ANNALES DE L'I. H. P.

BRUNO DE FINETTI

La prévision : ses lois logiques, ses sources subjectives

Annales de l'I. H. P., tome 7, n° 1 (1937), p. 1-68.
<http://www.numdam.org/item?id=AIHP_1937__7_1_1_0>

La prévision : ses lois logiques, ses sources subjectives

par

Bruno de FINETTI.

AVANT-PROPOS.

Dans les conférences que j'ai eu l'honneur de faire à l'Institut Henri Poincaré les 2, 3, 8, 9 et 10 mai 1935, et dont les pages qui suivent reproduisent le texte, je me suis proposé de donner une vue d'ensemble sur deux sujets dont je me suis particulièrement occupé, et d'éclaircir la liaison étroite qui les unit. Il s'agit, d'une part, de la définition de la probabilité (que je considère comme une entité purement subjective) et de la signification de ses lois, et, d'autre part, des notions et de la théorie des événements et nombres aléatoires « équivalents »; le lien entre les deux questions réside en ce que cette dernière théorie constitue la solution, pour le cas le plus typique, du problème du raisonnement inductif selon la conception subjective de la probabilité (et éclaircit ainsi, en général, la façon dont le problème de l'induction est posé). D'ailleurs, même s'il n'en était pas ainsi, autrement dit même si l'on n'acceptait pas le point de vue subjectif que nous avons adopté dans la première partie, la seconde n'en conserverait pas moins sa validité et constituerait un chapitre intéressant du calcul des probabilités.

L'exposé est divisé en six chapitres, dont les deux premiers ont pour objet la première question, les deux suivants la seconde, et les deux derniers l'examen des conclusions qu'on peut en tirer. La plupart des questions traitées ont été développées, tantôt en détail, tantôt succinctement, mais toujours d'une façon fragmentaire [1], dans mes

[1] Un exposé plus complet de mon point de vue, sous forme d'essai purement critique et philosophique, sans formules, constitue l'ouvrage [32].

travaux antérieurs; parmi eux, ceux qui ont trait aux questions étudiées ou effleurées dans ces conférences se trouvent indiqués dans la liste bibliographique ([1]).

Pour de plus amples détails concernant la matière de chacun des chapitres, je renvoie aux publications suivantes :

Chapitre I. — *La logique du probable :* [26], [34].
Chapitre II. — *L'évaluation d'une probabilité :* [49], [65], [70].
Chapitre III. — *Événements équivalents :* [29], [40].
Chapitre IV. — *Nombres aléatoires équivalents :* [46], [47], [48].
Chap. V. — *Considérations sur la notion d'équivalence :* [51], [62].
Chapitre VI. — *Observation et prévision :* [32], [56], [62].

Chacun de ces chapitres constitue une des cinq conférences ([2]), à l'exception des Chapitres IV et V qui correspondent à la quatrième, et dont le texte a été amplifié pour mieux éclaircir la notion utilisée d'intégration dans l'espace fonctionnel. Le texte des autres conférences n'a pas subi de modifications essentielles, à part plusieurs améliorations, entre autres au début du Chapitre III, où, pour plus de clarté, l'exposition a été complètement remaniée. Pour ces retouches j'ai profité des précieux conseils de MM. Fréchet et Darmois, qui ont bien voulu assister à ces conférences, et de M. Castelnuovo, qui a lu plusieurs fois le manuscrit et les modifications successives; la rédaction du texte a été revue par mon collègue M. V. Carmona et par M. Al. Proca, qui m'ont suggéré un certain nombre d'heureuses variantes de forme. Pour leur aimable aide j'ai le devoir de leur exprimer ici ma sincère reconnaissance. Enfin, je ne veux pas terminer ces lignes sans renouveler mes remerciements à M. le Directeur et aux membres du Comité de Direction de l'Institut Henri Poincaré pour le grand honneur qu'ils m'ont fait en m'invitant à donner ces conférences à Paris.

Trieste, le 19 décembre 1936, XV.

([1]) Voir page 66; les nombres en chiffres *gras* renvoient toujours à cette liste (chiffres romains pour les mémoires d'autres Auteurs; chiffres arabes pour ceux de l'auteur, rangés par ordre chronologique général).

([2]) Leurs titres étaient ceux des six chapitres, à l'exclusion du Chapitre V.

INTRODUCTION.

Henri Poincaré, le savant immortel auquel cet Institut est dédié, et qui a vivifié avec ses idées géniales tant de branches des mathématiques, est sans doute aussi le penseur qui a attribué à la théorie des probabilités le domaine d'application le plus vaste et un rôle tout à fait essentiel dans la philosophie scientifique. « Les faits prévus, dit-il, ne peuvent être que probables. Si solidement assise que puisse nous paraître une prévision, nous ne sommes jamais absolument sûrs que l'expérience ne la démentira pas ». Le calcul des probabilités repose sur « un instinct obscur, dont nous ne pouvons nous passer; sans lui la Science serait impossible, sans lui nous ne pourrions ni découvrir une loi, ni l'appliquer ». « A ce compte, toutes les sciences ne seraient que des applications inconscientes du calcul des probabilités; condamner ce calcul, ce serait condamner la Science tout entière » (1).

Ainsi, les questions de principe relatives à la signification et à la valeur de la probabilité, cessent d'être isolées dans une branche particulière des mathématiques, et acquièrent l'importance de problèmes fondamentaux du point de vue gnoséologique.

De pareilles questions admettent évidemment autant de réponses différentes qu'il y a d'attitudes philosophiques possibles; en donner une, ne signifie pas dire quelque chose qui puisse convaincre et satisfaire tout le monde, mais la connaissance d'un point de vue donné peut néanmoins être intéressante et utile même à ceux qui ne sauraient le partager. Le point de vue que j'ai l'honneur d'exposer ici peut être considéré comme la solution extrême du côté du subjectivisme; le lien unissant les diverses recherches que je me propose de résumer est constitué en effet par le principal but commun qu'elles poursuivent, à côté d'autres objectifs plus immédiats et plus concrets; ce but est celui de ramener dans le cadre de la conception subjective et d'expliquer *même* les questions qui semblent la démentir, et qui sont couramment invoquées contre elle. La première conférence aura pour objet de montrer

(1) [XXVIII], p. 213, 216, 217.

comment les lois logiques de la théorie des probabilités peuvent être rigoureusement établies en se plaçant au point de vue subjectif; dans les autres on verra comment, tout en refusant d'admettre l'existence d'une signification et d'une valeur objective, on peut se rendre compte des raisons, elles-mêmes subjectives, pour lesquelles dans une foule de problèmes les jugements subjectifs de divers individus normaux ne diffèrent pas essentiellement les uns des autres, ou même coïncident exactement. Les cas les plus simples feront l'objet de la deuxième conférence; les conférences suivantes seront consacrées à la question la plus délicate de cette étude, savoir à l'explication subjective de l'usage que nous faisons dans nos prévisions futures des résultats de l'observation, de l'expérience du passé.

Ce point de vue n'est qu'un des points de vue possibles, mais je ne serais pas tout à fait sincère si je n'ajoutais qu'il est le seul qui ne soit pas en conflit avec les exigences logiques de mon esprit. Si je ne veux pas en conclure qu'il est « juste », c'est que je sais bien que, aussi paradoxal que cela puisse paraître, rien n'est plus subjectif et personnel que cet « instinct de ce qui est logique » chez les divers mathématiciens, lorsqu'il s'agit de l'appliquer à des questions de principe.

CHAPITRE I.

LA LOGIQUE DU PROBABLE.

Considérons la notion de probabilité telle qu'elle est conçue par nous tous dans la vie quotidienne. Considérons un événement bien déterminé, et supposons que nous ne sachions pas à l'avance s'il se produira ou non; le doute dans lequel nous sommes au sujet de son apparition est susceptible d'une comparaison, et, par conséquent, d'une graduation. Si l'on admet uniquement : 1° qu'un événement incertain ne peut nous paraître que : (a) aussi probable, (b) plus probable et (c) moins probable qu'un autre; 2° qu'un événement incertain nous semble toujours plus probable qu'un événement impossible et moins probable qu'un événement certain; et qu'enfin, 3° un événement E' ne peut qu'apparaître plus probable qu'un autre E'', lorsqu'on juge E' plus probable qu'un troisième E estimé lui-même plus probable que E'' (propriété transitive), il suffira d'ajouter à ces trois axiomes évidemment banals, un quatrième, lui-même de nature purement qualitative, pour pouvoir construire rigoureusement toute la théorie des probabilités. Ce quatrième axiome nous dit que les inégalités se conservent dans la somme logique : si E est un événement incompatible avec E_1 et avec E_2, alors $E_1 + E$ sera plus ou moins, ou également probable que $E_2 + E$ selon que E_1 est plus ou moins, ou également probable que E_2; plus généralement, on en déduit alors (¹) que deux inégalités, comme

E_1 est plus probable que E_2,
E'_1 est plus probable que E'_2,

peuvent être ajoutées en donnant

$E_1 + E'_1$ est plus probable que $E_2 + E'_2$,

pourvu que les événements à ajouter soient incompatibles entre eux (E_1 avec E'_1, E_2 avec E'_2). On peut alors démontrer que, lorsqu'il s'agit d'événements pour lesquels on connaît une subdivision en cas possibles que nous jugeons également probables, la comparaison entre leurs probabilités peut être ramenée à la comparaison purement arithmétique des rapports entre les nombres des cas favorables et des cas possibles (non pas que le jugement ait alors une valeur objective, mais parce que tout ce qu'il y a de substantiel et donc de subjectif, est déjà inclus dans le jugement que les cas constituant la subdivision sont également probables). Ce rapport peut être alors choisi comme indice approprié pour mesurer une probabilité, et appliqué en général, même en dehors des cas où l'on peut effectivement employer le critère qui nous y conduit. Dans les autres cas, on pourra évaluer cet indice par comparaison : ce sera en fait un nombre, univoquement déterminé, tel qu'à des valeurs plus grandes ou plus petites que celle de ce nombre lui-même, correspondront des événements respectivement plus probables ou moins probables que l'événement considéré. On arrive ainsi, tout en partant d'un système d'axiomes purement qualitatifs, à une mesure quantitative de la probabilité, et ensuite au théorème des probabilités totales, ce qui permet de construire tout le calcul des probabilités [pour

---

(¹) *Voir* [34], p. 321, note (¹).

ce qui est des probabilités subordonnées, il faut cependant introduire un cinquième axiome : *voir* note (1), p. 14].

On peut cependant donner aussi une définition quantitative, numérique, directe du degré de probabilité attribué par un individu donné à un événement déterminé, de telle façon que toute la théorie des probabilités puisse se déduire immédiatement d'une condition très naturelle ayant une signification évidente. Il s'agit simplement de préciser mathématiquement l'idée banale et évidente que le degré de probabilité attribué par un individu à un événement donné est révélé par les conditions dans lesquelles il serait disposé de parier sur cet événement (1). Le procédé axiomatique dont nous venons d'indiquer plus haut les lignes générales présente l'avantage de permettre une analyse plus poussée et plus détaillée des concepts fondamentaux, de partir uniquement de notions qualitatives, d'éliminer la notion de « monnaie », étrangère à la question, mais qui est nécessaire pour parler de paris; cependant, une fois démontré que l'on peut écarter la défiance que fait naître le caractère un peu trop concret et peut-être artificiel de la définition fondée sur les paris, le second procédé s'avère préférable, surtout pour sa clarté.

Supposons qu'un individu soit obligé d'évaluer le prix p pour lequel il serait disposé d'échanger la possession d'une somme quelconque S (positive ou négative) subordonnée à l'arrivée d'un événement donné, E, avec la possession de la somme pS; nous dirons par définition que ce nombre p est la mesure du degré de probabilité attribué par l'individu considéré à l'événement E, ou, plus simplement, que p est la probabilité de E (selon l'individu considéré; cette précision pourra d'ailleurs être sous-entendue s'il n'y a pas d'ambiguïté).

Spécifions de plus que, dans la terminologie que je crois convenable de suivre, un événement est toujours un fait singulier; si l'on a à considérer plusieurs épreuves, nous ne dirons jamais « épreuves d'un même événement », mais « épreuves d'un même phénomène », et chaque

(1) Déjà Bertrand ([1], p. 24), en partant de cette observation, avait donné quelques exemples de probabilités subjectives, mais uniquement dans le but de leur opposer d'autres « probabilités objectives ». La théorie subjective a été développée d'après le schéma des paris comme dans cet exposé (Chap. I et II) dans mon premier Mémoire de 1928 sur ce sujet, non publié sous sa forme initiale, mais résumé ou développé partiellement dans [27], [34], [35], etc.

« épreuve » sera un « événement ». L'essentiel n'est pas, évidemment, le choix des termes : il s'agit de préciser que, selon nous, on n'a pas le droit de parler de la « probabilité d'un événement » si l'on entend par « événement » ce que nous appelons « phénomène » ; on ne peut le faire que s'il s'agit d'une « épreuve » bien déterminée (1).

Ceci posé, lorsqu'un individu a évalué les probabilités de certains événements, deux cas peuvent se présenter : ou bien il est possible de parier avec lui en s'assurant de gagner à coup sûr, ou bien cette possibilité n'existe pas. Dans le premier cas on doit dire évidemment que l'évaluation de la probabilité donnée par cet individu contient une incohérence, une contradiction intrinsèque ; dans l'autre cas, nous dirons que l'individu est cohérent. C'est précisément cette condition de *cohérence* qui constitue le seul principe d'où l'on puisse déduire tout le calcul des probabilités : ce calcul apparaît alors comme l'ensemble des règles auxquelles l'évaluation subjective des probabilités de divers événements par un même individu doive être assujettie si l'on ne veut pas qu'il y ait entre elles une contradiction fondamentale.

Voyons comment on démontre, par cette voie, le théorème des probabilités totales : c'est un résultat important par lui-même et aussi pour éclaircir le point de vue suivi. Soient $E_1, E_2, \ldots, E_n$ des événements incompatibles dont un (et un seul) doive se vérifier (nous dirons : une classe *complète* d'événements incompatibles), et soient $p_1, p_2, \ldots, p_n$ leurs probabilités évaluées par un individu donné ; si l'on fixe les enjeux (positifs ou négatifs) $S_1, S_2, \ldots, S_n$, les gains dans les n cas possibles seront les différences entre l'enjeu du pari gagné et la somme des n mises payées

$$G_h = S_h - \sum_1^n {}_i \, p_i S_i.$$

En considérant les S_h comme inconnues, on a un système d'équations linéaires de déterminant

$$\begin{vmatrix} 1-p_1 & -p_2 & \ldots & -p_n \\ -p_1 & 1-p_2 & \ldots & -p_n \\ \ldots & \ldots & \ldots & \ldots \\ -p_1 & -p_2 & \ldots & 1-p_n \end{vmatrix} = 1-(p_1+p_2+\ldots+p_n);$$

(1) Ce même point de vue a été soutenu par von Kries [XIX]; *voir* [65], [70], et pour le point de vue contraire [XXV].

si ce déterminant n'est pas nul, on peut fixer le S_h de façon que les G_h aient des valeurs arbitraires, en particulier toutes positives, contrairement à la condition de cohérence; par conséquent, la cohérence oblige à imposer la condition $p_1+p_2+\ldots+p_n=1$. Cette condition nécessaire pour réaliser la cohérence est aussi suffisante, parce que, si elle est satisfaite, on a identiquement (quels que soient les enjeux S_h)

$$\sum_1^n {}_h p_h G_h = 0,$$

et les G_h ne peuvent jamais, par conséquent, être tous positifs.

On a ainsi le théorème des probabilités totales sous la forme suivante : *dans une classe complète d'événements incompatibles, la somme des probabilités doit être égale à* 1. La forme plus générale : *la probabilité de la somme logique de n événements incompatibles est la somme de leurs probabilités*, n'en est que le corollaire immédiat.

Cependant, avons-nous ajouté, la condition est aussi suffisante; il est utile d'éclaircir quelque peu le sens de cette affirmation, car, à travers un cas concret, on pourra faire clairement ressortir la distinction, fondamentale pour ce point de vue, entre la logique du probable et les jugements de probabilité. En disant que la condition est suffisante, on entend que, une classe complète d'événements incompatibles $E_1, E_2, \ldots, E_n$ étant donnée, toutes les évaluations de probabilité qui attribuent à $p_1, p_2, \ldots, p_n$ des valeurs quelconques non négatives et de somme égale à l'unité sont des évaluations admissibles : chacune de ces évaluations correspond à une opinion cohérente, à une opinion légitime en soi, et chaque individu est libre d'adopter celle de ces opinions qu'il préfère, ou, pour mieux dire, celle qu'il *sent*. Le meilleur exemple est celui d'un championnat où le spectateur attribue à chacune des équipes une probabilité de vaincre plus ou moins grande d'après son propre jugement; la théorie ne peut repousser à priori aucun de ces jugements, sauf si la somme des probabilités attribuées à chaque équipe n'est pas égale à l'unité. Cet arbitraire que quiconque admettrait dans le cas cité, existe même, d'après la conception que nous avons soutenue, dans tous les autres domaines, y compris ceux, déterminés plus ou moins vaguement, dans lesquels les diverses conceptions objectives affirment leur validité.

A cause de cet arbitraire, l'objet du calcul des probabilités n'est donc plus une seule fonction $\mathbf{P}(E)$ des événements E, c'est-à-dire leur probabilité considérée comme quelque chose d'objectivement déterminé, mais l'ensemble de toutes les fonctions $\mathbf{P}(E)$ correspondant à des opinions admissibles. Et lorsqu'on demande de *calculer* la probabilité $\mathbf{P}(E)$ d'un événement E, l'énoncé du problème est toujours à préciser en ce sens : calculer la valeur qu'on est obligé d'attribuer à l'événement E, si l'on veut rester dans le domaine de la cohérence, après avoir évalué d'une façon déterminée les probabilités des événements constituant une certaine classe $\mathcal{E}$. Mathématiquement, on suppose la fonction $\mathbf{P}$ connue sur l'ensemble $\mathcal{E}$, et l'on demande quelle est la valeur unique, ou quelles sont les valeurs que l'on peut attribuer à $\mathbf{P}(E)$ sans que ce prolongement de $\mathbf{P}$ fasse apparaître une incohérence.

Il est intéressant de poser en général la question suivante : quels sont les événements E pour lesquels la probabilité est déterminée par la connaissance des probabilités attribuées aux événements d'une certaine classe $\mathcal{E}$? On est amené ainsi à introduire la notion (que je crois nouvelle) « d'événements linéairement indépendants » [26]. Soient $E_1, E_2, \ldots, E_n$ les événements de $\mathcal{E}$; parmi ces n événements, certains pourront se produire, d'autres non; les sous-classes d'une classe de n éléments étant au nombre de 2^n (la classe totale et la classe vide incluses), il y aura, au plus, 2^n cas possibles $C_1, C_2, \ldots, C_s (s \leqq 2^n)$ qu'on appelle, d'après Boole, les « constituants ». (« au plus », car un certain nombre de combinaisons pourraient s'avérer impossibles) (¹). Formellement, les C_h sont les événements qu'on obtient à partir du produit logique $E_1 . E_2 . \ldots, E_n$ en y remplaçant un groupe quelconque des E_i par les événements contraires (négations) $\sim E_i$ (ou, avec une notation abrégée, $\overline{E}_i$). Les constituants forment une classe complète d'événements incompatibles; les E_i sont des sommes de constituants, et les événements qui sont sommes de constituants sont les seuls événements logiquement dépendants des E_i, c'est-à-dire tels qu'on peut toujours dire s'ils sont vrais ou faux lorsque l'on sait, pour chaque événement $E_1, \ldots, E_n$, s'il est vrai ou faux.

Donner la probabilité d'un événement E signifie donner la somme

---

(¹) Ces notions se trouvent appliquées au calcul des probabilités dans *Medolaghi* [XXIV].

des probabilités de ses constituants

$$c_{i_1} + c_{i_2} + \ldots + c_{i_h} = p_i;$$

les probabilités de $E_1, \ldots, E_n$ étant fixées, on a n équations de ce type, lesquelles forment, avec l'équation $c_1 + c_2 + \ldots + c_s = 1$, un système de $n+1$ équations linéaires entre les probabilités c_h des constituants. On voit que, E étant un événement logiquement dépendant de $E_1, \ldots, E_n$, donc une somme de constituants $E = C_{h_1} + C_{h_2} + \ldots + C_{h_k}$, sa probabilité

$$p = c_{h_1} + c_{h_2} + \ldots + c_{h_k}$$

est univoquement déterminée lorsque cette équation est linéairement dépendante du système précédent; ce fait, observons-le, ne dépend pas de la fonction $\mathbf{P}$, mais seulement de la classe $\mathcal{E}$ et de l'événement E et peut donc être exprimé en disant que E est *linéairement dépendant* de $\mathcal{E}$, ou — ce qui revient au même si les $\mathcal{E}$ sont linéairement indépendants — que $E_1, E_2, \ldots, E_n$, E sont reliés linéairement entre eux.

La notion de « dépendance linéaire » ainsi définie pour des événements est parfaitement analogue à la notion géométrique bien connue, et jouit des mêmes propriétés; au lieu de démontrer directement ce fait, on peut le rendre tout de suite évident en introduisant une représentation géométrique qui fait correspondre à chaque événement un point, et, à la notion de « dépendance linéaire » logique, la notion de dépendance linéaire géométrique. Il s'agit de la représentation suivante : les constituants C_h sont représentés par les sommets A_h d'un simplexe dans l'espace à $(s-1)$ dimensions, l'événement somme de k constituants par le centre de gravité des k sommets correspondants affecté d'une masse k, et donc l'événement certain (somme de tous les s constituants) par le centre O du simplexe, affecté d'une masse s.

Cette représentation géométrique permet de caractériser d'une façon qui fait image l'ensemble de toutes les évaluations de probabilité possibles. On a vu qu'une fonction de probabilité $\mathbf{P}(E)$ est complètement déterminée lorsqu'on en donne les valeurs relatives aux constituants, $c_1 = \mathbf{P}(C_1)$, $c_2 = \mathbf{P}(C_2), \ldots, c_s = \mathbf{P}(C_s)$, valeurs qui doivent être non négatives et de somme égale à l'unité. Considérons maintenant la fonction linéaire f qui prend, sur les sommets A_h, les valeurs c_h; au point A, centre de gravité de $A_{h_1}, A_{h_2}, \ldots, A_{h_k}$, elle prendra évidemment la valeur $f(\mathrm{A}) = \frac{1}{k}(c_{h_1} + c_{h_2} + \ldots + c_{h_k})$, tandis que la probabilité $\mathbf{P}(\mathrm{E})$ de l'événement E, somme logique des constituants C_{h_1}, $\mathrm{C}_{h_2}, \ldots, \mathrm{C}_{h_k}$, sera $c_{h_1} + c_{h_2} + c_{h_k}$. On a donc, en général, $\mathbf{P}(\mathrm{E}) = k.f(\mathrm{A})$: la probabilité d'un événement E est la valeur de f dans son point représentatif A multipliée par la masse k, produit que l'on peut définir comme valeur de f pour le point A doué d'une masse k, en écrivant $\mathbf{P}(\mathrm{E}) = f(k.\mathrm{A})$. Le centre O correspondant à l'événement certain, on a en particulier $1 = f(s.\mathrm{O}) = s.f(\mathrm{O})$, c'est-à-dire $f(\mathrm{O}) = \frac{1}{s}$.

On voit ainsi tout de suite que les lois de probabilité possibles correspondent à toutes les fonctions linéaires de l'espace, qui sont non négatives sur le simplexe et ont la valeur $\frac{1}{s}$ à l'origine; une telle fonction f étant caractérisée par l'hyperplan $f = 0$, les lois de probabilité correspondent biunivoquement aux hyperplans qui ne coupent pas le simplexe. On peut noter que la probabilité $\mathbf{P}(\mathrm{E}) = f(k.\mathrm{A})$ est le moment du point doué de masse $k\mathrm{A}$ (distance $\times$ masse) par rapport à l'hyperplan $f = 0$ (en prenant comme unité le moment de $s\mathrm{O}$); si, en particulier, les s constituants sont également probables, l'hyperplan va à l'infini.

En donnant la valeur qu'elle prend sur un certain groupe de points, une fonction linéaire f est définie pour tous les points linéairement dépendants, mais elle reste indéterminée pour les points linéairement indépendants : le rapprochement avec la définition donnée des événements linéairement dépendants montre donc bien, comme nous l'avions annoncé, que dépendance et indépendance linéaire des événements signifie dépendance et indépendance des points correspondants dans la représentation géométrique donnée. On peut en déduire, d'une façon plus intuitive que par la voie directe, les deux critères suivants qui caractérisent la dépendance linéaire des événements. Dans le système de coordonnées barycentriques où $x_i = 1$, $x_j = 0\,(i \neq j)$ représente le point A_i, les coordonnées du centre de gravité de $\mathrm{A}_{h_1}, \mathrm{A}_{h_2}, \ldots, \mathrm{A}_{h_k}$ affecté d'une masse k seront

$$x_{h_1} = x_{h_2} = \ldots = x_{h_k} = 1, \qquad x_j = 0 \qquad (j \neq h_1, h_2, \ldots, h_k);$$

les sommes de constituants peuvent être représentées ainsi par un symbole de s chiffres, 1 ou 0 (par exemple la somme $\mathrm{C}_1 + \mathrm{C}_3$ par $10100\ldots0$). Des événements sont linéairement dépendants lorsque la matrice des coordonnées de leurs points représentatifs et du centre O est nulle, les lignes de cette matrice étant les symboles précédents relatifs aux événements considérés, et, — pour la dernière ligne qui est exclusivement constituée par des 1, — à l'événement certain. L'autre condition dit que des événements sont linéairement dépendants lorsque l'on peut attribuer à chacun d'eux un coefficient, de telle façon que, dans tous les cas possibles, la somme des coefficients des événements vérifiés ait toujours la même valeur. Si, en effet, les points correspondants aux événements donnés et le point O sont linéairement dépendants, il est possible d'exprimer O linéairement au moyen des autres, et cela signifie qu'il existe une combinaison de paris sur ces événements équivalente à un pari sur un événement certain, comme l'exprime la condition énoncée.

Une évaluation de probabilité peut être représentée non seulement par l'hyperplan $f = 0$ mais aussi par un point non extérieur au simplexe, conjugué à l'hyperplan [1], et défini comme le centre de gravité des s sommets affectés de masses proportionnelles aux probabilités des événements (constituants) qu'ils représentent. Cette représentation est utile parce que le simplexe donne ainsi une image intuitive de l'espace des lois de probabilité, et surtout parce que les relations linéaires sont conservées. Les ∞^{s-1} évaluations de probabilité admissibles peuvent en effet être combinées linéairement : si $\mathbf{P}_1, \mathbf{P}_2, \ldots; \mathbf{P}_m$ sont des fonctions de probabilité, $\mathbf{P} = \Sigma\lambda_i \mathbf{P}_i$ ($\lambda_i \geqq 0$, $\Sigma\lambda_i = 1$) l'est aussi, et le point représentatif de $\mathbf{P}$ est donné par la même relation, c'est-à-dire il est le centre de gravité des points représentatifs de $\mathbf{P}_1, \ldots, \mathbf{P}_m$ avec les masses $\lambda_1, \ldots, \lambda_m$; les évaluations de probabilité admissibles constituent donc, comme les points non extérieurs au simplexe, un ensemble convexe clos. Cette simple remarque permet de compléter tout de suite nos résultats, en précisant l'indétermination dont reste affectée la valeur de la probabilité d'un événement linéairement indépendant de certains autres, après

---

[1] Dans la polarité $f\binom{x}{y} = \Sigma x_i y_i = 0$ (coordonnées barycentriques). Il est commode ici, ayant à employer des notions métriques, de considérer le simplexe comme équilatéral. On peut préciser alors qu'il s'agit de la polarité par rapport à l'hypersphère imaginaire $\Sigma x_i^2 = 0$, et qu'elle fait correspondre à un point A quelconque l'hyperplan orthogonal à la droite AO et passant par le point A′ correspondant à A dans une inversion de centre O. Avec la notation vectorielle, l'hyperplan est le lieu des points Q tels que le produit intérieur $(A - O) \times (Q - O)$ donne $-R^2$, avec $R = \frac{l}{\sqrt{2s}}$, l = longueur de chaque arête du simplexe.

que la probabilité de ceux-ci a été fixée. Il suffit de noter que, en fixant la valeur de la probabilité de certains événements, on impose des conditions linéaires à la fonction **P**; les fonctions **P** encore admissibles constituent donc, elles aussi, un ensemble convexe clos. D'où l'on tire immédiatement la conclusion importante que, lorsque la probabilité p d'un événement E n'est pas univoquement déterminée par les données, les valeurs admissibles sont tous les nombres d'un intervalle clos $p' \leqq p \leqq p''$. Si E′ et E″ sont respectivement la somme de tous les constituants contenus dans E ou compatibles avec E, p' sera la plus petite valeur admissible pour la probabilité de E′, et p'' la plus grande pour E″.

Lorsque les événements considérés sont en nombre infini, notre définition n'introduit aucune difficulté nouvelle : **P** est une fonction de probabilité pour la classe infinie d'événements $\mathcal{E}$ lorsqu'elle l'est pour toute sous-classe finie de $\mathcal{E}$. Cette conclusion implique que le théorème des probabilités totales ne peut être étendu au cas d'une infinité, même dénombrable, d'événements (¹); une discussion sur ce sujet nous entraînerait trop loin.

Nous devons encore nous occuper de la définition des probabilités subordonnées et de la démonstration du théorème des probabilités composées. Soient deux événements E′ et E″; nous pouvons parier sur E′ et subordonner ce pari à E″ : si E″ ne se vérifie pas, le pari sera annulé; si E″ se vérifie, il sera gagné ou perdu selon que E′ sera ou ne sera pas vérifié. On peut considérer alors les « événements subordonnés » (ou « triévénements »), qui sont des événements d'une logique à trois valeurs; ce « triévénement » « E′ subordonné à E″ », $\frac{E'}{E''}$, est l'entité logique capable des trois valeurs : *vrai*, si E″ et E′ sont vrais; *faux*, si E″ est vrai et E′ faux; *nul*, si E″ est faux. On voit que deux triévénements $\frac{E'_1}{E''_1}$ et $\frac{E'_2}{E''_2}$ sont égaux si $E''_1 = E''_2$ et $E'_1 E''_1 = E'_2 E''_2$; on dira que $\frac{E'}{E''}$ est écrit sous la forme *normale* si $E' \supset E''$, et l'on voit que tout triévénement $\frac{E'}{E''}$ peut être écrit d'une seule façon sous la forme normale qui est $\frac{E'E''}{E''}$. On peut établir pour les triévénements une logique à trois valeurs parfaitement ana-

(¹) *Voir* [16], [24], [X], [28], [XI], [64].

logue à la logique ordinaire [64], mais cela n'est pas nécessaire pour le but que nous poursuivons.

Définissons la probabilité p de E′ subordonnément à E″ par la même condition relative aux paris, mais dans les cas où l'on convient que le pari reste sans effet si E″ n'a pas lieu. Le pari peut alors donner trois résultats différents : si S est l'enjeu, la mise à verser sera pS, et le gain $(1-p)S$, $-pS$ ou 0 selon que $\frac{E'}{E''}$ sera vrai, faux ou nul, car dans le premier cas on gagne l'enjeu et on perd la mise, dans le deuxième on perd la mise, dans le dernier la mise est remboursée (si $S<0$ ces considérations demeurent inchangées; il n'y a qu'à échanger la terminologie de débit et crédit). Supposons que $E'\supset E''$, et soient p' et p'' les probabilités de E′ et E″ : démontrons que, pour la cohérence, on doit avoir $p'=pp''$. Si l'on fait trois paris : sur E′ avec l'enjeu S′, sur E″ avec l'enjeu S″, sur $\frac{E'}{E''}$ avec l'enjeu S, le gain total correspond, dans les trois cas possibles, à

$$
\begin{array}{ll}
E' : & G_1=(1-p')S'+(1-p'')S''+(1-p)S; \\
E'' \text{ non } E' : & G_2=-p'S'+(1-p'')S''-pS; \\
\text{non } E'' : & G_3=-p'S'-p''S''.
\end{array}
$$

Si le déterminant

$$\begin{vmatrix} 1-p' & 1-p'' & 1-p \\ -p' & 1-p'' & -p \\ -p' & -p'' & 0 \end{vmatrix} = p'-pp''$$

ne s'annule pas, on peut fixer S, S′, S″ de façon que les G aient des valeurs arbitraires, en particulier toutes positives, et cela impliquerait un défaut de cohérence. Donc $p'=pp''$, et, en général, si E′ n'implique pas E″, cela est vrai en considérant non plus E′ mais E′ E″ : on a ainsi le théorème des probabilités composées [1]

$$\mathbf{P}(E'.E'')=\mathbf{P}(E').\mathbf{P}\left(\frac{E''}{E'}\right). \tag{1}$$

---

[1] Ce résultat qui, dans le schéma des paris, peut être déduit comme nous l'avons vu de la définition de cohérence, est susceptible lui aussi d'être exprimé sous une forme purement qualitative, à savoir la suivante, qui peut s'ajouter comme cinquième axiome

La condition n'est pas seulement nécessaire, mais aussi suffisante, dans le même sens que dans le cas du théorème des probabilités totales. Selon qu'un individu évalue $\mathbf{P}\left(\frac{E'}{E''}\right)$ plus grand, ou plus petit, ou égal à $\mathbf{P}(E')$, on dira qu'il juge les deux événements E′ et E″ comme étant en *corrélation positive* ou *négative*, ou comme *indépendants :* il en résulte ainsi que la notion même d'indépendance ou de dépendance de deux événements n'a qu'une signification subjective, relative à la fonction particulière **P** qui représente l'opinion d'un individu déterminé.

On dira que $E_1, E_2, \ldots, E_n$ constituent une classe d'événements indépendants si chacun d'eux est indépendant d'un produit quelconque de plusieurs autres de ces événements (l'indépendance deux à deux ne suffit naturellement pas); dans ce cas la probabilité d'un produit logique est le produit des probabilités, et, les constituants mêmes étant des produits logiques, la probabilité d'un événement quelconque logiquement dépendant de $E_1, \ldots, E_n$ sera donnée par une fonction algébrique de p_1, $p_2, \ldots, p_n$.

On obtient comme corollaire immédiat de (1), le théorème de Bayes sous la forme [1]

$$\mathbf{P}\left(\frac{E''}{E'}\right) = \mathbf{P}(E'')\frac{\mathbf{P}\left(\frac{E'}{E''}\right)}{\mathbf{P}(E')}, \tag{2}$$

qui se prête à l'énoncé suivant, particulièrement significatif : *la probabilité de* E′, *subordonnée à* E″, *se modifie dans le même sens et dans la même mesure que la probabilité de* E″ *considérée comme subordonnée à* E′.

Dans ce qui précède je n'ai fait que résumer, d'une façon rapide et incomplète, des idées et des résultats avec le but d'éclaircir ce qu'on doit entendre, en se plaçant du point de vue subjectif, par « lois logiques de la probabilité », et comment on peut les démontrer. Ces lois sont les conditions qui caractérisent les opinions cohérentes, admissibles en soi, et les distinguent des autres, qui sont intrinsèquement contradictoires.

---

aux quatre précédents (*voir* p. 4-5) : Si E′ et E″ sont contenus dans E, $\frac{E'}{E}$ est plus ou moins ou également probable que $\frac{E''}{E}$ selon que E′ est plus ou moins ou également probable que E″.

(1) On le trouve aussi exprimé sous cette même forme dans Kolmogoroff [XVII].

Le choix de l'une d'entre elles parmi toutes les autres n'a plus rien d'objectif, et n'entre pas dans la logique du probable; nous nous occuperons de ce problème dans les chapitres suivants.

CHAPITRE II.

L'ÉVALUATION D'UNE PROBABILITÉ.

La notion de probabilité, telle que nous l'avons décrite, est sans doute la plus voisine de celle de « l'homme de la rue »; mieux encore, c'est exactement celle qu'il applique tous les jours dans ses jugements pratiques. Pourquoi la Science devrait-elle s'en éloigner? Quelle signification plus adéquate devrait-elle lui découvrir?

On pourrait croire tout d'abord que la probabilité, dans le sens habituel, ne peut pas constituer l'objet d'une théorie mathématique. Cependant nous avons vu que les règles du calcul des probabilités, conçues comme les conditions nécessaires à assurer la cohérence entre les évaluations de probabilité d'un même individu, peuvent au contraire être développées et démontrées rigoureusement. Elles ne constituent alors, en effet, que l'expression précise des règles de la logique du probable qui sont appliquées d'une manière inconsciente, qualitativement sinon numériquement, par tous les hommes dans toutes les circonstances de leur vie.

On peut douter encore que cette conception, qui laisse tout à fait libre chaque individu d'évaluer les probabilités comme il le croit pourvu seulement que la condition de cohérence soit satisfaite, suffise pour rendre compte des concordances plus ou moins strictes qu'on observe entre les jugements des divers individus ainsi qu'entre les prévisions et les résultats observés. Y a-t-il donc, parmi l'infinité d'évaluations parfaitement admissibles en elles-mêmes, une évaluation particulière qu'on puisse qualifier, dans un sens d'ailleurs encore inconnu, d'*objectivement juste ?* Ou, du moins, peut-on se demander si une évaluation donnée est plus juste qu'une autre?

Deux sont les procédés d'où l'on a cru pouvoir déduire une signification objective de la probabilité : d'une part le schéma des cas également probables, et d'autre part la considération des fréquences.

Effectivement, c'est sur ces deux procédés que s'appuie généralement l'évaluation d'une probabilité dans les cas où normalement les opinions de la plupart des individus coïncident. Cependant, ces mêmes procédés ne nous obligent nullement à admettre l'existence d'une probabilité objective : au contraire, si l'on veut forcer leur signification pour arriver à une telle conclusion, on rencontre des difficultés bien connues, qui disparaissent lorsqu'on devient un peu moins exigeant, c'est-à-dire lorsqu'on ne cherche pas à éliminer mais à préciser ce qu'il y a de subjectif dans tout cela. En d'autres termes, il s'agit de considérer la coïncidence des opinions comme un fait psychologique; les raisons de ce fait pourront alors conserver leur nature subjective, qu'on ne peut laisser de côté sans soulever une foule de questions dont le sens n'est pas très clair. Ainsi dans le cas des jeux de hasard qui sont l'origine du calcul des probabilités, il n'y a aucune difficulté à comprendre et trouver bien naturel que les hommes soient généralement amenés, par des considérations de symétrie plus ou moins précises, mais sans aucun doute très spontanées, à attribuer des probabilités égales aux divers cas possibles. On peut alors justifier immédiatement la définition classique de la probabilité fondée sur le rapport du nombre des cas favorables au nombre des cas possibles : en effet, si l'on a une classe complète de n événements incompatibles, et si on les juge également probables, en vertu du théorème des probabilités totales, chacun d'eux aura nécessairement la probabilité $p = \frac{1}{n}$ et la somme de m d'entre eux la probabilité $\frac{m}{n}$. On obtient ainsi un critère très commode et très puissant : non seulement parce qu'il nous donne la possibilité de calculer facilement la probabilité lorsqu'on trouve une subdivision en cas, jugés également probables, mais encore parce qu'il fournit en général une méthode pour évaluer par comparaison une probabilité quelconque, en ramenant l'évaluation quantitative directe à des jugements purement qualitatifs (égalité ou inégalité de deux probabilités). Cependant ce critère n'est applicable que dans l'hypothèse que l'individu qui évalue les probabilités juge également probables les cas considérés; cela est encore dû à un jugement subjectif, que les considérations habituelles de symétrie que nous avons rappelées peuvent expliquer dans ses raisons psychologiques, mais qu'elles ne peuvent nullement transformer en quelque chose d'objectif. Si, par exemple, on voulait démontrer que l'évaluation, où toutes les probabilités sont jugées égales, est la seule « juste », et que si un individu ne la partage pas il « se trompe », on devrait d'abord expliquer ce qu'on veut dire en disant qu'un individu qui évalue une probabilité juge « juste » ou qu'il « se trompe », ensuite on devrait démontrer que les considérations de symétrie citées impliquent nécessairement que l'on doive accepter l'hypothèse des probabilités égales si l'on ne veut pas « se tromper ». Or, un événement quelconque ne peut qu'arriver ou ne pas arriver, et ni dans un cas ni dans l'autre on ne peut décider quel était le degré de doute avec lequel il était « raisonnable » ou « juste » de l'atteindre avant de savoir s'il était réalisé ou non.

Considérons maintenant l'autre critère, celui des fréquences. Il s'agit d'expliquer sa valeur du point de vue subjectif, et de démontrer précisément qu'il conserve tout son contenu, mais que, comme le critère précédent et comme tout autre critère possible, il est incapable de nous mener hors du champ des jugements subjectifs, dont il ne peut nous offrir qu'une analyse psychologique plus poussée. Dans le cas des fréquences, cette analyse se divise en deux parties : une partie élémentaire, renfermant les relations entre évaluations de probabilités et prévisions de fréquences futures; une seconde partie plus délicate, concernant les relations entre l'observation de fréquences passées et la prévision des fréquences futures. Pour le moment, nous nous bornerons à la première question, en admettant comme un fait psychologique connu, dont les raisons seront analysées ensuite, qu'on prévoit généralement des fréquences voisines de celles qui ont été observées.

La relation cherchée entre les évaluations de probabilités et les prévisions de fréquences est donnée par le théorème suivant. Soient $E_1, E_2, \ldots, E_n$ des événements quelconques (¹). Attribuons les valeurs

---

(¹) Pour éviter un malentendu possible, dû à la divergence entre notre conception et celle communément admise, il ne sera pas peut-être inutile de rappeler que, dans notre terminologie, un « événement » est toujours un fait singulier bien déterminé. Ce que l'on appelle parfois *répétitions* ou *épreuves* d'un même événement, sont pour nous autant d'événements distincts. Elles auront, en général, des caractères communs ou symétriques qui rendront naturel de leur attribuer des probabilités égales, mais nous n'admettons aucune raison à priori qui nous empêche en principe d'attribuer à chacune des épreuves $E_1, \ldots, E_n$ des probabilités $p_1, \ldots, p_n$ différentes et absolument quelconques. En principe, il n'y a aucune différence, pour nous, entre ce cas et celui de n événements qui ne présentent entre eux aucune analogie; l'analogie qui suggère la dénomination d'« épreuves d'un même événement » (nous dirions : « d'un même phé-

$p_1, p_2, \ldots, p_n$ à leurs probabilités, et les valeurs $\omega_0, \omega_1, \omega_2, \ldots, \omega_n$ aux probabilités pour que zéro, ou un seul, ou deux, etc., ou enfin, tous ces événements se vérifient (évidemment $\omega_0 + \omega_1 + \omega_2 + \ldots + \omega_n = 1$). Pour la cohérence on doit avoir :

$$p_1 + p_2 + \ldots + p_n = 0 \,.\, \omega_0 + 1 \,.\, \omega_1 + 2 \,.\, \omega_2 + \ldots + n \,.\, \omega_n$$

ou simplement

$$\bar{p} = \bar{f}, \tag{3}$$

si $\bar{p}$ indique la moyenne arithmétique des p_i, et $\bar{f}$ l'espérance mathématique de la fréquence (c'est-à-dire du nombre aléatoire qui prend les valeurs $\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$ selon que $0, 1, 2, \ldots, n$ des E_i se vérifient); notons à cet égard que la notion d'espérance mathématique a, elle aussi, une signification subjective, puisqu'elle n'est définie qu'en relation au jugement considéré, qui attribue à ces $n + 1$ cas possibles les probabilités ω_h.

Cette relation se simplifie encore dans des cas particuliers : si la fréquence est connue, le second membre représente tout simplement cette valeur de la fréquence; si l'on juge que les n événements sont également probables, le premier membre n'est autre chose que la valeur commune de ces probabilités. Commençons par le cas où ces simplifications ont lieu toutes les deux : on a n événements, on sait que m se sont vérifiés ou devront l'être, mais on ignore lesquels, et l'on juge également probable que l'un quelconque de ces événements doive se vérifier. La seule évaluation possible de la probabilité conduit dans ce cas à la valeur $p = \frac{m}{n}$. Si $m = 1$ on retombe dans le cas de n possibilités incompatibles également probables.

Si, dans le même cas où la fréquence est connue d'avance, notre jugement n'est pas aussi simple, la relation nous est encore très utile pour l'évaluation des n probabilités, car, en connaissant quelle doit en être la moyenne arithmétique, on a une indication globale sur leur ordre de grandeur commun, et l'on n'a qu'à s'arranger pour augmenter certains

---

nomène ») n'a rien d'essentiel, mais tout au plus une valeur pour l'influence qu'elle peut exercer sur notre jugement psychologique dans le sens de nous faire attribuer des probabilités égales ou à peu près égales aux divers événements.

termes et en diminuer d'autres jusqu'à ce que les rapports entre les diverses probabilités correspondent à notre jugement subjectif sur l'inégalité de leurs chances respectives. Comme exemple typique, on peut considérer un scrutin secret : on sait que pour les n votants $A_1, A_2, \ldots, A_n$, on a eu m bulletins favorables; on pourra évaluer alors les probabilités $p_1, p_2, \ldots, p_n$ pour que les divers votants aient donné un vote favorable, d'après l'idée que l'on a de leurs opinions; en tout cas cette évaluation doit être faite de telle façon que la moyenne arithmétique des p_i soit $\frac{m}{n}$.

Lorsque la fréquence n'est pas connue, l'égalité relie deux membres qui dépendent tous les deux d'un jugement de probabilité : l'évaluation des probabilités p_i n'est plus liée, à travers leur moyenne, à une donnée objective mais à l'évaluation d'autres probabilités, les probabilités ω_h des diverses fréquences. Toutefois, on a l'avantage de ne pas devoir évaluer exactement tous les ω_h pour appliquer la dite relation à l'évaluation des probabilités p_i; il suffira en effet d'une estimation très vague et de nature qualitative des ω_h, pour évaluer $\bar{f}$ avec assez de précision. Il suffit par exemple de juger « peu probable » que la fréquence s'éloigne sensiblement d'une certaine valeur a, ce qui revient à estimer très petite la somme de tous les ω_h pour lesquels $\left|\frac{h}{n} - a\right|$ n'est pas petit, pour donner approximativement $\bar{f} \cong a$.

Une fois $\bar{f}$ évalué, rien n'est à changer à ce qu'on a dit précédemment concernant le cas où la fréquence est connue : si les n événements sont jugés également probables, leur probabilité commune est $p = \bar{f}$; s'il n'en est pas ainsi certaines probabilités seront à augmenter ou à diminuer en s'arrangeant pour que leur moyenne arithmétique soit $p = \bar{f}$.

C'est ainsi qu'on évalue couramment les probabilités dans la plupart des problèmes pratiques. Soit par exemple à évaluer la probabilité pour qu'un individu donné, nous dirons Monsieur A, meure au cours de l'année. Si l'on voulait estimer directement à quelles conditions le pari, ou mieux l'assurance, comme on préférerait dire dans ce cas, nous semble équitable, cette évaluation nous apparaîtrait entachée d'une très grande incertitude; l'application du critère décrit facilite très fortement l'estimation. Pour cela, on doit considérer d'autres événements, par exemple le décès, pendant une année, d'individus du même âge que

Monsieur A et vivant dans le même pays. Supposons que l'on prévoie que parmi ces individus 13 pour 1000 à peu près mourront en une année; si en particulier toutes les probabilités sont jugées égales, leur valeur commune est alors $p = 0{,}013$, et la probabilité de décès pour Monsieur A est 0,013; si en général on connaît des raisons qui font varier d'un individu à l'autre les chances que nous attribuons à leur décès, cette valeur moyenne 0,013 nous donne au moins une base, de laquelle nous pourrons nous écarter dans un sens ou dans l'autre en tenant compte des caractéristiques qui différencient Monsieur A des autres individus.

Ce procédé comprend trois phases distinctes et successives : la première consiste en le choix d'une classe d'événements contenant celui que nous voulons considérer; la deuxième est la prévision de la fréquence; la troisième la comparaison entre les probabilités des événements singuliers et l'évaluation. Quelques observations à ce propos sont nécessaires pour éclaircir la signification et la valeur qui sont attribuées à ces considérations du point de vue subjectif, et qui s'éloignent essentiellement des opinions courantes. Ce n'est d'ailleurs qu'en raison de la nécessité de donner des éclaircissements sur ce point avant de continuer, qu'il était indispensable de s'arrêter quelque peu sur une question tellement élémentaire.

Le choix d'une classe d'événements est en soi arbitraire; si l'on choisit des événements « semblables », ce n'est que pour rendre plus facile l'application du procédé, c'est-à-dire la prévision de la fréquence et la comparaison des diverses probabilités, mais cette restriction n'est nullement essentielle, et, même si on l'admet, elle n'a qu'un sens très vague. Dans l'exemple précédent on pouvait considérer, non les individus du même âge et du même pays, mais ceux de la même profession, de même taille, de même profession et ville, etc., et, dans tous ces cas, on observerait une « similitude » assez sensible. Rien n'empêcherait toutefois à priori de grouper l'événement qui nous intéresse avec d'autres événements quelconques, en considérant par exemple la mort de Monsieur A dans l'année, comme *un sinistre* par rapport à toutes les assurances de la Compagnie auprès de laquelle il est assuré, y compris les assurances incendie, transport et autres, et, d'un certain point de vue, on pourrait encore soutenir que ces événements sont « semblables ».

C'est pourquoi nous évitons les expressions telles que « épreuves d'un même événement », « événements qui peuvent se répéter », etc., et, en général, toutes les considérations de fréquence, qui présupposent une classification conçue comme rigide et essentielle des événements, dans des classes ou collections ou séries. Toute classification de ce genre n'a qu'une fonction auxiliaire et une valeur arbitraire.

La prévision de la fréquence est fondée généralement sur l'hypothèse que sa valeur reste à peu près constante : dans notre exemple, la conviction que la proportion de décès sera de 13 pour 1000 pourrait avoir son origine dans l'observation qu'au cours des dernières années la mortalité des individus de conditions analogues était de 13 pour 1000 environ. Les raisons qui peuvent justifier cette façon de prévoir, ne pourront être analysées que plus loin; pour le moment, il suffit de constater qu'effectivement notre intuition nous amène à juger ainsi. Remarquons qu'une telle prévision est en général d'autant plus difficile que la classe considérée est plus étroite.

La comparaison des diverses probabilités est par contre d'autant plus difficile que les événements sont plus nombreux, et moins homogènes : la difficulté est évidemment réduite au minimum lorsque les événements nous apparaissent également probables. En pratique, on doit chercher à concilier le mieux possible ces exigences opposées, en vue de la meilleure application de ces deux parties du procédé : c'est seulement en fonction de ces exigences que la classe des événements à considérer pourra être choisie d'une façon plus ou moins appropriée.

Une image rendra ces considérations encore plus évidentes. Si l'on doit estimer l'épaisseur d'une feuille de papier, on pourra plus facilement y parvenir en estimant d'abord l'épaisseur d'un paquet de n feuilles dans lequel elle est insérée et ensuite la mesure dans laquelle les diverses feuilles ont la même épaisseur. L'épaisseur peut être évaluée d'autant plus facilement que le paquet est plus gros; la difficulté de la comparaison successive des feuilles est par contre diminuée si l'on amincit le paquet en conservant seulement les feuilles jugées d'épaisseur à peu près égale à celle de la feuille qui nous intéresse.

Ainsi le critère basé sur la notion de fréquence se réduit, comme celui des événements également probables, à une méthode pratique pour ramener certaines évaluations subjectives de probabilités à d'autres évaluations, elles-mêmes subjectives, mais préférables, soit parce que plus accessibles à une estimation directe, soit parce qu'une estimation plus grossière ou même de nature purement qualitative suffit pour les conclusions envisagées. A priori, lorsqu'on accepte le point de vue subjectif, telle doit être effectivement la signification et la valeur d'un critère quelconque.

Dans le cas des prévisions de fréquences, on ne fait que ramener l'évaluation des p_i à celle des ω_h et à une comparaison entre les p_i; l'estimation des ω_h n'a pas besoin d'ailleurs d'atteindre plus qu'une approximation très grossière, telle qu'elle suffise pour déterminer les p_i assez précisément. Il faut remarquer toutefois que cette « prévision de la fréquence » n'est autre chose qu'une évaluation des ω_h : elle n'est nullement une prophétie, qu'on pourrait dire juste si la fréquence s'avère effectivement égale ou voisine de $\bar{f}$, et fausse dans le cas contraire. Toutes les fréquences $\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$ sont possibles, et, quelle que soit la fréquence réalisée, rien ne pourra nous donner raison ou tort si notre opinion actuelle consiste à attribuer à ces $n+1$ cas des probabilités ω_h, conduisant à une certaine valeur

$$\bar{p} = \bar{f} = \frac{\omega_1 + 2\omega_2 + 3\omega_3 + \ldots + n\omega_n}{n}. \tag{3}$$

On croit souvent pouvoir échapper à ces objections en observant que l'impossibilité de préciser les relations entre probabilités et fréquences est analogue à l'impossibilité pratique qu'on rencontre dans toutes les sciences expérimentales à relier exactement les notions abstraites de la théorie avec les réalités empiriques (1). L'analogie est, à mon avis, illusoire : dans les autres sciences on a une théorie qui affirme et prévoit avec certitude et exactitude ce qui devrait arriver si elle était tout à fait exacte; dans le calcul des probabilités c'est la théorie elle-même qui nous oblige à admettre la possibilité de toutes les fréquences. Dans les autres sciences, l'incertitude découle bien donc de la liaison imparfaite entre la théorie et les faits; dans notre cas, au contraire, elle ne peut avoir son origine dans cette liaison, mais dans le sein de la théorie elle-même [32], [65], [IX]. Aucune relation entre probabilités et fréquences n'a de caractère empirique, car la fréquence observée, quelle qu'elle soit,

---

(1) Ce point de vue est soutenu, avec des variantes plus ou moins profondes, dans la plupart des traités modernes, entre autres ceux de Castelnuovo [VI], Fréchet-Halbwachs [XII], Lévy [XX], von Mises [XXV].

est toujours compatible avec toutes les opinions concernant les probabilités respectives; ces opinions, par conséquent, ne peuvent être ni confirmées ni démenties, étant donné qu'elles ne contiennent aucune affirmation catégorique, comme par exemple : tel événement *doit* se vérifier ou *ne peut pas* se vérifier.

Cette dernière considération peut paraître bien étrange, si l'on pense que la prévision d'une fréquence future est généralement fondée sur l'observation du passé; on dira, « nous corrigeons » nos opinions initiales si « l'expérience les dément it ». Ce procédé tellement naturel et instinctif n'est-il donc pas justifié ? Oui; seulement la manière de le formuler n'est pas exacte, ou, pour mieux dire, n'a pas de sens. Il ne s'agit pas de « corriger » des opinions qui ont été « démenties » : il s'agit tout simplement de substituer à l'évaluation initiale de la probabilité la valeur subordonnée à l'apparition des faits qui ont été en effet observés; cette probabilité est une chose tout à fait distincte de l'autre, et leurs valeurs peuvent très bien ne pas coïncider sans que cela puisse être interprété comme la « correction d'une opinion démentie ».

L'explication de l'influence exercée par l'expérience sur nos prévisions futures, développée d'après les idées que nous venons d'exposer, constitue le point que nous avions laissé de côté dans l'analyse du critère basé sur les fréquences. Ce développement fera l'objet des chapitres suivants, dans lesquels nous ferons une étude plus développée du cas le plus typique sous ce rapport : le cas des événements équivalents, et, en général, des nombres ou éléments aléatoires équivalents quelconques. Cette étude est importante pour le développement de la conception subjective, mais j'espère que l'aspect mathématique présentera lui-même quelque intérêt, indépendamment de l'interprétation philosophique; en effet, les événements et les nombres aléatoires équivalents sont caractérisés par des conditions simples et significatives qui peuvent justifier par elles-mêmes une étude approfondie des problèmes qui se posent à leur égard.

CHAPITRE III.

ÉVÉNEMENTS ÉQUIVALENTS.

Pourquoi sommes-nous donc poussés, dans la plupart des problèmes, à évaluer une probabilité d'après l'observation d'une fréquence ? Il s'agit là de la question des relations entre l'observation des fréquences passées et la prévision de fréquences futures, que nous avions laissée en suspens, mais qui s'est présentée de nouveau sous une forme quelque peu modifiée lorsque nous nous sommes demandé si une prévision de fréquence peut, en un certain sens, être confirmée ou démentie par l'expérience. La question que nous nous posons maintenant renferme en réalité le problème même du raisonnement par induction. Ce problème essentiel qui n'a jamais reçu de solution satisfaisante jusqu'à présent, pourra-t-il en recevoir une si l'on emploie la conception subjective de la probabilité et la théorie que nous avons esquissée ?

Pour mieux fixer les idées envisageons un exemple concret ou plutôt une interprétation concrète du problème, qui n'en restreint nullement la généralité. Supposons qu'on joue à pile ou face avec une pièce d'apparence irrégulière. Les probabilités d'obtenir « face » au premier, deuxième, ..., $h^{\text{ième}}$ coup, c'est-à-dire les probabilités $\mathbf{P}(E_1)$, $\mathbf{P}(E_2)$, ..., $\mathbf{P}(E_h)$, ..., des événements $E_1, E_2, \ldots, E_h, \ldots$, consistant dans l'arrivée de « face » aux différents coups, ne pourront être évaluées qu'en chiffrant à priori l'influence de l'irrégularité apparente de la pièce.

On nous objectera sans doute, que pour y arriver, c'est-à-dire pour obtenir les probabilités « correctes » des épreuves futures, nous pourrons utiliser les résultats acquis dans les épreuves antérieures : c'est, en effet, dans ce sens que — selon l'interprétation courante — nous « corrigerions » l'évaluation de $\mathbf{P}(E_{n+1})$ d'après l'observation des épreuves qui ont amené, ou non, E_1, E_2, ..., E_n. Une telle interprétation nous semble inacceptable, non seulement parce qu'elle présuppose l'existence objective de probabilités inconnues, mais aussi parce qu'elle ne peut même pas être formulée correctement : en effet, la probabilité de E_{n+1} évaluée après la connaissance d'un certain résultat A des n épreuves antérieures, n'est plus $\mathbf{P}(E_{n+1})$ mais $\mathbf{P}\left(\frac{E_{n+1}}{A}\right)$. On aura d'une façon précise

$$A = E_{i_1}E_{i_2}\ldots E_{i_r}\overline{E}_{j_1}\overline{E}_{j_2}\ldots\overline{E}_{j_s} \qquad (r+s=n),$$

le résultat A consistant en ce que les r coups i_1, i_2, ..., i_r donnent « face » et les autres s coups j_1, j_2, ..., j_s donnent « pile » : A est donc un des constituants formés avec E_1, E_2, ..., E_n. Mais alors, s'il s'agit d'une probabilité subordonnée, nous pouvons appliquer le théorème des probabilités composées et l'interprétation des résultats qui en découlent constituera notre justification du raisonnement inductif.

En général, nous avons

$$\mathbf{P}\left(\frac{E_{n+1}}{A}\right) = \frac{\mathbf{P}(A.E_{n+1})}{\mathbf{P}(A)}; \tag{4}$$

notre explication du raisonnement par induction n'est, au fond, rien d'autre que ce qu'exprime cette formule, à savoir : la probabilité de E_{n+1} évaluée lorsqu'on connaît le résultat A de E_1, ..., E_n n'est pas un élément de nature essentiellement nouvelle (justifiant l'introduction d'un terme nouveau, par exemple « probabilité statistique » ou *à posteriori*). Cette probabilité n'est pas indépendante de la « probabilité à priori » et ne la remplace pas; elle découle en fait du même jugement à priori, en en retranchant seulement, pour ainsi dire, les composantes du doute relatives aux épreuves dont le résultat a été acquis.

Pour éviter dans ce qui va suivre des interprétations erronées, il convient, avant tout, de rappeler encore une fois le sens que nous attribuons dans ce travail à un certain nombre de termes. Considérons d'abord une classe d'événements (comme, dans l'exemple, les divers coups de pile ou face). Nous dirons parfois qu'ils constituent des *épreuves* d'un phénomène donné; cela servira d'ailleurs uniquement à rappeler qu'on a presque toujours intérêt à appliquer les raisonnements qui vont suivre au cas où les événements considérés sont des événements *d'un même type*, ou qui ont des caractères *analogues*, sans qu'on attache une signification intrinsèque ou une valeur précise quelconque à cette caractéristique extérieure dont la définition comporte évidemment une très large part d'arbitraire. Nos raisonnements ne feront intervenir que les événements, c'est-à-dire les épreuves, prises chacune individuellement : l'analogie des événements n'entre pas en ligne de compte par elle-même, mais seulement dans le sens et la mesure où elle peut avoir influencé d'une façon quelconque le jugement d'un individu sur les probabilités considérées.

Il est évident qu'en posant le problème comme nous l'avons fait, il nous sera impossible de *démontrer* la validité du principe d'induction, c'est-à-dire du principe d'après lequel la valeur de la probabilité devrait être voisine de la fréquence observée $\left[\text{par exemple, dans le cas précédent : } \mathbf{P}\left(\frac{E_{n+1}}{A}\right) \simeq \frac{r}{n}\right]$. Ce principe ne pourra être justifié que dans des cas particuliers, ce qui n'est pas dû d'ailleurs à une insuffisance de la méthode suivie, mais correspond logiquement et nécessairement aux exigences essentielles de notre point de vue. En effet, la probabilité étant purement subjective, rien ne pourrait nous obliger à la prendre voisine de la fréquence; tout ce qu'on peut démontrer est qu'une telle évaluation découle d'une manière cohérente de notre jugement initial lorsque celui-ci satisfait à certaines conditions tout à fait naturelles et suffisamment étendues.

Nous nous bornerons dans ce qui suit aux conditions les plus simples qui définissent les événements que nous appelons équivalents, et que nous exposerons, pour fixer les idées, sur l'exemple déjà mentionné; nos résultats n'en auront pas moins une portée tout à fait générale.

Il s'agit d'évaluer les probabilités de tous les résultats possibles sur les n premières épreuves (n quelconque). Ces résultats possibles sont au nombre de 2^n dont $\binom{n}{n} = 1$ consistant en la répétition de « face » n fois, $\binom{n}{n-1} = n$ avec $(n-1)$ fois « face » et une fois « pile », ..., et, en général $\binom{n}{r}$ avec r fois « face » et $n-r$ fois « pile ». Si nous désignons $\omega_r^{(n)}$ la probabilité pour que sur les n coups on obtienne, dans un ordre quelconque, r fois « face » et $n-r$ fois « pile », $\omega_r^{(n)}$ sera la somme des probabilités des $\binom{n}{r}$ façons distinctes dont on peut réaliser ce résultat; la moyenne de ces probabilités élémentaires sera donc $\omega_r^{(n)} : \binom{n}{r}$.
Ayant groupé les 2^n résultats de cette façon, nous pouvons distinguer utilement, bien que d'une manière arbitraire, deux éléments de variation de la probabilité : on a d'abord une probabilité moyenne plus ou moins grande pour chaque fréquence, et ensuite une répartition plus ou moins uniforme des probabilités $\omega_r^{(n)}$ entre les divers résultats d'égale fréquence

et ne différant l'un de l'autre que par l'ordre de la succession des épreuves favorables et défavorables. En général, on admettra des probabilités différentes dépendant de l'ordre, soit qu'on suppose qu'un coup a une influence sur celui qui le suit immédiatement, soit qu'on suppose l'existence de circonstances extérieures variables, etc.; néanmoins il est particulièrement intéressant d'étudier le cas où la probabilité ne dépend pas de l'ordre des épreuves. Dans ce cas tous les résultats comportant une même fréquence $\frac{r}{n}$ sur n coups ont donc une même probabilité, qui est $\omega_r^{(n)} : \binom{n}{r}$; si cette condition est satisfaite, nous dirons que les événements de la classe considérée — comme les différents coups dans l'exemple envisagé du jeu de pile ou face — sont *équivalents* (par rapport à notre jugement de probabilité). On verra mieux encore combien cette condition est simple et à quel point sa signification est naturelle, lorsque nous l'aurons exprimée sous d'autres formes dont les unes sembleront au premier abord plus larges, et les autres plus restrictives.

Il est presque évident que la définition de l'équivalence entraîne le résultat suivant : la probabilité pour que n épreuves déterminées aient toutes un résultat favorable est toujours la même, quel que soit l'n-uple choisi : cette probabilité sera donc égale à $\omega_n = \omega_n^{(n)}$, puisque les n premiers cas constituent un n-uple particulier. Réciproquement, si cette propriété subsiste, les événements sont équivalents, car, comme on le démontrera sous peu, il en découle que tous les résultats comportant, sur n épreuves, r favorables et s défavorables, ont la même probabilité, à savoir :

$$(5) \qquad \omega_r^{(n)} : \binom{n}{r} = (-1)^s \Delta^s \omega_r.$$

On a déjà obtenu ainsi une autre conclusion : la probabilité pour que r épreuves soient favorables et s défavorables sera toujours $\omega_r^{(n)} : \binom{n}{r}$ (avec $n = r + s$), non seulement s'il s'agit des n premières épreuves dans l'ordre admis au début, mais aussi dans le cas d'épreuves quelconques.

On peut énoncer une autre condition équivalente à la définition donnée : la probabilité d'une épreuve E quelconque, subordonnée à l'hypothèse A d'un résultat respectivement favorable et défavorable dans r et s autres épreuves déterminées, ne dépend pas des événements choisis, mais simplement de r et s (ou de r et de $n = r + s$). Si, en effet,

$$\mathbf{P}(\mathrm{A}) = \omega_r^{(n)} : \binom{n}{r} \qquad \text{et} \qquad \mathbf{P}(\mathrm{A.E}) = \omega_{r+1}^{(n+1)} : \binom{n+1}{r+1},$$

on aura

$$\mathbf{P}\left(\frac{\mathrm{E}}{\mathrm{A}}\right) = \frac{r+1}{n+1}\left(\omega_{r+1}^{(n+1)} : \omega_r^{(n)}\right) = p_r^{(n)}, \tag{6}$$

fonction de n et de r seulement; si, inversement, on suppose $\mathbf{P}\left(\frac{\mathrm{E}}{\mathrm{A}}\right) = p_r^{(n)}$, fonction de n et r, il résulte évidemment que pour tout n-uple la probabilité pour que toutes les épreuves soient favorables sera

$$\omega_n = p_0^{(0)} \cdot p_1^{(1)} \cdot \ldots \cdot p_{n-1}^{(n-1)}. \tag{7}$$

En général, on voit facilement que, dans le cas d'événements équivalents, tout problème de probabilités relatif à $\mathrm{E}_{i_1}, \mathrm{E}_{i_2}, \ldots, \mathrm{E}_{i_n}$ ne dépend pas du choix des indices (distincts) $i_1, \ldots, i_n$, mais uniquement des probabilités $\omega_0, \omega_1, \ldots, \omega_n$. Ce fait justifie bien la dénomination d'« événements équivalents » que nous avons introduite : lorsque la condition indiquée est satisfaite, tout problème est en effet parfaitement déterminé si on l'énonce pour des événements *génériques*. Ce même fait rend très naturelle l'extension de la notion d'équivalence au domaine plus vaste des nombres aléatoires : nous dirons que $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n, \ldots$ sont des nombres aléatoires équivalents s'ils jouent un rôle symétrique par rapport à tout problème de probabilité, ou, en d'autres termes, si la probabilité que $\mathrm{X}_{i_1}, \mathrm{X}_{i_2}, \ldots, \mathrm{X}_{i_n}$ satisfassent à une condition donnée est toujours la même quel que soit le choix des indices (distincts) $i_1, \ldots, i_n$. Comme pour les *événements* équivalents, tout problème de probabilité est parfaitement déterminé lorsqu'on l'énonce pour des nombres aléatoires *génériques;* en particulier si $\mathrm{X}_1, \ldots, \mathrm{X}_n, \ldots$ sont des nombres aléatoires équivalents, les événements $\mathrm{E}_i = (\mathrm{X}_i \leq x)$ (x nombre fixe quelconque) ou plus généralement $\mathrm{E}_i = (\mathrm{X}_i \,\varepsilon\, \mathrm{I})$ (I ensemble de nombres quelconque) sont équivalents. Cette propriété nous sera très utile, ainsi que la suivante : l'espérance mathématique d'une fonction quelconque de n nombres aléatoires équivalents reste invariable lorsqu'on change l'n-uple choisi; en particulier, on aura des valeurs $m_1, m_2, \ldots, m_k, \ldots$ telles que $\mathcal{M}(\mathrm{X}_i) = m_1$, quel que soit i, $\mathcal{M}(\mathrm{X}_i\mathrm{X}_j) = m_2$ quels que soient i et j ($i \neq j$), et en général $\mathcal{M}(\mathrm{X}_{i_1}, \mathrm{X}_{i_2}, \ldots, \mathrm{X}_{i_k}) = m_k$ quels que soient $i_1, \ldots, i_k$ (distincts !). Cette observation a été faite par

M. Khintchine [1], qui l'a exploitée pour simplifier la démonstration de plusieurs des résultats que j'avais établis pour les événements équivalents ; j'ai utilisé ensuite de cette idée dans l'étude des nombres aléatoires équivalents et je m'en servirai également dans cet exposé.

On peut, en effet, faire entrer comme cas particulier l'étude des événements équivalents dans l'étude des nombres aléatoires équivalents, en observant que les événements E_i ne sont équivalents que s'il en est ainsi de leurs « indicateurs », c'est-à-dire des nombres aléatoires X_i tels que $X_i = 1$ ou $X_i = 0$, selon que E_i se vérifie ou non. A propos de ces « indicateurs », mentionnons quelques-unes de leurs propriétés simples qui constituent la raison de leur utilité.

L'indicateur de $\overline{E}_i$ est $1 - X_i$, celui de $E_i E_j$ est $X_i X_j$, celui de $E_i + E_j$ est $1 - (1 - X_i)(1 - X_j) = X_i + X_j - X_i X_j$ [ce n'est que dans le cas d'événements incompatibles ($X_i X_j = 0$) qu'on a simplement $X_i + X_j$]. L'indicateur de $E_{i_1}, E_{i_2}, \ldots, E_{i_r}. \overline{E}_{j_1}, \overline{E}_{j_2}, \ldots, \overline{E}_{j_s}$ est donc

$$X_{i_1} X_{i_2} \ldots X_{i_r} (1 - X_{j_1})(1 - X_{j_2}) \ldots (1 - X_{j_s})$$
$$= X_{i_1} X_{i_2} \ldots X_{i_r} - \sum_1^s {}_h X_{i_1} X_{i_2} \ldots X_{i_r} X_{j_h} + \sum_1^s {}_{hk} X_{i_1} X_{i_2} \ldots X_{i_r} X_{j_h} X_{j_k} - \ldots \pm X_1 X_2 \ldots X_n.$$

L'espérance mathématique de l'indicateur n'est que la probabilité de l'événement correspondant; ainsi la possibilité de transformer les opérations logiques sur les événements en opérations arithmétiques sur les indicateurs facilite grandement la résolution d'un certain nombre de problèmes. On déduit tout de suite, en particulier, la formule (5) annoncée pour $\omega_r^{(n)}$ dans le cas des événements équivalents : si le produit de h épreuves a toujours la probabilité ω_h, la probabilité $\omega_r^{(n)} : \binom{n}{r}$ de $E_{i_1}, E_{i_2}, \ldots, E_{i_r}. \overline{E}_{j_1}, \overline{E}_{j_2}, \ldots, \overline{E}_{j_s}$ se déduira du développement ci-dessus de l'indicateur de cet événement, et l'on obtiendra

$$(5) \quad \omega_r^{(n)} : \binom{n}{r} = \omega_r - \binom{s}{1} \omega_{r+1} + \binom{s}{2} \omega_{r+2} - \ldots + (-1)^s \omega_{r+s} = (-1)^s \Delta^s \omega_r.$$

En posant $\omega_0 = 1$, la formule reste vraie pour $r = 0$.

En laissant de côté pour le moment les questions de principe qui nous ont guidés jusqu'ici, nous développerons maintenant l'étude des

[1] [XV], voir aussi [XVI].

événements et nombres aléatoires équivalents, en démontrant avant tout que la loi des grands nombres et même la loi forte des grands nombres sont valables pour les nombres aléatoires équivalents X_i, et que la loi de probabilité de la moyenne Y_n de n parmi les nombres X_i, tend vers une loi limite lorsque n croît indéfiniment. Il suffit même, dans la démonstration, de supposer

$$\mathfrak{M}(X_i) = m_1, \qquad \mathfrak{M}(X_i^2) = \mu_2, \qquad \mathfrak{M}(X_i X_j) = m_2,$$

pour tous les i et j ($i \neq j$), condition qui est bien moins restrictive que celle de l'équivalence. Remarquons encore qu'il suffit de considérer explicitement les nombres aléatoires, le cas des événements y étant inclus par la considération des « indicateurs »; une moyenne Y_n est identique, dans ce cas, à la fréquence sur n épreuves.

La « loi des grands nombres » consiste en la propriété suivante : *si Y_h et Y_k sont respectivement les moyennes de h et de k nombres X_i (les deux moyennes pourront contenir des termes communs ou non), la probabilité pour que $|Y_h - Y_k| > \varepsilon$ ($\varepsilon > 0$ quelconque) sera aussi petite que l'on voudra pourvu que h et k soient suffisamment grands;* cela découle immédiatement du calcul de l'espérance mathématique de $(Y_h - Y_k)^2$, à savoir :

$$(8) \qquad \mathfrak{M}(Y_h - Y_k)^2 = \frac{h+k-2r}{hk}(\mu_2 - m_2)$$
$$= \left(\frac{1}{h} + \frac{1}{k} - \frac{2r}{hk}\right)(\mu_2 - m_2) \leqq \left(\frac{1}{h} + \frac{1}{k}\right)(\mu_2 - m_2),$$

où r est le nombre des termes communs, c'est-à-dire des X_i qui entrent aussi bien dans Y_h que dans Y_k. S'il s'agit en particulier de moyennes « successives », c'est-à-dire si tous les termes de la première figurent dans l'autre, comme par exemple si

$$Y_h = \frac{1}{h}(X_1 + X_2 + \ldots + X_h), \qquad Y_k = \frac{1}{k}(X_1 + X_2 + \ldots + X_k) \qquad (h < k),$$

on aura $r = h$, et

$$(9) \qquad \mathfrak{M}(Y_h - Y_k)^2 = \left(\frac{1}{h} - \frac{1}{k}\right)(\mu_2 - m_2).$$

Lorsqu'on considère des moyennes successives, on a en plus le résultat suivant qui constitue la loi forte des grands nombres : *ε et θ*

étant donnés, il suffit de choisir h suffisamment grand pour que la probabilité de trouver les moyennes successives $Y_{h+1}, Y_{h+2}, \ldots, Y_{h+q}$ *comprises toutes entre* $Y_h - \varepsilon$ *et* $Y_h + \varepsilon$, *diffère de l'unité d'une quantité plus petite que* θ, *q étant aussi grand que l'on veut.* Si l'on admet que la probabilité pour que toutes les inégalités

$$|Y_h - Y_{h+i}| < \varepsilon \qquad (i = 1, 2, 3, \ldots)$$

soient vraies est égale à la limite de la probabilité analogue pour $i = 1, 2, \ldots, q$ lorsque $q \to \infty$, on peut dire que *toutes* les moyennes $Y_{h+i} (i = 1, 2, 3, \ldots)$ tomberont entre $Y_h - \varepsilon$ et $Y_h + \varepsilon$, sauf dans un cas de probabilité $< \theta$; je préfère cependant éviter ce genre d'énoncés, car ils supposent essentiellement l'extension du théorème des probabilités totales au cas d'une infinité d'événements, et cette extension n'est pas admissible, du moins selon mon point de vue (*voir* p. 13).

La démonstration de la loi forte des grands nombres peut être achevée facilement, en considérant d'abord l'oscillation entre les termes Y_{h+i} avec l'indice $(h + i)$ carré, et ensuite les oscillations dans les segments compris entre deux indices carrés successifs. Si les Y à indices carrés considérés tout d'abord ne diffèrent pas entre eux de plus de $\frac{\varepsilon}{3}$, et les Y à indices compris entre deux nombres carrés successifs ne s'écartent également pas plus de $\frac{\varepsilon}{3}$, les écarts entre les Y_{h+i} ne pourront visiblement dépasser ε. Or, il suffit d'appliquer l'inégalité de Bienaymé-Tchebycheff pour parvenir à majorer la probabilité d'une exception à l'une de ces limitations partielles ([1]), et les probabilités correspondantes résultent inférieures à $36(\mu_2 - m_2)\,\varepsilon^{-2} \sum_s^\infty{}_i\, i^{-2}$ ($s =$ partie entière de $\sqrt{h}$); la probabilité d'une exception à l'une ou l'autre des limitations partielles, ne pourra donc dépasser

$$72(\mu_2 - m_2)\,\varepsilon^{-2} \sum_s^\infty i^{-2}.$$

([1]) La formule trouvée $\mathfrak{M}(Y_h - Y_k)^2 = \frac{h-k}{hk}(\mu_2 - m_2)$ donne, par l'inégalité de Bienaymé-Tchebycheff, $\mathbf{P}\{|Y_h - Y_k| < \varepsilon\} < \frac{1}{\varepsilon^2}(\mu_2 - m_2)\frac{h-k}{hk}$; en appliquant le théorème des probabilités totales ainsi qu'il est indiqué dans ma note [47], on tire de cette expression les résultats ci-après.

Cette valeur ne dépend pas de q et tend vers zéro lorsque $s \to \infty$ (c'est-à-dire : $h \to \infty$); la loi forte des grands nombres est ainsi démontrée.

Du fait que la loi des grands nombres subsiste, découle facilement l'autre résultat énoncé, à savoir que *la loi de probabilité* $\Phi_n(\xi) = \mathbf{P}(Y_n \leqq \xi)$ *tend vers une loi-limite pour* $n \to \infty$. Si la probabilité pour que $|Y_h - Y_k| > \varepsilon$ est plus petite que θ, la probabilité pour que $Y_h < \xi$ et $Y_k > \xi + \varepsilon$, sera à plus forte raison plus petite que θ et l'on aura $\Phi_h(\xi) < \Phi_k(\xi + \varepsilon) + \theta$, et d'une façon analogue $\Phi_h(\xi) > \Phi_k(\xi - \varepsilon) - \theta$. Comme ε et θ peuvent être choisis aussi petits que l'on veut, il s'ensuit qu'il existe une loi-limite $\Phi(\xi)$ telle que $\lim\limits_{n \to \infty} \Phi_n(\xi) = \Phi(\xi)$ sauf peut-être pour les points de discontinuité [1].

Si en particulier les nombres aléatoires $X_1, X_2, \ldots, X_n, \ldots$ sont les indicateurs d'épreuves équivalentes d'un phénomène donné, c'est-à-dire s'ils correspondent à des événements équivalents $E_1, E_2, \ldots, E_n, \ldots$, les hypothèses seront satisfaites; il suffirait même que les événements fussent également probables $[\mathbf{P}(E_i) = \mathfrak{M}(X_i) = m_1 = \omega_1]$ et ayant la même corrélation deux à deux $[\mathbf{P}(E_i E_j) = \mathfrak{M}(X_i X_j) = m_2 = \omega_2]$. Remarquons que pour les indicateurs on a $X^2 = X$ (car $0^2 = 0$, $1^2 = 1$) ainsi que $\mu_2 = m_1 = \omega_1$. Pour Y_h, fréquence sur h épreuves, on a donc

$$(10) \qquad \mathfrak{M}(Y_h) = \omega_1, \qquad \mathfrak{M}(Y_h - Y_k)^2 = \left(\frac{1}{h} + \frac{1}{k} - \frac{2r}{hk}\right)(\omega_1 - \omega_2),$$

et les résultats démontrés signifient simplement que les fréquences sur deux groupes suffisamment nombreux d'épreuves sont, presque sûrement, très voisines [même s'il s'agit de deux groupes disjoints ($r = 0$); s'il y a des épreuves communes ($r > 0$), encore mieux]. Ces mêmes résultats signifient encore que les fréquences successives dans une même suite d'épreuves oscillent presque sûrement entre une quantité moindre qu'un ε donné, à partir d'un rang h suffisamment grand, quel que soit le nombre d'épreuves ultérieures; et enfin qu'il existe une loi de proba-

---

[1] Remarquons que, si les X_i sont équivalents, la loi $\Phi_n(\xi)$ est la même pour toutes les moyennes Y_n de n termes; on a alors une suite de fonctions Φ_n dépendant seulement de n et tendant vers Φ; avec les hypothèses moins restrictives que suppose la démonstration, deux moyennes Y'_n et Y''_n formées de termes distincts pourront avoir deux lois Φ'_n et Φ''_n différentes, mais le résultat subsistera dans le sens que toutes les $\Phi_n(\xi)$ relatives à la moyenne de n termes quelconques différeront très peu de $\Phi(\xi)$ (et donc entre elles) si n est suffisamment grand.

bilité $\Phi(\xi)$ ne différant que très peu de celle d'une fréquence Y_h pour h très grand.

Pour déterminer complètement la loi-limite $\Phi(\xi)$, la connaissance de m_1, m_2, μ_2 ne suffit évidemment plus, sauf dans le cas limite de la non-corrélation deux à deux, $m_2 = m_1^2$, dans lequel $\Phi(\xi)$ est dégénérée et se réduit à la loi où toute la probabilité est concentrée en un point $\xi = m_1$. Dans ce cas, la loi des grands nombres et la loi forte se réduisent aux lois de Bernoulli et de Cantelli [III], [V], d'après lesquelles l'écart entre Y_h et la valeur m_1, *fixée d'avance*, tend stochastiquement vers zéro, et d'une façon « forte ». Dans le cas général d'une classe de nombres aléatoires équivalents, Φ est déterminée par la connaissance de la suite $m_1 m_2 \ldots m_n \ldots$ tout entière, car ces valeurs sont les *moments* relatifs à la loi Φ :

$$m_n = \int_0^1 \xi^n \, d\Phi(\xi), \tag{11}$$

et donc

$$\psi(t) = \sum_0^\infty{}_n \frac{i^n t^n}{n!} m_n \tag{12}$$

est la fonction caractéristique de Φ.

En effet,

$$Y_h^n = \frac{1}{h^n}(X_1 + X_2 + \ldots + X_h)^n = \frac{1}{h^n}\sum X_{i_1} X_{i_2} \ldots X_{i_n};$$

parmi les h^n produits, il y en a $h(h-1)\ldots(h-n+1)$ qui sont formés par des facteurs distincts; les produits contenant un même terme plus d'une fois constituent donc une fraction de plus en plus négligeable lorsque h augmente, ainsi que

$$\mathfrak{M}(Y_h^n) = \frac{h(h-1)\ldots(h-n+1)}{h^n} m_n + O\left(\frac{1}{h}\right) \to m_n \qquad (h \to \infty). \tag{13}$$

Si, en particulier, les X_i sont les indicateurs d'épreuves équivalentes d'un phénomène, Y_h la fréquence sur h épreuves, m_n est la probabilité ω_n pour que n épreuves aient toutes un résultat favorable : telle est donc la moyenne de la $n^{\text{ième}}$ puissance de la fréquence sur un nombre très grand d'épreuves. La fonction caractéristique de $\Phi(\xi)$ est

$$\psi(t) = \sum_0^\infty{}_n \frac{i^n t^n}{n!} \omega_n \tag{14}$$

et l'on a

$$(15) \qquad \Phi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{it} - e^{it\xi}}{it}\psi(t)\,dt$$

car — les Y_h ayant la signification de fréquences — la distribution de probabilité ne peut être comprise qu'entre o et 1, et donc $\Phi(-1) = 0$. La fonction caractéristique de $\Phi_h(\xi)$ est

$$(16) \qquad \psi_h(t) = \Omega_h\left(e^{i\frac{t}{h}}\right),$$

où Ω_h est le polynome

$$(17) \qquad \Omega_h(z) = \sum_{0}^{h}{}_k \binom{h}{k} \omega_k (z-1)^k,$$

et $\psi_h(t)$ converge uniformément vers $\psi(t)$. Ce fait peut être prouvé directement; c'est par cette voie que j'avais développé systématiquement l'étude des événements équivalents dans mes premiers travaux [29], [40], et démontré l'existence de la loi-limite Φ et de ψ, que j'appelais la « fonction caractéristique du phénomène » (1).

Donner la loi limite Φ, ou la fonction caractéristique ψ, équivaut donc, comme on le voit, à donner la suite des ω_n; cela suffit par conséquent pour déterminer la probabilité de tout problème relatif à des événements équivalents. Tout problème se ramène, en effet, dans le cas des événements équivalents, aux probabilités $\omega_r^{(n)}$ pour que, sur n épreuves, un nombre r quelconque soient favorables; on a (en posant $s = n - r$)

$$(18) \qquad \omega_r^{(n)} = (-1)^s \Delta^s \omega_r = \binom{n}{r}\int_0^1 \xi^r (1-\xi)^s \, d\Phi(\xi),$$

et une formule analogue ayant la même signification est valable aussi pour le cas général. Soit, en effet, $\mathbf{P}_\xi(\mathrm{E})$ la probabilité qu'on attribuerait à un événement générique E si l'on considérait les événements $E_1, \ldots, E_n, \ldots$ indépendants et également probables avec probabilité ξ; la probabilité $\mathbf{P}(\mathrm{E})$ du même événement générique, les E_i étant équivalents

(1) J'avais réservé alors le nom de « phénomène » au cas des épreuves équivalentes; je crois maintenant préférable d'employer ce mot dans le sens qu'on lui donne couramment, et spécifier le cas échéant, qu'il s agit d'un phénomène dont les épreuves sont jugées équivalentes.

avec loi-limite $\Phi(\xi)$, est

$$\mathbf{P}(\mathrm{E}) = \int_0^1 \mathbf{P}_\xi(\mathrm{E})\, d\Phi(\xi) \quad (^1). \tag{19}$$

On peut exprimer ce fait en disant que les lois de probabilité $\mathbf{P}$ correspondant au cas des événements équivalents, sont des combinaisons linéaires des lois $\mathbf{P}_\xi$ correspondant au cas des événements indépendants équiprobables, les poids dans la combinaison linéaire étant exprimés par $\Phi(\xi)$.

Cette conclusion met en évidence un fait intéressant qui rapproche notre cas d'un schéma bien connu, avec lequel même il coïncide du point de vue formel. Si l'on a un phénomène à épreuves équivalentes, et si Φ est la loi-limite des fréquences, on peut imaginer facilement un schéma qui donne pour tout problème les mêmes probabilités relatives à ce phénomène ; il suffit de considérer un nombre aléatoire X dont la loi de probabilité est Φ et des événements qui, conformément à l'hypothèse $\mathrm{X} = \xi$ (ξ valeur quelconque entre 0 et 1) sont indépendants avec une probabilité $p = \xi$; les épreuves d'un phénomène ainsi construit seront toujours des événements équivalents. Nous analyserons plus loin la signification de ce résultat, après avoir examiné son extension aux nombres aléatoires équivalents. Pour le moment, bornons-nous à en déduire le résultat suivant : pour que Φ puisse représenter la loi-limite correspondant à une classe d'événements équivalents, il est nécessaire et suffisant que la distribution soit comprise entre 0 et 1 [donc que $\Phi(-\varepsilon) = 0$, $\Phi(1+\varepsilon) = 1$ lorsque $\varepsilon > 0$]; en d'autres termes, il faut que les ω_h soient les moments d'une distribution comprise entre 0 et 1, ou encore que $(-1)^s \Delta^s \omega_r \geqq 0$ $(r, s = 1, 2, \ldots)$ comme il résulte de l'expression de $\omega_r^{(n)}$.

Si l'on ne connaît que les probabilités des diverses fréquences sur n épreuves, $\omega_0^{(n)}, \omega_1^{(n)}, \omega_2^{(n)}, \ldots, \omega_n^{(n)}$, la condition pour qu'il puisse exister un phénomène à épreuves équivalentes pour lequel les $\omega_r^{(n)}$ aient les valeurs données, sera évidemment que les $\omega_1, \omega_2, \ldots, \omega_n$ correspondants soient les premiers n moments d'une distribution sur $(0, 1)$;

---

(1) Il est clair que le cas particulier précédent — formule (18) — s'obtient en posant E = « sur n épreuves (données), r résultent favorables »; alors, en effet,

$$\mathbf{P}_\xi(\mathrm{E}) = \binom{n}{r} \xi^r (1-\xi)^{n-r}, \qquad \mathbf{P}(\mathrm{E}) = \omega_r^{(n)}.$$

ces ω_h peuvent être déterminés en fonction des $\omega_r^{(n)}$ par la formule

$$\omega_h = \sum_h^n {}_r\,\omega_r^{(n)} \frac{r!(n-h)!}{n!(r-h)!}\,; \tag{20}$$

enfin la condition pour que $\omega_1 \ldots \omega_n$ soient les premiers n moments d'une distribution sur $(0, 1)$ est que toutes les racines du polynome

$$f(\xi) = \begin{vmatrix} 1 & \xi & \xi^2 & \ldots & \xi^k \\ \omega_0 & \omega_1 & \omega_2 & \ldots & \omega_k \\ \omega_1 & \omega_2 & \omega_3 & \ldots & \omega_{k+1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \omega_{k-1} & \omega_k & \omega_{k+1} & \ldots & \omega_{2k-1} \end{vmatrix} \quad \text{si } n = 2k-1, \tag{21}$$

$$f(\xi) = \begin{vmatrix} 1 & \xi & \xi^2 & \ldots & \xi^k \\ \omega_1 & \omega_2 & \omega_3 & \ldots & \omega_{k+1} \\ \omega_2 & \omega_3 & \omega_4 & \ldots & \omega_{k+2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \omega_k & \omega_{k+1} & \omega_{k+2} & \ldots & \omega_{2k} \end{vmatrix} \quad \text{si } n = 2k \tag{22}$$

se trouvent dans l'intervalle $(0, 1)$ limites comprises (1).

CHAPITRE IV.

NOMBRES ALÉATOIRES ÉQUIVALENTS.

Ainsi que nous l'avons vu, dans un problème quelconque concernant les événements équivalents $E_1, E_2, \ldots, E_n$, la probabilité sera complètement déterminée, soit par la suite des probabilités ω_n, soit par la loi-limite de la fréquence $\Phi(\xi)$ [ou, ce qui revient au même, par la fonction caractéristique $\psi(t)$ correspondante]. Nous avons ainsi complètement caractérisé les familles d'événements équivalents, et nous avons particulièrement mis en lumière la signification essentielle de $\Phi(\xi)$ qui se rattache au résultat fondamental démontré : les lois de probabilité **P**, correspondant au cas d'équivalence, sont les combinaisons linéaires des

(1) Ce résultat découle de Castelnuovo [VII] (*voir* aussi [VIII]), comme nous l'avons noté dans [29].

lois de probabilité $\mathbf{P}_\xi$ correspondant au cas d'indépendance et équiprobabilité (probabilité $= \xi$). On a, en effet,

$$\mathbf{P}(E) = \int \mathbf{P}_\xi(E)\, d\Phi(\xi), \tag{19}$$

et $d\Phi(\xi)$ représente bien la distribution des poids dans la combinaison linéaire.

Nous allons étendre ce résultat fondamental au cas des nombres aléatoires équivalents, pour lesquels par contre nous n'avons démontré jusqu'ici que des théorèmes préliminaires, qui nous ont servi plutôt à en faire découler certains résultats concernant les événements eux-mêmes, qu'à résoudre le problème analogue, c'est-à-dire à caractériser complètement les familles de nombres aléatoires équivalents.

Considérons donc maintenant le cas des nombres aléatoires équivalents, et prenons un exemple pour plus de clarté. Dans l'étude des événements équivalents, nous avons pris comme exemple le cas d'un jeu de pile ou face; admettons maintenant que les $X_1, X_2, \ldots, X_n$ représentent des mesures d'une même grandeur; il suffira que les conditions dans lesquelles sont effectuées les mesures ne présentent aucune dissymétrie apparente qui puisse justifier une dissymétrie dans notre évaluation des probabilités, pour qu'on puisse les considérer comme des nombres aléatoires équivalents.

La transposition à ce cas de notre conclusion antérieure sera moins aisée évidemment que dans le cas des événements, un *nombre aléatoire* n'étant plus caractérisé, du point de vue de la probabilité, par un nombre (probabilité) comme les événements, mais par une fonction (fonction de répartition, par exemple, ou fonction caractéristique, etc.). Le cas d'indépendance et équiprobabilité correspond donc à l'hypothèse de l'indépendance des nombres aléatoires X_i et de l'existence d'une fonction de répartition commune $V(x)$; en appelant $\mathbf{P}_V(E)$ la probabilité attribuable à un événement générique E lorsque l'on considère les X_i comme indépendants et ayant tous la même fonction de répartition V, les combinaisons linéaires seront les lois du type

$$\mathbf{P}(E) = \Sigma c_i \mathbf{P}_{V_i}(E)$$

(avec les poids $c_i > 0$, $\Sigma c_i = 1$); à la limite

$$\mathbf{P}(E) = \int \mathbf{P}_V(E)\, d\mathcal{F}(V), \tag{23}$$

l'intégrale étant étendue à l'espace fonctionnel des fonctions de répartition, et la distribution des poids étant caractérisée par le fonctionnel $\mathcal{F}(V)$, d'une manière qui sera précisée par la suite. Même avant de connaître la signification exacte de cette intégration, on se rend compte immédiatement que si $\mathbf{P}(E)$ est une combinaison linéaire des $\mathbf{P}_V(E)$, on sera dans le cas d'équivalence : il suffit de remarquer que, chaque $\mathbf{P}_V(E)$ donnant la même probabilité aux événements définis d'une façon symétrique (¹) par rapport à $X_1, \ldots, X_n, \ldots$, la même condition sera nécessairement satisfaite par toute combinaison linéaire $\mathbf{P}(E)$; il s'agit donc seulement de prouver la réciproque, c'est-à-dire de démontrer que, dans le cas d'équivalence, $\mathbf{P}(E)$ est nécessairement de la forme $\int \mathbf{P}_V(E)\, d\,\mathcal{F}(V)$(²). La définition de l'intégrale

$$\int f(V)\, d\mathcal{F}(V)$$

que nous devons introduire dans l'espace fonctionnel n'est que la généralisation tout à fait naturelle de l'intégrale de Stieltjes-Riemann (³) : en subdivisant d'une façon quelconque l'espace des fonctions de répartition en un nombre fini de domaines partiels, on considère les expressions $\Sigma \overline{f}_i c_i$ et $\Sigma \underline{f}_i c_i$ où c_i est le poids d'une générique de ces parties, et $\overline{f}_i$, $\underline{f}_i$ respectivement la borne supérieure et inférieure des valeurs prises par

---

(¹) Sont symétriques, en ce sens, par exemple, l'événement E = « le point de coordonnées $(X_1, X_2, \ldots, X_n)$ tombera dans le domaine D » (de l'espace euclidien à n dimensions) et les événements consistant dans la même éventualité pour un des $n!$ points $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$ correspondant aux $n!$ permutations de coordonnées ; en particulier :

(D rectangle) :

$$E = « a_h < X_h < b_h \quad (h = 1, 2, \ldots, n) »$$

et les

$$« a_h < X_{i_h} < b_h \quad (h = 1, 2, \ldots, n) » ;$$

(D sphérique) :

$$E = « \Sigma(X_h - a_h)^2 < \rho^2 » \quad \text{et les} \quad « \Sigma(X_{i_h} - a_h)^2 < \rho^2 » ;$$

(D demi-espace) :

$$E = « \Sigma a_h X_h > a » \quad \text{et les} \quad « \Sigma a_h X_{i_h} > a », \quad \ldots.$$

(²) On peut admettre ce résultat et omettre les développements suivants qui ont pour but de le démontrer et le préciser (jusqu'à la fin du Chapitre IV), sans préjudice pour une vue d'ensemble sur la thèse soutenue dans ces conférences.

(³) Pour les raisons qui nous font estimer l'intégrale de Stieltjes-Riemann mieux appropriée au calcul des probabilités, *voir* [58], et aussi [64].

la fonction f dans ces domaines. La borne inférieure de $\Sigma \overline{f}_i c_i$ et la borne supérieure de $\Sigma \underline{f}_i c_i$, lorsqu'on change la subdivision de toutes les façons possibles, sont respectivement l'intégrale supérieure et inférieure de f, étendues à l'espace fonctionnel des fonctions de répartition par rapport à la distribution de poids $\mathcal{F}$; lorsqu'elles coïncident, leur valeur commune est précisément l'intégrale $\int f(\mathrm{V})\, d\mathcal{F}(\mathrm{V})$ que nous allons examiner de plus près dans ce qui suit.

Nous allons montrer que, dans les conditions qui nous intéressent, cette intégrale existe, et que, pour en déterminer la valeur, il suffit de connaître le poids pour des domaines fonctionnels très simples des fonctions de répartition. Supposons d'abord que $f(\mathrm{V})$ ne dépende que des valeurs

$$y_1 = \mathrm{V}(x_1), \quad y_2 = \mathrm{V}(x_2), \quad \ldots, \quad y_s = \mathrm{V}(x_s)$$

que la fonction V prend sur un ensemble fini donné d'abscisses x_1, x_2, ..., x_s; il en est ainsi de la probabilité $f(\mathrm{V})$ pour que, n nombres aléatoires suivant la loi V tombent dans un domaine D *rectangle*, à savoir que le premier tombe entre x_1 et x'_1, le deuxième entre x_2 et x'_2, ..., le dernier entre x_n et x'_n. Cette probabilité est [1]

$$(24)\qquad \begin{aligned} f(\mathrm{V}) &= [\mathrm{V}(x'_1) - \mathrm{V}(x_1)][\mathrm{V}(x'_2) - \mathrm{V}(x_2)] \ldots [\mathrm{V}(x'_n) - \mathrm{V}(x_n)] \\ &= (y'_1 - y_1)(y'_2 - y_2) \ldots (y'_n - y_n) \qquad (s = 2n). \end{aligned}$$

Il est évident que, pour déterminer l'intégrale d'une telle fonction, il suffit de connaître le poids des domaines fonctionnels définis seulement par les ordonnées $y_1 \ldots y_s$ correspondant aux abscisses $x_1 \ldots x_s$, c'est-à-dire des domaines de l'espace à s dimensions défini par $y_1 \ldots y_s$; si f est une fonction continue des y_i, il suffira même de connaître le poids des domaines définis par les inégalités $y_i < a_i$ $(i = 1, 2, 3, \ldots, s)$, domaines dont la signification est la suivante : ils comprennent les fonc-

(1) Il n'est pas nécessaire de s'occuper particulièrement des points de discontinuité : en effet, une fonction V déterminée est continue presque partout (mieux encore : partout, sauf, au plus, sur un ensemble dénombrable), et de même dans l'intégration, le poids de l'ensemble de fonctions de répartition ayant x pour point de discontinuité est toujours nul, sauf, au plus, pour un ensemble dénombrable de points x; il suffit d'observer que ces points sont les points de discontinuité de $\Phi(x) = \int \mathrm{V}(x)\, d\mathcal{F}(\mathrm{V})$, et que $\Phi(x)$ est fonction de répartition.

tions de répartition V dont la courbe représentative $y = V(x)$ reste au-dessous de chacun des s points (x_i, a_i). Soit $\Phi(x)$ la courbe en escalier dont les points (x_i, a_i) sont les sommets inférieurs; la condition énoncée peut s'exprimer par $V(x) < \Phi(x)$ [pour tout x (1)], et le poids de l'ensemble des fonctions de répartition V telles que $V(x) < \Phi(x)$ sera désigné par $\mathcal{F}(\Phi)$, en donnant ainsi une signification concrète à $\mathcal{F}$, qui représentait jusqu'ici une distribution de poids d'une façon purement symbolique. L'intégrale $\int f(V)\,d\mathcal{F}(V)$ n'est dans le cas envisagé que l'intégrale ordinaire de Stieltjes-Riemann dans l'espace à s dimensions. Si $f(V)$ ne dépend pas seulement des ordonnées de $V(x)$ pour un groupe fini d'abscisses $x_1 \ldots x_s$, nous considérerons le cas où il est possible d'approcher $f(V)$, par excès et par défaut, au moyen de fonctions du type précédent, et de telle façon que la valeur de l'intégrale soit univoquement déterminée par ses valeurs approchées par excès et par défaut. En d'autres mots, il faudra que, ε étant fixé arbitrairement, on puisse trouver deux fonctions $f'(V)$ et $f''(V)$ dépendant d'un nombre fini de valeurs $V(x_i)$, telles que

$$f'(V) \leqq f(V) \leqq f''(V) \qquad \text{et} \qquad \int f'(V)\,d\mathcal{F}(V) > \int f''(V)\,d\mathcal{F}(V) - \varepsilon.$$

Revenons au cas de n nombres aléatoires indépendants obéissant à la loi $V(x)$: si $f(V)$ est la probabilité pour que le point $(X_1, X_2, \ldots, X_n)$ tombe dans un domaine D qui ne se réduit pas à une somme de domaines rectangles, f' et f'' pourront représenter des probabilités analogues pour des domaines D′ contenu dans D, et D″ contenant D, formés d'une somme de domaines rectangles.

Nous n'avons pas besoin de poursuivre jusqu'au bout l'analyse des conditions d'intégrabilité ; nous nous contenterons d'avoir montré qu'elles sont satisfaites dans des conditions suffisamment larges, qui comprennent tous les cas intéressants à considérer. Reprenons maintenant le problème concernant nos nombres aléatoires équivalents $X_1, X_2, \ldots, X_n, \ldots$, pour démontrer l'existence de la fonctionnelle $\mathcal{F}$ ayant une signification analogue au $\Phi(\xi)$ des événements équivalents. Soit V la fonction en escalier dont les sommets inférieurs

---

(1) On conviendra toujours qu'une inégalité comme $f(x) < g(x)$ entre deux fonctions signifie qu'elle subsiste pour tout x (à moins que l'on n'ait précisé dans le cas particulier envisagé qu'il s'agit d'une valeur x bien déterminée).

sont les s points

$$(x_i, y_i) \qquad (i = 1, 2, \ldots, s;\ \ x_{i+1} > x_i;\ \ y_{i+1} > y_i);$$

nous désignerons par $\mathcal{F}_h(\mathrm{V})$ la probabilité pour que, sur h nombres X_1, $X_2, \ldots, X_h$, hy_1 au plus dépassent x_1, hy_2 au plus dépassent $x_2, \ldots, hy_s$ au plus dépassent x_s, ou, en d'autres termes, la probabilité

$$\mathbf{P}\, G_h(x) \leqq V(x)$$

pour que la fonction de répartition $G_h(x)$ des valeurs de X_1, $X_2, \ldots, X_h$ ne dépasse jamais $V(x)$. Précisément la fonction $G_h(x)$ est la « fonction de répartition observée » résultant de l'observation de $X_1 \ldots X_h$; elle représente la courbe en escalier dont l'ordonnée est nulle à gauche du plus petit des h nombres $X_1 \ldots X_h$, égale à $\frac{1}{h}$ entre le plus petit et le deuxième, égale ensuite à $\frac{2}{h}, \frac{3}{h}, \ldots, \frac{(h-1)}{h}$, et enfin égale à l'unité à droite du plus grand des h nombres considérés. Les sauts de $G_h(x)$ ont donc lieu aux points de l'axe des abscisses qui correspondent aux valeurs X_i; avant de connaître ces valeurs, $G_h(x)$ est aléatoire car ces abscisses sont aléatoires.

On démontre sans difficulté, en l'étendant au cas des nombres aléatoires équivalents, un théorème donné par M. Glivenko [1] dans le cas des nombres aléatoires indépendants : il est très probable que, pour h et k suffisamment grands, $G_h(x)$ et $G_k(x)$ diffèrent très peu, et, dans le cas d'une suite de moyennes *successives* $G_h(x)$, $G_{h+1}(x)$, ..., nous avons également une convergence stochastique forte. En divisant l'axe des x par un nombre fini suffisamment grand de points $x_1 \ldots x_s$, la démonstration peut être ramenée à celle donnée pour les propriétés analogues dans le cas d'événements équivalents. Pour un x donné, $G_h(x)$ et $G_k(x)$ donnent en effet respectivement les fréquences Y_h, Y_k pour h et k épreuves de la suite d'événements équivalents $E_i = (X_i < x)$; l'écart entre $G_h(x)$ et $G_k(x)$ a donc une valeur moyenne quadratique inférieure à $\sqrt{\frac{1}{h} + \frac{1}{k}}$ [*voir* formule (10)], et la probabilité pour qu'il dépasse ε peut être rendue aussi petite que l'on voudra en choisissant h et k au delà d'un N suffisamment grand. En prenant N tel que la proba-

---

(1) [XIII], *voir* aussi Kolmogoroff [XVIII], et [45].

bilité d'un écart plus grand que ε soit $> \frac{\theta}{s}$ pour chacune des abscisses

$$x = x_1,\ x_2,\ \ldots,\ x_s,$$

on déduit que, sauf dans des cas de probabilité totale $< \theta$, les deux fonctions $G_h(x)$ et $G_k(x)$ ne différeront pas de plus de ε pour aucune des abscisses $x_1 \ldots x_s$.

Dans ces conditions, la probabilité $\mathcal{F}_k(V - \varepsilon)$ pour que $G_k(x)$ ne dépasse pas la courbe en escalier $V(x) - \varepsilon$ pour aucun x, c'est-à-dire la probabilité d'avoir

$$G_k(x_i) < y_i - \varepsilon \qquad (i = 1, 2, \ldots, s)$$

ne peut par conséquent être supérieure à $\mathcal{F}_h(V) + \theta$, car, pour vérifier les conditions imposées, il faut que $G_h(x)$ ne dépasse $V(x)$ pour aucun x, ou bien qu'on ait $G_h(x) - G_k(x) > \varepsilon$ pour une au moins des abscisses $x_1 \ldots x_s$. On a ainsi

$$\mathcal{F}_k(V - \varepsilon) - \theta \leqq \mathcal{F}_h(V) \leqq \mathcal{F}_k(V + \varepsilon) + \theta \tag{25}$$

(la deuxième inégalité peut être démontrée de la même façon); en définissant la convergence d'une façon appropriée [comme pour les fonctions de répartition (¹)], on en conclut que $\mathcal{F}_h \to \mathcal{F}$; c'est la fonctionnelle $\mathcal{F}$ qui permet de caractériser la famille de nombres aléatoires équivalents envisagée.

Pour démontrer la formule fondamentale

$$\mathbf{P}(E) = \int \mathbf{P}_V(E)\, d\mathcal{F}(V), \tag{23}$$

remarquons que l'on a, pour tout h,

$$\mathbf{P}(E) = \int \mathbf{P}_{h,V}(E)\, d\mathcal{F}_h(V) \tag{26}$$

si $\mathbf{P}_{h,V}(E)$ est la probabilité de E, subordonnée à l'hypothèse

$$G_h(x) = V(x).$$

Si l'événement E dépend des n premiers nombres aléatoires $X_1, \ldots, X_n$ (pour fixer les idées par un exemple simple, imaginons que l'événement E consiste en ce que X_1 soit compris entre a_1 et b_1, X_2 entre a_2

(¹) *Voir* Lévy [XX], p. 194.

et b_2, ..., X_n entre a_n et b_n), il faudra naturellement supposer $h \geqq n$; si h est très grand par rapport à n, il est clair que $\mathbf{P}_{h,V}(E) \cong \mathbf{P}_V(E)$, car la probabilité $\mathbf{P}_{h,V}(E)$ s'obtient en supposant $X_1 \ldots X_n$ choisis au hasard, simultanément (c'est-à-dire sans répétition) entre les h valeurs où $G_h = V$ est discontinue, tandis que $\mathbf{P}_V(E)$ est la probabilité analogue, en prenant cette fois-ci les combinaisons compatibles avec la notion d'indépendance. Le fait de comprendre ou d'exclure les répétitions a une influence de plus en plus négligeable lorsque $h \to \infty$; donc $\mathbf{P}_{h,V}(E) \to \mathbf{P}_V(E)$. Cette relation et la relation $\mathcal{F}_h(V) \to \mathcal{F}(V)$ permettent de démontrer que

$$\mathbf{P}(E) = \int \mathbf{P}_{h,V}(E)\, d\mathcal{F}_h(V) = \int \mathbf{P}_V(E)\, d\mathcal{F}(V).$$

Considérons un type particulier d'événements E, ce qui nous permettra d'analyser la relation entre la distribution fonctionnelle donnée par $\mathcal{F}$, relativement aux nombres aléatoires équivalents X_i et les distributions linéaires $\Phi_x(\xi)$, c'est-à-dire les lois-limites $\Phi(\xi)$, relatives aux événements $E_i = (X_i < x)$. Un événement E appartient au type particulier envisagé s'il exprime une condition dépendant seulement du fait que certains des nombres aléatoires $X_1 \ldots X_n$ sont inférieurs ou supérieurs [1] à un nombre x unique donné; par exemple, E = « X_1, X_3, X_8 sont $> x$, X_2 et X_7 sont $< x$ », E = « parmi les nombres X_2, X_3, X_9, X_{12}, il y en a trois qui sont $> x$ et un seul $< x$ », E = « dans la suite $X_1 X_2 \ldots X_{100}$ il n'y a pas plus que trois nombres consécutifs $> x$ », etc. En d'autres termes, l'événement E est une combinaison logique des $E_i = (X_i < x)$ pour un x unique donné.

La théorie des événements équivalents nous apprend que la probabilité de tout événement E de ce type est complètement déterminée par la connaissance de $\Phi_x(\xi)$, et nous pouvons d'ailleurs l'exprimer à l'aide de $\mathcal{F}(V)$: nous pourrons donc exprimer $\Phi_x(\xi)$ au moyen de $\mathcal{F}(V)$, et l'on a précisément

$$\Phi_x(\xi) = \int_{V(x) < \xi} d\mathcal{F}(V), \qquad d\Phi_x(\xi) = \int_{\xi < V(x) < \xi + d\xi} d\mathcal{F}(V). \tag{27}$$

[1] Le cas de l'égalité a une probabilité nulle, sauf pour des valeurs particulières de x en nombre fini ou dénombrable tout au plus; nous négligeons ce cas pour plus de simplicité, en observant d'ailleurs qu'il n'implique aucune difficulté essentielle, mais seulement la considération de deux valeurs distinctes de la fonction de répartition à gauche et à droite de x.

Soit en effet $E^{(n)}$ l'événement consistant en ce que la fréquence des valeurs $< x$ sur les n premières épreuves $X_1 \ldots X_n$ ne dépasse pas ξ; par définition $\Phi_x(\xi) = \lim \mathbf{P}(E^{(n)})$, et d'ailleurs $\mathbf{P}_V(E^{(n)})$, pour n très grand, est très voisin [1] de o si $V(x) > \xi$, de 1 si $V(x) < \xi$; on aura donc

$$\begin{aligned}(27)\qquad \Phi_x(\xi) &= \lim \int \mathbf{P}_V(E^{(n)})\, d\mathcal{F}(V) \\ &= \int_{V(x)<\xi} 1 . d\mathcal{F}(V) + \int_{V(x)>\xi} 0 . d\mathcal{F}(V) = \int_{V(x)<\xi} d\mathcal{F}(V).\end{aligned}$$

On peut en déduire ou bien obtenir directemeut le résultat suivant : $\omega_r^{(n)}(x)$, la probabilité que r parmi n nombres aléatoires $X_{i_1} \ldots X_{i_n}$ fixés d'avance soient $< x$, est égale à

$$(28)\qquad \omega_r^{(n)}(x) = \binom{n}{r} \int [V(x)]^r [1 - V(x)]^{n-r}\, d\mathcal{F}(V),$$

et en particulier pour $r = n$:

$$(29)\qquad \omega_n(x) = \int [V(x)]^n\, d\mathcal{F}(V), \qquad \omega_1(x) = \int V(x)\, d\mathcal{F}(V);$$

cette dernière formule donne en particulier la probabilité pour qu'un nombre X_i générique, mais fixé, soit inférieur à x, c'est-à-dire la fonction de répartition attribuée à chacun des X_i avant de commencer les épreuves.

Jusqu'ici, dans $\Phi_x(\xi)$, $\omega_r^{(n)}(x)$, $\omega_n(x)$, nous n'avons considéré x que comme un paramètre qui détermine les événements $E_i = (X_i < x)$, mais qui ne varie pas; si, par contre, on considère maintenant ces expressions comme des fonctions de x, on pourra faire certaines remarques qui font apparaître ces expressions sous un jour nouveau. Considérons n parmi les nombres aléatoires donnés : $X_1, X_2, \ldots, X_n$; $\omega_n^{(n)}(x)$ est la probabilité pour qu'aucun de ces n nombres ne dépasse x, et constitue donc la fonction de répartition de la valeur maxima entre $X_1 \ldots X_n$; $\omega_n^{(n)}(x) + \omega_{n-1}^{(n)}(x)$ est d'une façon analogue la fonction de répartition de l'avant-dernier, par ordre croissant, des nombres $X_1 \ldots X_n$;

---

[1] N'oublions pas que $\mathbf{P}_V(E^{(n)})$ est simplement la probabilité d'une fréquence $< \xi$ sur n épreuves indépendantes, avec probabilité constante $p = V(x)$, et a donc la valeur

$$\sum_{\nu < \xi n} \binom{n}{\nu} p^\nu q^{n-\nu} \qquad [q = 1 - p = 1 - V(x)].$$

$\omega_n^{(n)}(x) + \omega_{n-1}^{(n)}(x) + \ldots + \omega_r^{(n)}(x)$ celle du $r^{\text{ième}}$; enfin

$$\omega_n^{(n)}(x) + \ldots + \omega_1^{(n)}(x) = 1 - \omega_0^{(n)}(x)$$

celle du plus petit des X_i. Comme le montre l'identité

$$\omega_1^{(n)}(x) + 2\omega_2^{(n)}(x) + 3\omega_3^{(n)}(x) + \ldots + n\omega_n^{(n)}(x) = n\omega_1(x), \tag{30}$$

la moyenne de ces n fonctions de répartition est $\omega_1(x)$, c'est-à-dire la fonction de répartition d'un quelconque des X_i : ce fait est bien naturel, car, d'après la définition de l'équivalence, chaque nombre X_i a la même probabilité $\frac{1}{n}$ d'être le plus petit, ou le deuxième, ..., le plus grand, et, en général, toutes les permutations des X_i ont la même probabilité $\frac{1}{n!}$ de les disposer par ordre croissant (s'il existe une probabilité différente de zéro pour que les n valeurs ne soient pas distinctes, les modifications à apporter à ces énoncés sont évidentes).

Il existe une relation étroite entre les fonctions de répartition d'un nombre de rang déterminé et la fonction $\Phi_x(\xi)$: par définition, $\Phi_x(\xi)$ est la valeur limite, pour $n \to \infty$, de la probabilité pour que sur n nombres $X_1 \ldots X_n$ il y en ait au plus ξn qui soient $< x$; cette probabilité est égale à $\sum_r \omega_r^{(n)}(x)$, la somme étant étendue aux indices $r < \xi n$. Mais cette somme est la fonction de répartition de celui parmi les nombres $X_1 \ldots X_n$ qui occupe le rang « partie entière de ξn » lorsqu'on les a rangés par ordre croissant : en considérant ξ fixe, $\Phi_x(\xi)$ est donc la fonction de répartition de la valeur de rang $\simeq \xi n$ sur un nombre n très grand de nombres aléatoires donnés.

On voit sans peine que $\Phi_x(\xi)$ est une fonction jamais décroissante de ξ et x, que $\Phi = 0$ si $\xi < 0$ et $\Phi = 1$ si $\xi > 1$ (Φ n'est donc défini substantiellement que sur la bande $0 \leqq \xi \leqq 1$), et enfin que $\Phi \to 0$ et $\Phi \to 1$ respectivement pour $x \to -\infty$ et $x \to +\infty$. Inversement, chaque fonction $\Phi_x(\xi)$ douée de ces propriétés peut être associée d'une infinité de manières à une loi de probabilité de nombres aléatoires équivalents ; une telle fonction $\Phi_x(\xi)$ étant donnée, on peut toujours construire une distribution de poids $\mathcal{F}(V)$ dans l'espace fonctionnel de telle façon que la formule (27) soit vérifiée. La façon la plus simple de le faire est la suivante : soit $V_\lambda(x) = \xi$ l'équation explicite de la ligne de niveau $\Phi_x(\xi) = \lambda$, qui représente, comme il résulte des propriétés de $\Phi_x(\xi)$, une fonction de répartition, et définissons la distribution $\mathcal{F}(V)$ en attribuant le poids $\lambda' - \lambda$ ($\lambda' > \lambda$) à l'ensemble des $V(x)$ tels que

$$V_\lambda(x) < V(x) < V_{\lambda'}(x)$$

pour tout x; de cette façon l'intégration dans l'espace fonctionnel se réduit à une intégrale simple :

$$(31) \qquad \int f(V)\, d\mathcal{F}(V) = \int_0^1 f(V_\lambda)\, d\lambda.$$

Nous avons, par exemple,

$$(32) \qquad \omega_n(x) = \int_0^1 [V_\lambda(x)]^n\, d\lambda = \int_0^1 \xi^n\, d\Phi_x(\xi),$$

ce qui suffit à démontrer que la distribution obtenue satisfait bien à la condition voulue; cela résulte d'ailleurs directement du calcul de

$$(33) \quad \Phi_x(\xi) = \int_{V(x)<\xi} d\mathcal{F}(V) = \int_{V_\lambda(x)<\xi} d\lambda = \{\text{la valeur de } \lambda \text{ pour laquelle } V_\lambda(x) = \xi\}.$$

Cependant, il existe toujours une infinité d'autres distributions $\mathcal{F}(V)$ correspondant à la même fonction $\Phi_x(\xi)$: il suffit d'observer par exemple que si l'on pose d'une façon quelconque

$$\Phi_x(\xi) = c_1 \Phi_x^{(1)}(\xi) + c_2 \Phi_x^{(2)}(\xi) + \ldots + c_k \Phi_x^{(k)}(\xi)$$

avec $c_i > 0$, $\Sigma c_i = 1$, les $\Phi_x^{(i)}(\xi)$ satifaisant aux mêmes propriétés que pour Φ, et si l'on introduit de même $V_\lambda^{(i)}(x)$ on aura toujours

$$(34) \qquad \omega_n(x) = \sum c_i \int_0^1 [V_\lambda^{(i)}(x)]^n\, d\lambda.$$

La fonction $\Phi_x(\xi)$ caractérise donc bien toutes les familles d'événements équivalents $E_i = (X_i < x)$ pour x quelconque, mais cela ne suffit pas dans les problèmes où l'interdépendance de ces diverses familles entre en jeu : la connaissance complète de la distribution $\mathcal{F}(V)$ dans l'espace fonctionnel est alors indispensable.

On remarquera encore que si l'on considérait des « éléments aléatoires équivalents » d'un espace quelconque, on arriverait à des résultats parfaitement analogues : implicitement, nous avons déjà considéré en effet des fonctions aléatoires équivalentes, car, par exemple, les $G_h(x)$, fonctions de répartition de $X_{i_1}, X_{i_2}, \ldots, X_{i_h}$, constituent, en considérant tous les groupes $i_1 i_2 \ldots i_h$ possibles, une famille de fonctions aléatoires toutes équivalentes. Le résultat général qui a été établi pour les événements et pour les nombres aléatoires, et qu'on pourrait démontrer pour des éléments aléatoires d'un espace quelconque, s'énonce en disant que *les lois de probabilité d'une classe d'éléments aléatoires équivalents sont les « moyennes » entre des lois de probabilité relatives au cas d'indépendance.*

CHAPITRE V.

CONSIDÉRATIONS SUR LA NOTION D'ÉQUIVALENCE.

Nous avons ainsi établi la notion générale d'équivalence stochastique, et obtenu le résultat fondamental qui permet de caractériser les lois de probabilité correspondant au cas d'équivalence, comme des combinaisons linéaires de lois correspondant au cas d'indépendance et d'équiprobabilité. Essayons maintenant de montrer comment cela serait interprété d'après les conceptions courantes, exposer les raisons qui nous empêchent de partager une telle opinion, et expliquer la signification de la notion d'équivalence ainsi que le rôle qu'elle est appelée à jouer à notre point de vue dans le calcul des probabilités.

Considérons par exemple une urne de composition inconnue et tirons-en des boules. Le rapport du nombre des boules blanches au nombre total, peut avoir diverses valeurs possibles ρ_i, auxquelles on attribue certaines probabilités c_i. Ces tirages sont des événements équivalents, et la probabilité d'un certain fait est $\mathbf{P}(E) = \Sigma_i c_i \mathbf{P}_i(E)$, s'il y a une probabilité $\mathbf{P}_i(E)$ correspondant à la composition ρ_i; vice-versa, comme nous l'avons déjà remarqué, si les E_i sont équivalents, la loi $\mathbf{P}(E)$ est toujours de la forme $\mathbf{P}(E) = \Sigma_i c_i \mathbf{P}_i(E)$ $(c_i > 0, \Sigma_i c_i = 1)$ ou de la forme limite $\mathbf{P}(E) = \int \mathbf{P}_\xi(E)\, d\Phi(\xi)$ (intégrales de Stieltjes), où $\mathbf{P}_\xi$ est la loi de probabilité correspondant à l'hypothèse d'indépendance et de probabilité constante ξ. Les « événements équivalents » correspondent donc à ceux qu'on qualifierait ordinairement d' « événements indépendants à probabilité constante p mais inconnue », et $\Phi(\xi)$ serait « la probabilité pour que cette « probabilité inconnue » p soit plus petite que ξ ». D'une façon analogue, les « nombres aléatoires équivalents » correspondent à ceux qu'on appellerait « nombres aléatoires indépendants avec loi de probabilité $G(x)$ constante mais inconnue », et la fonctionnelle $\mathcal{F}$ donnerait « la distribution de probabilité de cette loi inconnue », $\mathcal{F}(V)$ étant la probabilité pour que $G(x) \leqq V(x)$ [1] (pour V en escalier, avec sauts en nombre fini). Mais alors, dira-t-on, pourquoi partir d'une définition nouvelle et faire tant d'efforts pour conclure qu'elle ne caractérise rien d'autre que ce cas bien connu?

Ce n'est pas sans raison que nous nous sommes vus obligés à procéder de cette façon. La définition ancienne ne pouvait en fait être dépouillée de son caractère, pour ainsi dire, « métaphysique » : on était obligé de supposer que, au delà de la loi de probabilité correspondant à notre jugement, il devait y en avoir une autre, inconnue, correspondant à quelque chose de réel, et que les diverses hypothèses sur cette loi inconnue — d'après laquelle les diverses épreuves ne seraient plus dépendantes mais indépendantes — constitueraient des *événements* dont on pourrait considérer la probabilité. De notre point de vue, ces phrases sont complètement dénuées de sens, et on n'en a jamais donné de justification qui puisse paraître satisfaisante, même par rapport à un point de vue différent. Si nous considérons le cas de l'urne dont la composition est inconnue, nous pourrons sans aucun doute parler de la probabilité des diverses compositions et des probabilités subordonnées à une de telles compositions; en effet, l'affirmation qu'il y a tant de boules blanches et tant de noires dans l'urne exprime un fait objectif qu'on peut vérifier directement, et la probabilité subordonnée à un événement objectif donné a bien été définie : si, par contre, on joue à pile ou face avec une pièce d'apparence irrégulière, comme dans l'exemple du Chapitre III, on n'a pas le droit de considérer comme hypothèses distinctes la supposition que cette imperfection ait une influence plus ou moins sensible sur la « probabilité inconnue », car cette « probabilité inconnue » ne peut pas être définie, et les hypothèses que l'on voudrait introduire de cette façon n'ont aucune signification objective. La différence entre ces deux cas est essentielle et l'on ne peut pas la négliger; on ne peut pas, dans le deuxième cas, reprendre « par analogie » des raisonnements qui étaient valables dans le premier pour des raisons qui ne subsistent plus dans

---

(1) *Voir* note (1), page 41.

l'autre. Si, après de nombreux tirages, la fréquence observée des boules blanches est f, pourquoi attribuerions-nous une valeur voisine de f à la probabilité pour que la boule soit blanche dans les tirages qui vont suivre? On peut répondre qu'après l'observation d'une telle fréquence, nous attribuons une valeur très grande à la probabilité pour que le nombre des boules blanches constitue à peu près la fraction f du total, et de plus, en admettant que cette fraction soit ρ, nous jugeons que les tirages sont indépendants et ont tous la même probabilité $p = \rho$. Cette explication est parfaitement satisfaisante même du point de vue subjectif, et ne diffère pas formellement de celle qu'on donne d'habitude et qui se réduit au fond au théorème des probabilités composées. Mais, dans le cas précédent de pile ou face, il en va autrement : les termes correspondants qui rendraient possible la traduction de ce raisonnement n'existent pas; si, néanmoins, nous voulons raisonner d'une façon identique et rigoureuse dans les deux cas, il faudra d'abord chercher quels sont les éléments communs qui les caractérisent l'un et l'autre et quels sont ceux qui les différencient.

Le résultat auquel nous sommes parvenus nous donne la réponse cherchée, qui est très simple et très satisfaisante : la définition nébuleuse et inexacte d' « événements indépendants avec probabilité fixe mais inconnue » doit être remplacée par celle d' « événements équivalents ». Cette réponse fournit une condition qui s'applique directement aux évaluations des probabilités d'événements individuels et ne se heurte donc à aucune des difficultés que la conception subjective se propose d'éliminer; elle constitue une condition très naturelle et très claire, de caractère purement qualitatif, se réduisant à exiger que certains événements soient jugés également probables, ou, d'une façon plus précise que toutes les combinaisons de n événements $E_{i_1}, \ldots, E_{i_n}$ aient toujours la même probabilité, quel que soit le choix ou l'ordre des E_i. La même condition simple de « symétrie » par rapport à nos jugements de probabilité, définit les nombres aléatoires équivalents et peut définir en général les éléments aléatoires équivalents dans un espace quelconque; elle conduit dans tous les cas aux mêmes conclusions pratiques : une expérience assez riche nous amène toujours à considérer comme probables des fréquences ou distributions futures voisines de celles observées.

Après la démonstration de l'existence d'une loi-limite, ce fait peut être expliqué par un raisonnement presque parallèle à celui que l'on emploie ordinairement, lorsqu'on tient compte de la « probabilité inconnue », mais qui ne donne pas prise aux mêmes critiques. A la « probabilité inconnue », substituons la fréquence sur les premières N épreuves, avec N assez grand pour que la loi de probabilité correspondante coïncide pratiquement avec la loi-limite : $\Phi_N(\xi) \cong \Phi(\xi)$, et pour que le nombre des épreuves dont on a à s'occuper soit négligeable par rapport à N; de la sorte $\binom{R}{r}\binom{S}{s} : \binom{N}{n}$ (probabilité pour que, sur n épreuves choisies au hasard parmi les N, dont R sont favorables et S défavorables, il y en ait r de favorables et s défavorables) est pratiquement égal à $\binom{n}{r}\xi^r(1-\xi)^s$ avec $\xi = \frac{R}{N}$, $1 - \xi = \frac{S}{N}$. On peut raisonner alors comme suit : considérer comme hypothèses possibles les $N+1$ fréquences possibles sur N épreuves, leurs probabilités étant les $\omega_h^{(N)}$ $(h = 0, 1, \ldots, N)$, et constater que les hypothèses pour lesquelles $\frac{R}{N}$ est voisin de $\frac{r}{n}$ sont précisément celles d'après lesquelles la probabilité d'une fréquence $\frac{r}{n}$ sur n épreuves est la plus forte, et, par conséquent [1], celles dont la probabilité subordonnée à l'observation de la fréquence $\frac{r}{n}$ sur n épreuves est le plus fortement augmentée par rapport à la probabilité non subordonnée, conclure finalement que les hypothèses les plus voisines du résultat observé prennent une importance de plus en plus prépondérante lorsque le nombre des observations croît, et que cela nous conduit nécessairement à approcher notre prévision de l'observation. On démontre d'une façon plus précise que la loi-limite $\overline{\Phi}$ subordonnée à l'observation de la fréquence $\frac{r}{n}$ sur n épreuves, est telle que

$$(35) \qquad d\overline{\Phi}(\xi) = \alpha\xi^r(1-\xi)^s d\Phi(\xi) \qquad \left[\alpha \text{ tel que} \int_0^1 \alpha\xi^r(1-\xi)^s d\Phi = 1\right]$$

et que la fonction caractéristique correspondante est

$$(36) \qquad \overline{\psi}(t) = \alpha D^r(i - D)^s \psi(t) \qquad \left[\alpha \text{ tel que } \overline{\psi}(0) = 1;\ D = \frac{d}{dt},\ i = \sqrt{-1}\right].$$

En particulier, la probabilité d'une épreuve ultérieure, subordonnée à

[1] *Voir* formule (2), p. 15, et explications relatives.

cette hypothèse, sera

$$p_r^{(n)} = \int \xi\, d\overline{\Phi} = \int \xi . \alpha \xi^r (1-\xi)^s d\Phi, \tag{37}$$

c'est-à-dire la moyenne des ξ de $(0,1)$ avec les poids $\alpha\xi^r(1-\xi)^s d\Phi$ au lieu des poids $d\Phi$; les ξ autour du maximum $\xi = \frac{r}{n}$ de $\xi^r(1-\xi)^s$ sont évidemment de plus en plus renforcés.

En fonction des $\omega_h^{(n)}$ on tire facilement

$$p_r^{(n)} = \frac{r+1}{n+1}\,\frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} = \frac{r+1}{n+2+(s+1)\left(\dfrac{\omega_r^{(n+1)}}{\omega_{r+1}^{(n+1)}} - 1\right)}; \tag{38}$$

il est intéressant de voir que cette formule — qui permet déjà d'expliquer, bien qu'incomplètement, l'influence de la fréquence observée sur l'évaluation d'une probabilité — peut être déduite d'une façon directe et très élémentaire de la définition d' « événements équivalents ».

De cette définition on déduit en effet que

$$\frac{\omega_r^{(n)}}{\binom{n}{r}} = \frac{\omega_r^{(n+1)}}{\binom{n+1}{r}} + \frac{\omega_{r+1}^{(n+1)}}{\binom{n+1}{r+1}}, \tag{39}$$

car le premier membre exprime la probabilité que r données sur n épreuves soient favorables, et le deuxième donne la somme des probabilités pour que la dite combinaison se vérifie avec, respectivement, un résultat favorable ou défavorable dans la $(n+1)$-ième épreuve. En simplifiant, on a

$$\omega_r^{(n)} = \frac{s+1}{n+1}\omega_r^{(n+1)} + \frac{r+1}{n+1}\omega_{r+1}^{(n+1)} \tag{40}$$

et à l'aide de cette identité on obtient

$$(38)\quad p_r^{(n)} = \frac{\omega_{r+1}^{(n+1)}}{\binom{n+1}{r+1}} : \frac{\omega_r^{(n)}}{\binom{n}{r}} = \frac{r+1}{n+1}\,\frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} = \frac{(r+1)\omega_{r+1}^{(n+1)}}{(s+1)\omega_r^{(n+1)} + (r+1)\omega_{r+1}^{(n+1)}}$$

$$= \frac{r+1}{n+2+(s+1)\left(\dfrac{\omega_r^{(n+1)}}{\omega_{r+1}^{(n+1)}} - 1\right)} \qquad \text{C. Q. F. D}$$

Cette formule acquiert une signification particulière dans le cas de Laplace, où les $\omega_r^{(n)}$ ne dépendent pas de r, et où l'on a donc

$\omega_r^{(n)} = \frac{1}{n+1}$ (on peut vérifier immédiatement que cette hypothèse est admissible et qu'elle correspond à la loi-limite homogène : $\Phi(\xi) = \xi$, $d\Phi(\xi) = d\xi$, pour $0 \leqq \xi \leqq 1$). Dans le cas de Laplace on a simplement, comme on le sait bien, $p_r^{(n)} = \frac{r+1}{n+2}$, car l'autre terme du dénominateur s'évanouit; le résultat est encore le même si $\omega_{r+1}^{(n+1)} = \omega_r^{(n+1)}$; si par contre $\omega_{r+1}^{(n+1)}$ est plus grand ou plus petit que $\omega_r^{(n+1)}$, la probabilité $p_r^{(n)}$ sera respectivement plus grande ou plus petite que dans le cas de Laplace. En tous cas, $p_r^{(n)}$ est voisin de $\frac{r+1}{n+2}$, donc voisin de la fréquence $\frac{r}{n}$, si le rapport $\frac{\omega_r^{(n+1)}}{\omega_{r+1}^{(n+1)}}$ diffère peu de l'unité; il suffit donc d'admettre cette condition pour justifier rapidement l'influence de l'observation sur la prévision, dans le cas d'événements équivalents.

En ce qui concerne les nombres aléatoires équivalents ou les éléments équivalents quelconques, les résultats et les démonstrations seraient parfaitement analogues; pour ceux-ci de même que pour les événements, la théorie subjective résout ainsi complètement le problème de l'induction dans le cas de l'équivalence, correspondant au cas qui est le plus ordinairement considéré, et conduit aux mêmes conclusions généralement admises ou démontrées au moyen de raisonnements vagues et imprécis.

Toutes les autres hypothèses peuvent cependant être étudiées au même titre et par le même procédé. On pourrait ne pas exclure du tout à priori une influence de l'ordre des épreuves, et attribuer par conséquent des probabilités plus ou moins différentes entre elles aux $\binom{n}{r}$ diverses combinaisons de r résultats favorables et s défavorables sur les $n = r + s$ premières épreuves. Il y aurait alors un nombre de degrés de liberté et donc une complication bien plus grande, mais rien ne serait à changer dans la conception et la position problème, qui demeurent celles exposées au début du Chapitre III, avant de restreindre nos considérations au cas des événements équivalents, et qui sont substantiellement condensées dans la formule (4).

L'influence de l'ordre sur l'évaluation des probabilités de A et $A.E_{n+1}$ (*voir*, p. 26) ne modifie pas, en effet, la façon dont le problème est posé et résolu d'après la conception subjective; on sera seulement amené, dans le cas général, à tenir compte de circonstances qui dans le cas de l'équivalence sont à négliger par définition. On pourra tenir compte en effet non plus seulement de la fréquence observée, mais aussi de régularités ou de tendances à certaines régularités que l'observation peut révéler. Supposons, par exemple, que les n premières épreuves donnent alternativement un résultat favorable et un autre défavorable. Dans le cas de l'équivalence, notre prévision pour le coup suivant sera la même après ces n épreuves qu'après toute autre expérience de même fréquence égale à $1/2$ (¹), mais avec une succession tout à fait irrégulière des différents résultats : c'est, en effet, l'absence de toute influence de l'ordre sur les jugements d'un certain individu qui caractérise par définition les événements qu'il considère « équivalents ». Dans le cas où les événements ne seraient pas conçus comme équivalents, nous serions par contre amenés en général à modifier de façon bien différente nos prévisions après n coups à résultats alternés, qu'après n coups irrégulièrement disposés de même fréquence égale à $1/2$; l'attitude la plus naturelle consistera notamment à prévoir que le prochain coup aura une grande probabilité de présenter un résultat opposé au précédent.

Il serait sans doute possible et intéressant d'étudier cette influence de l'ordre dans des hypothèses simples, par des généralisations plus ou moins larges du cas de l'équivalence et des développements qui s'y rattachent (²), mais cette étude est encore à faire. Ce qui est essentiel par rapport à notre conception — et c'est sur cela que nous voulons insister encore quelque peu — nous est déjà appris par la théorie des événements équivalents : quelle que soit l'influence de l'observation sur la prévision future, elle n'implique et ne signifie nullement que nous *corrigeons* l'évaluation primitive de la probabilité $\mathbf{P}(E_{n+1})$ qui a été *démentie* par l'expérience, en lui substituant une autre $\mathbf{P}^*(E_{n+1})$ qui est *conforme* à cette expérience et donc probablement plus *voisine de la probabilité réelle;* au contraire, elle se manifeste seulement dans le sens que, lorsque l'expérience nous apprend le résultat A des n premières épreuves, notre jugement sera exprimé non plus par la probabilité $\mathbf{P}(E_{n+1})$, mais par la probabilité $\mathbf{P}\left(\frac{E_{n+1}}{A}\right)$, à savoir celle que notre opinion initiale attribuait

(¹) Pour que cela soit exact, supposons $n =$ pair.

(²) On pourait en premier lieu considérer le cas des classes d'événements qui se comporteraient par rapport aux « chaînes » d'ordre $1, 2, \ldots, m, \ldots$, de la même façon que les classes d'événements équivalents par rapport aux classes d'événements équiprobables et indépendants.

déjà à l'événement E_{n+1} considéré comme dépendant de l'éventualité A. Rien donc de cette opinion initiale n'est répudié ou corrigé : ce n'est pas la fonction **P** qui a été modifiée (remplacée par une autre $\mathbf{P}^*$), mais bien l'argument E_{n+1} qui a été remplacé par $\frac{E_{n+1}}{A}$, et c'est précisément pour demeurer fidèles à l'opinion initiale (telle qu'elle se manifeste dans le choix de la fonction **P**) et cohérents dans notre jugement que nos prévisions varient lorsqu'un changement a lieu dans des circonstances connues.

De la même façon, celui qui possède le n° 2374 d'une loterie avec 10000 billets, s'attribuera au début une probabilité 1/10000 de gagner le premier lot, mais évaluera successivement sa probabilité à 1/1000, 1/100, 1/10, 0, s'il assiste à l'extraction des chiffres successifs qui donnent — par exemple — le n° 2379. A chaque instant son jugement est parfaitement cohérent, et il n'aurait aucune raison de se dire à chaque tirage que la précédente évaluation de probabilité n'était pas juste (du moment où elle était faite). En fin de compte, chaque évaluation de probabilité différente de 0 et 1 sera sûrement abandonnée, car un événement (bien déterminé) peut seulement se vérifier ou ne pas se vérifier; une évaluation de probabilité n'a de sens que tant et jusqu'au moment où un individu ne connaît pas le résultat de l'événement envisagé; avant qu'il ne connaisse ce résultat (et qu'il soit ainsi amené à la valeur définitive 0 ou 1), il pourra apprendre successivement plusieurs circonstances qui feront modifier son jugement, tantôt dans un sens et tantôt dans l'autre, sans qu'il s'agisse là d'une correction ou d'un rejet. C'est de façon parfaitement identique que nous envisageons l'influence de l'observation sur la prévision dans le cas général des jugements fondés sur l'expérience.

C'est ainsi que lorsqu'on adopte le point de vue subjectif, le problème de l'induction reçoit une réponse qui est naturellement subjective mais parfaitement logique en soi, tandis qu'au contraire, lorsqu'on prétend *éliminer* les facteurs subjectifs, on réussit seulement (c'est, du moins, mon opinion) à les *cacher*, plus ou moins habilement, en n'échappant jamais à une erreur de logique. Il est vrai que dans plusieurs cas — comme par exemple dans l'hypothèse de l'équivalence — ces facteurs subjectifs n'ont jamais une influence trop prononcée pourvu que l'expérience soit assez riche; cette circonstance est très importante car elle explique bien comment dans certaines conditions il se produit une concordance plus ou moins étroite entre les prévisions de divers individus (1), mais elle montre aussi bien que des opinions discordantes sont toujours légitimes. Cela ne porte donc nullement atteinte au caractère purement subjectif de l'ensemble de la théorie des probabilités.

Nous reviendrons dans le prochain et dernier chapitre sur ces questions très générales de principe, après un résumé de tout ce que nous avons dit et des considérations ultérieures qui mettront mieux en lumière leurs raisons d'être et leur portée.

CHAPITRE VI.

OBSERVATION ET PRÉVISION.

Le besoin de clarté dans la pensée scientifique et philosophique ne nous est jamais apparu aussi essentiel qu'aujourd'hui : l'analyse critique la plus poussée des concepts les plus évidemment intuitifs ne peut plus être considérée comme un jeu pour sophistes, mais comme une des questions qui touchent le plus directement le progrès de la Science. A chacune de nos affirmations, une question surgit invariablement dans notre esprit : cette affirmation a-t-elle effectivement un sens? Pour ne donner qu'un exemple, nous savons bien que la notion de simultanéité semblait, il n'y a pas très longtemps encore, parfaitement claire et sûre, à tel point qu'on avait même cru pouvoir considérer le temps comme une notion donnée à priori. Pourquoi ne le croyons-nous plus aujourd'hui? Parce que nous avons appris à connaître la nécessité de ne concevoir une notion qu'au point de vue qu'on peut appeler « opératif » (2) : toute notion n'est qu'un mot dépourvu de sens tant qu'on ne sait pas comment vérifier pratiquement un énoncé quelconque

---

(1) C'est le même point de vue sur lequel Poincaré a plusieurs fois insisté (et duquel s'inspirent ses exemples bien connus de la roulette, du battage des cartes, de la distribution des petites planètes, etc.) (*voir*, par exemple [XVIII]); la seule différence est dans le fait que nous n'admettons pas de concevoir que notre opinion initiale concerne des « lois inconnues ».

(2) *Voir*, par exemple, Bridgman, [II], et notamment les paragraphes *Der operative Charakter der Begriffe* et *Allgemeine Bemerkungen zum operativen Gesichtspunkt* (p. 3-22).

où cette notion intervient; dans l'exemple donné plus haut, cette vérification pratique nous est fournie par le procédé d'Einstein relatif aux signaux lumineux. Une évolution analogue a déjà eu lieu depuis longtemps dans les sciences mathématiques : autrefois, par exemple, le problème de savoir si l'on a $1-1+1-1+\ldots=1/2$ ou non, était considéré sous un aspect nébuleux, mystérieux, métaphysique; il a suffi de définir ce qu'on veut entendre par « limite » (par exemple limite ordinaire, limite dans le sens de Cesàro) et toutes les obscurités se sont évanouies.

Il est tout à fait naturel que ce même besoin de clarté soit profondément ressenti dans le domaine des probabilités, soit parce que cette notion est très intéressante du point de vue mathématique comme du point de vue expérimental, soit parce qu'elle semble peut-être plus rebelle à tout effort de précision. En ce qui concerne le côté mathématique de la question, les opinions ne semblent pas par trop discordantes : formellement ([1]), la théorie des probabilités est la théorie des fonctions additives et non négatives d'événements; les opinions divergent seulement sur un point : à savoir sur la question si ces fonctions ont besoin d'être simplement ou complètement additives (c'est-à-dire additives seulement sur les ensembles finis ou bien sur les ensembles dénombrables). Cependant, le côté vraiment essentiel du problème est naturellement la question de la signification et de la valeur de la notion de probabilité, et sur ce terrain les opinions divergent fortement. Deux points de vue nettement opposés sont possibles : le premier, le plus communément admis, considère les éléments subjectifs de la notion naïve de probabilité qui intervient dans notre vie quotidienne, comme des éléments dangereux qu'il convient d'éliminer pour que la notion de probabilité puisse atteindre un niveau vraiment scientifique; le point de vue opposé considère par contre les éléments subjectifs comme essentiels, et estime qu'on ne peut les éliminer sans enlever à la notion et à la théorie des probabilités toute raison d'être. La différence entre les deux points de vue est également très nette du côté philosophique : selon l'un, la probabilité est un élément qui fait partie du monde physique et existe en dehors de nous; selon l'autre, elle n'exprime que

---

([1]) *Voir*, par exemple, Cantelli [IV], [V], Kolmogoroff [XVII], Lomnicki [XXI], etc.

l'opinion d'un individu, et ne peut pas avoir de signification sinon en fonction de celui-ci.

L'un et l'autre de ces deux points de vue cherchent à donner un sens bien défini aux affirmations sur les probabilités : seuls les domaines dans lesquels ces concepts devraient recevoir un sens sont complètement différents. Donner un sens vérifiable aux affirmations de probabilité dans le monde extérieur, voudrait dire ne pas les considérer comme quelque chose d'effectivement nouveau, mais comme des affirmations particulières concernant le monde physique, par exemple, des affirmations sur la limite de certaines fréquences. Cependant, si l'on voulait interpréter les exigences du point de vue opératif dans le seul cadre du monde extérieur, d'une façon que l'on pourrait appeler positiviste, je crois que le but de rendre claires toutes nos notions ne pourrait jamais être complètement atteint. Nous sommes parfois amenés à faire une affirmation qui a un sens purement subjectif, et cela est bien légitime; mais, si l'on cherche à la remplacer ensuite par quelque chose d'objectif, on ne fait pas un progrès, mais bien une erreur. Mieux qu'en cherchant de tout ramener à l'objectif, on pourrait atteindre la clarté en réduisant systématiquement tout concept au subjectif; la valeur d'un concept résulterait alors de l'analyse des raisons profondes et essentielles qui nous ont poussés, peut-être inconsciemment, à l'introduire, et qui nous fournissent l'explication de son utilité.

Ce point accepté, il ne devrait pas être difficile de se persuader que la définition basée sur la fréquence-limite [1] est loin d'éclaircir la notion de probabilité; en effet, même si l'on accepte une telle définition, on n'appliquera pas pour cela le calcul des probabilités dans le but de connaître certaines valeurs limites de fréquences : ce but sera toujours celui de juger comme plus ou moins vraisemblable l'arrivée de certains faits, plus ou moins complexes, mais vérifiables dans un temps fini : ce sont ces événements seuls qui nous intéressent, et à leur égard la dite définition ne nous apprend rien. Nous n'appliquons la notion de probabilité que pour faire des prévisions vraisemblables : si je veux justifier par l'observation pratique d'une fréquence, la conviction qu'une fré-

---

[1] *Voir*, par exemple, von Mises [XXV], [XXVI], Reichenbach [XXIX], [XXX], [XXXI]; *voir* encore Dörge [IX] où cette conception a été modifiée à la suite de critiques qui sont justifiées aussi d'après notre point de vue.

quence voisine apparaîtra dans un certain groupe d'épreuves ultérieures, et si, pour cela, je procède en estimant d'abord que la fréquence-limite sera voisine de la fréquence observée, et ensuite qu'il est raisonnable d'attendre une fréquence voisine de la fréquence-limite, je ne fais qu'introduire une notion intermédiaire mystérieuse à travers laquelle les prémisses et la conclusion sont reliées indirectement par deux jugements subjectifs au lieu de l'être directement par un jugement subjectif unique [62], [65]. Il ne me sert donc à rien de donner le nom de probabilité à la fréquence-limite ou à n'importe quelle autre entité objective si le lien de ces considérations avec les jugements subjectifs qui s'y appuient demeure subjectif. Il vaut mieux alors chercher à analyser directement l'élément subjectif dans lequel est ancrée directement la notion de probabilité : c'est cette voie que j'ai suivie.

Je connais très bien les doutes que fait surgir couramment ce point de vue, et c'est pour cette raison que je me suis proposé d'exposer aussi clairement qu'il m'était possible, de quelle manière viennent à être conçus et se posent, suivant la conception subjective, les problèmes pour lesquels les objections habituelles affectent la forme la plus frappante.

Il y a trois objections essentielles : on doute que la conception subjective permette de définir la probabilité, de démontrer les lois logiques qui la régissent et enfin d'expliquer et justifier les applications qu'on en a fait aux problèmes les plus divers. Résumons rapidement nos réponses à ces trois objections.

La définition de la probabilité que nous avons donnée est bien irréprochable du point de vue opératif, pourvu qu'on admette que ce dernier s'applique également dans le domaine psychologique. Le schéma des paris donne en principe une méthode de mesure expérimentale directe du degré de doute relatif à un événement donné; si l'application pratique se heurte parfois à une indétermination de ce degré subjectif de doute qu'on voudrait mesurer, cela n'est qu'une conséquence de ce degré restreint d'idéalisation sans laquelle il serait toujours impossible d'atteindre la précision et de développer une théorie quelconque. L'indétermination est sans doute plus forte ici que dans les sciences physiques à cause du fait que la grandeur à mesurer est subjective, mais la différence n'est pas essentielle; une autre définition, également subjective, et très analogue, qu'il est peut-être utile de rapprocher de la définition subjective de la probabilité, est celle de l'ophélimite de Vilfredo Pareto [XXVII], qui, en la faisant découler de la considération des « courbes d'indifférence », a bien appliqué avec perspicacité le « point de vue opératif » à des faits psychologiques.

Le fait qu'une estimation directe ne soit pas toujours possible, constitue d'ailleurs la raison de l'utilité des règles logiques de la probabilité : leur but pratique n'est que de ramener une évaluation directe peu accessible, à d'autres au moyen desquelles la détermination est plus aisée et plus précise. En adoptant la définition subjective, ces règles logiques découlent avec rigueur et facilité d'une seule condition très naturelle : celle de *cohérence*, qui nous oblige à prendre garde en évaluant les probabilités de ne pas donner la possibilité à un adversaire qui parie contre nous de gagner sûrement, par une judicieuse combinaison de ses mises sur les divers événements, quelle que soit l'éventualité qui se réalise. Les théorèmes fondamentaux (probabilités totales, probabilités composées) ne sont que des corollaires immédiats de cette condition fondamentale. On constate même qu'on pourrait éliminer tout ce qui est quantitatif, soit dans la condition de cohérence, soit dans la définition même de la probabilité, pour n'en conserver que le côté purement qualitatif de la définition (inégalité entre deux probabilités) et de la condition de cohérence (petit nombre d'axiomes très simples). L'application de ces règles logiques se réduit en tout cas à distinguer si, les probabilités étant évaluées arbitrairement pour des événements d'une classe $\mathcal{E}$ en respectant la cohérence, les probabilités d'autres événements sont univoquement déterminées par la condition de cohérence, ou bien s'il existe une limitation, ou enfin si n'importe quelle valeur demeure admissible.

Du point de vue logique, la théorie des probabilités ne serait qu'une logique polyvalente avec une échelle continue de modalités (1), superposée à une logique à deux valeurs. Cela veut dire que, pour chaque événement, on n'admet que deux *résultats* possibles (trois pour les « événements subordonnés », mais cela n'a qu'une signification for-

(1) [62]; selon les opinions de Lukasiewicz, [XXII], Mazurkiewicz, [XXIII], Reichenbach, [XXIX], il s'agirait aussi d'une logique à une échelle continue de modalités, laquelle ne serait pas cependant conçue comme superposée à une logique à deux valeurs. Des critiques, auxquelles se prête ce point de vue, sont développées dans Hosiasson [XIV].

melle); l'infinité de modalités intermédiaires ne découle pas d'une insuffisance de la logique à deux valeurs à cet égard, mais sert seulement à mesurer notre doute lorsque nous ne savons pas encore laquelle des deux modalités objectives est juste.

L'explication subjective des applications les plus importantes du calcul des probabilités constitue le problème le plus délicat. On n'aura pas de difficulté à admettre que l'explication subjective soit la seule applicable dans le cas des prévisions pratiques (résultats sportifs, faits météorologiques, événements politiques, etc.) qu'ordinairement on ne fait pas entrer dans le cadre de la théorie des probabilités, alors élargi. Il sera par contre plus difficile de convenir que cette même explication donne effectivement la raison de la valeur plus scientifique et plus profonde que l'on attribue à la notion de probabilité dans certains domaines classiques, et l'on émettra des doutes sur la possibilité qu'elle offre d'unifier les diverses conceptions de la probabilité appropriées aux divers domaines envisagés et qu'on croyait devoir introduire jusqu'à présent. Notre point de vue reste dans tous ces cas le même : *montrer qu'il y a des raisons psychologiques assez profondes pour rendre très naturelle la concordance exacte ou approchée qu'on observe entre les opinions des divers individus, mais qu'il n'y a pas de raisons rationnelles, positives, métaphysiques, qui puissent enlever à ce fait le caractère d'une simple concordance d'opinions subjectives.*

Le cas des jeux de hasard ne conduit qu'à observer comment les caractères de symétrie présentés par les divers « cas possibles » peuvent forcer notre esprit à les juger également probables, mais non à *imposer* logiquement une telle évaluation de la probabilité. Le cas de la fréquence a besoin par contre d'une analyse soignée, ce qui nous a conduits à des développements mathématiques assez étendus. Des raisonnements analogues, et à peine plus délicats que ceux relatifs aux jeux de hasard, suffisent pour expliquer le lien entre l'évaluation des probabilités de certains événements et la prévision du nombre d'entre eux qui se vérifieront, c'est-à-dire le lien entre l'évaluation des probabilités et la prévision des fréquences. La question essentielle, et la seule qui soit d'ailleurs un peu moins élémentaire, est la justification et l'explication des raisons pour lesquelles dans la prévision d'une fréquence on est généralement guidé ou du moins influencé, par l'observation des fréquences dans le passé. Il s'agissait de démontrer qu'on n'a pas besoin d'admettre, comme on le fait couramment, que *la probabilité d'un phénomène a une valeur déterminée* et qu'il suffit de parvenir à la connaître; au contraire, on peut poser la question d'une façon qui a un sens parfaitement clair du point de vue subjectif, en distinguant d'une part la probabilité d'une épreuve considérée comme isolée, et d'autre part la probabilité de la même épreuve précédée par d'autres dont on suppose par hypothèse qu'on connaît le résultat.

Nous avons étudié le cas où dans l'évaluation des probabilités subordonnées on n'est influencé que par la fréquence observée. On peut caractériser ce cas d'une façon équivalente, mais plus intuitive et plus simple comme le cas où les diverses épreuves du phénomène considéré — ou, en général, les événements considérés — jouent un rôle symétrique par rapport à tout problème de probabilité; ou bien encore : le cas où la probabilité pour que r épreuves données aient un résultat favorable et s autres un résultat défavorable, ne dépend que de r et s, ou enfin : le cas où la probabilité, pour que n épreuves aient toutes un résultat favorable, est la même, quelle que ce soit l'n-uple choisi. Ces conditions, qui définissent les « événements équivalents », ont une signification immédiate et très claire du point de vue de la théorie subjective des probabilités, et il y a de nombreux cas pratiques où elles se présentent spontanément à notre esprit. Cela suffit pour expliquer alors notre foi en une stabilité de la fréquence, car, avec cette hypothèse, la probabilité d'une épreuve ultérieure subordonnée à l'observation d'une certaine fréquence, tend à coïncider avec la valeur même de celle-ci. Il existe cependant un cas particulier, celui de l'indépendance, pour lequel l'influence de l'observation passée est rigoureusement nulle. Ce cas constitue une exception; dans tous les autres cas l'influence des résultats acquis tend à prévaloir lorsque le nombre des cas observés augmente, mais naturellement d'une façon qui n'est nullement uniforme (on peut avoir par exemple des évaluations très voisines de celles qui correspondent au cas de l'indépendance, pour lequel la dite influence est nulle). Cette évaluation *subjective* joue donc un rôle essentiel; la condition d' « équivalence » elle-même n'a par ailleurs qu'une valeur *subjective*.

Ce raisonnement ne s'applique pas seulement aux fréquences : dans le cas des nombres aléatoires équivalents (et, en général, dans le cas des éléments aléatoires d'un espace quelconque) on peut justifier d'une façon et avec des réserves analogues, une certaine stabilité dans la distribution des valeurs. Le problème même de la péréquation devrait être étudié de ce point de vue [1] : la courbe de distribution ajustée serait alors la loi de probabilité, subordonnée à l'observation des valeurs effectivement observées; cette courbe dépendrait d'une opinion subjective, mais d'autant moins que l'expérience serait plus riche; on pourrait, par contre, voir dans cette conception la vraie raison des procédés de péréquation et des conditions de « régularité » et de « proximité de l'allure observée » qui les inspirent. De ce point de vue, ces conditions ne sont plus des conditions arbitraires ou formelles, mais des conséquences de théorèmes sur les nombres aléatoires équivalents et des tendances naturelles à notre esprit dans l'évaluation des probabilités de différentes allures.

Ce que j'ai dit et démontré pour le cas de l'équivalence, peut être répété évidemment, avec les modifications nécessaires, pour des conditions moins simples et moins typiques auxquelles il a été simplement fait allusion (fin Chap. V). Le sens des conclusions est toujours le même : l'observation ne peut pas confirmer ou démentir une opinion, qui est et ne peut être autre chose qu'une opinion, donc ni vraie, ni fausse; l'observation peut seulement nous donner des renseignements qui sont susceptibles d'*influencer* notre opinion. Le sens de cette affirmation est très précis : elle signifie qu'à la probabilité d'un fait subordonné à ces renseignements — probabilité bien distincte de celle du même fait non subordonné à d'autres — nous pouvons attribuer effectivement une valeur différente.

Ainsi, je crois avoir réussi, sinon à persuader ceux qui sont loin d'accepter le point de vue subjectif, du moins à prouver que ce point de vue donne une réponse irréprochable à toutes les questions, et qu'il permet de les réunir en quelque sorte en une conception cohérente unique. Certains esprits — convaincus par ailleurs que la théorie subjective des probabilités constitue une conception cohérente, complète, et parfaitement acceptable en elle-même — refuseront à s'y rallier pour des raisons d'ordre philosophique. On peut, en effet, penser que les concepts scientifiques devraient avoir toujours une signification réelle, que la Science doit s'occuper exclusivement de réalités, et que le point de

[1] [56]; le point de vue de Poincaré, [XXVIII], p. 237-239, est très analogue.

vue subjectif conduirait à s'éloigner chaque jour davantage de ce principe avec l'application de plus en plus étendue des probabilités aux sciences physiques : non seulement cette branche particulière des mathématiques que constitue le calcul des probabilités, non seulement ses applications aux jeux et à la statistique, mais aussi une partie chaque jour plus vaste des conceptions physiques cesserait de correspondre à une réalité objective. Et l'on pourrait dire que les lois déterministes du type classique et cette espèce de succédané (1) que constituent les lois « statistiques », n'auront même plus cette caractéristique essentielle commune qui les reliait jusqu'ici, à savoir la connexion avec la réalité. N'y aurait-il donc pas là un abîme infranchissable séparant ces deux types de lois qui coexistent aujourd'hui en physique?

Pour franchir cet abîme, le point de vue adopté jusqu'ici nous conduit tout naturellement à une solution qui est exactement l'opposée de celle qu'on envisage habituellement : au lieu d'étendre le caractère de réalité des lois classiques aux lois de probabilité, on peut essayer au contraire de faire participer ces mêmes lois classiques au caractère subjectif des lois statistiques. J'ai déjà cité cette phrase de Poincaré : « si solidement assise que puisse nous paraître une prévision, nous ne sommes jamais absolument sûrs que l'expérience ne la démentira pas ». Les lois n'ont de valeur que pour nous-mêmes, en cela que — et seulement dans le sens que — nous estimons *très peu probable*, d'après l'expérience et l'analyse scientifique de ses résultats, qu'une « loi » soit démentie par la vérification d'un événement contredisant un résultat qu'elle avait prévu. Les lois rigides ne sont *prouvées* par l'expérience que dans le sens de la vérification d'un accord entre elles et un certain nombre de faits. Se demander si ces faits sont vérifiés *parce que la loi est vraie*, ou se demander si la loi véritable n'est pas *différente de celle que nous possédons* avec laquelle elle ne coïnciderait que dans ces cas particuliers, ou enfin se demander si la loi *n'existe pas*, sont des questions qui, précisément du point de vue opératif, n'ont aucune espèce de sens. Il y a toujours une infinité d'explications possibles pour un même groupe d'observations : si nous en choisissons une, et si nous énonçons une loi, ce ne pourra être que pour des raisons subjectives qui nous la font considérer digne de confiance. Les lois rigides sont formulées et acceptées

---

(1) Pour certains côtés de nos opinions sur cette question, on peut voir aussi [36].

par notre esprit pour les mêmes raisons qui nous conduisent à formuler et accepter un jugement de probabilité quelconque : la seule différence consiste en la probabilité très élevée que nous attribuons, dans le cas des lois rigides, à l'accord exact avec les faits expérimentaux. La probabilité est si élevée qu'on peut l'appeler « certitude pratiquement absolue », ou, simplement, « certitude », en sous-entendant toutefois le reste qui est néanmoins essentiel du point de vue philosophique et logique.

La notion de « cause » dépend ainsi de la notion de probabilité, et découle donc elle aussi de la même source subjective que tous les jugements de probabilité [32] : cette explication semble constituer la véritable traduction logique de la conception de « cause » préconisée par David Hume, que je considère comme le plus haut sommet qui ait été atteint par la philosophie. La théorie subjective des probabilités pourra ainsi ouvrir le champ de la Science à cette conception, dont la signification et la valeur semblent ne pas avoir été suffisamment comprises ni appréciées jusqu'ici.

C'est pour ces raisons que la théorie des probabilités ne doit pas être considérée comme une théorie auxiliaire pour les branches où la Science n'a pas encore découvert le mécanisme déterministe qui « doit » exister; elle doit être considérée par contre comme constituant les prémisses logiques de tout raisonnement par induction. De même que la logique ordinaire à deux valeurs est l'instrument nécessaire de tout raisonnement où intervient simplement le fait qu'un événement arrive ou n'arrive pas, ainsi la logique du probable, logique à une échelle continue de valeurs, est l'instrument nécessaire de tous les raisonnements où intervient, visible ou caché, le degré de doute, le jugement de certitude pratique ou d'impossibilité pratique, enfin l'estimation de vraisemblance d'un événement quelconque. Tout ce qui ne se réduit pas à une simple constatation, à une vérité historique isolée, tout ce qui nous conseille pour l'avenir, la croyance même qu'en sortant de notre chambre nous verrons comme les autres jours les mêmes rues et les mêmes maisons à leurs mêmes places, tout cela constitue un jugement de probabilité, qui se base, peut-être inconsciemment et indistinctement, sur les principes du calcul des probabilités. Ce calcul constitue donc le fondement de la plus grande partie de notre pensée, et nous pouvons bien répéter avec Poincaré : « sans lui la Science serait impossible ».

LISTE BIBLIOGRAPHIQUE.

[I]. BERTRAND (J.). — *Calcul des probabilités*, Paris, 1889.

[II]. BRIDGMAN (P. W.). — *Die Logik der heutigen Physik* (trad. allemande par M. W. KRAMPF), Max Hueber, ed. München, 1932.

[III]. CANTELLI (F. P.). — *Sulla probabilità come limite della frequenza* (*Rend. R. Acc. Naz. Lincei*, 5e série, vol. XXVI, genn. 1917).

[IV]. — *Una teoria astratta del calcolo delle probabilità* (*Giorn. Ist. Ital. Attuari*, A. III, n° 2, 1932).

[V]. — *Considérations sur la convergence dans le calcul des probabilités* (*Ann. Inst. H. Poincaré*, vol. V, fasc. 1, 1935).

[VI]. CASTELNUOVO (G.). — *Calcolo delle probabilità*, Zanichelli, ed., 1925.

[VII]. — *Sul problema dei momenti* (*Giorn. Ist. Ital. Attuari*, A. I, n° 2, 1930).

[VIII]. — *Sur quelques problèmes se rattachant au calcul des probabilités* (*Ann. Inst. H. Poincaré*, vol. III, fasc. 4, 1933).

[IX]. DÖRGE (K.). — *Ueber das Anwendungsproblem der Wahrscheinlichkeitsrechnung und das Induktionsproblem* (*Deutsche Math.*, Bd. 1, 1936).

[X], [XI]. FRÉCHET (R.-M.). — *Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements* (*Rend. R. Ist. Lombardo*, 2e série, vol. LXIII, première Note, fasc. 11-15; seconde Note, fasc. 16-18, 1930).

[XII]. FRÉCHET (R.-M.) et HALBWACHS (M.). — *Le calcul des probabilités à la portée de tous*, Dunod, 1924.

[XIII]. GLIVENKO (V.). — *Sulla determinazione empirica delle leggi di probabilità* (*Giorn. Ist. Ital. Attuari*, A. IV, n° 1, 1933).

[XIV]. HOSIASSON (J.). — *La théorie des probabilités est-elle une logique généralisée?* (*Actual. Scient. Industr.*, n° 391, Hermann, 1936).

[XV]. KHINTCHINE (A.). — *Sur les classes d'événements équivalents* (*Rec. Math. Moscou*, t. 39, n° 3, 1932).

[XVI]. — *Remarques sur les suites d'événements obéissant à la loi des grands nombres* (*ibid.*).

[XVII]. KOLMOGOROFF (A.). — *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Ergebn. Math. usw.*, Bd. II, Heft 3, Springer, Berlin, 1933).

[XVIII]. — *Sulla determinazione empirica di una legge di distribuzione* (*Giorn. Ist. Ital. Attuari*, A. IV, n° 1, 1933).

[XIX]. VON KRIES (J.). — *Die Prinzipien der Wahrscheinlichkeitsrechnung*, Tübingen, 1886.

[XX]. LÉVY (P.). — *Calcul des probabilités*, Paris, Gauthier-Villars, 1925.

[XXI]. LOMNICKI (A.). — *Nouveaux fondements de la théorie des probabilités* (*Fund. Math.*, t. 4, 1923).

[XXII]. LUKASIEWICZ (J.). — *Philosophische Bemerkungen zu mehrwertigen Systemen der Aussagenkalküls* (*C. R. Soc. Sciences Varsovie*, 1930).

[XXIII]. MAZURKIEWICZ (S.). — *Zur Axiomatik der Wahrscheinlichkeitsrechnung* (*S. P. Towarzystwa Naukowego Warszawskiego*, 1932).

[XXIV]. Medolaghi (P.). — *La logica matematica e il calcolo delle probabilità* (*Boll. Ass. Ital. Attuari*, n° 18, 1907).

[XXV]. von Mises (R.). — *Wahrscheinlichkeit, Statistik und Wahrheit* (*Schriften zur wiss. Weltauffassung*, Bd. III, Wien, Springer, 1928).

[XXVI]. — *Théorie des probabilités : fondemements et applications* (*Ann. Inst. H. Poincaré*, vol. III, fasc. 2, 1932).

[XXVII]. Pareto (V.). — *Manuel d'Économie Politique*, Paris, Ciard et Brière, 1909.

[XXVIII]. Poincaré (H.). — *La Science et l'Hypothèse* (Chap. XI : *Le calcul des probabilités*, Paris, Flammarion éd., 1906).

[XXIX]. Reichenbach (H.). — *Wahrscheinlichkeitslehre*, Leiden, Sijthoff, 1935.

[XXX]. — *Die Induktion als Methode der wissenschaftlichen Erkenntnis* (*Actual. Scient. Industr.*, n° 391, Paris, Hermann, 1936).

[XXXI]. — *Wahrscheinlichkeitslogik als Form wissenschaftlichen Denkens* (*ibid.*).

---

[16]. *Sui passaggi al limite nel calcolo delle probabilità* (*Rend. R. Ist. Lombardo*, 2e série, vol. LXIII, fasc. 2-5, 1930).

[24]. *A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili* (*ibid.*, fasc. 11-15, 1930).

[26]. *Problemi determinati e indeterminati nel calcolo delle probabilità* (*Rend. R. Acc. Naz. Lincei*, 6e série, vol. XII, 2e sem., fasc. 7-8, ott. 1930).

[27]. *Fondamenti logici del ragionamento probabilistico* (*Boll. Un. Mat. Ital.*, A. IX, n° 5, dic. 1930).

[28]. *Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali* (*Rend. R. Ist. Lombardo*, 2e série, vol. LXIII, fasc. 16-18, 1930).

[29]. *Funzione caratteristica di un fenomeno aleatorio* (*Memorie R. Acc. Naz. Lincei*, 6e série, vol. IV, fasc. 5, 1930).

[32]. *Probabilismo : Saggio critico sulla teoria delle probabilità e sul valore della scienza* (*Biblioteca di Filosofia diretta da Antonio Aliotta*, Napoli, Perrella ed., 1931).

[34]. *Sul significato soggettivo della probabilità* (*Fund. Math.*, t. 17, Warszawa, 1931).

[35] *Sui fondamenti logici del ragionamento probabilistico* (*Atti Soc. Ital, Progr. Scienze*, Riunione Bolzano-Trento del 1930, vol. II, Roma, 1931).

[36]. *Le leggi differenziali e la rinunzia al determinismo* (*Rend. Semin. Mat- R. Univ. Roma*, 2e série, vol. VII, 1931).

[38]. *Probabilità fuori dagli schemi di urne* (*Period. di Mat.*, 6e série, vol. XII, n° 1, 1932).

[40]. *Funzione caratteristica di un fenomeno aleatorio* (*Atti Congr. Int. Matem., Bologna*, 1928, t. 6, Bologna, Zanichelli ed., 1932).

[45]. *Sull'approssimazione empirica di una legge di probabilità* (*Giorn. Ist. Ital. Attuari*, A. 4, n° 3, 1933).

[46 à 48]. *Classi di numeri aleatori equivalenti. La legge dei grandi numeri nel caso dei numeri aleatori equivalenti. Sulla legge di distrubuzione dei valori in una successione di numeri aleatori equivalenti* (*Rend. R. Acc. Naz. Lincei*, 6e série, vol. XVIII, 2e sem., fasc. 3-8, 1933).

[49]. *Sul concetto di probabilità* (*Riv. Ital. Statist. ecc.*, A. V, n° 4, 1933).

[51]. *Indipendenza stocastica ed equivalenza stocastica* (*Atti Soc. Ital. Progr. Scienze*, Riunione Bari del 1933, vol. II, Roma, 1934).

[R. 2]. Compte rendu de Reichenbach H. [XXIX] (*Zentralblatt für Mathematik und ihre Grenzgebiete*, Bd. 10, Heft 8, 1935).

[56]. *Il problema della perequazione* (*Atti Soc. Ital. Progr. Scienze*, Riunione Napoli del 1934, vol. II, Roma, 1935).

[58]. (En collaboration avec M. M. Jacob). — *Sull'integrale di Stieltjes-Riemann* (*Giorn. Ist. Ital. Attuari*, A. VI, n° 4, 1935).

[62]. *La logique de la probabilité* (communication au Congr. Int. Philosophie Scient., Paris 1935) (*Actual. Scient. Industr.*, n° 391, Hermann, 1936).

[64]. *Les probabilités nulles* (*Bull. Sci. Math.*, 1936).

[65]. *Statistica e probabilità nella concezione di R. von Mises* (*Suppl. Statist. Nuovi Probl. ecc.*, A. II, n° 3, 1936).

[70]. *Riflessioni teoriche sulle assicurazioni elementari* (sera présenté comme communication sur le sixième sujet au Congrès International d'Actuaires, Paris, 1937).