ANNALES DE L'I. H. P.

RICHARD VON MISES

Théorie des probabilités. Fondements et applications

Annales de l'I. H. P., tome 3, nº 2 (1932), p. 137-190
<http://www.numdam.org/item?id=AIHP_1932__3_2_137_0>

Théorie des probabilités. Fondements et applications

PAR

RICHARD von MISES

Les fondements de la théorie

Mesdames, Messieurs,

Au début des conférences que j'aurai l'honneur de donner ici, dans votre bel Institut, je tiens à vous dire combien je suis heureux de me trouver en face d'un public aussi distingué que compétent dans la matière que je vais traiter. En effet, si on parcourt l'histoire de la théorie des probabilités on y rencontre presque à chaque pas des idées fondamentales d'origine française. Ne citons que Blaise PASCAL et Pierre de FERMAT, les premiers fondateurs du calcul, puis le grand Pierre Simon, Marquis de LAPLACE qui érigea l'édifice impérissable de la « Théorie analytique », enfin Siméon Denis POISSON, l'auteur de la fameuse loi des grands nombres. Et ces grands maîtres du dix-huitième furent suivis jusqu'à nos jours par une longue série de célèbres géomètres qui apportèrent chacun des importants compléments à la théorie classique, série qui culminait en HENRI POINCARÉ. C'est un risque bien grand que d'essayer d'apporter des modifications à une théorie si généralement admise et soutenue par de si grandes autorités. Mais je suis sûr qu'une nouvelle manière de concevoir les lois et les faits fondamentaux des probabilités ne sera pas adoptée par le monde mathématique avant d'avoir trouvé l'approbation des savants français auxquels je vais soumettre, dans ces conférences, mes idées principales.

On connaît bien les grandes difficultés qui s'opposent à une définition exacte de la notion de probabilité. Citons par exemple Henri POINCARÉ qui, en abordant son cours célèbre, s'exprime ainsi : « On ne peut guère donner une définition satisfaisante de la Probabilité. On dit généralement... etc. » Voilà une chose bien étrange : une théorie mathématique qui commence avec un « on dit ». Et les dernières conclusions du premier chapitre de POINCARÉ sont : « Il y a là quelque chose de mystérieux, d'inaccessible au mathématicien. » On ne saurait nier qu'une fondation un peu plus approfondie ne soit désirable pour une théorie qui, sans doute, fait partie des sciences mathématiques ; surtout dans nos temps où la recherche des principes, la recherche des fondements et des axiomes domine presque dans toutes les branches de notre science. Je vais essayer aujourd'hui de vous donner une esquisse rapide d'une nouvelle théorie que j'ai développée dans plusieurs travaux depuis 1919. Dans les conférences prochaines j'ai l'intention à donner quelques détails et applications diverses de ces nouveaux principes.

I

Envisageons pour un moment la *définition classique* dont je viens de parler. Elle définit la probabilité comme le rapport du nombre des cas favorable à un événement, au nombre total des cas « également possibles ». Que signifient ces mots : également possibles » ? La possibilité n'est pas mesurable, elle n'a pas des degrés comme la longueur ou la température. Sans doute l'expression « également possibles » n'est qu'une équivoque pour « également probable », de sorte que la définition contient un *cercle vicieux* partiel ; elle ne fait que ramener le cas général où les probabilités de plusieurs événements possibles sont différentes, au cas spécial où elles ont toutes la même valeur. Pour utiliser la définition il faudrait avoir par avance une définition suffisante pour le cas des probabilités égales entre elles, ou, comme nous dirons, pour le cas d'une répartition uniforme des probabilités. Il en serait de même, si on voulait en mécanique définir la vitesse en général en la ramenant au cas simple d'une vitesse uniforme, mais sans définir celle-ci au préalable. Dans la théorie que je vais développer, l'énoncé classique sur le rapport du nombre des cas favorables, etc., obtiendra la place légitime d'un *théorème spécial* qui nous permet de résoudre certains problèmes du calcul des probabilités, mais qui ne joue nullement le rôle d'une définition.

Constatons de plus que le fait de ramener toute probabilité au cas où tous les événements possibles ont la même probabilité, n'est pas toujours praticable. Nous disons qu'un dé est bon pour le jeu si les six points qui peuvent sortir sont également probables. Nous en déduisons que la probabilité pour qu'un nombre pair se produise, est $\frac{3}{6} = \frac{1}{2}$. Mais si un dé est faux, qu'il soit falsifié ou abîmé pour une raison quelconque, les six cas possibles n'étant plus également possibles, est-ce qu'il n'existe plus une certaine probabilité $\left(\text{différente de } \frac{6}{1}\right)$ d'amener le point 6 ? Est-ce que le théorème d'addition, qui nous donne la probabilité d'un nombre pair égale à la somme des probabilités pour les points 2, 4, 6, n'est plus juste ? Et nous avons la même difficulté en matière d'assurances sur la vie. Quand nous disons qu'une certaine probabilité de survie annuelle est 0,991, où sont là les cas également probables ? Y en a-t-il mille dont 991 favorables, ou dix mille dont 9910 favorables ? J'insiste sur ce point, car je trouve ici le côté le plus faible de toute la théorie usuelle : on démontre tous les théorèmes, soit la loi d'addition ou de multiplication, soit la loi des grands nombres etc., en supposant qu'il s'agisse, en dernière ligne, d'une énumération de cas également probables, et puis *on applique ces théorèmes à des problèmes de toute autre nature.* C'est une exigence inévitable pour une théorie qui prétend être une théorie mathématique, que ses démonstrations soient basées sur des définitions univoques et que ses énoncés aient un sens parfaitement défini.

Pour parvenir à une telle théorie il faut d'abord fixer l'objet dont il s'agit dans le calcul des probabilités. Il y a à distinguer trois groupes de problèmes : les questions du jeu de hasard, les questions de statistique générale (y compris les assurances sur la vie, etc.), enfin les questions posées par la physique moderne. En tout cas on a affaire à une longue suite d'événements ou de phénomènes, à des *expériences ou observations en masse.* Dans les jeux de hasard une expérience individuelle consiste par exemple dans le processus suivant : un dé est mis dans un cornet, on secoue le cornet, on jette le dé sur la table et on lit le chiffre apparaissant en haut. Une longue suite de telles épreuves est un objet particulier de la théorie des probabilités. En affaire de statistique générale on considère comme expérience individuelle par exemple l'examen de l'année du décès d'une certaine personne ; une série de telles recherches, concernant les personnes nées dans un certain endroit et à une certaine époque, constitue un « phénomène en masse » qui peut servir comme point de départ pour un problème de probabilités. Enfin c'est la même chose, en troisième lieu, en physique mathématique, quand il s'agit par exemple des différentes vitesses d'un grand nombre de molécules gazeuses ; ici, l'épreuve individuelle est la mesure des trois composantes de la vitesse d'une certaine molécule à un moment donné. Nous constatons que, dans tous ces cas, le résultat d'une expérience considérée est *un chiffre* ou un *groupe de chiffres*, on peut aussi dire, *un point dans l'espace à r dimensions*. Nous nous servirons aussi du mot « élément » pour une expérience individuelle, et nommerons le résultat de l'expérience « *caractère distinctif* » de l'élément. La notion de probabilité que nous allons établir, est étroitement liée à la notion de *fréquence relative* avec laquelle un certain caractère distinctif se présente dans la suite des résultats des observations.

Généralement dans les cours de probabilités, arrivé à un certain point des développements, on introduit une notion particulière de probabilité dite « statistique. » Selon la conception usuelle, celle-ci diffère de la probabilité mathématique, à laquelle on a affaire dans les jeux de hasard, comme un terrain triangulaire diffère d'un triangle géométrique. Mais il faut admettre qu'une série de jeux de pile ou face est aussi un phénomène réel, comparable à un triangle concret. Et jamais en géométrie on ne considère deux définitions distinctes de la longueur ou de l'aire du triangle ou d'une figure quelconque, l'une valable dans la théorie, l'autre pour l'expérience. Tous les théorèmes de géométrie sont énoncés et prouvés pour les figures abstraites, et rigoureusement définies, et puis on les applique — avec une certaine approximation — aux figures concrètes, aux corps réels, sans donner une nouvelle définition de ces objets. A mon avis la théorie des probabilités est une *discipline tout à fait du même genre que la géométrie ou la mécanique rationelle.* Elle a pour but la description systématique de certains phénomènes réels et se sert d'une construction idéalisée des notions et définitions abstraites. Je veux maintenant donner une telle construction idéale pour le cas des phénomènes de statistiques ou de probabilité.

II

Considérons d'abord *le cas le plus simple* qui se présente dans le jeu de pile ou face ou dans le jeu de rouge et noir ou dans la probabilité de survie annuelle. Dans tous ces exemples le résultat d'une expérience individuelle ou le caractère distinctif d'un élément de la suite considérée est toujours l'un des deux signes : pile ou face, rouge ou noir, survie ou décès. Nous y substituerons, pour abréger, les chiffres zéro et un, et parlerons généralement *d'une alternative* : « *zéro ou un* ». La notion idéalisée que nous formons d'un tel jeu ou d'un tel « phénomène en masse » le plus simple est la suivante : nous envisageons une *suite infinie* d'expériences dont chacune ait pour résultat l'un des deux chiffres 0 ou 1 ; le rapport du nombre n_1 des « 1 » parmi les n premiers résultats, au nombre n, *tend vers une limite* p, si le nombre n croît indéfiniment ; en plus la correspondence entre la valeur (0 ou 1) d'un résultat et la place où ce résultat se trouve dans la suite infinie, remplit une certaine *condition d'* « *arégularité* » ou de « *fortuité* » que nous allons étudier à l'instant. Nous nommerons une telle suite infinie, obéissant aux deux conditions citées un « *collectif le plus simple* » ou une alternative, et la limite p la *probabilité du chiffre 1* (ou de l'événement représenté par ce chiffre) *dans le collectif considéré*. L'expression « probabilité » sans cette adjonction « dans le collectif considéré » n'existe pas pour nous. Il n'y a pas une probabilité de survie pour un individu déterminé N. N., il n'y a qu'une probabilité de survie pour une certaine *classe* d'hommes, constituant la base d'un collectif, par exemple pour les hommes de quarante ans, nés en France, de profession non-dangeureuse. Chaque individu appartenant à des classes diverses, on ne peut jamais parler d'une probabilité de survie pour l'individu lui-même, sans se rapporter à un collectif bien défini.

Passons maintenant à une recherche plus détaillée de la notion *d'arégularité ou de fortuité*. On comprend qu'il faut distinguer, dans les problèmes du calcul des probabilités, entre la suite des pairs et impairs dans l'ordre naturel des nombres, et par exemple une suite représentée par les résultats successifs du jeu de pile ou face. Dans ces deux cas nous pouvons admettre que la limite du rapport $\frac{n_1}{n}$ (nombre des éléments d'un des deux genres au nombre total des éléments) existe et a la valeur $\frac{1}{2}$.

Mais nous réservons le nom de probabilité pour la limite *du deuxième cas,* caractérisé par la propriété suivante : On peut constater qu'en ne considérant que les résultats de quatre en quatre, on arrive dans le cas de pile ou face à une suite exactement de même caractère, avec le rapport $\frac{1}{2}$ des résultats « pile », tandis que dans le cas des nombres naturels la série incomplète ne contiendra que des pairs *ou* des impairs exclusivement. Pour définir exactement les séries qui forment le vrai objet du calcul des probabilités nous introduisons une notion nouvelle, celle du « *choix de position.* »

Envisageons une alternative, c'est-à-dire une suite infinie d'expériences dont chacune ait pour résultat 0 et 1. Nous disons qu'on effectue un « choix de position » sur cette suite en fixant une liste des nombres entiers croissants, $\alpha_1, \alpha_2, \alpha_3, \cdots$ de telle manière que le fait qu'un certain nombre α appartienne ou non à la liste ne dépende pas des résultats des expériences numérotées $\alpha, \alpha + 1, \alpha + 2, \alpha + 3, \cdots$ On peut prendre par exemple pour établir un « choix de position » une loi arithmétique quelconque ; nous prendrons les éléments de cinq en cinq, ou bien ceux dont le numéro est une puissance de 10 ou un nombre premier, etc. Mais on établit aussi un « choix de position » en utilisant les résultats qui précédent l'élément ; par exemple : prenons chaque élément précédé d'un zéro ou d'une paire des zéros etc. Dans l'argot des habitués de Monte-Carlo un choix de position s'appelle un «*système à jouer*». Le fait fondamental qu'aucun système à jouer ne puisse altérer les chances d'un joueur, devient une base de notre théorie des probabilités par l'introduction de l'axiome suivant : Dans chaque suite partielle des expériences qu'on obtient par un choix de position effectué sur la suite totale, la limite du rapport n'_1/n' (nombre des résultats « 1 » parmi les n' premiers éléments choisis, au nombre total n') *existe encore et a la même valeur p que dans la suite primordiale.* Quelle que soit la complication du système imaginé, il suffit d'après cette hypothèse de continuer le jeu assez longtemps pour que le rapport des gains et des échecs parmi les parties choisies tende vers la même limite, vers laquelle il tend dans la suite complète des jeux. Cet *axiome de fortuité* ou de *l'impossibilité d'un système à jouer,* fait une partie essentielle de notre théorie. Permettez-moi maintenant de résumer la définition du collectif le plus simple ou de l'alternative et de la probabilité dans une alternative.

Nous nommons « alternative » ou « collectif le plus simple » une suite infinie d'expériences dont chacune ait pour résultat zéro ou un, la correspondance entre les résultats et l'ordre des expériences satisfaisant aux deux conditions suivantes :

1° : Le rapport du nombre n_1 des résultats égaux à un, au nombre total n des expériences tend vers une limite déterminée p, lorsque n augmente indéfiniment ;

2° : Si on effectue un choix de position sur la suite donnée, le rapport analogue n_1'/n', ne tenant compte que des expériences choisies, tend vers la même limite p. — Cette limite, indifférente à un choix de position, s'appelle *la probabilité de l'événement considérée, dans le collectif envisagé.*

III

Avec cette définition les principales difficultés de la notion de probabilité sont surmontées. Revenons maintenant au cas général, envisagé déjà plus haut, d'une suite d'expériences dont les résultats sont donnés par des chiffres quelconques ou par des groupes de chiffres (au lieu de zéro et un dans le cas du collectif le plus simple). En tout cas on peut représenter le résultat d'une telle expérience — ou le « caractère distinctif » d'un élément du collectif général — par un point dans l'espace à r dimensions. L'ensemble des points considérés si on parcourt tous les éléments de la suite infinie des expériences, sera appellé son *ensemble caractéristique* ; dans le cas le plus simple d'une alternative l'ensemble caractéristique se réduit à l'ensemble de deux points avec les coordonnées 0 et 1. Dans le jeu de hasard avec deux dés on a un ensemble caractéristique composé de 36 points rangés dans un carré. Dans une recherche statistique sur la longueur des feuilles d'arbres l'ensemble caractéristique est un certain intervalle de l'axe réel. Dans les applications bien connues du calcul des probabilités au tir on considère comme ensemble caractéristique l'aire de la cible, c'est-à-dire un espace continu à deux dimensions.

Désignons par E l'ensemble caractéristique d'une certaine suite infinie d'expériences et par A une portion quelconque de E. Soit n_A le nombre de tels éléments parmi les n premiers éléments de la suite dont le caractère distinctif (ou le point représentant le résultat de l'expérience) appartient à A. Nous demandons d'abord que le rapport n_A/n tende vers une limite déterminée p_A, le nombre n augmentant indéfiniment. Soit B une autre portion de E, sans points communs avec A, et p_B la limite du rapport n_B/n, le nombre n_B étant défini d'une manière analogue à n_A. Admettons maintenant que p_A et p_B ne s'évanouissement pas tous les deux, de sorte que le quotient p_A/p_B soit défini. Si nous ne tenons compte que des éléments de la suite primordiale dont les points représentatifs appartiennent soit à A soit à B, nous avons affaire à une suite d'expériences dont les résultats sont caractérisés par un des deux signes A et B. L'existence de la limite du rapport $n_A/(n_A + n_B)$ est garantie déjà par la première hypothèse ci-dessus, et sa valeur est $p_A/(p_A + p_B)$. Nous demandons maintenant en plus que la suite partielle avec les caractères distinctifs A et B satisfasse aussi à la deuxième condition, formulée en haut pour une alternative : Si on effectue un choix de position sur les $(n_A + n_B)$ éléments il faut que le rapport du nombre des éléments avec le caractère A au nombre total des éléments retenus ait une limite égale à la limite de $\frac{n_A}{n_A + n_B}$, c'est-à-dire égale à : $\frac{p_A}{p_A + p_B}$.

En résumant tout ce qui a été dit jusqu'ici on peut formuler la définition du collectif et de la probabilité dans le cas le plus général comme il suit :

Un collectif est une suite infinie d'expériences dont les résultats sont représentés par certains points dans un espace à r dimensions, la correspondence entre les résultats et l'ordre des expériences satisfaisant aux deux conditions suivantes :

1° Soit A une portion quelconque de l'ensemble caractéristique ; le rapport $\frac{n_A}{n}$ du nombre n_A de telles expériences parmi les n premières dont les résultats appartiennent à A, au nombre total n des expériences, tend vers une limite déterminée, n augmentant infiniment ;

2° Soient A et B deux portions non-vides de l'ensemble caractéristique sans points communs, n_A et n_B les nombres de telles expériences parmi les n premières dont les résultats appartiennent à A ou B respectivement, les limites p_A et p_B ne s'évanouissant pas toutes les deux ; si on effectue un choix de position sur les $(n_A + n_B)$ expériences de sorte qu'il ne reste que n'_A et n'_B, le rapport $\frac{n'_A}{n'_A + n'_B}$ tend vers une limite égale à la limite du rapport $\frac{n_A}{n_A + n_B}$ ou à $\frac{p_A}{p_A + p_B}$. Ces deux conditions remplies, la limite p_A s'appelle *la probabilité d'un résultat appartenant à l'ensemble A, dans le collectif considéré.*

IV

Dans notre définition générale de la probabilité nous n'avons pas fait de distinction entre le cas des probabilités *discontinues* et des probabilités dites continues ou *géométriques*. En effet, cette différence n'est pas essentielle du point de vue de la notion de probabilité, elle ne touche qu'aux détails du calcul dans des problèmes spéciaux. On pourrait, en principe, développer toute la théorie en ne considérant que des *fonctions d'ensemble additifs* p_A qui définissent la répartition de la probabilité sur les portions de l'ensemble caractéristique. Mais pour l'analyse des problèmes classiques il faut étudier deux types spéciaux de répartition. L'un est le cas des probabilités discontinues ou arithmétiques. L'ensemble caractéristique ne contient qu'un *nombre fini de points*, et à chacun d'eux correspond une probabilité finie. L'autre cas spécial est celui des probabilités continues ou géométriques. L'ensemble caractéristique est un *intervalle* dans l'espace à r dimensions, et dans chaque point C existe la limite du quotient p_A/F, pour F évanouissant, où A désigne une portion d'espace enfermant le point C et F le volume de A. Nous appellons cette limite une « *densité de probabilité* » pour la bien distinguer de la probabilité elle-même. Une répartition arithmétique est donnée par les probabilités ponctuelles, correspondant aux points en nombre fini de l'ensemble caractéristique. Une répartition géométrique sera connue lorsqu'on donne la valeur de la densité de probabilité pour tous les points de l'ensemble caractéristique.

Pour le calcul des probabilités il est avantageux dans tous les cas de se servir de la *fonction de répartition* que je vais définir comme suit. En nous bornant au cas d'une seule dimension nous considérons la fonction $f(x)$ où f signifie la probabilité pour que le caractère distinctif dans le collectif soit *inférieur ou égal à x*. Évidemment *$f(x)$ est une fonction monotone, non-décroissante qui commence par la valeur* 0 *pour* $-\infty$ *et* *finit par la valeur* 1 *pour* $+\infty$. Si $f(x)$ est dérivable nous avons le cas de la probabilité géométrique, si $f(x)$ n'a que des points de croissance en nombre fini c'est le cas de la probabilité arithmétique. Pour définir la « moyenne » de la répartition, la « déviation » (carré de l'écart moyen) et les autres « moments » d'une répartition on peut se servir des *intégrales de Stieltjes* :

$$a = \int x df(x), \qquad \int x^2 df(x), \qquad s^2 = \int (x-a)^2 df(x).$$

Ces intégrales se réduisent à des intégrales riemanniennes ou à des sommes finies dans les cas spéciaux des probabilités géométriques ou arithmétiques respectivement. Je n'insiste pas sur la généralisation pour le cas de plusieurs variables et sur d'autres détails purement mathématiques qui ne touchent pas les questions des fondements ou des principes généraux.

Mais permettez-moi d'ajouter une remarque. On trouve souvent exprimée l'opinion que, dans le cas des probabilités géométriques, il existe une difficulté spéciale à définir les cas également possibles, à cause des différentes mesures possibles dans un espace continu. Cette opinion n'est pas fondée. Du point de vue de ma théorie il n'y a aucune différence à cet égard entre les cas des probabilités géométriques et arithmétiques. Nous y reviendrons tout de suite.

V

La définition du collectif et de la probabilité dans un collectif, établie ci-dessus, n'est que le commencement, le point de départ de la théorie rationnelle des probabilités. Son point essentiel, la base principale de tous les théorèmes et toutes les recherches spéciales c'est l'introduction et la définition des *quatre opérations simples* ou opérations fondamentales qu'on peut exécuter sur des collectifs donnés. Avant d'entrer dans une étude de ces opérations je veux préciser le problème le plus général de la théorie des probabilités, et voici comment. Par des opérations bien définies et se ramenant à quatre opérations simples, *on peut déduire d'un collectif donné* ou de plusieurs collectifs donnés *un nouveau collectif*. Tout géométriques ou arithmétiques — ou bien dans une autre hypothèse quelconque sur les probabilités primordiales.

La théorie que j'ai l'honneur de vous développer ici est basée, comme je l'ai déjà dit, sur le fait qu'il n'y existe que *quatre* opérations simples et leurs combinaisons, pour déduire de nouveaux collectifs de collectifs donnés. Selon que le nombre des opérations simples à effectuer dans un certain cas est fini ou non, on peut distinguer les *problèmes finis* ou algébriques des problèmes *infinis* (dénombrables) ou transcendants. Un exemple de problème infini se présente dans l'énoncé du théorème de Bernoulli ou de la loi des grands nombres, ou plus généralement dans les théorèmes fondamentaux que j'ai l'intention de développer dans ma prochaine conférence. Passons maintenant à une étude rapide des quatre opérations simples.

VI

On peut d'abord parvenir à un nouveau collectif, partant d'un collectif donné, en effectuant un choix de position sur le collectif donné. Soient $e_1, e_2, e_3, \cdots$ les éléments et $x_1, x_2, x_3, \cdots$ les caractères distinctifs de ces éléments (c'est-à-dire les résultats des expériences représentées par les lettres $e_1, e_2, e_3, \cdots$ respectivement. On établit une série $\alpha_1, \alpha_2, \alpha_3, \cdots$ de nombres naturels croissants, satisfaisant aux conditions mentionnées ci-haut, et on considère la suite des éléments $e_{\alpha_1}, e_{\alpha_2}, e_{\alpha_3}, \cdots$ avec les valeurs non-altérées $x_{\alpha_1}, x_{\alpha_2}, x_{\alpha_3}, \cdots$ On démontre facilement que cette suite est un nouveau collectif d'après la définition établie, et qu'il a la même répartition que le collectif primordial — selon la deuxième condition à remplir par un collectif (axiome d'arégularité). La première opération simple, — que nous nommons « *sélection* », — *laisse la répartition inchangée.*

La deuxième opération simple est le « *mélange* ». En l'effectuant on retient tous les éléments $e_1, e_2, e_3, \cdots$ du collectif donné C, sans en supprimer un seul, mais on fait varier les valeurs des caractères distinctifs $x_1, x_2, x_3, \cdots$. On déduit les nouvelles valeurs $x'_1, x'_2, x'_3, \cdots$ des caractères distinctifs dans le collectif déduit C', en effectuant une *transformation multivoque*, de telle sorte qu'à une valeur de x' corresle problème du calcul des probabilités consiste en ceci : *Trouver la fonction de répartition dans le collectif déduit, étant données les fonctions de répartition dans les collectifs initiaux.*

Des questions comme les suivantes : Quelle est la probabilité d'amener le point « 6 » avec un certain dé ?, ou : Quelle est la probabilité de la survie annuelle pour une certaine classe d'hommes ?, ne sont pas des questions du calcul des probabilités. C'est la même chose pour la géométrie qui ne nous donne pas la valeur de la distance de deux points particuliers ; mais nous permet de calculer une distance si certaines distances ou autres grandeurs géométriques sont données. De même la mécanique rationelle ne répond pas à la question : Quel est le mouvement d'une certaine pierre en ce moment ? Elle nous apprend à calculer la vitesse d'une pierre pour le temps t, si la vitesse pour un temps précédent et la force motrice sont données. Combien de malentendus séculaires, combien de recherches inutiles, soi-disant philosophiques, ont leur racine dans ce point sur lequel on ne peut pas assez insister. Personne ne doute que la mécanique rationnelle ne fournisse que des *relations* entre des grandeurs mécaniques, entre les vitesses, les coordonnées etc., des différents corps. Mais dans la théorie des probabilités on s'oppose habituellement à cette restriction naturelle. On n'aime pas admettre que, comme dans toutes les sciences exactes, chaque problème y présente deux côtés tout à fait différents : L'un est *l'établissement des données* du problème, et cela est une affaire essentiellement expérimentale, comme la mesure d'une longueur en géométrie, ou d'une vitesse initiale en mécanique, enfin d'une probabilité dans un collectif primordial en calcul des probabilités. L'autre côté du problème est *la recherche théorique* qui déduit des résultats, utilisant les grandeurs données et se servant des méthodes de la logique ou de la mathématique pure. Quand on ne considère que cette deuxième partie, c'est-à-dire la recherche déductive, on n'a pas besoin de s'intéresser aux sources d'où proviennent les données. Le géomètre qui déduit la règle à calculer le troisième côté d'un triangle, les deux autres et un angle étant donnés, ne pose jamais la question : D'où connaît-on les longueurs des deux côtés supposés donnés ? On ne pourra jamais comprendre la théorie rationnelle des probabilités avant d'être convaincu *que la recherche des probabilités primordiales n'est ni un devoir ni un but de la théorie mathématique des probabilités.* Naturellement on s'occupe souvent des problèmes où les répartitions primordiales ne sont établies que par une hypothèse, soit arbitraire ou une hypothèse basée sur des certaines considérations de physique théorique. Mais jamais les théorèmes et les équations du calcul des probabilités ne fournissent autre chose que des *relations* entre les probabilités supposées données et les probabilités déduites. Ajoutons que l'établissement des données peut consister dans la constatation que certaines parties de l'espace caractéristique sont également probables, soit dans le cas des probabilités pondent plusieurs valeurs de x. Par exemple dans le jeu de roulette on remplace les valeurs o, 1, 2, $\cdots$, 36 par les trois valeurs o, 1 et 2 en faisant correspondre à la nouvelle valeur $x' = 1$ les valeurs 1, 3, 5, $\cdots$ 35, à la nouvelle valeur $x' = 2$ les nombres pairs du collectif primordial et en ajoutant $x' = 0$ si $x = 0$. Évidemment la répartition dans le nouveau collectif ne contient que trois probabilités distinctes, à savoir les probabilités d'amener zéro ou d'amener un nombre impair ou un nombre pair. On pourra dire aussi que l'ensemble caractéristique du collectif primordial C a été subdivisé en plusieurs parties et qu'à tous les x qui appartiennent à une même partie correspond le même x'. De cette façon on a effectué un certain « mélange » ou « confusion » parmi les éléments du collectif primordial, en ne tenant plus compte des différences entre les valeurs confondues de x qui appartiennent à un même ensemble partiel (ou correspondent à une même valeur de x'). Si l'ensemble caractéristique de C a plusieurs dimensions on peut admettre que C' en a moins. On déduit facilement de notre définition de la probabilité la règle pour calculer la répartition dans le collectif C', déduit par l'opération d'un « mélange », la répartition dans C étant supposée donnée. D'après la définition les probabilités dans C, à savoir $p_0, p_1, p_2, \cdots, p_{36}$ au cas du jeu de roulette, sont les limites des fréquences relatives avec lesquelles les nombres o, 1, 2, $\cdots$, 36 respectivement sortent dans le jeu. D'autre part la probabilité p'_1 dans C' est, selon la définition, la limite de la fréquence relative d'amener un nombre impair quelconque. Évidemment il ne faut que sommer les $p_1, p_3, \cdots, p_{35}$ pour parvenir à la valeur cherchée de p'_1.

De cette façon nous avons retrouvé la loi bien connue des *probabilités totales* ou loi d'addition des probabilités. Mais au lieu de parler de « plusieures manières selon lesquelles un événement peut se produire, manières qui ne peuvent arriver simultanément » nous n'utilisons que des termes mathématiques bien précisés, comme : ensemble de points, subdivision de l'ensemble caractéristique en ensemble partiels, sommation d'une fonction sur les points de l'ensemble partiel. Au surplus notre déduction de la loi est indépendante de la supposition que les probabilités $p_1, p_2, p_3, \cdots$ sont égales entre elles ou qu'il s'agit de cas également probables. La déduction est juste aussi pour une roulette mal ajustée ou pour un jeu quelconque avec des chances quelconques pour les événements particuliers. Et comment veut-on éviter un malentendu comme celui qui se présente dans l'énoncé habituel de la loi de sommation si on le prend mot à mot : L'événement de gagner cent francs à la roulette, le 30 novembre 1931 à midi, peut se produire de deux manières différentes, en mettant 100 francs sur la couleur rouge à Monte-Carlo ou en mettant sur la même couleur en même temps à Nice ; les deux manières sont incompatibles parce qu'on ne peut pas être présent à Nice et à Monte-Carlo simultanément. La probabilité de chaque manière est $\frac{1}{2}$, alors on déduit de la loi de sommation dans son énoncé habituel que la probabilité de gagner cent francs est l'unité. Si on essaie de trouver un énoncé qui ne permette pas une telle application insensée de la loi de sommation on est conduit inévitablement à la définition du collectif etc.

Mentionnons encore un cas particulier, intéressant du point de vue des principes de la théorie des probabilités. Dans les jeux de hasard on rencontre souvent des collectifs dont l'ensemble caractéristique ne comprend qu'un nombre fini m de points et où toutes les probabilités sont égales entre elles. Prenons comme exemple d'une « répartition uniforme » le jeu avec deux dés corrects, où l'on a 36 points ou 36 événements possibles, tous avec la même probabilité $1/36$. En général, si m est le nombre de points caractéristiques dans un collectif C à répartition uniforme, la probabilité de chaque point est $1/m$. Si nous déduisons maintenant par l'opération de mélange un nouveau collectif C', en partageant l'ensemble de m points en deux ensembles partiels comprenant m' et m'' points, il faut sommer m' fois la valeur $1/m$ pour parvenir à la probabilité p' que l'événement, réunissant les m' premiers cas possibles, se produise. Ainsi on arrive à la formule

$$p' = \frac{m'}{m},$$

et voilà retrouvé le fameux « rapport du nombre des cas favorables au nombre des cas également possibles ». Dans cette formule s'exprime un théorème très spécial, un théorème tout à fait précis et complètement démontrable — mais pas une définition.

Passons maintenant à la troisième opération simple qui nous permettra de résoudre certains problèmes d'une autre catégorie. Nous désignons cette opération par le terme *départagement* et appellons la règle de calcul introduite par cette opération la « *loi de division* ». Je vais expliquer d'abord le sens de l'opération, en donnant un exemple simple. Reprenons le jeu de roulette avec ses 37 points caractéristiques, numérotés de 0 à 36, dont chacun possède une certaine probabilité p_0 à p_{36}. Supposons maintenant *qu'il soit déjà connu* que, à un certain moment, un nombre impair se soit produit. On demande : Quelle est la probabilité que ce nombre impair soit 17 ? Pour formuler cette question d'une manière générale, il faut envisager un collectif primordial C dont l'ensemble caractéristique E comprend plus de deux points, et départager un ensemble partiel E′ comprenant au moins deux points. Si nous supprimons tous les éléments de C dont les points caractéristiques n'appartiennent pas à E′, nous arrivons à un nouveau collectif C′ avec l'ensemble caractéristique E′. On déduit facilement que, x désignant un point de E′, la probabilité p' de x dans le nouveau collectif C′ est proportionnelle à la probabilité p de x dans C. Puisque la somme de toutes les probabilités dans un collectif est égale à l'unité on a la formule

$$p' = \frac{p}{P},$$

où le dénominateur P désigne la somme (ou l'intégrale) de toutes les probabilités p pour les points de E′, ou plus brièvement, la probabilité de E′ dans C. Il est évident que la déduction de ce résultat est indépendante de la supposition que tous les points de E ont la même probabilité, mais la formule est applicable au cas d'une répartition uniforme dans C. Dans ce cas le quotient p : P peut être interprété comme rapport du nombre des cas favorables au nombre total des cas également possibles — en accord avec la « définition » classique.

Dans la théorie traditionnelle la loi de division n'a pas trouvé de place à côté des lois familières de sommation et de multiplication. Dès les commencements de la théorie des probabilités on a pris l'habitude de parler d'une probabilité des « causes ou des hypothèses » comme d'une notion opposée à celle des probabilités des « faits ». Je ne puis admettre à cette conception. A mon avis il est indispensable d'éliminer d'une théorie mathématique des notions indéterminées qui manquent de clarté comme la confrontation des « faits » et des

« causes ». Il ne s'agit, dans tous les problèmes que nous traitons, que *d'une même probabilité*, limite de la fréquence relative avec laquelle un certain résultat se produit dans une certaine suite d'expériences. Nous reviendrons encore une fois sur la question actuelle quand nous traiterons certains exemples d'opérations combinées.

Les trois opérations simples considérées jusqu'à présent nous mettent en état de déduire un nouveau collectif quand on suppose un seul collectif primordial donné. La quatrième et dernière opération simple nous permet de passer à un nouveau collectif C en partant de deux collectifs donnés C′ et C″. Nous appellons cette opération la *composition* de deux collectifs et la règle de calcul à laquelle elle donne lieu, *la loi de multiplication* ou de composition. Voici l'établissement de la composition de deux collectifs dans le cas le plus simple. Soient $e'_1, e'_2, e'_3, \cdots$ les éléments, $x'_1, x'_2, x'_3, \cdots$ les caractères distinctifs du collectif C′, enfin $e''_1, e''_2, e''_3, \cdots$, $x''_1, x''_2, x''_3, \cdots$ les éléments et les caractères distinctifs de C″. Nous constituons un nouveau collectif C dont le premier élément e_1, est la combinaison des deux expériences e'_1 et e''_1. Le caractère distinctif de e_1 est le point avec les coordonnées x'_1 et x''_1 dans un espace à deux dimensions. De la même façon on définit les éléments $e_2, e_3, \cdots$ comme les combinaisons de e'_2 et e''_2, de e'_3 et e''_3, etc. On en déduit facilement que, si la nouvelle suite $e_1, e_2, e_3, \cdots$ avec les caractères distinctifs $x'_1x''_1, x'_2x''_2, x'_3x''_3, \cdots$ satisfait aux conditions d'un collectif, la probabilité d'une combinaison quelconque de deux valeurs x', x'' est le *produit* des probabilités de x' dans C′ et de x'' dans C″. La question de savoir si les conditions sont remplies ou non, est très délicate et demande une étude spéciale qu'il me faut supprimer ici. Je me borne à constater qu'il faut introduire une certaine distinction des collectifs *composables* ou *non-composables*, un discernement qui ne coïncide pas avec la distinction habituelle d'indépendance ou de dépendance. On peut composer deux collectifs dépendant l'un de l'autre et on arrive à la règle connue de multiplication des probabilités non-indépendantes. — En général l'ensemble caractéristique d'un collectif composé possède $(r' + r'')$ dimensions si les collectifs primordiaux dont il est déduit, ont des espaces caractéristiques à r' et r'' dimensions respectivement.

Permettez-moi de donner une petite table des quatre opérations simples.

1e opération : *Sélection* (des éléments). La répartition reste inchangée.

2e opération : *Mélange* (des caractères distinctifs). Règle de l'addition ou de probabilité totale.

3e opération : *Départagement* (d'une partie de l'ensemble caractéristique). Règle de division.

4e opération : *Composition* (de deux ou plusieurs collectifs). Règle de multiplication ou de probabilité composée.

VII

Ces quatre opérations et les règles de calcul auxquelles elles conduisent comprennent ce qu'on peut appeller les « fondements » de la théorie des probabilités. Je refuse strictement de parler de « conventions », nécessaires pour le traitement des problèmes particuliers. Tout problème n'est qu'une certaine combinaison des opérations simples, et la résolution d'un problème se ramène toujours à une suite d'applications des quatre règles établies. Dans chaque recherche particulière entrent comme données les répartitions (c'est-à-dire les probabilités) des collectifs primordiaux. J'ai développé ce système du calcul des probabilités dans mon « Cours des probabilités » qui vient de paraître l'année dernière (1), en donnant des exemples soigneusement choisis de tous les domaines d'application, jeu de hasard, statistique générale, physique théorique. Aujourd'hui je me bornerai à discuter un ou deux cas caractéristiques pour en expliquer quelques détails ; dans mes deux conférences suivantes j'espère pouvoir traiter certains problèmes d'importance plus générale.

Prenons comme premier exemple, tout à fait simple, la question posée par le Chevalier de MÉRÉ, et résolue, avec certaines restrictions, par FERMAT et PASCAL : Quelle est la probabilité d'amener au moins une fois « six » en jettant quatre fois un dé, ou d'amener au moins une fois douze en jettant 24 fois une paire de dés ? Dans la théorie classique on commence par la recherche des cas également possibles et on compte ces cas d'une certaine manière ; de cette façon, le résultat obtenu n'est applicable qu'au cas d'un dé correct, pour lequel les proba-

(1) R. v. MISES, *Vorlesungen aus dem Gebiete der angewandten Mathematik*, Bd. I. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Leipzig und Wien 1931, p. 574.

bilités d'apparition des faces sont égales entre elles. Notre théorie prend pour point de départ le collectif primordial dont les éléments e sont les coups successifs d'un dé quelconque. La répartition dans ce collectif C est donnée par les six probabilités $p_1, p_2, \cdots, p_6$ dont la somme est l'unité. Nous passons du collectif donné C par l'opération de sélection, effectuée quatre fois, à quatre nouveaux collectifs C_1, C_2, C_3, C_4, dont les éléments sont respectivement $e_1, e_5, e_9, \cdots$; $e_2, e_6, e_{10}, \cdots$; $e_3, e_7, e_{11}, \cdots$; $e_4, e_8, e_{12}, \cdots$; d'après la loi qui domine la première opération simple ou la sélection, les probabilités $p_1, p_2, \cdots, p_6$ sont valables aussi dans $C_1, C_2, \cdots, C_4$. On déduit maintenant, en applicant la deuxième opération simple ou mélange, quatre nouveaux collectifs $C'_1, C'_2, \cdots, C'_4$ dont chacun n'est qu'une alternative : Les points « un », « deux », « trois », « quatre » et « cinq » dans les C y sont confondus et remplacés tous par le nouveau caractère distinctif « zéro », tandis qu'on met le numéro « 1 » pour le point « six ». D'après la loi d'addition, déduite pour l'opération de mélange, on a dans les quatre collectifs $C'_1, \cdots, C'_4$ les probabilités

$$p'_1 = p_6 \qquad \text{et} \qquad p'_0 = p_1 + p_2 + p_3 + p_4 + p_5 = 1 - p_6.$$

Enfin il faut effectuer trois fois une composition de deux collectifs, à savoir de C'_1 avec C'_2, puis du résultat avec C'_3 et du nouveau résultat avec C'_4. Suivant les règles de composition, le collectif résultant de toutes ces opérations a pour ensemble caractéristique les $2^4 = 16$ points dans l'espace à quatre dimensions, les points dont les coordonnées sont les combinaisons des zéros et unités. Pour chaque point on obtient la probabilité en multipliant le nombre correspondant des facteurs p'_0 et p'_1. La réponse à la question posée par MÉRÉ est donnée par la somme des probabilités de tous les points excepté celui dont les coordonnées sont 1, 1, 1, 1, puisque c'est le seul qui correspond à un coup (des quatre dés) sans le point « six ». La règle de multiplication, appliquée trois fois, donne pour ce point la probabilité $p_1'^4 = (1 - p_6)^4$ et ainsi la somme cherchée est égale à $p' = 1 - (1 - p_6^4)^4$.

Dans toute cette recherche aucune allusion n'a été faite à l'existence des cas également possibles. Si l'on désire appliquer le résultat obtenu au jeu avec un dé correct on mettra $p_6 = 1/6$ et obtiendra $p' = 1 - (1 - 1/6)^4 = 0,516$. — Pour le deuxième problème de 24 coups avec une paire de dés on arrive de même façon à la formule

$$p' = 1 - (1 - p_6 q_6)^{24},$$

où p_6, q_6 sont les probabilités — supposées données — d'amener « six » avec le premier ou le deuxième dé respectivement. Pour deux dés corrects on mettra $p_6 = q_6 = 1/6$ et aura $p' = 0{,}491$. On sait bien que ces deux chiffres sont en bon accord avec les observations qu'avait faites le Chevalier de MÉRÉ avant que la théorie fût abordée par les mathématiciens. Je cite ce fait, car il nous montre de nouveau que dans les applications de la théorie des probabilités on ne doit envisager que les phénomènes réels ainsi qu'on fait en mécanique, en thermodynamique ou dans toute autre branche de la physique.

Envisageons encore un deuxième exemple où l'opération de départagement ou la règle de division intervient également. Soit un nombre m d'urnes différentes contenant chacune des boules blanches et des boules noires et soient p_ν, q_ν respectivement les probabilités de tirer une boule blanche ou noire de la $\nu^{ème}$ urne. Les valeurs p_ν, q_ν, pour lesquelles on a $p_\nu + q_\nu = 1$, représentent les répartitions connues dans m collectifs primordiaux, $C_1, C_2, \cdots, C_m$. En outre nous supposons donné un collectif C dont l'élément consiste dans le procédé de faire sortir une certaine boule des m urnes et dont la répartition est définie par les probabilités $p'_1, p'_2, \cdots, p'_m$ dont la somme $p'_1 + p'_2 + \cdots + p'_m = 1$. Les questions à résoudre sont les suivantes : On a tiré une fois et amené une boule blanche ; quelle est la probabilité pour que la boule vienne de sortir de l'urne avec le numéro ν ? On a effectué n tirages sur une même urne et en amené x boules blanches, quelle est la probabilité que les tirages se soient faits dans l'urne avec le numéro ν ?

Pour passer des $(m + 1)$ collectifs primordiaux donnés au collectif dérivé correspondant aux questions posées, on commence par composer chacun des collectifs $C_1, C_2, \cdots C_m$ avec le collectif C. Le produit $p_\nu p'_\nu$, donne la probabilité pour que, dans une expérience composée, l'urne numérotée ν sorte et dans cette urne une boule blanche soit tirée. L'ensemble caractéristique du collectif déduit par composition comprend $2m$ points, d'un espace à 2 dimensions, $0{,}1$; $0{,}2$; $\cdots$ $0{,}m$ et $1{,}1$; $1{,}2$; $\cdots$ $1{,}m$; si on désigne par 0 et 1 une boule noire ou blanche respectivement. Maintenant on effectue une opération simple de troisième espèce en départageant, comme ensemble partiel, les points avec l'abscisse 1 qui donnent pour probabilité totale $P = p_1 p'_1 + p_2 p'_2 \cdots p_m p'_m$. D'après la régle de division il suit que la probabilité pour que la boule blanche tirée soit sortie de l'urne avec le numéro ν est :

$$\frac{p_\nu p_\nu'}{\mathrm{P}} = \frac{p_\nu p_\nu'}{p_1 p_1' + p_2 p_2' + \cdots p_\nu p_\nu'}.$$

Comme le précédent, ce résultat est indépendant de la supposition que des cas également possibles existent : en outre il est identique à la formule connue sous le nom de formule de BAYES.

Pour répondre à la deuxième question il faut d'abord résoudre le problème de BERNOULLI, c'est-à-dire chercher la probabilité pour que, dans n répétitions d'une même expérience, un événement se produise x fois. Dans ce but on opère $(n - 1)$ fois une composition de deux collectifs pour parvenir à un collectif avec un caractère distinctif à n dimensions, puis on effectue le mélange qui réunit tous les points pour lesquels la somme des n coordonnées (égales à 0 ou 1) soit x. En utilisant les lois de multiplication et d'addition on arrive à la formule bien connue

$$C_n^x \cdot p_v^x (1 - p_v)^{n-x}.$$

En remplaçant p_v par cette expression dans la formule précédente, on obtient la réponse cherchée.

$$\frac{p_v^x (1 - p_v)^{n-x} \cdot p_v'}{\sum\limits_v p_v^x (1 - p_x)^{n-x} \cdot p_v'},$$

On voit de nouveau, dans cet exemple, comment des problèmes plus ou moins compliqués se ramènent à une combinaison successive d'opérations simples.

VIII.

J'ai fini la légère esquisse que je m'étais proposé de vous donner aujourd'hui, esquisse d'une nouvelle méthode de concevoir les idées et de traiter les problèmes de la théorie des probabilités. Permettez-moi de n'ajouter que deux remarques concernant les relations de ma méthode avec certaines idées modernes qu'on rencontre dans les publications mathématiques récentes.

Il faut mentionner d'abord les recherches abordées par M. CANTELLI, continuées par M. Fréchet et quelques autres qui s'occupent d'une nouvelle notion de « variable aléatoire » et essayent de définir une certaine « convergence au sens du calcul des probabilités ». Mais qu'est-ce que c'est qu'une variable aléatoire ? Les auteurs ne donnent pas une définition exacte de cette notion qui entre essentiellement dans toutes leurs déductions, ils prétendent qu'elle soit connue *a priori*. En vérité il ne s'agit pas d'une classe spéciale des variables indépendantes, mais plutôt d'une classe spéciale des *fonctions*, à savoir des fonctions de répartition. Tous les théorèmes déduits dans cet ordre d'idées sont des théorèmes concernant les fonctions monotones, croissantes de 0 à 1, et représentant les répartitions dans certains collectifs. La variable aléatoire est le *caractère distinctif* dans un collectif, et il n'existe aucune autre possibilité d'introduire cette notion que de commencer par la définition du collectif. Si on essaie de compléter les propositions citées, dans le sens d'une fondation suffisante des notions appliquées on arrive inévitablement à la théorie établie dans ma conférence. Quant aux théorèmes sur la convergence « au sens du calcul des probabilités » il y est question de certains théorèmes de limites concernant les fonctions de répartition. Chacun de ces théorèmes se traduit d'une manière simple par un énoncé *sur la limite d'une certaine suite des fonctions de répartition*, la notion de limite prise au sens généralement adopté dans l'analyse.

Une autre classe de problèmes traitée dans de récentes publications touche aux idées établies par M. Borel. Il s'agit d'une sorte de probabilité purement mathématique, c'est-à-dire des probabilités dont l'objet n'appartient pas au monde physique comme le jeu de hasard, mais à l'arithmétique même. On parle par exemple de la probabilité pour qu'un certain chiffre, 3 par exemple, apparaisse dans le développement d'un certain nombre décimal. Il est entendu qu'ici le mot « probabilité » ne signifie que la limite de la fréquence relative avec laquelle le chiffre se présente dans la fraction décimale infinie. La définition classique qui parle du rapport des cas favorables aux cas également possibles est supprimée, en fait, bien qu'on ne le mentionne *pas expressément*. Mais la différence essentielle entre ce point de vue et ma théorie se manifeste dans le rôle que joue la notion d'irrégularité. Pour appliquer quelques règles du calcul des probabilités classiques aux nombres décimaux il faut présumer que ce développement possède une certaine irrégularité. Je cite par exemple la notion de « nombre entièrement normal », introduite par M. BOREL. On y suppose que chaque suite de n chiffres, (soit par exemple la suite 1, 2, 7, avec $n = 3$) se présente à la limite avec la fréquence 10^{-n} $\left(\text{donc } \frac{1}{1000}\right)$, si on considère tous les groupes de n chiffres successifs dans le développement. On démontre d'une manière élégante que l'ensemble des nombres normaux a pour limite l'unité, sans qu'on puisse décider si un nombre donné quelconque, par exemple le nombre π, est normal ou non. Mais quoi qu'il en soit il y a ici une base exacte pour en déduire certains théorèmes mathématiques concernant les nombres normaux. Ce que je prétends est que cette base et ces déductions n'ont rien à faire avec les problèmes et l'objet du calcul des probabilités au sens habituel, et voici pourquoi. La définition du nombre normal établit une certaine irrégularité dans la succession des chiffres des développements considérés. Si on prend par exemple pour choix de position chaque $n^{\text{ième}}$ chiffre, on pourra démontrer que la fréquence relative des zéros reste inchangée dans l'ensemble partiel des chiffres. Mais on peut choisir aussi les nombres entiers les plus grands compris dans $\pi, \pi^2, \pi^3, \cdots$, ou la suite des nombres premiers $3, 5, 7, 11, \cdots$ et maintenant on ne sait plus si dans ce choix la fréquence des zéros n'a pas une toute autre valeur. Si on ne supposait pour une suite de chiffres formant un collectif, que l'irrégularité restreinte qui définit un nombre normal, on ne pourrait plus affirmer qu'aucune martingale, aucun système à jouer ne soit possible. Au contraire, si on donne un nombre normal quelconque et si l'on suppose qu'il représente la suite des résultats d'un jeu de hasard ou d'un collectif (dont l'ensemble caractéristique est formé par les nombres 0,1,2,... 9) il est immédiatement possible de construire un choix de position qui altère les chances du joueur.

Ces remarques ne tendent pas à déprécier les recherches vraiment élégantes et très intéressantes qui concernent les nombres normaux et des problèmes semblables. Mais il me semble utile de délimiter la tâche et la portée de l'une et l'autre théorie. Dans mes prochaines conférences je me propose de compléter les résultats acquis aujourd'hui en signalant leurs diverses applications.

Les théorèmes fondamentaux

Dans ma première conférence j'ai eu l'honneur de développer une nouvelle théorie des probabilités qui prétend être une théorie rationnelle du même genre que la géométrie ou la mécanique rationnelle. Cela veut dire : de même que la mécanique donne une représentation idéalisée des phénomènes concrets du mouvement et de l'équilibre, la théorie rationnelle des probabilités veut fournir une construction abstraite, idéalisée ou mathématisée, de certains phénomènes observables, à savoir des « phénomènes en masse » ou des suites d'expériences répétées. Aujourd'hui je me propose d'établir quelques théorèmes analytiques que j'appelle les « théorèmes fondamentaux du calcul des probabilités » et qui touchent aux énoncés bien connus de la loi de Bernoulli, loi des grands nombres, loi de Laplace, de Bayes etc. A mon avis, tous ces théorèmes parviennent, dans ma théorie, à une précision qui leur est inaccessible sans l'aide des notions établies dans les développements précédents. Surtout les lois des grands nombres — nous parviendrons à deux énoncés distincts au lieu d'un seul — sont étroitement liées aux définitions fondamentales qui forment la base de ma théorie.

I

Permettez-moi d'abord de résumer en peu de mots les traits principaux de la théorie que j'ai déjà exposée. Il y a cinq points essentiels.

1. *Introduction de la notion de collectif.* — Avant de parler d'une probabilité au sens rationel il faut, dans chaque cas particulier, établir un collectif, c'est-à-dire une suite infinie d'expériences ou d'observations précisément définies, suite qui satisfasse aux deux conditions suivantes.

Première condition à remplir par un collectif : Soit A un ensemble partiel de l'ensemble caractéristique, c'est-à-dire une portion quelconque des résultats possibles des observations, et n_A le nombre de telles expériences parmi les n premières dont les résultats appartiennent à A ; il existe alors une limite p_A de la fréquence relative $n_A : n$, lorsque le nombre n augmente indéfiniment.

Deuxième condition à remplir par un collectif : Si on effectue un « choix de position » sur les éléments de la suite, c'est-à-dire un choix indépendant du résultat de l'expérience choisie, la fréquence relative des expériences appartenant à la portion A (si l'on ne tient compte que des éléments choisis) ne change pas et reste à la limite égale à p_A.

2. *Définition de la probabilité mathématique* : La limite p_A de la fréquence relative $n_A : n$, limite indépendante du choix de position, s'appelle la probabilité des résultats A (ou des évènements) représentés par A) dans le collectif considéré. Une probabilité sans cette adjonction « dans le collectif considéré » n'existe pas pour nous.

3. *Le but général de la théorie des probabilités* consiste en ceci : Au moyen de certaines opérations, on peut déduire, d'un collectif donné ou de plusieurs collectifs donnés un nouveau collectif ; la théorie des probabilités nous apprend à calculer la répartition des probabilités dans le collectif final, à partir des répartitions dans les collectifs initiaux.

Remarquons encore que la théorie mathématique des probabilités n'a pas à s'occuper de la recherche des probabilités dans les collectifs primordiaux qui servent comme point de départ dans un problème déterminé. Cette recherche est une affaire essentiellement expérimentale, comme l'établissement des vitesses initiales ou des forces motrices dans un problème de mécanique, etc.

4. *Caractère général des répartitions.* — Dans tous ces raisonnements, aucune distinction n'est faite entre le cas des probabilités discontinues ou arithmétiques et celui des probabilités continues ou géométriques. Au contraire nos définitions comprennent le cas le plus général d'une répartition quelconque.

5. *Les quatre opérations simples.* Tout procédé qui permet de déduire un nouveau collectif des collectifs donnés se ramène à une combinaison de quatre opérations simples ou fondamentales, à savoir : *a*). La *sélection* ou choix de position qui ne change pas la répartition ; *b*) Le *mélange* qui conduit à la loi d'addition des probabilités ; *c*). Le *départagement,* fournissant la règle de division ; enfin *d*) La *composition* de deux collec-

tifs qui conuit à la règle de multiplication. Chaque problème en théorie des probabilités se résout par une combinaison de ces opérations simples en nombre fini ou infini.

Ajoutons que ces règles s'appliquent de la même manière aux trois types de questions jeux de hasard, la statistique générale et l'assurance sur la vie, enfin les questions posées par la physique moderne.

II

Les théorèmes de limite ou théorèmes fondamentaux dont je vais maintenant vous donner un bref exposé, sont étroitement liés, ainsi que je l'ai déjà dit, aux fondements même de ma théorie. En suivant mes développements, sans doute vous êtes-vous déjà posé la question : la définition d'une probabilité comme limite d'une fréquence relative, n'implique-t-elle pas une *anticipation* de la loi des grands nombres ? Ou plutôt une *contradiction*, puisqu'on y suppose sûr ce que la loi de Bernoulli ne donne que comme vraisemblable ? J'espère vous convaincre que la théorie, basée sur l'idée de la fréquence relative, est libre de toute contradiction et que, au contraire, le *vrai sens* de la dite loi ne se découvre qu'au moyen de cette théorie. Pour traiter en général ces questions et différentes autres du même genre il faut d'abord développer quelques questions analytiques. Je vais les donner sans insister sur la démonstration détaillée.

Commençons par établir un simple théorème sur la *limite d'un produit de fonctions*. Soit $f_1(x)$, $f_2(x)$, $f_3(x)$,... une suite infinie de fonctions d'une variable réelle, les f étant complexes ou réels. Supposons que, pour $x = a$, les fonctions f satisfassent aux conditions suivantes :

$$(1) \quad f_v(a) = 1, \quad f_v'(a) = 0, \quad f_v''(a) = -r_v^2 < 0\,; \quad r^2 \leqq r_v^2 \leqq R^2.$$

les troisièmes dérivées au plus étant uniformément bornées au voisinage de a, et

$$(1') \qquad |f_v'''(x)| < c \quad \text{pour} \quad |x - a| < \varepsilon.$$

Si nous posons

$$(2) \qquad r_1^2 + r^2 + \cdots r_n^2 = s_n^2 \cdot \qquad u = s_n(x - a),$$

on a le théorème suivant, facile à démontrer :

$$(3) \quad \lim_{n \to \infty} \left[f_1\left(a + \frac{u}{s_n}\right) \cdot f_2\left(a + \frac{u}{s_n}\right) \cdots f_n\left(a + \frac{u}{s_n}\right) \right] = e^{-\frac{u^2}{2}}, \qquad |u| \leqq \mathrm{U}.$$

J'ai publié la démonstration de ce théorème et de tout ce qui suit en 1919 dans le 4e volume de la « Mathematische Zeitschrift ». On trouvera aussi la plupart des démonstrations dans mon « Cours des probabilités » qui vient de paraître.

Supposons maintenant que la valeur absolue de f_v, hors de $x = a$, reste toujours au-dessous de l'unité, et qu'il existe un $\eta > 0$ pour tout ε :

$$(1'') \qquad |f_v(x)| < 1 - \eta \qquad \text{pour} \qquad |x - a| > \varepsilon,$$

et que enfin les f_v s'évanouissent assez rapidement à l'infini.

$$(1''') \qquad |f_v(x)| = o(x^{-k}) \qquad k > 0.$$

Sous ces conditions on peut démontrer qu'au surplus

$$(4) \quad \lim_{n \to \infty} \int_{-\infty}^{x} f_1\left(a + \frac{u}{s_n}\right) f_2\left(a + \frac{u}{s_n}\right) \cdots f_n\left(a + \frac{u}{s_n}\right) \psi(u) du = \int_{-\infty}^{x} e^{-\frac{u^2}{2}} \psi(u) du,$$

pour $|\psi|$ borné. La convergence vers la limite est uniforme pour toute valeur de x entre $-\infty$ et $+\infty$. — Remarquons entre parenthèse que l'équation (4) est la vraie source de la formule de Stirling qui s'en déduit facilement.

Nous appliquerons ces théorèmes (3) et (4) concernant le produit d'un nombre infini de fonctions, aux questions de la théorie des probabilités, aussi bien directement, qu'en utilisant une transformation de Laplace. Envisageons d'abord, en nous bornant au cas d'une seule dimension, une série de fonctions de répartition $P_1(x), P_2(x), P_3(x), \cdots$. Les $P_v(x)$ sont réels, monotones, non-décroissants, et satisfont aux conditions

$$(5) \qquad P_v(-\infty) = 0, \qquad P_v(+\infty) = 1.$$

On appelle la « *Transformée de Laplace* » d'une fonction de répartition $P_v(x)$ la fonction complexe, définie par l'intégrale de Stieltjes :

$$(6) \qquad \mathfrak{P}_v(z) = \int_{-\infty}^{+\infty} e^{xzi} dP_v(x).$$

Il est avantageux de modifier cette définition. Nous désignons par a_v la valeur moyenne de la répartition $P_v(x)$, c'est-à-dire

$$a_v = \int_{-\infty}^{\infty} x dP_v(x), \tag{7}$$

et nous posons

$$f_v(z) = \int_{-\infty}^{\infty} e^{(x - a_v)zi} dP_v(x), \qquad v = 1, 2, 3, \cdots \tag{8}$$

Ces fonctions f_v ont pour $a = 0$ toutes les qualités exigées dans (1) et (1') :

$$f_v(0) = 1, \quad f_v'(0) = \int (x - a_v) dP_v(x) = 0, \quad f_v''(0) = -\int (x - a_v)^2 dP_r(r).$$

On y voit que la deuxième dérivée de f_v est égale, au signe près à la déviation de la répartition $P_v(x)$.

Pour remplir la condition concernant la troisième dérivée de f_v il faut et il suffit que les moments absolus du troisième degré soient bornés pour la suite $P_1(x), P_2(x), P_3(x), \cdots$.

Jusqu'ici distinction n'est pas faite entre les cas des probabilités arithmétiques et géométriques. Si nous supposons maintenant que $P_v(x)$ soit la fonction de répartition d'un *collectif arithmétique*, avec les caractères distinctifs $1, 2, 3, \cdots m$ et les probabilités correspondantes $p_v(1), p_v(2), \cdots, p_v(m)$ [de sorte que par exemple $P_v(x)$ est égal à $p_v(1) + p_v(2)$ pour $2 \leqq x < 3$] on arrive au lieu de (8) à :

$$f_v(z) = \sum_{x=1}^{m} e^{(x - a_v)zi} p_v(x). \tag{8'}$$

Comme *inversion* de cette formule on déduit d'une manière élémentaire :

$$p_v(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_v(z) e^{(a_v - x)zi} dz, \qquad x = 1, 2, \cdots m. \tag{9'}$$

D'ailleurs, en supposant qu'il s'agisse d'un *collectif géométrique*, de sorte que $P_v(x)$ ait une dérivée continue $p_v(x)$ on peut remplacer (8) par

$$f_v(z) = \int_{-\infty}^{\infty} e^{(x - a_v)zi} p_v(x) dx, \tag{8''}$$

et sous la condition que $p(x)$ soit à variation bornée, le théorème de Fourier nous donne l'inversion :

$$(9'') \qquad p_v(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_v(z)e^{(a_v - x)zi}dz.$$

III

Les formules (9') et (9'') servent à résoudre un problème très important du calcul des probabilités. Nous introduisons sous le nom de « *sommation de deux collectifs* » une opération non-simple qui est combinée de la *composition* de deux collectifs et un certain *mélange*. La composition des collectifs avec les probabilités $p_1(x)$ et $p_2(x)$ amène la répartition à deux dimensions $p_1(x)$, $p_2(y)$ — les p signifiant une probabilité ponctuelle dans le cas arithmétique ou une densité au cas géométrique. En effectuant un mélange on passe à la répartition à une seule dimension

$$(10) \quad q_2(x) = \sum_z p_1(z)p_2(x - z) \qquad \text{ou} \qquad q_2(x) = \int_{-\infty}^{\infty} p_1(z)p_2(x - z)dz.$$

Evidemment $q_2(x)$ fournit la probabilité pour que la *somme des caractères distinctifs* du premier et du deuxième collectif soit égale à x. On déduit facilement de (10) que la valeur moyenne b_2 de $q_2(x)$ et la déviation s_2^2 de $q_2(x)$ soient respectivement égales à $a_1 + a_2$ et à $r_1^2 + r_2^2$, (r_v^2 = déviation de P_v) et que la transformée de Laplace pour la répartition $Q_2(x)$ soit

$$(11) \qquad g_2(z) = \int_{-\infty}^{\infty} e^{(x - b_n)zi}dQ_2(x) = f_1(z) \cdot f_2(z).$$

De la même manière on arrive, pour la « *sommation des n collectifs* » $P_1(x)$, $P_2(x)$, $\cdots$ $P_n(x)$, à une fonction de répartition $Q_n(x)$ dont la transformée de Laplace $g_n(z)$ est donnée par

$$(12) \qquad g_n(z) = f_1(z)f_2(z)\cdots f_n(z).$$

En appliquant les formules inverses (9') et (9'') on a maintenant pour la probabilité (ou la densité de probabilité) dans le collectif déduit par sommation des n collectifs donnés :

$$(12') \qquad q_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_1(z)f_2(z)\cdots f_n(z)e^{(b_n - x)zi}dz.$$

(Pour le cas des probabilités arithmétiques il faut poser ici $f_v(z) = 0$ pour $|z| > \pi$). Prenons maintenant

$$z = \frac{u'}{s_n}, \qquad x = b_n + s_n\sqrt{2}\,u,$$

$$b_n = a_1 + a_2 + \cdots a_n$$
$$s_n^2 = r_1^2 + r_2^2 + \cdots r_n^2,$$

et utilisons le théorème (4) sur la limite d'un produit de fonctions. On en déduit facilement

$$(13) \qquad \lim_{n \to \infty} s_n\sqrt{2}\,q_n(b_n + s_n\sqrt{2}\,u) = \frac{1}{\sqrt{\pi}}e^{-u^2}, \; |u| < U$$

ou :

$$(13') \qquad q_n(x) \backsim \frac{1}{s_n\sqrt{2\pi}}e^{-\frac{(x-b_n)^2}{2s_n^2}}.$$

Pour que (13) et (13') soient exactes il faut vérifier encore les conditions à remplir par les fonctions $f_v(x)$ hors du point $x = 0$, à savoir : la valeur absolue de $f_v(x)$ doit rester au-dessous de l'unité et s'évanouir assez rapidement (par exemple comme une puissance négative de x) pour x infini. Dans le cas des probabilités géométriques où $p_v(x)$ et $q_v(x)$ signifient des densités de probabilité, les conditions sont remplies, si on suppose que les *variations totales* des fonctions $p_1(x), p_2(x), p_3(x), \cdots$ possèdent une *borne supérieure*. Dans le cas des probabilités arithmétiques où $p_v(1), p_v(2), \cdots$ et de même $q_v(1), q_v(2), \cdots$ signifient les probabilités des caractères distinctifs $1, 2, \cdots$, il faut soumettre les $p_v(x)$ à une autre *condition supplémentaire*. Il suffit de supposer que pour chaque v (excepté au plus un nombre fini des v) il existe un x tel que $p_v(x)$ et $p_v(x+1)$ soient tous deux différents de zéro. On tomberait en défaut par exemple si tous les $p_v(x)$ s'évanouissaient pour les x impairs.

Résumons maintenant les résultats auxquels nous sommes parvenus ; il y en a deux :

1) *La sommation des n collectifs avec les probabilités arithmétiques $p_1(x)$, $p_2(x)$, $\cdots p_n(x)$ ($x = 0, 1, 2, \cdots m$) conduit pour n infini à la loi de Gauss pour la probabilité ponctuelle $q_v(x)$ ($x = 0, 1, 2, \cdots nm$) sous les conditions suivantes :*

a) La somme des déviations r_v^2 des répartitions primordiales tend vers l'infini avec n ;

b) Les moments absolus de troisième degrés possèdent une borne supérieure ;

c) Dans chaque répartition primordiale — excepté au plus un nombre fini d'entre elles — il existe au moins une paire de nombres immédiatement voisins pour lesquels la probabilité p_v ne s'évanouit pas.

2) *La sommation des n collectifs avec les densités de probabilité $p_1(x)$, $p_2(x) \cdots p_n(x)$ conduit pour n infini à la fonction de Gauss pour la densité de probabilité $q_n(x)$ sous les conditions suivantes :*

a) et b) comme au premier cas ;

c) Les variations totales des densités primordiales $p_v(x)$ possèdent une borne supérieure.

Ajoutons la remarque suivante : Dans les deux cas on peut, partant des théorèmes établis, en déduire une conséquence bien connue. Si on n'envisage, dans le collectif résultant, que la *fonction de somme* ou fonction de répartition $Q_n(x)$ — la somme ou l'intégrale indéfinie de $q_n(x)$ dans le premier ou deuxième cas respectivement. — il suit de (13')

$$Q_n(x_2) - Q_n(x_1) \backsim \frac{1}{\sqrt{\pi}} \int_{u_1}^{u_2} e^{-u^2} du, \tag{14}$$

avec

$$u_1 = \frac{x_1 - b_n}{s_n\sqrt{2}}, \qquad u_2 = \frac{x_2 - b_n}{s_n\sqrt{2}}.$$

On sait bien que cette équation (14) est également valable sous des conditions beaucoup plus générales. Il ne faut point supposer que les répartitions données soient arithmétiques ou géométriques. Les beaux travaux de M. Lindenberg par exemple nous montrent que les deux conditions *a*) et *b*), mentionnées plus haut, suffisent déjà pour que (14) soit justifiée. Mais notre équation (13) qui fournit deux énoncés spéciaux, *pour les probabilités ponctuelles et les densités de probabilité respectivement*, perdrait son sens si on ne supposait pas les collectifs donnés arithmétiques ou géométriques et elle tomberait en défaut si les conditions *c*) (ou des conditions semblables) n'étaient pas remplies.

J'appelle « *premier théorème fondamental du calcul des probabilités* » l'énoncé contenu dans les équations (13) et (14) et sa généralisation pour le cas des plusieurs dimensions. On peut le formuler, en supprimant quelques détails, comme il suit : *La probabilité pour que la somme des n variables indépendantes ait la valeur x est donnée, pour n infiniment grand, par une loi de Gauss dont la moyenne et la déviation se calculent par l'addition des moyennes et des déviations des répartitions données.* La « précision » de la fonction de Gauss, égale à la moitié de la valeur réciproque de la déviation, tend vers zéro avec n croissant indéfiniment. — Avant d'entrer dans une discussion de ce théorème je passe à la démonstration du deuxième.

IV

Nous avons défini, plus haut, une certaine opération non-simple sous le nom de « sommation de n collectifs », et l'étude de cette opération pour le cas $n \to \infty$ nous a conduit au premier théorème fondamental. Je vais introduire maintenant une autre opération que je nommerai « *conclusion a posteriori* » ou, plus complètement, « *conclusion tirée de n répétitions d'une expérience* ». Le problème qui, pour n infini, fournira le deuxième théorème fondamental est une généralisation du problème de Bayes.

Imaginons un ensemble d'urnes arrangées sur une table, dont chacune soit caractérisée par les valeurs spéciales d'un groupe de r chiffres, $x_1, x_2, \cdots x_r$. Soit donnée comme répartition primordiale une densité $p(x_1, x_2, \cdots x_r)$, signifiant la probabilité de saisir une urne avec les caractères distinctifs $x_1, x_2, \cdots x_r$. (On appelle souvent ce p la probabilité *a priori*.) Supposons que chaque urne contienne des billets portant les chiffres $a_1, a_2, a_3, \cdots a_{r+1}$, et que la variable x_k signifie, pour chaque urne, la probabilité d'en tirer un billet avec le numéro $a_k (k = 1, 2, \cdots r + 1)$, d'où il suit que $x_1 + x_2 + \cdots x_{r+1} = 1$. Supposons enfin qu'on ait tiré n fois dans une même urne et qu'on y ait trouvé $\alpha_1 n$ fois le numéro a_1, $\alpha_2 n$ fois le numéro $a_2, \cdots, \alpha_{r+1} n$ fois le numéro a_{r+1}, ainsi que $\alpha_1 + \alpha_2 + \cdots \alpha_{r+1} = 1$. On pose la question : Quelle est la probabilité pour que l'urne qui a donné le résultat mentionné soit une urne avec les caractères distinctifs $x_1, x_2, \cdots x_r$?

Évidemment on a affaire dans notre problème à un collectif qui est composé des deux suivants : La première composante a pour élément le procédé de saisir une certaine des urnes arrangées sur la table, et sa répartition est donnée par la densité $p(x_1, x_2, \cdots x_r)$. La deuxième composante a pour élément le tirage n fois répété dans une même urne, et sa répartition est définie par les probabilités arithmétiques, proportionelles à

$$\left[x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_{r+1}^{\alpha_{r+1}}\right]^n. \tag{15}$$

Dans ce deuxième collectif les $\alpha_1, \alpha_2, \cdots \alpha_{r+1}$ jouent le rôle des caractères distinctifs, les $x_1, x_2 \ldots$ sont des paramètres donnés. D'après la règle de multiplication le produit de $p(x_1, x_2, \cdots x_r)$ avec l'expression précédente fournit — abstraction faite d'un facteur constant — la probabilité pour l'élément composé dont le caractère distinctif comprend les $2r$ variables x_k et α_k. Enfin il faut effectuer un *départagement* : On connaît déjà le résultat du tirage n fois répété, et on ne cherche que la probabilité « *a posteriori* » $q(x_1, x_2, \cdots x_r)$ des valeurs des x_n. Suivant la règle de division que j'ai établie dans ma première conférence il faudrait diviser le produit de $p(x_1, x_2, \cdots x_r)$ et l'expression (15) par la somme de ces produits pour toutes les valeurs possibles de $x_1, x_2, \cdots x_r$, c'est-à-dire, dans notre cas, par l'intégrale à r dimensions du produit, prise sur le cube $x_k = 0$ jusque $x_k = 1$. Cependant, le diviseur étant une constante par rapport aux $x_1, x_2, \cdots x_r$, on peut écrire :

$$q(x_1, x_2, \cdots x_r) = C \cdot p(x_1, x_2, \cdots x_r)\left[x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_{r+1}^{\alpha_{r+1}}\right]^n, \tag{16}$$

en ajoutant que la constante C soit déterminée à l'aide de l'équation

$$\int q(x_1, x_2, \cdots x_r) dx_1 dx_2 \cdots dx_r = 1, \tag{16'}$$

l'intégrale étant prise sur le cube mentionné.

Au second membre de l'équation (16) il faut remplacer x_{r+1} et α_{r+1} par

$$x_{r+1} = 1 - x_1 - x_2 - \cdots x_r \quad \text{et} \quad \alpha_{r+1} = 1 - \alpha_1 - \alpha_2 - \cdots \alpha_r; \tag{16''}$$

on obtient ainsi un produit de n fonctions, égales entre elles, fonctions des variables $x_1, x_2, \cdots x_r$. Notre théorème (3) sur la limite d'un produit de fonctions, généralisé pour le cas de plusieurs variables indépendantes, nous fournit — je supprime quelques transformations détaillées — le résultat asymptotique

$$(17) \qquad q(x_1, x_2, \cdots x_r) \sim \sqrt{\left(\frac{n}{2\pi}\right)^r \frac{1}{\alpha_1 \alpha_2 \cdots \alpha_{r+1}}}\, e^{-\frac{n}{2}\sum_{\nu=1}^{r+1} \frac{(x_1 - \alpha_2)^2}{\alpha_\nu}}.$$

On y voit que *l'influence de la probabilité, dite « a priori »* $p(x_1, x_2, \cdots x_r)$, *s'évanouit pour n infini.* La probabilité $q(x_1, x_2, \cdots x_r)$, trouvée par « conclusion *a posteriori* », ne dépend que des fréquences relatives $\alpha_1, \alpha_2, \cdots \alpha_r$ avec lesquelles les nombres $a_1, a_2, \cdots a_r$ se sont produits parmi les n tirages effectués.

On peut regarder déjà l'énoncé de l'équation (17) comme le deuxième théorème fondamental, mais il est avantageux d'en déduire une conclusion plus concise qui lui est essentiellement équivalente. Nous opérons *encore un mélange* sur le collectif considéré actuellement, en posant la question : Quelle est la probabilité pour que la moyenne ou « l'espérance mathématique » du nombre tiré, c'est-à-dire, la variable

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots \alpha_{r+1} x_{r+1}.$$

soit située près d'une valeur x ? La réponse se trouve, d'après la règle d'addition, par l'intégration de $q(x_1, x_2, \cdots, x_z)$ sur l'espace à r dimensions, terminé par les deux hyperplans infiniment voisins

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots \alpha_{r+1} x_{r+1} = x, \quad \text{et} \quad \alpha_1 x_1 + \alpha_2 x_2 + \cdots \alpha_{r+1} x_{r+1} = x + dx.$$

Je donne le résultat sans démonstration. La densité cherchée $q(x)$, pour n infini, s'écrit en formule asymptotique

$$(18) \qquad q_n(x) \sim \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n(x-\alpha)^2}{2\sigma^2}},$$

où α et σ^2 signifient la moyenne arithmétique et la déviation (le carré de l'écart quadratique) des n résultats du tirage,

$$(18') \quad \left\{ \begin{aligned} &\alpha = \alpha_1 a_1 + \alpha_2 a_2 + \cdots \alpha_{r+1} a_{r+1} \\ \text{et } &\sigma^2 = \alpha_1 (a_1 - \alpha)^2 + \alpha_2 (a_2 - \alpha)^2 + \cdots \alpha_{r+1} (a_{r+1} - \alpha)^2. \end{aligned} \right.$$

L'analogie des équations (13') et (18) est évidente. C'est pourquoi il convient de les rapprocher sous les noms de premier et deuxième théorème fondamental du calcul des probabilités. Permettez-moi de formuler encore une fois le résultat compris dans (18), en supprimant quelques détails : *Si dans un tirage n fois répété sur une même urne ou, autrement dit, si dans l'observation des n premiers éléments d'un collectif quelconque, n nombres sont apparus avec la moyenne* α *et la déviation (carré de l'écart quadratique)* σ^2, *alors la probabilité pour que l'espérance mathématique du tirage soit située près d'une valeur x, est donnée, pour n infini, par une loi de Gauss dont la moyenne est* α *et la déviation* $\sigma^2 : n$. — La précision tend vers l'infini avec n, l'influence de la probabilité « *à priori* » s'évanouit. — L'équation (18) est une conclusion directe de notre théorème sur le produit de n fonctions tandis que le premier théorème fondamental en était déduit au moyen d'une transformation de Laplace.

V

Je passe maintenant à une brève discussion des deux théorèmes fondamentaux, en me proposant surtout d'éclaircir les relations qu'ils ont avec l'axiome principal de ma théorie qui affirme l'existence d'une limite de la fréquence relative dans la suite des éléments d'un collectif.

Quant au premier théorème fondamental, concernant la « sommation de n collectifs » il suffira maintenant de se borner au cas le plus simple où les collectifs considérés ne sont que des alternatives, égales entre elles, avec les probabilités particulières $p_v(0) = p_v(1) = \frac{1}{2}$. La somme des n caractères distinctifs y est identique au nombre des sorties « 1 ». Pour fixer les idées on envisagera un jeu de pile ou face où le numéro « 1 » représente le résultat « face ». La formule (13') fournit la probabilité — pour n infini — pour que dans n répétitions du jeu l'évènement « face » se produise x fois. Pour le moment, seul nous intéresse comme conséquence particulière de l'équation intégrée (14), un résultat qui fut déjà démontré par Jacques BERNOULLI avant la découverte de l'équation générale. *La probabilité pour que le rapport x : n soit situé entre les limites* $\frac{1}{2} - \varepsilon$ *et* $\frac{1}{2} + \varepsilon$, *quelque petit soit* ε, *tend vers l'unité, quand n augmente infiniment.*

On est bien tenté de croire que ce théorème de Bernoulli est en contradiction avec l'axiome de l'existence d'une limite de la fréquence relative. Mais cette opinion n'est basée que sur une conception erronée du mot « probable ». Il est bien entendu que, dans un énoncé de notre théorie, le mot « probabilité » ne pourra jamais signifier autre chose qu'une limite de fréquence relative. Le vrai sens de la proposition citée de Bernoulli, envisagée du point de vue de notre définition de la probabilité, est le suivant : On considère une série de n épreuves successives comme *un élément individuel* d'un collectif ; soit $x : n$ son caractère distinctif x signifiant le nombre de résultats « 1 » ; alors, dans une suite infinie de tels éléments la fréquence relative des éléments pour lesquels le rapport $x :$ n est situé dans les limites $\frac{1}{2} - \varepsilon$, $\frac{1}{2} + \varepsilon$, ne diffère que très peu de l'unité pour n assez grand. Voilà deux propositions distinctes, mais compatibles, concernant une série infinie de zéros et de uns : L'une (ma définition de la probabilité) prétend que, en parcourant *toute la série*, on y rencontre, au total, à la limite, à peu près *autant de zéros que de uns*. L'autre (le théorème de Bernoulli) prétend que, en fixant un nombre n et réunissant chaque n éléments successifs de la série dans un groupe, enfin en parcourant *tous ces groupes*, on trouve, pour n assez grand, *dans la majorité de ces groupes* approximativement autant de zéros que de uns.

Il va sans dire que les deux énoncés sont compatibles ; pour la suite alternant régulièrement 0, 1, 0, 1, 0, 1,... par exemple ils sont valables tous les deux. Mais il ne sera pas inutile de donner un exemple d'une suite *pour laquelle la première proposition* (l'existence de la limite) *est remplie, tandis que la loi de Bernoulli tombe en défaut*. Considérons une table mathématique quelconque, par exemple la table des racines carrées des nombres entiers successifs, en imaginant que la table soit continuée à l'infini. Chaque fois que la cinquième décimale sera un nombre pair nous la remplacerons par un « zéro », et si elle est un nombre impair par « un ». Pour cette suite infinie on démontre facilement que la fréquence relative des uns et des zéros tend vers la limite $\frac{1}{2}$. Mais on sait encore beaucoup plus sur la constitution de cette table. Envisageons le nombre $x = 10^{12}$ dont la racine $\sqrt{x} = 10^6$. Si on passe au nombre voisin $x + 1$, la racine n'augmente que d'une quantité d'environ $\frac{1}{2}10^{-6}$. Ainsi, la cinquième décimale du développement de $\sqrt{x}$ reste la même pour environ vingt nombres successifs. Au voiisinage de $x = 10^{14}$ la cinquième décimale de $\sqrt{x}$ ne change qu'après environ cent pas, c'est-à-dire, on trouve dans cet endroit de la table une suite continue de zéros ou de uns se succédant sans interruption environ deux cents fois. La constitution de la suite considérée de uns et de zéros est donc telle que des groupes infiniment croissants et ne renfermant que des uns ou des zéros exclusivement, alternent. Alors, si on fixe un nombre n, si grand qu'il soit, pour la longueur des groupes, on trouvera, en parcourant toute la série infinie, que presque tous les groupes ne comprennent que des uns ou des zéros exclusivement et *que presque aucun groupe ne contient à peu près autant de zéros que de uns.* Voilà un exemple caractéristique d'une suite de zéros et de uns où la valeur $\frac{1}{2}$ pour la limite de la fréquence relative des uns se produit tout à fait autrement qu'au cas du jeu de pile ou face pour lequel le théorème de Bernoulli est juste. Tandis que, dans le collectif, la compensation des zéros et des uns se réalise au dedans de chaque groupe assez long des éléments successifs, au contraire, dans le cas de la table des racines carrés les uns et les zéros ne se compensent que par la succession alternative *des groupes purs d'une sorte et de l'autre,* la longueur des groupes croissant infiniment.

Naturellement la suite considérée, provenant de la table des racines carrées des nombres entiers, ne possède pas les qualités d'un collectif. Elle ne satisfait qu'à la première des deux conditions, établies dans ma première conférence, mais pas à la deuxième : l'irrégularité. Ainsi il devient évident que la loi de Bernoulli — conséquence spéciale du premier théorème fondamental — n'est pas démontrable dans une théorie qui est basée sur l'existence d'une limite de la fréquence relative sans exiger une irrégularité assez étendue. — D'ailleurs on peut démontrer que, de la loi de Bernoulli supposée juste, découle l'existence d'une limite de la fréquence relative.

On sait que la loi de Bernoulli fût généralisée par POISSON qui considéra une infinité d'alternatives dont les probabilités changeaient de l'une à l'autre. Si l'on admet le premier théorème fondamental on peut en déduire une loi encore plus générale que celle de Poisson. Voilà l'énoncé de ce que j'appelle la *première loi des grands nombres,* conséquence immédiate du premier théorème fondamental trouvée pour la première fois par TCHEBYCHEF : *La probabilité pour que la moyenne arithmétique de n variables dont chacune est soumise à une répartition quelconque, soit située* *dans des limites aussi étroites qu'on veut au voisinage de la valeur de son espérance mathématique, tend vers l'unité, quand n augmente indéfiniment*; il suffit ici de supposer que la somme des déviations des n répartitions particulières croît vers l'infini avec n.

VI

Évidemment la loi citée plus haut de Tchebychef n'exprime que le fait que, dans la formule du premier théorème fondamental, la précision de la fonction de Gauss tend vers l'infini. D'une manière tout à fait analogue on peut déduire en partant du deuxième théorème fondamental un énoncé parallèle que je nommerai *la deuxième loi des grands nombres*. J'en donnerai d'abord l'énoncé, en me bornant, comme précédemment, au cas d'une seule dimension : *Si l'observation des n premiers éléments d'un collectif donne comme moyenne arithmétique des n résultats la valeur x, alors la probabilité pour que l'espérance mathématique du résultat soit située dans des limites aussi étroites qu'on veut autour de x, tend vers l'unité, pour n assez grand, l'influence de la « probabilité a priori » s'évanouissant en même temps.*

Appliquons cette proposition, pour mieux comprendre son sens, au cas du jeu de pile ou face. Dans ce but il faut s'imaginer qu'on a à sa disposition un nombre de pièces de monnaie différentes, et que, pour chacune d'elles, on a les chances d'amener soit le côté « face » (ou le résultat « 1 ») soit le côté « pile ». Maintenant on suit le processus suivant : On saisit au hasard une pièce, on la jette n fois sur la table et on note le nombre relatif des cas où « face » est sortie, (c'est la moyenne arithmétique des résultats si on les représente par 0 et 1). En répétant ce processus un nombre infini de fois et en ne regardant que les séries où la moyenne arithmétique constatée est exactement x, notre théorème prétend que, dans l'énorme majorité de ces cas la pièce utilisée sera telle que la probabilité d'amener « face » soit située près de x.

On voit que la deuxième loi des grands nombres, malgré son analogie étroite à la première, décrit un état de choses tout à fait différent. Il semble aussi qu'on n'exprime que très imparfaitement les relations entre les deux lois en nommant l'une l'inversion de l'autre. Tandis que la première ne touche qu'à une seule série infinie d'expériences, effectuées avec une seule pièce de monnaie, la deuxième fournit un énoncé concernant des pièces différentes l'une de l'autre, et s'appliquant à une infinité de séries de longueur finie n qu'on effectue, chacune, avec une autre pièce.

Résumons maintenant en quelques mots le contenu essentiel des quatre propositions analogues mais bien distinctes, dont il s'agit ici : l'axiome sur l'existence d'une limite de la fréquence relative, l'axiome sur l'irrégularité, ainsi que la première et de la deuxième loi des grands nombres. Nous envisageons d'abord une série de n épreuves répétées avec une pièce de monnaie ; soit x le rapport du nombre des résultats « face » à n.

1) Le premier axiome de la théorie des probabilités prétend que si on continue les épreuves en augmentant n à l'infini, le rapport x tend vers une limite déterminée p.

2) Le deuxième axiome prétend que si on effectue un choix de position et si on ne tient compte, en calculant le rapport x, que des éléments choisis, la limite reste invariable.

Ces deux axiomes admis, on peut démontrer les lois suivantes :

3) La première loi, des grands nombres : Si on répète pour n fixé et en utilisant la même pièce, la série de n épreuves à l'infini, le rapport x sera situé, pour n assez grand, très près de la valeur p dans l'énorme majorité des séries.

4) La deuxième loi des grands nombres : Si on répète pour n fixé, la série de n épreuves à l'infini, en utilisant des pièces différentes, et si on ne retient que les séries où le rapport est égal à x, alors, pour n assez grand, dans l'énorme majorité de ces séries il s'agira effectivement d'une pièce pour laquelle la continuation infinie des jeux conduirait à une limite très voisine de x.

Avec cette confrontation, je crois avoir démontré ce que je me suis proposé : que la théorie des probabilités basée sur la notion de fréquence relative est libre de toute contradiction, et qu'elle contribue, au contraire, à jeter un peu de lumière sur des questions qui restent, dans la théorie classique, au moins peu claires.

Quelques remarques sur certaines questions de physique statistique

Parmi les questions principales qu'on discute actuellement dans la théorie des probabilités il n'y en a pas de plus intéressantes que celles qui touchent aux fondements même de la physique moderne. On sait bien que, depuis plus d'un demi siècle, des méthodes statistiques sont généralement admises dans plusieurs chapitres de la physique. C'est surtout BOLTZMANN à Vienne qui, en 1866, remplaça le deuxième théorème fondamental de la théorie de la chaleur par un énoncé essentiellement statistique. Au lieu de dire que, dans tout système déterminé, l'entropie augmente dans tous les cas, on se borne maintenant à admettre qu'elle n'augmente que dans l'énorme majorité des cas, mais qu'elle peut aussi diminuer ; ou, plus exactement, que des diminutions sensibles ont une probabilité minime. Dès le temps de BOLTZMANN, les physiciens se sont accoutumés peu à peu à l'idée que le hasard ou la probabilité entre dans leur science où jadis ne dominait que la certitude. Je n'insisterai pas sur une discussion de la question, quelle que soit son importance pour savoir, si l'introduction de la notion de probabilité dans la physique est définitive ou si elle n'est qu'un moyen provisoire, un auxiliaire pour échapper à certaines difficultés de calcul. Je ne veux pas, non plus, m'occuper à confronter l'interprétation des événements naturels, au sens de la physique classique, avec une théorie physique basée sur l'idée de hasard. Ce que je me propose pour aujourd'hui c'est de traiter, à l'aide des certains résultats de l'algèbre et de l'analyse, un problème qui se pose dans toutes applications de la statistique aux questions physiques, et qu'on pourrait appeler la *recherche statistique des événements successivement enchaînés*, ou rattachés l'un à l'autre.

I

Permettez-moi d'analyser ce problème en rappellant l'application la plus ancienne de la statistique en physique, c'est-à-dire la théorie cinétique des gaz dans sa première forme. On y part des hypothèses suivantes : Chaque molécule possède à un temps donné une vitesse déterminée qui est représentée par un point dans « l'espace des vitesses ». Cet espace est subdivisé en m portions dont chacune a le même volume. A chacune de ces portions correspond une certaine valeur de l'énergie $c_1, c_2, \cdots c_m$. Enfin on introduit comme hypothèse fondamentale une certaine répartition donnée dans le collectif primordial : *Les probabilités* pour que le point représentatif d'une molécule soit situé dans l'un quelconque des volumes cités *sont égales entre elles*, donc chacune égale à $\frac{1}{m}$.

A partir de ces données on déduit d'une manière connue, par une combinaison des opérations simples, les résultats suivants pour le *collectif dérivé* dont l'élément consiste à constater combien de molécules, dans l'ensemble de n molécules, se trouvent dans chacun des m volumes différents. On trouve que la probabilité pour que x_1 molécules aient leurs points représentatifs dans le premier, x_2 molécules dans le deuxième, ...x_m molécules dans le m-ième volume, est égale à

$$p_n(x) = \frac{n!}{x_1!\, x_2! \cdots x_m!}\left(\frac{1}{m}\right)^n. \tag{1}$$

Nous désignons cette probabilité par $p_n(x)$ en mettant x comme abréviation pour un système de nombres entiers non-négatifs $x_1, x_2, \cdots x_m$ avec la somme

$$x_1 + x_2 + \cdots x_m = n. \tag{2}$$

Nous appellerons un tel système de nombres x une « distribution » (distribution des molécules sur les valeurs de l'énergie) en réservant le mot « répartition », comme jusqu'à présent, pour désigner l'ensemble des probabilités dans d'un collectif. On peut maintenant évaluer la distribution la plus probable sous la condition que l'énergie totale E soit constante,

$$c_1x_1 + c_2x_2 + \cdots c_mx_m = \mathrm{E} = \text{const.}, \tag{3}$$

c'est-à-dire chercher le système des nombres x, satisfaisant aux équations (2) et (3), pour lequel l'expression (1) atteint son maximum. Ce fut Cl. MAXWELL et plus tard L. BOLTZMANN qui donnèrent, pour n très grand, le résultat devenu classique. La distribution la plus probable est déterminée par

$$(4) \qquad b_k = Ce^{-\gamma c_k}, \qquad k = 1, 2, \cdots m,$$

où C et γ sont des constantes, à éliminer à l'aide des conditions (2) et (3). Au surplus on sait que la valeur de $p_n(x)$ pour $x_k = b_k$ surpasse *énormément* toutes les autres probabilités, de même que, pour n très grand, il est « à peu près sûr » que la distribution des molécules soit celle définie par (4) ou une distribution approchée.

Jusqu'ici tout est correct et il n'y a pas d'objection à faire. Mais les physiciens ne se bornent pas au résultat obtenu par cette déduction, ils ajoutent une proposition qui ressemble un peu à la conséquence démontrée mais qui la dépasse. Voilà ce qu'ils prétendent : Si on abandonne l'ensemble de molécules à lui-même, pour un temps assez long, *on trouvera les molécules presque toujours* dans une distribution identique à (4) ou très voisine de (4). En autres termes, on transfère la répartition des probabilités, évaluée d'après des méthodes du calcul des probabilités, *à la succession des états successifs d'un gaz abandonné à lui-même.* Évidemment les suppositions nécessaires pour définir un collectif ne sont pas remplies si on envisage les états se succédant dans des intervalles de temps très petits. Du point de vue classique il faut donc admettre que l'état actuel des vitesses détermine d'une manière univoque l'état suivant. A l'époque de BOLTZMANN, lorsqu'on ne doutait pas que la mécanique classique ne fut applicable au mouvement des molécules gazeuses, il apparut une difficulté insurmontable. On chercha à se tirer d'affaire par l'introduction d'une nouvelle hypothèse établie exprès dans ce but.

Vous connaissez sans doute l'hypothèse dite « *ergodique* » qui prétend que chaque système mécanique passe, en temps infini, par tous les états qui correspondent à la valeur constante de l'énergie totale. Concédons — quoiqu'il y ait beaucoup d'objections — que, à l'aide de cette hypothèse, on réussisse à démontrer la proposition citée sur la succession des états d'un système. Cependant, on ne peut pas y voir une solution satisfaisante du problème. D'abord en 1912 déjà M. PLANCHEREL avait démontré que l'hypothèse ergodique dans sa forme primordiale est strictement contradictoire. Un système ne peut qu'arriver *au voisinage* de chaque phase compatible avec la valeur initiale de l'énergie bien que le voisinage soit infiniment étroit. Mais cette proposition quasi-ergodique ne permet plus la conclusion à laquelle on est conduite facilement en partant de l'hypothèse ergodique stricte. De plus, toute la manière de concevoir ces idées a profondément changé depuis le temps de BOLTZMANN. On se défie aujourd'hui de la validité de la mécanique classique quand il s'agit de molécules ou de corps aussi petits que celles-ci. Mais si les équations différentielles de la mécanique analytique, au moins dans leur sens classique, ne sont plus valables la base de l'hypothèse ergodique s'évanouit. De quel côté qu'on regarde les choses, il faut admettre qu'il y a ici une difficulté qui, dans la théorie développée jusqu'à présent, n'était pas complètement surmontée.

Ajoutons encore que ces difficultés apparaissent dans *toutes* les applications de la théorie des probabilités à la physique, et non seulement au cas de la théorie cinétique des gaz. On a toujours affaire à une série d'états successifs d'un système ou, ce qui revient au même, à une série d'expériences qui se succèdent dans le temps. En général il est inadmissible de considérer ces états successifs comme des épreuves indépendantes l'une de l'autre, formant la suite des éléments d'un collectif. Si nous envisageons par exemple le mouvement Brownien, il est clair que les positions des particules à deux instants successifs ne peuvent pas différer autant que diffèrent leurs positions en deux instants séparés par un intervalle de temps assez long. Dans ce cas, le principe fondamental de toute théorie des probabilités, *l'impossibilité de trouver un système à jouer*, ou l'insensibilité de la limite de la fréquence relative à un choix de position, quelconque est inapplicable. Néanmoins on réussit souvent, sans doute, à trouver une solution en appliquant certaines règles usuelles du calcul des probabilités. Il me semble utile d'essayer d'analyser de plus près cette difficulté générale. Nous y arriverons, en partant d'un point de vue basé sur notre théorie des probabilités, à l'aide de quelques considérations algébriques empruntées de la théorie des matrices quadratiques.

II

Voici la proposition essentielle qui nous conduira au but. Nous supposons que le système considéré — un ensemble de molécules et de particules en mouvement Brownien etc., — soit susceptible de certaines phases, numérotées de 1, 2, 3, ... m. Le système, arrivé à la phase y, possédera une probabilité dépendant de y, *pour parvenir*, durant un intervalle fixe de temps, *à la phase* x. Désignons par $p(x\,;y)$ cette *probabilité de passage* et remarquons que

$$\sum_{x=1}^{m} p(x\,;y) = 1. \tag{5}$$

Les valeurs des $p(x;y)$, non-négatives et assujetties à vérifier l'équation (5), représentent m collectifs primordiaux avec m répartitions, données par les colonnes de la matrice quadratique $p(x\,;y)$. — Au surplus soit une répartition $q_0(x)$ donnée, signifiant la probabilité pour que le système se trouve dans la phase x *au commencement du temps considéré*. Nous cherchons à répondre aux deux questions suivantes : Quelle est la probabilité $q_n(x)$ pour que le système se trouve, après n intervalles fixes de temps, dans la phase x ? Et que peut-on dire sur la limite de $q_n(x)$ pour n infini ?

La première de ces questions se résout d'une manière élémentaire par une combinaison de deux opérations simples. Il faut, d'abord, *composer* les répartitions données $q_0(x)$ et $p(x;y)$. Le produit $q_0(z).\ p(x;z)$ nous donne la probabilité pour que le système se trouve à l'instant initial en z, et ensuite en x. Ajoutons un *mélange*, en confondant toutes les valeurs de z de 1 à m ; nous parvenons par sommation à la probabilité

$$q_1(x) = \sum_{z=1}^{m} q_0(z) \cdot p(x\,;z), \tag{6}$$

pour que le système se trouve, en x, après un intervalle quel que soit son point de départ. De la même manière on arrive à l'équation de récurrence

$$q_n(x) = \sum_{z=1}^{m} q_{n-1}(z) \cdot p(x\,;z). \tag{7}$$

On peut donner une autre forme aux équations (6) et (7) en utilisant la notion et les règles du calcul des « *matrices quadratiques* » et de leur multiplication avec un « *vecteur* » à m dimensions. Désignons par $\boldsymbol{q_n}$

le vecteur dont les composants sont $q_n(1)$, $q_n(2) \cdots q_n(m)$ et par P la matrice

$$(8) \qquad P = \begin{pmatrix} p(1\,;1) & p(1\,;2) & \cdots & p(1\,;m) \\ p(2\,;1) & p(2\,;2) & \cdots & p(2\,;m) \\ \cdots & \cdots & \cdots & \cdots \\ p(m\,;1) & p(m\,;2) & \cdots & p(m\,;m) \end{pmatrix}.$$

L'équation (7) s'écrit maintenant

$$(9) \qquad q_n = P \cdot q_{n-1},$$

d'où il suit immédiatement que

$$(9') \qquad q_n = P^n q_0.$$

La recherche des probabilités successives $q_1(x)$, $q_2(x), \cdots q_n(x)$ revient alors à une étude des *puissances successives d'une matrice quadratique.*

La matrice (8) est caractérisée par deux propriétés essentielles : ses éléments sont non-négatifs, et la somme de chaque colonne des éléments est égale à l'unité. Des résultats importants sur les matrices avec éléments non-négatifs ont été établis par M. FROBENIUS en 1912. Pour étudier les puissances P^n il faut, d'après FROBENIUS, distinguer les cas où P est décomposable ou non décomposable. Aux termes de notre problème la décomposabilité signifie que les m phases possibles du système se divisent en deux ou plusieurs parties de telle sorte qu'il *n'existent pas de probabilités de passage d'une partie à l'autre.* De plus, au cas où la matrice est partiellement décomposable il y a des groupes de phases pour lesquels il n'existe exclusivement que des probabilités de départ ou des probabilités d'arrivée. Une matrice dont tous les éléments sont différents de zéro est toujours non-décomposable. Des deux exemples suivants, symbolisés par

$$(10) \qquad \begin{pmatrix} A & O \\ O & B \end{pmatrix} \qquad \begin{pmatrix} A & C \\ O & B \end{pmatrix},$$

le premier donne le cas de décomposabilité totale, l'autre le cas d'une décomposabilité partielle où A est un groupe de phases sans probabilités de départ et B un groupe sans probabilités d'arrivée.

Avant d'aborder le deuxième problème mentionné plus haut, la recherche de la limite de $q_n(x)$ pour n infini, il est avantageux de s'occuper de la question préliminaire suivante : Existe-t-il des répartitions q (x) qui, *une fois établies, se maintiennent indéfiniment* ? Il s'agit ,comme on voit, des solutions de l'équation homogène

$$q = \mathrm{P} \cdot q, \tag{11}$$

ou,

$$q(x) = \sum_{z=1}^{m} q(z)\mathrm{P}(x\,;\,z). \tag{11'}$$

Je cite le résultat essentiel, en supprimant la démonstration : Pour Pnon-décomposable, c'est-à-dire s'il y a, pour toute division des m phases en deux parties, des probabilités de passage d'une partie à l'autre et réciproquement, il existe toujours *une solution univoque* de l'équation (11) et tous les $q(x)$ y sont positifs. Pour le cas d'une matrice décomposable, pour un groupe de phases sans probabilités d'arrivée il n'existe comme solution stationnaire que q $(x) = 0$.

Je passe maintenant à l'étude de la limite de $q_n(x)$ pour n infini. On se demandera si, pour n croîssant, $q_n(x)$ ne s'approche de la valeur stationnaire $q(x)$ ou, ce que revient au même, si les puissances successives P^n ne tendent pas vers une matrice-limite dont toutes les colonnes soient égales à q (x). Le résultat n'est pas si simple qu'on pourrait le présumer. Même pour une matrice non-décomposable la convergence de $q_n(x)$ vers la fonction stationnaire $q(x)$ dépend d'une condition additionnelle. Il faut et il suffit, pour que $q_n(x)$ tende vers $q(x)$, *qu'au moins un élément de la diagonale principale*, qui se trouve soit dans P ou dans une certaine puissance P^n, non décomposable, ne s'évanouisse pas. Un élément $p(x\,;\,x)$ de la diagonale principale dans P indique une probabilité d'état «stationnaire», un tel élément dans P^n indique une probabilité de « retoua ». Il est remarquable que dans le cas de la convrgence la limite $q(x)$ de $q_n(x)$ est tout-à-fait *indépendante de la probabilité initiale* $q_0(x)$; nous reviendrons encore plus loin sur ce fait important. De plus, si la matrice P est *symétrique*, c'est-à-dire $p(x\,;\,y) = p(y\,;\,x)$, la solution des équations (11') où la somme des coefficients d'une ligne du second membre devient égale à 1, n'est autre que $q(1) = q(2) = \cdots q\ (m) = \frac{1}{m}$ et c'est là une *répartition uniforme*. Signalons donc le résultat suivant, qui est fondamental pour les questions de physique mentionnées plus haut.

Un système quelconque est susceptible de m phases différentes, les probabilités de passage satisfaisant aux conditions suivantes : a) Pour chaque division des m phases en deux parties il y a des probabilités de passage d'une partie à l'autre ; b) Pour une phase au moins il existe une probabilité d'y demeurer ; c) La probabilité du passage de x à y est égale à la probabilité du passage inverse. *Sous ces conditions la probabilité* $q_n(x)$ *pour que le système se trouve dans la phase* x *après* n *passages, tend vers* $\frac{1}{m}$, *quelque soit la probabilité initiale* $q_0(x)$.

Les différents cas de non-convergence de la fonction $q_n(x)$ présentent un grand intérêt du point de vue mathématique. On peut démontrer que dans le cas général, pour n très grand, $q_n(x)$ s'approche d'une fonction *périodique de* n. La durée d'une période est toujours inférieure ou égale à m. Si l signifie la longueur de la période, la puissance P^l se décompose en l parties dont chacune possède toutes les propriétés d'une matrice de puissances convergentes. Donc, dans la suite infinie P^1, P^2, P^3,... chacune des l parties converge vers une matrice dont toutes les colonnes sont égales entre elles. Il en suitt que dans $q_n(x)$ pour n infini, les rapports relatifs des valeurs appartenant à une même partie convergent vers des limites indépendantes de $q_0(x)$, tandis que les sommes des valeurs pour chaque groupe particulier dépendent de la probabilité initiale $q_0(x)$. Ces faits ne touchent pas essentiellement les applications aux problèmes de physique puisqu'on est encore libre de fixer l'unité du temps ou, autrement dit, l'intervalle d'un passage. On peut partir par exemple d'intervalles qui correspondent à la longueur l de la période.

III

Revenons au cas de convergence de la matrice des probabilités de passage et cherchons à tirer des conclusions utiles pour les problèmes de physique statistique. La probabilité pour que le système se trouve au commencement en x_0 puis dans n intervalles successifs en x_1, x_2, $\cdots x_n$ est évidemment égale au produit

$$q_0(x_0)p(x_1\,;\,x_0)p(x_2\,;\,x_1)\cdots p(x_n\,;\,x_{n-1}). \tag{12}$$

Signalons une propriété importante de cette fonction. Soit r un nombre entier au-dessous de n. Nous prenons la somme de toutes les expressions (12) pour $x_r = x$ fixé et $x_0, x_1, \cdots x_{r-1}, x_{r+1}, \cdots x_n$ quelconques. Je prétends que cette somme est égale à $q_r(x)$. En effet, la sommation sur x_0 donne

$$\sum_{x_0} q_0(x_0)p(x_1 ; x_0) = q_1(x_1),$$

d'après la règle générale (7). Si on continue à sommer sur x_1, on obtient de la même manière $q_2(x_2)$ etc., jusqu'à la somme, prise sur $x_0, x_1, \ldots x_{r-1}$, égale à $q_r(x_r)$. D'ailleurs le dernier facteur $p(x_n ; x_{n-1})$, sommé sur x_n (variable qui n'apparaît qu'une fois dans tout le produit) donne 1, puis l'avant-dernier, sommé sur x_{n-1}, fournit du même 1 etc. Il s'en suit donc la proposition mentionnée :

$$(13) \qquad \sum_{(x_r = x)} q_0(x_0)p(x_1 ; x_0)p(x_2 ; x_1) \cdots p(x_n ; x_{n-1}) = q_r(x).$$

Abordons maintenant la question fondamentale de toute cette recherche. Nous observons le système pendant n intervalles successifs. S'il se trouve, durant ce temps, y fois dans la phase x nous appellerons le rapport $y : n$ la « durée (relative) de la présence en x ». Pour chaque x et n donnés il y aura une probabilité $\pi_n(z ; x)$, pour que la durée de la présence en x soit égale à z ou, que le système demeure pendant zn intervalles de temps en x. Je me propose de déterminer la répartition π_n, c'est-à-dire la fonction π_n de z pour x et n donnés. Commençons par calculer la *moyenne de cette répartition.*

Evidemment on trouve $\pi_n(z;x)$, en partant de l'expression (12), par un certain mélange. Il suffit de sommer tous les produits (12) correspondant aux combinaisons de $x_1, x_2, \ldots x_n$ dans lesquelles la valeur x se répète $y = zn$ fois. La moyenne de π_n est par définition la somme

$$(14) \qquad \sum_z z\,\pi_n(z ; x) = \frac{1}{n} \sum_{y=0}^{n} y\,\pi_n\left(\frac{y}{n} ; x\right).$$

Je prétends que la somme du deuxième membre de cette équation est égale à

$$(14') \qquad q_1(x) + q_2(x) + \cdots q_n(x).$$

En effet, à cause de la propriété du produit (12) signalée plus haut on peut remplacer $q_r(x)$ par la somme de tous des produits dans lesquels on a $x_r = x$. Alors la somme (14'), réduite à ses éléments, comprend tous les produits (12) dans lesquels figure au moins une fois la valeur x parmi les variables $x_1, x_2, \cdots x_n$. Mais un tel produit qui contient deux fois la valeur x par exemple, avec $x_3 = x$ et $x_7 = x$, est compté deux fois dans (14'), une fois comme élément additionnel de $q_3(x)$, l'autre fois comme élément additionnel de $q_7(x)$. Une combinaison $x_1, x_2, \cdots x_n$ comprenant y fois la valeur x est comptée exactement y fois dans (14'), de sorte que la proposition est démontrée et nous avons

$$(15) \qquad \text{Moy.}\ (z) \equiv \sum_z z\pi_n(z\,;\,x) = \frac{1}{n}[q_1(x) + q_2(x) + \cdots q_n(x)].$$

En supposant que $q_n(x)$ converge vers $q(x)$ pour n tendant vers l'infini, on a

$$(15') \qquad \lim_{n \to \infty} (\text{Moy.}\ (z)) = \lim_{n \to \infty} \frac{1}{n}[q_1(x) + q_2(x) + \cdots] = q(x).$$

La moyenne ou l'espérance mathématique de la durée relative de la présence en x est égale à la probabilité limite ou probabilité stationnaire de la phase x.

Passons maintenant à l'étude de la déviation ou de l'écart quadratique de la répartition π_n. En supprimant quelques détails je pars d'une propriété du produit (12) analogue à celle qui a été établie tout à l'heure et aussi facile à démontrer. La probabilité pour que, dans la suite $x_1, x_2, \cdots, x_n$ la valeur x se produise deux fois, savoir $x_r = x$ et $x_{r+s} = x (r, s, r + s$ nombres naturels au-dessous de $n)$, est égale à $q_r(x) \cdot p_s(x\,;\,x)$ où $p_s(x\,;\,x)$ signifie l'élément avec les indices x, x de la puissance P^s, ou la probabilité de retour à x dans s passages. Prenons maintenant la somme

$$(16) \qquad \sum_{r,s} q_r(x) p_s(x\,;\,x) \qquad \text{pour} \qquad r = 1, 2, \cdots n\,; \quad s = 1, 2, \cdots n - r,$$

au lieu de la somme (14'). En décomposant l'expression (16) en ses éléments de la forme (12) on voit que (16) comprend toutes les probabilités des combinaisons de $x_1, x_2, \cdots, x_n$ dans lesquelles la valeur x figure au moins deux fois. Mais si la valeur x dans une combinaison particulière se produit y fois ($y > 2$) la probabilité correspondante y est comptée exactement avec une multiplicité égala à

$$\frac{1}{2}y(y-1).$$

Il s'en suit

$$\frac{1}{2}\sum_{y} y(y-1)\pi_n\left(\frac{y}{n}; x\right) = \sum_{r=1}^{n}\sum_{s=1}^{n-r} q_r(x)p_s(x; x). \tag{17}$$

Servons nous d'une relation bien connue entre les deux premiers moments d'une répartition et sa déviation, à savoir

$$\text{Dév } (z) = \sum_{z} z^2\pi_n(z; x) - \left(\sum_{z} z\pi_n(z; x)\right)^2. \tag{18}$$

En faisant la moyenne de (15), (17) et (18) on déduit immédiatement

$$\text{Dév}\left(\frac{y}{n}\right) = \frac{2}{n^2}\sum_{r,s} q_r(x)p_s(x; x) + \frac{1}{n^2}\sum_{r} q_r(x) - \left[\frac{1}{n}\sum_{r} q_r(x)\right]^2. \tag{19}$$

Il a déjà été dit que — sous les conditions de convergence établies plus haut — pour n augmentant indéfiniment les puissances successives $\mathbf{P}^s$ tendent vers une matrice dont toutes les colonnes sont égales à $q(1), q(2), \cdots, q(m)$ où $q(x)$ signifie la répartition stationnaire. De ce fait, on dérive facilement

$$\lim_{n\to\infty} \frac{2}{n^2}\sum_{r,s} q_r(x)p_s(x; x) = [q(x)]^2.$$

La deuxième expression du second membre de (18') s'évanouissant pour n infini on arrive enfin à

$$\lim_{n\to\infty} \text{Dév}\left(\frac{y}{n}\right) = 0. \tag{19'}$$

Les équations (15') et (19') contiennent le résultat fondamental. Le fait qui se manifeste en (19') (les déviations d'une suite des répartitions tendent vers zéro), n'est autre que celui qu'on aime souvent énoncer à l'aide de la notion de « convergence au sens du calcul des probabilités ». C'est le même fait qui joue le rôle décisif dans la loi des grands nombres où l'on parle d'une probabilité tendant vers l'unité ou d'un évènement « presque sûr ». Pour ma part, je préfère la formule suivante. Nous avons démontré que :

Pour un système quelconque, susceptible de n phases différentes, il est presque sûr que la durée relative de la présence du système dans la phase x, durant n passages suscessifs, soit pour n très grand à peu près égale à la valeur q(x) de la probabilité stationnaire et tout à fait indépendante de la répartition initiale de la phase considérée x.

Ce théorème dont le sens précis est donné par les équations (15′) et (19′) fournit une base suffisante pour l'application du calcul des probabilités aux questions physiques. Les conditions de convergence auxquelles l'énoncé précédent fait allusion sont les suivantes : Il faut que la matrice des probabilités de passage soit non-décomposable et qu'elle contienne dans la diagonale principale au moins un élément différent de zéro.

Ajoutons maintenant l'hypothèse déjà mentionnée que la matrice P soit symétrique, c'est-à-dire que la probabilité du passage de x à y soit égale à celle du passage de y à x. Il suit, comme il a déjà été dit, que la solution stationnaire $q(x)$ devient égale à une constante, $q(x) = 1/m$. Nous arrivons ainsi au théorème suivant : *Si les m phases d'un système quelconque sont choisies de telle manière que les probabilités d'un passage et du passage inverse soient égales entre elles, il est presque sûr que, durant un temps assez long, le système demeurera dans chaque phase à peu près le même temps.* Voilà un théorème exactement démontré, apte à remplacer l'hypothèse ergodique non-démontrable et contradictoire à un certain égard. Au lieu de l'assertion déterminée : le système demeurera en chaque phase dont il est susceptible un certain temps, nous évitons d'affirmer qu'une certaine probabilité tend vers l'unité.

IV

Voyons maintenant comment s'applique ce théorème aux problèmes de la statistique physique. Nous reprenons l'exemple de la théorie cinétique des gaz, sans nous borner, maintenant, à la forme la plus ancienne de cette théorie. Généralement on y part d'une hypothèse qui établit certains cas « également possibles » ou « également probables ». On suppose par exemple que certains endroits de l'espace des phases soient également probables, c'est-à-dire que le système considéré ait la même probabilité pour que son point représentatif se trouve dans l'un ou l'autre de ces endroits. Mais que veut dire l'expression « cas également probables » ? Il y a ici une autre difficulté que celle relative aux jeux de hasard et que j'ai mentionnée dans ma première conférence. Si on dit que, pour un dé correct, les six points qui peuvent sortir, sont également probables, celà signifie — au moins du point de vue de ma théorie — que les limites des fréquences relatives correspondantes sont égales entre elles. On trouve ces fréquences relatives en effectuant une série assez étendue d'épreuves dont chacune consiste en un procédé défini d'une façon précise, à savoir en un coup de dés avec toutes les préparations nécessaires. Mais que peut-on dire si on essaye d'appliquer cette idée au cas d'un système de molécules ou de particules en mouvement brownien ; quelle est l'épreuve particulière constituant l'élément du collectif ?

Il s'agit par exemple d'une mesure des coordonnées et des composantes de la vitesse de certaines molécules. Il importe peu que la réalisation concrète de la mesure ne soit pas practicable ; il suffirait déjà de pouvoir définir exactement quelles coordonnées et quelles vitesses il faut prendre et comment les mesurer en principe. Mais ici s'élève la difficulté : Est-ce qu'il est permis de mesurer les vitesses, etc., à intervalles successifs, dans un gaz abandonné à lui-même ? Evidemment ces mesures successives ne sont aucunement analogues aux parties successives d'un jeu de hasard. Dans le jeu de dés il est essentiel que chaque fois, avant de jouer, le cornet soit secoué pour restituer, dans un certain sens, l'état primordial duquel on est parti au coup précédent. C'est la même chose pour le battage des cartes avant le jeu ou que le brassage des boules enfermées dans une urne avant chaque tirage. Il faudrait, pour que les mesures successives dans l'ensemble des molécules constituent un collectif, qu'avant chacun d'eux l'état de l'inorganisation moléculaire fût restitué de nouveau par une intervention extérieure, analogue au procédé du battage des cartes. Il est clair qu'un tel procédé ou une telle définition d'un collectif serait très artificielle et exposée à beaucoup d'objections. Ajoutons encore que dans la nouvelle tournure que donne M. FERMI aux hypothèses déterminant les cas également probables, ceux-ci perdent leur évidence ou leur caractère intuitif. Il faut donc admettre qu'une théorie qui se débarasserait de telles hypothèses, difficiles à préciser, sur les cas également probables aurait un grand avantage sur les autres.

On voit bien que, dans notre théorie, cette condition est remplie. Ici la notion difficile à définir de la « probabilité d'une certaine phase » n'existe plus. Nous n'avons affaire qu'aux $p(x\,;y)$, qui sont les *probabilités de passage* d'une phase à l'autre et aux $q_n(x)$, qui sont les *probabilités de l'arrivée* dans une phase. Ces deux probabilités sont facile à introduire à l'aide de certains collectifs dont les éléments consistent dans des épreuves tout à fait praticables. Il n'est plus nécessaire d'imaginer une intervention extérieure répétée, restituant l'état chaque fois à nouveau l'état initial. Les phases successives par lesquelles le système passe n'y *sont pas considérées comme des épreuves indépendantes ou comme des éléments d'un collectif* . Au contraire il est tenu compte, dès le début de l'enchaînement réciproque des phases parcourues. — Remarquons aussi que, dans les théories les plus modernes de la physique statistique ce sont justement les probabilités de passage qui présentent le plus d'intérêt.

Tous les résultats caractéristiques de la théorie cinétique des gaz ou du mouvement Brownien etc., se retrouvent aisément si on part — au lieu d'une hypothèse sur des cas également possibles — d'une division de l'espace des phases en domaines particuliers tels que les probabilités de passage d'un domaine à l'autre sont réversibles ou symétriques.

Je finirai maintenant en esquissant en quelques mots la suite des idées concernant la déduction du fameux théorème qu'on nomme le « *théorème H* », ou théorème sur *l'accroissement de l'entropie* dans un système déterminé.

Ici, on définit une certaine fonction H des coordonnées et des vitesses d'un système. Il importe peu que cette grandeur H se présente, abstraction faite d'un facteur constant, comme le logarithme naturel de la probabilité de la phase considérée. Pour nous il est essentiel que la fonction H atteigne son maximum pour la distribution de Maxwell et qu'elle possède des valeurs extrêmement grandes pour les phases qui corespondent, exactement ou à peu près, à cette distribution des vitesses. D'autre part, le nombre ou plutôt la mesure, de ces phases auxquelles correspondent les grandes valeurs de H, est très étendue en comparaison de toutes les autres phases. Si nous admettons la reversibilité des probabilités de passage pour les domaines particuliers dans l'espace des phases, nous avons le résultat — déduit des considérations précédentes — que la probabilité d'arrivée dans un endroit quelconque, après un temps assez long, (la valeur stationnaire $q(x)$) est égale pour tous ces domaines. Ils en résultent deux énoncés importants : Premièrement *la probabilité pour un système d'arriver, après un temps assez étendu, dans une phase correspondant à la distribution des vitesses de Maxwell ou une distribution approchée est très près de l'unité.* Et deuxièmement *la probabilité pour que le système se trouve, durant un assez long mouvement, presque tout le temps dans un état de Maxwell ou un état étroitement voisin ne diffère que très peu de l'unité.* Ces deux énoncés constituent le théorème H si on remplace les mots « état de Maxwell ou état voisin » par les mots « état avec la valeur maximum de H ou une valeur voisine de ce maximum ».

Souvent on préfère donner une autre forme à l'énoncé en question. On dit : La probabilité d'une augmentation de la valeur de H surpasse énormément la probabilité d'une diminution. Ou plus exactement : Si on considère tous les moments où la fonction H a une valeur donnée $H = H_1$ près de la valeur maximum, et les instants postérieurs après un intervalle fixe t, on trouvera dans l'énorme majorité des cas que la nouvelle valeur de H, disons H_2, est égale ou supérieure à H_1, et très rarement qu'elle en est inférieure. Cette proposition n'est juste que pour un intervalle fixe t assez étendu. En ce cas on peut démontrer que la suite des valeurs H possède toutes les propriétés de la suite des éléments d'un collectif. La répartition dans ce collectif procède par mélange de la répartition $q(x)$, de telle manière que la valeur maximum et les valeurs voisines de H comprennent presque toute la probabilité et les valeurs sensiblement inférieures n'ont qu'une probabilité minime. On démontre d'une manière tout à fait élémentaire que, pour un tel collectif, la probabilité de l'inégalité $H_2 \geqq H_1$ surpasse de beaucoup celle de l'inégalité inverse $H_2 < H_1$. Toutes les objections qui ont été faites au cours du développement historique de la théorie, contre l'admissibilité logique de cet énoncé, disparaissent si l'on s'appuie sur les notions et les définitions exactes de la théorie rationnelle des probabilités.

J'espère vous avoir donné une idée, au moins approchée, de mes efforts pour éliminer certaines difficultés qui s'opposaient, jusqu'à présent, à une fondation satisfaisante de la statistique physique. Je suis sûr que la voie suivie qui consiste à *concentrer l'intérêt sur la matrice des probabilités de passage* et d'éliminer les hypothèses sur les cas également possibles, conduira au but désiré.

Conférences faites à l'Institut HENRI-POINCARÉ en novembre 1931.

Manuscrit reçu le 1er décembre 1931.