabilités individuelles de ces suites, *et de plus* du degré de couplage entre elles.

Nous sommes à même maintenant d'écrire les tables de valeurs de la logique probabilitaire ; on les trouvera ci-après, tables II *a* et II *b*.

II. — *Table de valeurs de la logique probabilitaire.*

a. Négation.

$P(\varphi x_i)$	$P(\overline{\varphi x_i})$
p	$1 - p$

c. Opération de sélection dans les cas limites de 1 et 0.

$P(\varphi x_i)$	$P(\psi y_i)$	$P(\varphi x_i, \psi y_i)$
1	1	1
1	0	0
0	1	?
0	0	?

b. Disjonction, produit, implication, équivalence, sélection.

$P(\varphi x_i)$	$P(\psi y_i)$	$P(\varphi x_i, \psi y_i)$	$P(\varphi x_i \vee \psi y_i)$	$P(\varphi x_i . \psi y_i)$	$P(\varphi x_i \supset \psi y_i)$	$P(\varphi x_i \equiv \psi y_i)$	$P(\psi y_i, \varphi x_i)$
p	q	u	$p + q - p.u$	$p.u$	$1 - p + p.u$	$1 - p - q + 2p.u$	$\frac{p.u}{q}$

Quant aux valeurs de la disjonction et du produit, elles ont été développées dans notre calcul, de même que la valeur de la probabilité relative inverse qui est celle de la règle de Bayes. Les deux autres, les probabilités de l'implication et de l'équivalence, découlent des transformations logistiques suivantes :

$$[\varphi x_i \supset \psi y_i] \equiv [\overline{\varphi x_i} \vee \psi y_i] \tag{78}$$

$$[\varphi x_i \equiv \psi y_i] \equiv [\varphi x_i . \psi y_i \vee \overline{\varphi x_i} . \overline{\psi y_i}] \tag{79}$$

auxquelles nous appliquons les règles de notre calcul.

On voit que la construction des tables de vérité de la logique probabilitaire ne renferme aucun arbitraire; ces tables sont complètement déterminées par le calcul. Pour cette raison, il serait oiseux de critiquer nos tables de valeurs et de leur reprocher, comme certains logiciens, la particularité que la valeur de vérité de la connexion n'est plus déterminée par les valeurs de vérité des éléments. Ce serait là une critique insoutenable; la valeur d'un système logique ne réside pas dans des qualités d'élégance et d'harmonie, elle ne dépend que de l'applicabilité du système à la réalité. On peut, il est vrai, construire d'autres généralisations de la logique dépendant uniquement de deux paramètres; ces systèmes auront cependant tous une caractéristique commune, celle de ne pouvoir jamais être interprétés comme une logique de probabilité. La forme de la généralisation dont nous avons besoin ne dépend pas de nos désirs; il faut la construire en envisageant la nature des objets; et le calcul des probabilités nous impose des lois d'une forme si parfaitement déterminée que les traits principaux de la logique probabilitaire ne dépendent plus du tout de notre choix. La dépendance de trois paramètres, en particulier, est un trait caractéristique pour toute logique probabilitaire.

On comprend maintenant la raison pour laquelle la logique probabilitaire n'a pu être découverte en partant des tableaux de la logique à deux valeurs, par des considérations ne se rattachant pas à une conception logistique du calcul des probabilités. La logique à deux valeurs doit être considérée comme un cas particulier d'une logique plus générale, cas particulier dans lequel la structure de cette logique générale n'apparaît que sous forme dégénérée.

Pour passer à ce cas particulier, partons du fait que nos formules sont valables pour des séries finies aussi bien que pour des séries infinies. Si nous entendons par limite de fréquence simplement la fréquence de la série finie, nos axiomes sont strictement remplis pour ces séries. La proposition à deux valeurs de vérités peut être considérée comme une suite propositionnelle de longueur 1 ; nous obtiendrons donc la logique à deux valeurs si nous spécialisons nos formules pour le cas d'une suite propositionnelle à 1 terme seulement.

Dans ce cas, la fréquence est restreinte aux deux valeurs 1 et 0; mais une restriction supplémentaire intervient. Nous avons développé plus haut l'inégalité (14), que nous pouvons écrire sous la forme

$$\frac{p+q-1}{p} \leqq u \leqq \frac{q}{p}. \tag{80}$$

Si nous posons $p = 1$, elle prend la forme

$$q \leqq u \leqq q. \tag{81}$$

Il s'ensuit que pour le cas $p = 1$, u n'est plus un paramètre indépendant; pour ce cas nous avons $u = q$. Pour le cas $p = 0$, cependant, u reste indéterminé. Nous avons exprimé ce résultat dans la table IIc. En prenant en considération cette table IIc, et en développant les tables IIa et IIb pour les valeurs 1 et 0 de p et q, nous obtenons les tables Ia et Ib de la logique à deux valeurs. Il est intéressant de constater que les cas indéterminés de la table IIc, indiqués par le point d'interrogation, disparaissent dans les formules obtenues.

On voit donc que le passage à une suite de longueur 1 ne nous fournit pas seulement la restriction des valeurs de vérité aux deux valeurs 1 et 0, mais aussi automatiquement la dépendance de deux paramètres seulement. Ce trait caractéristique de l'ancienne logique doit donc être considéré comme un phénomène de dégénérescence, qui ne se produit que pour le cas particulier de deux valeurs de vérité.

Notre logique probabilitaire nous ouvre aussi la voie pour la définition des *modalités*, bien connues dans la logique ancienne, à savoir les concepts de *nécessité*, *possibilité*, *impossibilité*. Le symbole (x) signifiant : « pour tous les x », et $(\exists x)$ signifiant : « il existe un x », nous adopterons les définitions suivantes [1] :

$$
(82)\left\{\begin{array}{llll}
\text{Le cas} & (x)\,\varphi(x) & \text{sera appelé :} & \textit{nécessité de } \varphi(x)\,[\mathrm{N}], \\
\quad » & (\exists x)\,\varphi(x).(\exists x)\,\overline{\varphi(x)} & \quad » \quad : & \textit{possibilité de } \varphi(x)\,[\mathrm{P}_{os}], \\
\quad » & (x)\,\overline{\varphi(x)} & \quad » \quad : & \textit{impossibilité de } \varphi(x)\,[\mathfrak{I}].
\end{array}\right.
$$

Ces définitions peuvent être appelées les *définitions par extension* de ces concepts, c'est-à-dire les définitions auxquelles nous sommes conduits si nous n'envisageons que l'extension des concepts [2].

Ajoutons que notre définition de la possibilité diffère de la définition usuelle en ce qu'elle exclut le cas de la nécessité; ce n'est là qu'une convention destinée à faciliter nos raisonnements.

Pour des séries finies, nos définitions sont équivalentes aux suivantes :

$$
(83)\left\{\begin{array}{llll}
\text{Le cas} & \mathrm{P}(\varphi x_i)=1 & \text{est appelé :} & \textit{nécessité de } \varphi(x_i)\,[\mathrm{N}], \\
\quad » & 0<\mathrm{P}(\varphi x_i)<1 & \quad » \quad : & \textit{possibilité de } \varphi(x_i)\,[\mathrm{P}_{os}], \\
\quad » & \mathrm{P}(\varphi x_i)=0 & \quad » \quad : & \textit{impossibilité de } \varphi(x_i)\,[\mathfrak{I}].
\end{array}\right.
$$

En appliquant ces définitions à nos tables IIa et IIb, nous obtenons par un calcul simple les tables IIIa et IIIb. Pour indiquer la modalité de $\varphi(x_i)$, nous écrivons $\mathrm{M}(\varphi x_i)$. On peut vérifier très simplement que ces tables sont valables aussi pour des séries infinies.

Ces tables montrent que l'indétermination provenant de l'existence d'un troisième paramètre disparaît en général et ne se manifeste que dans la ligne du milieu, caractérisée par P_{os} dans les deux premières colonnes. Nous retrouvons ainsi une propriété bien connue des modalités qui, en effet, ne sont pas déterminées dans ce cas. Si, par exemple nous jouons à pile et face, nous avons : pile est possible, face est possible, mais pile *ou* face est nécessaire. Si nous jouons, au contraire, avec deux pièces de monnaie, nous avons : pile à la première est possible, pile à la seconde est possible, pile à la première *ou* pile à la seconde n'est que possible. On peut facilement construire des exemples semblables pour les autres connexions logiques.

---

(1) Pour simplifier l'écriture nous omettons dans ce qui suit les guillemets dans les termes « $\varphi(x)$ », « $(x)\,\varphi(x)$ », etc.

(2) On obtiendrait des *définitions par compréhension*, ou mieux, *syntactiques*, si l'on considérait la structure interne des propositions, par exemple la relation de déductibilité entre deux formules; nous n'envisageons pas ici ces définitions.

III. — *Table de la logique topologique à trois valeurs.*

a. Négation.

$M(\varphi x_i)$	$M(\overline{\psi x_i})$
N	$\mathfrak{I}$
P_{os}	P_{os}
$\mathfrak{I}$	N

b. Sélection, disjonction, produit, implication, équivalence.

$M(\varphi x_i)$	$M(\psi y_i)$	$M(\varphi x_i, \psi y_i)$	$M(\varphi x_i \vee \psi y_i)$	$M(\varphi x_i . \psi y_i)$	$M(\varphi x_i \supset \psi y_i)$	$M(\varphi x_i \equiv \psi y_i)$	$M(\psi y_i, \varphi x_i)$
N	N	N	N	N	N	N	N
N	P_{os}	P_{os}	N	P_{os}	P_{os}	P_{os}	N
N	$\mathfrak{I}$	$\mathfrak{I}$	N	$\mathfrak{I}$	$\mathfrak{I}$	$\mathfrak{I}$	?
P_{os}	N	N	N	P_{os}	N	P_{os}	P_{os}
P_{os}	P_{os}	$\mathfrak{I}$ P_{os} N	P_{os} N	$\mathfrak{I}$ P_{os}	P_{os} N	$\mathfrak{I}$ P_{os} N	$\mathfrak{I}$ P_{os} N
P_{os}	$\mathfrak{I}$	$\mathfrak{I}$	P_{os}	$\mathfrak{I}$	P_{os}	P_{os}	?
$\mathfrak{I}$	N	?	N	$\mathfrak{I}$	N	$\mathfrak{I}$	$\mathfrak{I}$
$\mathfrak{I}$	P_{os}	?	P_{os}	$\mathfrak{I}$	N	P_{os}	$\mathfrak{I}$
$\mathfrak{I}$	$\mathfrak{I}$	?	$\mathfrak{I}$	$\mathfrak{I}$	N	N	?

18. Passage de la logique probabilitaire à la logique à deux valeurs. — La question se pose de savoir si l'on peut transformer un système d'énoncés appartenant à la logique probabilitaire, donc des mises, en un système d'énoncés à deux valeurs. Cette possibilité existe, et il y a deux manières distinctes de la réaliser.

La première consiste en l'introduction d'une classification telle que les énoncés d'un degré élevé de probabilité soient appelés « vrais », ceux d'un degré de probabilité plus bas étant appelés « faux ». Ce passage est un *passage par division;* nous introduisons deux *valeurs de démarcation* p_1 et p_2 et convenons que

$$\text{(84)} \qquad \left\{\begin{array}{lll} \text{Si} & P(a) \geqq p_2, & a \text{ est } \textit{vrai;} \\ \text{Si} & p_1 < P(a) < p_2, & a \text{ est } \textit{indéterminé;} \\ \text{Si} & P(a) \leqq p_1, & a \text{ est } \textit{faux.} \end{array}\right.$$

Les énoncés indéterminés sont exclus, c'est-à-dire on ne les prend pas en considération; il nous reste un système d'énoncés à deux valeurs, précisément celui qui est le plus souvent employé en physique. Strictement parlant nous sommes en présence de mises possédant un poids, que nous remplaçons cependant au moyen du passage par division par des énoncés considérés comme vrais ou faux.

Il ne faut jamais perdre de vue cependant le caractère approximatif du procédé employé. Les lois de la logique à deux valeurs sont valables, en général, mais il y a des cas où elles ne peuvent plus être employées. Imaginons deux énoncés de degrés p et q de probabilité, tels que $p > p_2$ et $q > p_2$. Si, de plus, les énoncés sont indépendants, leur produit logique a une probabilité $p.q$; mais il se peut que $p.q < p_2$. Nous aurions donc deux énoncés dont chacun serait vrai, mais ayant une connexion par « et » qui ne serait pas vraie.

Le second passage à la logique à deux valeurs peut se faire par l'interprétation de la fréquence. Nous substituerons au poids de la mise la fréquence de propositions dans la suite propositionnelle correspondante, et nous serons conduits ainsi à des énoncés qui entrent dans le cadre de la logique à deux valeurs.

On a cru qu'à cause de la possibilité de ce passage notre logique de la probabilité n'est pas une *véritable* logique probabilitaire, mais seulement un système abrégé qui, avec l'introduction de la définition des termes, se transforme en une logique à deux valeurs. Cette conception est cependant insoutenable. La logique probabilitaire est un système relationnel $\mathcal{L}$ défini par les tables de valeurs, et différent de la logique à deux valeurs; si nous prenons des suites propositionnelles comme éléments, nous pouvons écrire

$$\mathcal{L}[(a_i), (b_i), \ldots]. \tag{85}$$

L'interprétation de la fréquence est comparable à l'introduction de nouvelles variables, et nous avons

$$\mathcal{L}_0[a_i, b_i, \ldots], \tag{86}$$

où $\mathcal{L}_0$ est la logique à deux valeurs. Le système $\mathcal{L}_0$ se rapporte cependant à d'autres éléments, et le système $\mathcal{L}$, lequel considère les suites propositionnelles comme un tout, reste la véritable logique probabilitaire.

On peut considérer ce procédé comme une représentation de la logique probabilitaire sur la logique à deux valeurs, à l'aide de l'introduction de nouveaux éléments. Ce procédé est comparable à la représentation de la géométrie non-euclidienne sur la géométrie euclidienne. Une possibilité analogue existe pour d'autres systèmes de la logique à plusieurs valeurs (¹), mais elle n'altère pas l'indépendance de la logique probabilitaire, en tant que système relationnel.

Comparé au passage par division, le passage par l'interprétation de la fréquence possède l'avantage de pouvoir être exact. Cependant, il n'est exact que si deux conditions sont remplies :

1° Les éléments nouveaux a_i, b_i, ... doivent être des propositions caractérisées strictement par deux valeurs.

2° La proposition concernant la fréquence dans la suite propositionnelle doit être caractérisée strictement par deux valeurs.

Ces deux conditions sont remplies par le calcul purement mathématique des probabilités; c'est pourquoi ce calcul peut être construit dans le cadre de la logique à deux valeurs. Quant à son application à la

(¹) Post, par exemple, a déjà introduit une telle interprétation de ses systèmes à plusieurs valeurs. Pour d'autres systèmes de logique plurivalente *voir* Reichenbach, *Wahrscheinlichkeitslehre*, p. 365; récemment, M. Zawirski, a publié un très intéressant rapport sous le titre : *Über das Verhältnis der mehrwertigen Logik zur Wahrscheinlichkeitslogik*, *Studia philosophica*, Varsovie, 1935, p. 407.

réalité, c'est-à-dire à des propositions physiques, ces deux conditions ne sont pas remplies; c'est pourquoi le passage indiqué ne constitue qu'une approximation, pour toutes les propositions de nature physique.

Nous pouvons décrire d'une façon plus précise le caractère de cette approximation. Strictement parlant, les propositions élémentaires a_i, b_i, ... sont des mises possédant chacune un poids; si nous parlons, non d'un poids, mais de vérité ou de fausseté, nous effectuons un passage par division. Le passage de $\mathcal{L}$ à $\mathcal{L}_0$, à l'aide de l'interprétation de la fréquence, présuppose donc un autre passage par division concernant les éléments du dénombrement. D'autre part, selon nos considérations sur l'impossibilité de déterminer la limite de la fréquence, l'énoncé sur la fréquence, lui aussi, n'est pas un énoncé vrai, mais une mise possédant un poids.

Pour cette raison, l'interprétation de la fréquence ne peut pas éliminer le caractère approximatif de cette logique. Cela ne signifie cependant pas qu'un pareil passage soit superflu; au contraire, il s'agit là d'un procédé permettant d'obtenir une approximation d'un degré très élevé, et qui joue, pour cette raison, un rôle prédominant parmi les méthodes utilisées en science.

Nous pourrions essayer de construire un système de la connaissance en attribuant à chaque proposition un poids; nous constaterions cependant que de cette manière on obtient des poids assez mal déterminés. Le procédé actuel de la science substitue à cette méthode directe une méthode indirecte, qu'on doit considérer comme un des procédés les plus pénétrants et les plus efficaces imaginés par l'esprit humain. Nous commençons par un passage par division, nous n'acceptons que les propositions ayant un poids élevé ou faible, et nous omettons systématiquement le domaine intermédiaire. En appliquant l'interprétation de la fréquence, nous construisons, à l'aide de procédés de dénombrement, le poids des propositions omises auparavant. Cependant cela ne constitue pas le seul but de nos opérations; nous pouvons même contrôler le poids des propositions acceptées au commencement, et, si nécessaire, les déplacer du domaine qu'ils occupaient à l'origine dans l'échelle des poids vers un autre. De cette manière, une proposition acceptée au début comme vraie peut être rangée plus tard dans le domaine de l'indéterminé ou être considérée comme fausse. Ce procédé ne comporte pas de contradiction, parce que le changement du caractère de vérité d'une des propositions élémentaires n'influence pas beaucoup la fréquence résultante. Nous insistons toujours sur le postulat que le poids apprécié doit être confirmé plus tard par le dénombrement d'autres propositions, dont le poids a été également apprécié. Les appréciations initiales sont donc soumises à un processus de « dissolution », guidé par l'interprétation de la fréquence. Ce processus de dissolution nous conduit à des appréciations nouvelles; l'amélioration qui accompagne ce procédé consiste en ce que l'appréciation individuelle devient de moins en moins importante, et que l'erreur éventuelle influence de moins en moins le système tout entier. Nous construirons donc, par l'action réciproque de la division et de l'interprétation de la fréquence, un système de poids beaucoup plus exact que celui obtenu par une appréciation directe de tous les poids considérés.

Dans ce procédé la fonction essentielle de l'interprétation de la fréquence devient manifeste. Quoique notre logique des propositions n'ait pas deux valeurs, mais comporte une échelle continue, nous n'avons pas besoin de *commencer* la connaissance avec la logique probabilitaire. Nous commençons par une logique approximative à deux valeurs, et nous développons l'échelle continue à l'aide de l'interprétation de la fréquence. La même méthode s'applique en sens contraire : si un énoncé de probabilité est donné, nous le vérifions, à l'aide de l'interprétation de la fréquence, en le réduisant à des énoncés d'une logique approximative à deux valeurs. Cette logique approximative à deux valeurs est supérieure à la logique probabilitaire initiale, parce qu'elle omet le domaine douteux intermédiaire des poids.

L'interprétation de la fréquence rend cette réduction possible, en résolvant des poids en des fréquences; elle nous permet de confirmer l'appréciation directe avec celle de poids de degré supérieur ou inférieur. L'interprétation de la fréquence nous dispense du maniement d'un système inutilisable directement.

On ne doit pas oublier, cependant, que la logique à deux valeurs reste toujours approximative. Le système de la connaissance est écrit dans la langue de la logique probabilitaire; la logique à deux valeurs est une langue de substitution applicable seulement dans le cadre d'une approximation. Une épistémologie qui perd de vue ce fait risque de s'égarer sur les hauteurs désertes d'une pure idéalisation.

19. Le problème de l'induction. — Le système de la logique probabilitaire est un système formel, tout comme le calcul des probabilités; la question de l'applicabilité se pose donc pour ce système aussi bien que pour le système mathématique. Elle trouve la même réponse : le système est valable si les suites physiques possèdent une limite de la fréquence, et si la valeur de cette limite correspond au moins approximativement à la fréquence observée dans la suite finie présentée par l'expérience. Il est évident que nous ne savons rien quant à cette correspondance; ce problème déjà discuté auparavant se pose donc aussi bien avant qu'après la construction du système formel de la logique probabilitaire; ce système ne peut donc être qu'un élément auxiliaire de notre analyse complète.

Une idée s'impose : l'hypothèse concernant la permanence approximative de la fréquence observée doit être interprétée comme une *mise*, et non pas comme une proposition vraie. Il reste à démontrer que cette mise est *la plus favorable*, seul argument qui puisse justifier son adoption. C'est sous cette forme que se pose pour nous le problème de l'induction.

Formulons d'abord le principe en question.

Règle de l'induction. — Si une section de n termes d'une série x_i présente la fréquence f^n pour les termes appartenant à la classe B, nous misons sur le fait que la fréquence, pour une prolongation de la série, converge vers une limite p à l'intérieur de $f^n \pm \delta$.

Le principe précédent peut être considéré comme une généralisation du principe d'induction de la philosophie traditionnelle appelé *inducto per enumerationem simplicem*. Ce principe plus restreint est formulé d'ordinaire comme suit : si nous avons constaté qu'un phénomène déterminé B a eu lieu n fois, nous admettrons qu'il arrivera toujours. Il est évident que cet énoncé constitue un cas particulier de notre principe général, cas où la fréquence f^n observée est égale à 1 ; la théorie du principe traditionnel est donc contenue dans la théorie de notre principe plus général. D'autre part, on voit bien que les objections critiques présentées contre le principe traditionnel, surtout de la part de Hume, s'appliquent également à notre principe.

Les objections de Hume peuvent être résumées en deux énoncés :

1° Le principe de l'induction ne se base pas sur une nécessité logique.

2° Le principe de l'induction ne peut être considéré comme fondé sur l'expérience; une pareille inférence serait elle-même une inférence inductive, et la démonstration constituerait donc un cercle vicieux.

C'est surtout la seconde objection qui renferme le paradoxe de l'induction. Le principe s'est avéré valable assez souvent et il est difficile de supprimer la conclusion : « il sera également valable à l'avenir »; néanmoins, la conscience logique s'oppose à juste titre à la contrainte psychologique qui nous amène à cette conclusion, au sujet de laquelle la logique prononce un *non-liquet.* C'est en cela que consiste la difficulté particulière de l'inférence inductive; cette inférence est la forme de toute inférence empirique, et sert de base à toute justification par expérience, — c'est pour cette raison qu'elle ne peut pas être justifiée elle-même de la même façon. Voilà l'origine de la nature unique du problème épistémologique de la probabilité. L'exemple de la géométrie qui nous a servi déjà fréquemment comme modèle des relations logiques discutées, n'est plus valable ici : la validité de la géométrie de l'espace physique peut être démontrée par l'expérience, parce que les inférences employées dans cette démonstration ne présupposent pas la validité de la géométrie, au moins dans les grandes dimensions. Elles présupposent cependant le principe de l'induction, celui-ci étant l'élément nécessaire de toute démonstration empirique. De la même façon, la démonstration empirique du principe de l'induction présupposerait ce principe; dans ce cas, ce fait logique implique donc un cercle vicieux, et la démonstration échoue.

On pourrait essayer une défense du principe basée sur certaines idées développées dans le calcul des probabilités. Il existe un théorème concernant l'inférence inductive qu'on peut résumer dans la forme suivante : étant donné une fréquence f^n dans une série de n termes, la probabilité u_n pour que la série converge vers une limite voisine de la valeur f^n est déterminée par certaines autres probabilités p_ρ; mais quelles que soient ces probabilités p_ρ données, u_n convergera vers 1 avec la prolongation de la série. On serait enclin à chercher la solution du problème de l'induction dans ce théorème; cependant, un examen critique nous montre que notre théorème ne nous tire pas d'affaire. D'abord, le théorème n'est pas valable sans certaines hypothèses restrictives concernant un caractère Bernoullien des séries en question; ensuite, toute la démonstration admet l'applicabilité des règles du calcul des probabilités, et comme celle-ci présuppose, d'après notre démonstration, l'inférence inductive, cette tentative de démonstration aboutirait, elle aussi, à un cercle vicieux. La justification du principe de l'induction ne peut pas être obtenue par la théorie des probabilités; celle-ci, au contraire, présuppose la validité du principe. On peut affirmer seulement que cette justification devra être obtenue à l'aide de concepts formés dans la théorie de la probabilité.

Il faut donc chercher une autre solution du problème, indépendante des règles du calcul des probabilités. Une telle solution existe, et elle est basée sur une idée assez simple.

Il y a une propriété de notre règle de l'induction qu'on peut considérer comme hors de doute. Si la série envisagée possède une limite, notre règle, appliquée à plusieurs reprises, doit nous conduire à une valeur approximative de la limite, après un nombre fini d'étapes. Ce théorème est une conséquence simple de la définition de la limite, laquelle implique qu'après un nombre fini de termes, les écarts doivent rester compris dans un petit intervalle δ. En misant toujours sur la dernière valeur de fréquence f_n, nous suivons peut-être une route en zigzag, mais qui doit aboutir à un certain moment à l'intervalle $f^n + \delta$ et y rester toujours — sous condition naturellement que la limite existe.

Mais d'où savons-nous cela ? A quoi bon notre règle si la limite n'existe pas ? Ou si la série est trop longue et ne converge qu'après un nombre de termes trop grand pour qu'on puisse y parvenir pendant la durée d'une vie humaine ?

Certainement, dans ce cas, notre règle de l'induction ne serait bonne à rien. Mais dans ce cas nous ne connaîtrions, non plus, aucun moyen de faire des prédictions. Nous ne pourrions donc rien faire, bien sûr. Mais il ne s'ensuit pas que dans la situation actuelle de l'homme de science, ou de l'homme de la rue, il vaille mieux renoncer au procédé inductif. Une telle conséquence serait raisonnable seulement si nous savions qu'il n'existe pas de limite de la série. Or, notre situation est différente : nous ne savons pas si une limite existe. C'est cette distinction qu'il importe de ne pas perdre de vue lorsqu'on analyse le problème de l'induction; *savoir qu'une limite n'existe pas* est bien différent de *ne pas savoir si une limite existe*. La disjonction soutenant qu'une limite existe, ou n'existe pas, ne se répète pas si nous passons des énoncés concernant *des objets*, à des énoncés concernant *notre connaissance des objets;* les deux cas « savoir qu'une limite existe » et « savoir qu'une limite n'existe pas » n'épuisent pas les possibilités; entre ces deux cas s'intercale le cas de « ne pas savoir si une limite existe ». C'est ce cas qui caractérise notre situation actuelle, car la décision concernant une supposition sur les phénomènes futurs ne peut être faite par rapport aux possibilités de la structure du monde, mais seulement par rapport à *nos connaissances* de cette structure.

C'est là, me semble-t-il, la tournure décisive qui rend possible une solution du problème de l'induction. Si l'on avait considéré ce problème comme insoluble, c'est qu'on s'était tenu à la supposition fausse qu'une justification de l'induction exigerait la démonstration que le procédé de l'induction doit mener au succès; mais ce serait trop demander. Ce qu'on peut soutenir c'est seulement qu'une justification serait impossible si l'on pouvait démontrer que le procédé de l'induction ne mène pas au succès; mais si la démonstration négative implique l'impossibilité de la justification, il ne s'ensuit pas que la justification implique la démonstration positive. Une justification de l'induction est possible si, au moins, le succès n'est pas exclu; elle ne suppose qu'une démonstration montrant que l'induction est un moyen d'amener le succès si un succès est réalisable.

Cette démonstration peut être donnée; elle se déduit de la définition de la limite. La règle de l'induction nous conduit à la valeur approximative si une limite existe, et si elle est pratiquement accessible aux êtres humains; le raisonnement que nous venons de donner s'applique tout aussi bien pour ce cas que nous appelerons le cas d'une *limite pratique* (1).

Si une limite pratique n'existe pas, il faut, il est vrai, renoncer à des prédictions; mais aussi longtemps que cette thèse négative n'est pas démontrée, il est préférable au moins de *tenter de prédire*. Rem-

(1) On voit facilement que, d'autre part, le postulat d'une limite pratique exige moins que celui de la limite mathématique en tant qu'il peut être satisfait aussi par des séries semi-convergentes, qui présentent approximativement les propriétés d'une limite dans les domaines accessibles aux êtres humains, mais qui divergent à l'infini. L'infini mathématique, pour ainsi dire, n'est pas l'affaire du physicien.

plissons donc la condition nécessaire, s'il nous est interdit de remplir la condition suffisante.

Cette réflexion est-elle juste ? Est-ce que la prédiction devient impossible si les séries n'ont pas de limites ? Je peux imaginer un clairvoyant qui désigne la série, terme par terme à l'avance, même s'il n'y a pas de limite. Où réside la nécessité de la règle de l'induction ?

Il est vrai qu'on peut imaginer un être doué d'une telle clairvoyance; mais si je voulais le contrôler, il ne me resterait autre chose à faire que de compter ses succès et de juger par cette statistique de ses qualités. Si la statistique montre que cet homme est bon prophète, il n'y a qu'à l'utiliser; mais il ne serait guère à conseiller de se fier à lui avant l'examen statistique. En d'autres termes, je jugerai des qualités du clairvoyant à l'aide de la règle de l'induction, et je dois le faire parce que je n'ai pas d'autre moyen sur lequel, et sur les qualités de prédictions duquel, je sache quelque chose. La règle de l'induction est notre *moyen primaire* pour obtenir des prédictions ; tous les autres moyens sont *secondaires* et dépendent d'un contrôle fait à l'aide d'inductions. Il y a beaucoup de ces moyens secondaires, et de meilleurs que la clairvoyance, bien sûr; nous en parlerons plus loin.

D'autre part, notre exemple montre que la théorie n'a pas besoin de supposer que *toutes* les séries des phénomènes naturels possèdent une limite de fréquence. Ce postulat n'aurait aucun sens; la série des phénomènes eux-mêmes ne détermine pas une fréquence; c'est seulement après l'introduction d'une *classification* qu'une fréquence est définie. La convergence de cette fréquence dépendra toujours du caractère de la classification. Il suffit de supposer qu'*il est possible d'introduire une classification* de façon que la fréquence de la série converge. Dans notre exemple, la fréquence de la classification initiale ne converge pas ; mais nous introduisons une nouvelle classification en comptant les événements où la prédiction est juste, et cette classification peut entraîner une convergence de la fréquence. Il existe des méthodes purement mathématiques, c'est-à-dire qui ne font pas usage d'un phénomène physique associé, pour l'introduction d'une nouvelle classification. Par exemple, on peut considérer la fréquence non convergente $f^n = f$ comme *marque continue* de l'élément y_n, et envisager le cas où il y a une fonction distributive $\varphi(f)$, telle que la série , du point de vue de cette classification, possède la nature d'une série primitive de probabilité [1]. A l'aide de pareils moyens on peut manier des séries ayant une fréquence non convergente; mais le maniement se fait toujours par réduction à d'autres classifications qui admettent l'usage du principe de l'induction.

Notre thèse doit donc être formulée comme suit : *la possibilité de prédire suppose la possibilité d'une classification telle que le procédé de l'induction répétée mène au succès.* Dans ce sens, l'applicabilité de la méthode de l'induction est la *condition nécessaire* de la possibilité de prédictions. On peut aussi dire : si des prédictions sont possibles, le procédé de l'induction est une *condition suffisante* pour les trouver. Il y en a peut-être d'autres, mais que nous ne connaissons pas; c'est pourquoi, *pour nous*, l'induction est la méthode nécessaire pour arriver à des prédictions.

Nous voyons maintenant pourquoi une pareille relation peut être maintenue : elle est la conséquence logique de la définition du terme « possibilité de prédictions ». C'est pourquoi nous pouvons démontrer la position unique du principe inductif à l'aide de relations tautologiques seules. Quoique l'inférence inductive ne soit pas une tautologie, la démonstration montrant qu'elle nous mène à la mise la plus favorable est basée seulement sur des tautologies. La conception formelle de la logique se trouvait, à l'égard du problème de l'induction, en présence du paradoxe qu'une inférence menant à quelque chose de nouveau doit être justifiée dans une conception de la logique n'admettant que des transformations vides, à savoir tautologiques : ce paradoxe se résout par la connaissance que ce « quelque chose de nouveau » fourni par l'inférence, n'est pas considéré comme une proposition vraie, mais comme la mise la plus favorable, et que la démonstration ne se dirige pas vers la vérité de la conclusion, mais vers la relation logique entre le procédé et le but de la connaissance.

Guidé par un instinct plutôt que par un argument logique, on pourrait soulever l'objection qu'il y a dans notre théorie un concept tel que « la condition nécessaire de la connaissance », concept qui depuis la philosophie de Kant est accompagné d'un arrière-goût déplaisant. Dans notre théorie, cependant, cette qualité ne dérive pas d'une structure

---

[1] J'ai employé pour ce type de séries le nom de séries réductibles; cf. *Erkenntnis*, 6, 1936, p. 36.

apriorique de la raison; elle provient d'autres réflexions. Qui veut quelque chose, doit dire ce qu'il veut; qui veut prédire, doit dire ce qu'il entend par « prédire ». Si nous essayons de trouver une définition de ce terme qui corresponde, au moins à un certain degré, à la pratique de la langue, la définition, quelle que soit la détermination plus précise, doit entraîner le postulat de l'existence de certaines séries possédant une limite de la fréquence. C'est de cette composante de la définition que se déduit le caractère du principe de l'induction comme condition nécessaire de la possibilité de prédire. L'application du principe de l'induction ne signifie donc aucune restriction, aucune renonciation à d'autres possibilités de prédictions; elle ne signifie que l'interprétation mathématique de ce que nous entendons par « possibilité de prédictions » proprement dite.

Revenons à notre démonstration du fait que la mise selon le principe de l'induction est la plus favorable. Nous avons déjà rencontré auparavant ce concept; mais le raisonnement qui nous avait conduit à parler alors du caractère le plus favorable de la mise était différent de l'argument que nous venons d'employer. Nous avions parlé du *poids* d'une mise, déterminé par la probabilité de la proposition si on la considère comme incorporée dans une série; et la mise la plus favorable était caractérisée par le poids le plus élevé. Cependant, en ce qui concerne la mise déterminée par le principe de l'induction, nous ne connaissons pas de poids coordonné; nous la désignons comme la plus favorable pour d'autres raisons qui dérivent de l'idée qu'il s'agit ici d'un procédé approximatif.

Cette différence concernant la méthode employée pour justifier la préférence de la mise, entraîne une différence concernant l'applicabilité de la mise. Si nous employons une mise possédant un poids apprécié, nous connaissons la qualité de notre mise, c'est-à-dire nous savons à peu près les fautes qui s'y rattachent inévitablement, Si nous employons au contraire une mise justifiée comme méthode approximative, nous ne savons rien sur le degré d'approximation atteint. En misant sur la supposition que la valeur observée de la fréquence est égale à la limite, nous ne savons pas les frontières entre lesquelles cette supposition est juste, et nous ne savons pas non plus, dans le procédé de répétition de la mise avec la prolongation de la série, combien d'étapes nous séparent de l'étape à partir de laquelle la fréquence ne sort plus d'un certain intervalle donné. C'est donc une *mise aveugle* que nous faisons; la meilleure qui nous soit accessible, il est vrai, mais nous restons dans l'ignorance quant à la qualité de cette meilleure supposition.

Il existe cependant une méthode de détermination du poids d'une pareille mise, par conséquent de transformation d'une mise aveugle en une mise avec un poids apprécié. Cette méthode fait usage du schéma de la grille des probabilités déjà étudié par nous auparavant, en combinaison avec une application répétée de mises aveugles.

Imaginons une série finie de fréquences observées

$$f^1,\ f^2,\ \ldots,\ f^n.$$

Selon la règle de l'induction, nous aurons à miser sur f^n comme valeur de la limite. Naturellement nous n'insisterons pas sur la valeur précise f^n, mais nous admettrons un écart δ; nous savons que plus δ est grand, plus notre pari sera certain. Soit $f^n = q$; notre mise prend la forme suivante : la valeur de la limite est à l'intérieur de l'intervalle $q \pm \delta$. Ou, pour employer un langage abrégé : la valeur de la limite est égale à $q \pm \delta$. C'est là une mise aveugle.

Après cette détermination, nous répétons l'observation avec d'autres séries obtenues par une méthode physique semblable; nous incorporons donc la série dans une grille

$$\left\{\begin{array}{cccc} f^{11} & f^{12} & \ldots & f^{1n} \\ f^{21} & f^{22} & \ldots & f^{2n} \\ \ldots & \ldots & \ldots & \ldots \\ f^{s1} & f^{s2} & \ldots & f^{sn} \end{array}\right. \tag{87}$$

Les valeurs f^{kn}, pour diverses valeurs k, ne seront pas égales; déterminons le nombre m de celles d'entre elles qui sont situées dans l'intervalle $q \pm \delta$. Le rapport

$$\frac{m}{s} = p \tag{88}$$

détermine la probabilité de second ordre pour qu'une des séries horizontales ait une limite $q \pm \delta$; ou plus exactement, en introduisant une erreur ε arbitrairement choisie, nous faisons la mise aveugle que cette probabilité de second ordre est égale à $p \pm \varepsilon$.

La probabilité p n'est pas encore le poids u de notre mise sur la limite q, dans un cas particulier. Cette probabilité u dépend encore de la fréquence f^{kn} observée dans ce cas particulier. C'est ici qu'intervient le théorème sur l'induction, déjà mentionné. A l'aide de la règle de Bayes, on détermine u en fonction des probabilités initiales p et de la fréquence f^{kn} observée. Cette formule est basée sur certaines hypothèses concernant la structure de la suite, qui permettent d'appliquer le théorème de Bernouilli. La question de savoir si ces hypothèses sont valables suppose d'autres mises, de façon que la détermination de u renferme une structure assez compliquée de mises enchaînées; toutefois, quelle que soit la forme définitive de cette relation, nous savons que u est une fonction croissante de p.

Il nous suffit donc de considérer la mise concernant p, et d'ajouter que la détermination de p, en combinaison avec d'autres mises, signifie la transformation d'une mise aveugle en une mise possédant un poids apprécié, égal à u.

La méthode de cette transformation est typiquement approximative. Au lieu de faire *une* seule mise aveugle, nous en faisons un grand nombre dans les directions horizontales et verticales de la grille; et nous arrivons à l'aide d'une statistique à une transformation des mises aveugles des directions horizontales en des mises possédant un poids apprécié. Notre procédé peut faire, à première vue, l'impression d'une prestidigitation; cependant, il s'agit là non seulement de la méthode utilisée actuellement par toutes les sciences qui emploient des statistiques, mais aussi d'un procédé auquel on peut donner une interprétation et une justification mathématiques. On peut démontrer les théorèmes suivants, valables si les séries en question ont des limites :

1. Si s est maintenu constant, il existe un $n = n_1$, tel que les mises primaires, concernant les suites horizontales, deviennent justes, et restent justes pour tous les n plus grands que n_1.

2. Si s est maintenu constant, il existe un $n = n_2$, tel que les mises secondaires deviennent justes, pour l'ensemble de s séries, et restent justes pour tous les n supérieurs à n_2.

A ce théorème s'ajoute l'inégalité

$$n_2 \leqq n_1. \tag{89}$$

Elle signifie, intuitivement parlant : les mises de second ordre peuvent être justes si les mises du premier ordre ne sont pas encore toutes justes. La raison de cette relation se trouve dans le fait qu'une petite erreur dans la statistique verticale ne change pas beaucoup le rapport $\frac{m}{s}$; les mises du second ordre sont, jusqu'à un certain degré, indépendantes des mises du premier ordre.

On peut démontrer deux autres théorèmes :

3. Il existe un $s = s_3$ et un $n = n_3$ tels que les mises du second ordre soient toutes justes, c'est-à-dire ne diffèrent pas de plus d'un écart ε donné de leurs valeurs pour $n = \infty$ et $s = \infty$.

Pour démontrer ce théorème, supposons que, dans un dénombrement s'effectuant dans le sens vertical et ayant pour éléments des suites horizontales, il existe une limite u' de la fréquence. Cela revient à dire qu'il existe un $s = s_3$, tel que la fréquence $\frac{m'}{s}$, définie de façon analogue à $\frac{m}{s}$ mais pour $n = \infty$, coïncide à une valeur près de l'intervalle $\pm \varepsilon$, avec la limite p' de cette même fréquence pour $s = \infty$. Il est donc bien nécessaire, en vertu du théorème 2, qu'à ce $s = s_3$ corresponde un $n = n_3$, tel que $\frac{m}{s}$ coïncide aussi à une valeur près de l'intervalle $\pm \varepsilon$, avec la limite p'.

4. En général il n'existe aucun $n = n_4$, tel que toutes les mises primaires soient justes, pour tous les s; mais s'il en existe, c'est-à-dire s'il s'agit du cas particulier de la convergence uniforme, nous aurons l'inégalité

$$n_3 \leqq n_4. \tag{90}$$

Ces quatre théorèmes mettent en évidence une différence profonde entre les mises primaires et secondaires, qui peut être interprétée comme une *supériorité* des mises secondaires sur les mises primaires. Nous résumons cette différence dans les propositions suivantes :

α. Les mises secondaires sont indépendantes des mises primaires individuelles; l'une des mises primaires peut être fausse sans que cela entraîne la fausseté des mises secondaires.

β. Si dans toutes les séries considérées les fréquences ont une limite, la valeur juste de toutes les mises secondaires peut être trouvée en un nombre fini d'étapes. On ne peut démontrer cependant un théorème analogue pour les mises primaires.

C'est là l'interprétation mathématique du passage des mises aveugles à des mises possédant un poids apprécié. Quoique ce passage introduise de nouvelles mises aveugles, sur un niveau logique plus élevé, il signifie un progrès logique : il implique une amélioration de la convergence.

L'application du procédé décrit consiste en ce que nous *corrigeons* les mises primaires après la détermination de leur poids. Si la probabilité du second ordre est petite, nous ne suivons plus la règle de l'induction par rapport à la mise primaire; nous préférerons une autre mise dont le poids est plus grand, déterminé par la méthode de la grille. Ainsi nous arrivons à remplacer la règle de l'induction par une autre; mais l'inférence qui mène à cette autre prescription est, de nouveau, une inférence inductive. L'application répétée du principe de l'induction peut nous conduire à faire tomber en désuétude ce principe même.

Il n'y a rien de contradictoire dans ce procédé. Son interprétation mathématique consiste en ce que, au lieu de miser sur f^n, nous misons sur

$$f^n + \sigma(f^n). \tag{91}$$

Si $\sigma(f^n)$ est une fonction qui tend vers zéro, quand n croît vers l'infini, la mise corrigée convergera asymptotiquement vers la mise originale. La méthode de la correction détermine la fonction σ, de façon que la valeur $f^n + \sigma(f^n)$ soit approximativement égale à la valeur p de la limite, dans des parties de la série où f^n est encore éloignée de cette valeur. Nous avions parlé plus haut d'un clairvoyant qui pourrait prédire la valeur de la limite de la fréquence; en fait, il ne ferait pas autre chose que donner une fonction σ de la forme décrite. Cependant nous ne savons pas d'avance si le σ donné par le clairvoyant ne diminue pas la valeur de la mise, et s'il remplit la condition de la convergence vers zéro; c'est pourquoi nous devons le contrôler d'abord à l'aide d'inductions.

Nous n'avons pas besoin, cependant, d'attendre un clairvoyant pour la détermination de fonctions pareilles. Les méthodes de la science peuvent être interprétées comme des procédés qui, à l'aide d'inductions enchaînées et transversales, produisent une convergence améliorée de nos prédictions : le schéma discuté de la grille est le schéma des méthodes scientifiques. Démontrons-le par un exemple.

Les physiciens ont essayé depuis longtemps de liquéfier le charbon; ils n'y ont pas réussi. Ils ont exposé le charbon à des températures de plus en plus élevées, mais le charbon ne fondait pas; la règle de l'induction fournirait donc la mise que le charbon ne fondra à aucune température. Les physiciens ont refusé cependant de faire une telle inférence, à cause d'expériences avec d'autres substances qui se sont comportées de façon semblable mais qui, pourtant, se sont finalement liquéfiées. Le schéma de cette inférence serait une grille; si nous désignons par B l'événement de la liquéfaction, le schéma des événements ordonnés par températures croissantes serait :

Charbon..........	$\bar{B}$	$\bar{B}$	$\bar{B}$	$\bar{B}$	$\bar{B}$	$\bar{B}$	...
Cuivre............	$\bar{B}$	$\bar{B}$	$\bar{B}$	$\bar{B}$	B	B	...
Fer...............	$\bar{B}$	$\bar{B}$	$\bar{B}$	$\bar{B}$	$\bar{B}$	B	...
....	..	..	..	..	..	..	...

La mise primaire dans la série du charbon, serait que B n'arrivera jamais. La mise secondaire nous montre que cette mise aveugle possède un poids très petit; nous la corrigeons donc et préférons miser, ici aussi, sur l'arrivée de B avec une température plus élevée encore.

Cette méthode de correction est la méthode de la science. Notre exemple concerne un cas particulier où la fréquence est égale à 1 ou 0; le cas plus général discuté dans le calcul des probabilités arrive également. Je ne peux pas entrer ici dans une analyse de la méthode scientifique, et démontrer son caractère probabilitaire par d'autres exemples (1); il nous suffira d'avoir étudié le schéma logique de ces opérations. Ce schéma logique est donné dans la logique probabilitaire.

Les propositions de la science sont donc des mises; la plupart d'entre elles sont des mises possédant un poids apprécié, mais les mises du dernier niveau sont des mises aveugles. Nous pouvons en déterminer le poids, mais la méthode de cette détermination introduit de nouveau des mises aveugles, sur un niveau plus élevé. La science considérée comme un tout est donc basée, en dernière instance, sur des mises aveugles.

---

(1) Je me refère pour cette analyse à mon livre *Experience and Prediction*, Chicago University Press 1938.

Nous n'avons donc aucune garantie pour le succès de cette méthode; si l'incertitude de l'induction individuelle est éliminée par l'appréciation d'un poids, elle est en même temps transférée au système tout entier, de telle façon que le caractère de pari reparaît pour la science considérée comme un tout. L'incertitude du succès ne doit cependant pas nous retenir de faire, au moins, *la tentative* de prédire, en choisissant les meilleures méthodes accessibles. Ici se répète la discussion faite à propos de la règle de l'induction. Le principe de l'induction, nous le savons, fournit la mise la plus favorable parce qu'il détermine la seule mise dont *nous sachions* qu'elle mène au succès, si ce succès est possible. Quant au système des inductions enchaînées, nous en savons davantage : nous savons qu'il est supérieur à l'induction isolée. L'ensemble du système nous conduira au succès plus vite qu'une induction singulière; et il peut nous apporter le succès même si l'induction singulière échoue. Cette différence logique, la supériorité du filet des inductions enchaînées, peut être démontrée par des considérations purement mathématiques, à savoir à l'aide des seules tautologies; notre préférence du système des inductions peut donc être justifiée sans faire appel à des hypothèses concernant la nature du monde. Il est bien remarquable qu'une telle démonstration puisse être donnée; quoique nous ne sachions pas si nos méthodes de prédiction auront du succès, nous pouvons établir un ordre entre elles et préciser que la meilleure est le système des inductions enchaînées. La justification de l'induction singulière se répète donc pour le système considéré comme un tout : le système des inductions enchaînées est la meilleure mise que nous connaissions en ce qui concerne l'avenir.

Voilà le rôle dominant dévolu au calcul des probabilités dans la construction de la science. D'une statistique des jeux de hasard il s'est développé, en passant par les étapes des statistiques sociales et physiques, jusqu'à un stade dans lequel il finit par englober toutes les méthodes inductives de la science; il est devenu de nos jours la logique de la science. Je ne sais si, peut être, la physique des quanta nous réserve encore des applications plus profondes du calcul des probabilités; mais la physique classique exige déjà, pour une compréhension et une justification de la méthode inductive, une interprétation dans le cadre de la logique des probabilités.

Les considérations précédentes nous ont éloignés des problèmes purement mathématiques; qu'il me soit permis d'indiquer en quelques mots les raisons qui m'ont déterminé à présenter ici mes idées sur les fondements logiques du calcul des probabilités.

La valeur de ces idées pour le mathématicien, me semble-t-il, consiste en ce qu'elles dessinent un cadre général englobant toutes les constructions mathématiques du calcul des probabilités. La solution des problèmes particuliers ne suppose pas, il est vrai, la connaissance de ce cadre logique; et les rendements admirables de la théorie mathématique des probabilités montrent que le mathématicien n'a pas besoin du logicien tant qu'il s'agit du développement des relations internes, de l'enchaînement judicieux d'inférences déductives conduisant à des formules qui embrassent une multitude de relations dans une expression simple. Mais ce que notre construction logique donne au mathématicien, c'est, me semble-t-il, une sorte de système de référence dans lequel chaque théorème mathématique trouve sa place indiquée par des coordonnées logiques, une sorte d'édifice logique renfermant des casiers pour chaque théorème spécial, ordonnés systématiquement. En particulier, le problème de l'irrégularité des suites trouve ici sa place logique comme cas spécial dans une théorie plus générale des séries embrassant tous les degrés d'ordre, à partir d'une régularité stricte jusqu'au cas extrême de l'irrégularité des jeux de hasard.

J'ai mis au premier plan de mes recherches, le problème épistémologique de la probabilité. La valeur de la construction présentée du système mathématique des probabilités me semble ressortir du fait qu'elle achemine vers une solution le problème épistémologique de ce concept fort discuté. Quoique cette solution nous montre que l'idée d'une connaissance vraie de la nature n'est qu'un fantôme, un rêve et un désir plutôt qu'une théorie sérieuse, elle nous conduit à une justification de l'usage des méthodes de la probabilité dans toutes leurs applications. Les concepts qui se présentent dans cette théorie de l'application ne sont pas moins « clairs et distincts » que ceux de la théorie mathématique; tout en n'arrivant pas à nous présenter l'énoncé probabilitaire comme *vrai*, ils nous démontrent que *miser selon les règles de la probabilité* est l'action la plus raisonnable qui nous reste en présence de l'inconnu de l'avenir. Cette démonstration se fait à l'aide de déductions strictes. S'il demeure un reste d'incertitude, d'indétermination dans les énoncés probabilitaires appliqués, ce ne sont pas les concepts qui manquent de détermination, mais leurs objets, les choses physiques, qui ne nous révèlent pas le secret de leur ordre avant leur entrée dans la sphère du présent.

Le droit du mathématicien à créer des constructions dans son monde d'idées est incontestable; il me semble néanmoins que la solution précédente du problème épistémologique des probabilités peut lui paraître valable. Si grande que soit la valeur de l'harmonie et de l'élégance des constructions mathématiques, il lui sera une satisfaction plus profonde encore de savoir que ses idées sont applicables à la réalité physique — et de comprendre pourquoi.

(D'après des conférences
données à l'Institut Henri Poincaré
en mai-juin 1937.)

FIN DU VOLUME VII.