ANNALES DE L'I. H. P.

J.A. VILLE

Leçons sur quelques aspects nouveaux de la théorie des probabilités

Annales de l'I. H. P., tome 14, n° 2 (1954), p. 61-143

<http://www.numdam.org/item?id=AIHP_1954__14_2_61_0>

Leçons sur quelques aspects nouveaux de la théorie des probabilités

par

J. A. VILLE,
Professeur à l'Institut de Statistique
de l'Université de Paris.

CHAPITRE I.

PROBABILITÉS ET FRÉQUENCES.

1.1 **Tirages équivalents.** — La plupart des raisonnements de probabilité consistent, pour étudier la nature aléatoire d'un événement, à dire : « il revient au même, du point de vue aléatoire, d'observer l'apparition de l'événement lui-même, ou d'observer les résultats de tel autre tirage au sort équivalent ».

Exemple 1. — Si par exemple l'événement E est le suivant :

« D'une urne, contenant deux boules blanches et une boule noire, on extrait une boule; on remet la boule dans l'urne, et l'on procède à une nouvelle extraction; on fait ainsi 12 tirages. On note le nombre de fois qu'une boule blanche est sortie; on considère que l'événement E s'est produit si la couleur blanche est apparue au moins 6 fois et au plus 10 fois ».

L'étude de l'événement E pourra se faire comme suit : observant que chaque tirage peut donner lieu à trois éventualités (première boule

blanche, deuxième boule blanche, boule noire), et que par conséquent l'ensemble des 12 tirages peut donner lieu à

$$3^{12} = 531441$$

éventualités, on peut, par la pensée, substituer à la suite des 12 tirages un tirage *unique*, qui serait le tirage d'un bulletin dans une urne contenant 531441 bulletins. Chacun de ces bulletins porterait alors la description d'une série de 12 tirages, sous la forme

$$BBNNBN\ldots BB,$$

c'est-à-dire une suite de 12 lettres B ou N.

Si l'on admet qu'il *revient au même* de procéder aux 12 tirages successifs dans l'urne à trois boules, ou au tirage unique dans l'urne aux 531441 bulletins, l'étude de l'événement E se ramène à un décompte, parmi les 531441 bulletins, de ceux décrivant un résultat associé à l'apparition de E, c'est-à-dire sur lesquels la lettre B figure 6, 7, 8, 9 ou 10 fois. Ce décompte relève uniquement de l'analyse combinatoire. Le calcul montre que les bulletins sur lesquels la lettre B figure 6, 7, 8, 9 ou 10 fois sont en nombres respectifs de

$$59136,\quad 101376,\quad 126720,\quad 112640,\quad 67584,$$

d'où il résulte que parmi les 531441 bulletins dont un est susceptible de sortie dans le tirage unique auquel nous nous sommes ramenés, il en est 467456 dont la sortie conclut à l'apparition de E. Nous considérons la fraction $\frac{467456}{531441}$ dont une valeur approchée est $\frac{88}{100}$. Cette fraction elle-même n'ayant pas une forme parlant à l'esprit, nous lui substituons sa valeur approchée, de sorte que nos considérations se ramènent en fin de compte à dire :

Il revient au même, du point de vue aléatoire, et en ce qui concerne l'événement E, de procéder au tirage avec l'urne, 12 fois de suite, ou de procéder à un tirage unique avec une urne contenant 100 bulletins dont 88 sont favorables à E.

Toutes les fois que l'on aura ramené l'étude d'un événement, dépendant d'un ou plusieurs tirages, à l'étude d'un autre tirage au sort, nous dirons que nous aurons employé des tirages équivalents. L'événement étudié devra avoir dans les deux cas même probabilité (ou des probabi-

lités très voisines) mais la manière de l'obtenir pourra être entièrement différente. Donnons d'autres exemples de tirages équivalents :

Exemple 2. — On sait que si l'on jette deux fois de suite une pièce de monnaie parfaitement symétrique, les quatre cas :

PP (Pile, Pile),
PF (Pile, Face),
FP (Face, Pile),
FF (Face, Face),

ont même probabilité $\frac{1}{4}$. Il revient donc au même de jeter deux fois de suite une pièce de monnaie symétrique, ou de jeter une seule fois un dé en forme de tétraèdre régulier, parfaitement équilibré, dont les quatre faces porteraient les indications PP, PF, FP, FF.

Exemple 3. — Soit à tirer au sort un lot, à attribuer à une obligation, parmi une série d'obligations à lots numérotées de 000 000 à 999 999. On peut procéder de deux manières différentes. On peut placer dans un million de petits étuis en cuivre des bulletins portant les numéros des obligations, brasser ces étuis et en tirer un au hasard. On peut également, avec dix boules seulement, procéder aux tirages successifs des six chiffres devant constituer le numéro du gagnant.

L'équivalence des tirages que nous venons d'envisager est évidente, dans la mesure où l'on admet qu'il est possible de parler de pièces parfaitement symétriques, de dés parfaitement équilibrés, de brassages parfaits. On remarquera que du point de vue pratique, des tirages équivalents peuvent poser des problèmes techniques tout à fait différents. Il est à la portée de tout le monde de jeter deux fois de suite une pièce de monnaie à peu près symétrique; il est plus difficile de se procurer un dé tétraédrique; quant au tirage d'un lot parmi un million, on voit que la confection, le remplissage et le brassage des petits étuis que nous avons supposés représentent un travail énorme, puisque avec des étuis minuscules pesant 1 g, le poids du matériel à manipuler représente 1 t. Il faut remarquer également que l'équivalence, évidente aux yeux d'un probabiliste, l'est moins aux yeux du profane. En ce qui concerne le lot parmi un million, certaines gens pensent que le numéro 000 000 a moins de chances de sortir que les autres, parce qu'il faut que la boule zéro sorte six fois de suite; si au contraire on emploie le système des

petits étuis, le numéro ooo ooo semble être mis « sur pied d'égalité » avec tous les autres numéros, parce que le tirage se fait d'un seul coup.

Il est des cas où le recours à un tirage équivalent est nécessaire pour donner un sens plus concret à un problème de probabilités. Soit par exemple le problème suivant :

Exemple 4. — On tire au hasard un nombre p compris entre o et 1 ; puis, on procède à n tirages pouvant donner lieu à l'apparition d'une certaine éventualité E. Quelle est la probabilité que l'éventualité E apparaisse r fois ? La probabilité cherchée, quand p est connu, est

$$\frac{n!}{r!\,(n-r)!}p^r q^{n-r} \qquad (q = 1-p).$$

Donc, étant donné la manière dont p est déterminé, cette probabilité est

$$\frac{n!}{r!\,(n-r)!}\int_0^1 p^r q^{n-r}\,dp = \frac{1}{n+1}.$$

La réalisation matérielle d'un pareil tirage n'est pas aisée, si l'on s'en tient aux termes de l'énoncé, qui suppose que *l'on commence* par tirer au sort un nombre compris entre o et 1. Ce tirage exige en théorie une infinité d'opérations; de toute manière, si l'on se contente d'une approximation, il y aura une opération délicate à faire. Si l'on suppose que n reste borné, par exemple inférieur à 100, on peut obtenir un résultat équivalent en procédant comme suit :

On rassemble un matériel consistant en une urne, 100 boules blanches et 100 boules noires. On met dans l'urne une boule blanche et une boule noire. On tire une des boules, et on la remet dans l'urne, en l'accompagnant d'une boule de même couleur. L'urne contient ainsi trois boules. On en tire une, et on la remet dans l'urne en ajoutant encore une boule de même couleur que la boule extraite. Le procédé se poursuit indéfiniment, le $n^{\text{ième}}$ tirage ayant lieu avec une urne contenant $n+1$ boules, mais dont la composition dépend des résultats des tirages antérieurs. On peut montrer que la suite des résultats de tirages ainsi faits est absolument identique, du point de vue des probabilités d'apparition, à la suite que l'on obtiendrait en suivant à la lettre les demandes de l'énoncé, c'est-à-dire en « tirant » d'abord p, à supposer que ce soit possible, et en laissant ensuite p constant.

L'intérêt de la notion de tirages équivalents, du point de vue de la théorie des probabilités, réside surtout dans le fait que les problèmes les plus divers peuvent se ramener, quant aux questions de principe, à des questions relatives au jeu de pile ou face. C'est là une circonstance connue depuis longtemps, mais qu'il paraît souhaitable de préciser avec quelques détails.

Dans ce qui suit, le jeu de pile ou face sera considéré comme servant de base à une espèce d' « étalon de probabilité », comparable, *mutatis mutandis*, aux étalons qui ont été établis pour les grandeurs physiques.

1.2. **Production de probabilités quelconques à partir d'un étalon produisant normalement la probabilité $\frac{1}{2}$.** — Nous supposons dans ce paragraphe que par un procédé quelconque, on a construit un appareil qui, à une cadence donnée, fournit des indications bivalentes. Nous supposons par exemple que cet appareil débite des valeurs

$$(1) \qquad X_1, \quad X_2, \quad X_n, \quad \ldots \quad \ldots \qquad (X_i = 0 \text{ ou } 1)$$

ne pouvant prendre que les valeurs 0 et 1, à raison d'une par seconde, à partir de sa mise en marche. Nous admettons que cet appareil comporte un mécanisme de tirage au sort, conçu de manière que les valeurs X_i soient indépendantes entre elles, et que les valeurs 0 et 1 soient équiprobables. Nous dirons que cet appareil est un étalon débitant des probabilités $\frac{1}{2}$ à raison d'une par seconde. Il existe des procédés pour construire de tels appareils. La question se pose maintenant, en supposant que l'on en possède un, de savoir si l'on peut lui faire débiter des probabilités différentes de $\frac{1}{2}$, et à quelle cadence. Précisons ce que nous entendons par cela. Le sens de cette affirmation est le suivant :

Faire débiter la probabilité $\frac{1}{3}$ à l'appareil consistera à manipuler les X_i de manière que l'étalon se comporte exactement comme une urne contenant trois boules identiques, et dans laquelle on ferait des tirages successifs.

1.2.1. Tirage d'un seul événement. — Occupons-nous tout d'abord de faire procéder par l'étalon à un seul tirage d'un événement E de probabilité p donnée, quelconque mais différente de $\frac{1}{2}$. Seule est donnée la

valeur de p; les combinaisons des X_i correspondant à E et celles correspondant à non E sont laissées à notre discrétion. Appelons $\bar{E}$ l'événement « non E ». Quelle que soit la manière dont nous traduisons la suite des X_i, à tout instant nous serons dans une des trois situations :

a. Les X_i ont montré que E s'est produit;

b. Les X_i ont montré que $\bar{E}$ s'est produit;

c. Les X_i n'ont pas encore décidé entre E et $\bar{E}$.

Il doit donc exister une fonction dépendant des X_i, soit

$$F(X_1, X_2, \ldots, X_n, \ldots), \tag{2}$$

ne pouvant prendre que les valeurs o et 1, telle que

$$\begin{cases} F(X_1, X_2, \ldots) = 1 & \text{signifie que E s'est produit;} \\ F(X_1, X_2, \ldots) = 0 & \text{» } \bar{E} \text{ » » .} \end{cases} \tag{3}$$

Tant que nous sommes dans le cas *c* la valeur de F reste indéterminée. Nous allons chercher à déterminer une telle fonction. On remarquera, d'après (3), que, la probabilité p de E étant imposée, nous devrons avoir [1]

$$\mathfrak{M}\, F(X_1, X_2, \ldots) = p. \tag{4}$$

Supposons $p > q$, pour fixer les idées, et choisissons F de la forme

$$F(X_1, X_2, \ldots, X_n, \ldots) = X_1 + (1 - X_1)G(X_2, X_3, \ldots). \tag{5}$$

G étant une fonction de même nature que F, c'est-à-dire ne prenant que les valeurs o et 1. L'équation (4) nous donne

$$\mathfrak{M}\, F = p = \frac{1}{2}[1 + \mathfrak{M}\, G],$$

d'où

$$\mathfrak{M}\, G = 2p - 1 = p - q = p_1 \qquad (p + q = 1). \tag{6}$$

b. Si nous avions été dans le cas où $p < q$, nous aurions posé

$$F(X_1, X_2, \ldots) = X_1[1 - G(X_2, X_3, \ldots)], \tag{7}$$

et nous aurions été amenés à

$$\mathfrak{M}\, G = q - p.$$

---

[1] Nous représentons par $\mathfrak{M}\, F$ la valeur moyenne de F.

De toute façon, partant d'un événement de probabilité p, on se trouve amené, au coup suivant, à déterminer un événement de probabilité $p_1 = |p - q|$. En procédant sur ce nouvel événement comme sur le premier, on se trouve amené à un événement de probabilité $p_2 = |p_1 - q_1|$, et ainsi de suite. Le procédé arrive à une fin si cette suite d'opérations amène à un événement de probabilité $\frac{1}{2}$; mais, en général, le procédé se poursuit indéfiniment. La fonction F que nous avons déterminée par le procédé ci-dessus l'est sans ambiguïté; on constate qu'il n'y a pas lieu d'attendre indéfiniment pour que la valeur de F soit déterminée. Si en effet F est de la forme (5), nous voyons que si $X_1 = 1$, $F = 1$, et qu'il est inutile de poursuivre le tirage. Si F est de la forme (7), c'est dans le cas où $X_1 = 0$ que sa valeur est déterminée dès le premier coup. Comme G peut être supposé formée de la même manière que F, ainsi que toutes les autres fonctions qui suivent, nous voyons qu'à chaque coup il y a une probabilité égale à $\frac{1}{2}$ pour que l'on soit fixé (si on ne l'a pas été auparavant) sur l'apparition de E ou de $\overline{E}$.

Le nombre moyen de coups nécessaires pour conclure à l'apparition de E ou $\overline{E}$ est ainsi :

$$1.\frac{1}{2} + 2.\frac{1}{4} + 3.\frac{1}{8} + \ldots = 2.$$

Nous constatons qu'il faut en moyenne deux coups, à un étalon conçu pour le tirage à pile ou face, pour procéder à un tirage avec probabilité p différente de $\frac{1}{2}$.

Exemple. — Pour éclairer le procédé général auquel nous venons de faire allusion, on peut expliciter F dans un cas simple, soit celui de $p = \frac{1}{3}$. On peut prendre pour F la fonction

$$F = X_1 - X_1X_2 + X_1X_2X_3 - X_1X_2X_3X_4 + \ldots.$$

La valeur moyenne de F est bien :

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \ldots = \frac{1}{3}.$$

Nous conclurons que E s'est produit dès que F aura pris la valeur 1, et que $\overline{E}$ s'est produit dès que F aura pris la valeur 0. D'après la nature

de F, on voit que la valeur de F est définitivement fixée dès que l'un des X_i est nul, et que par conséquent le nombre moyen de coups nécessaires pour que F prenne la valeur 0 ou 1, est 2. La probabilité que F ne soit jamais déterminée est nulle, puisque cette circonstance se produit seulement si tous les X_i ont pour valeur 1. La probabilité que la valeur de F se fixe sur 1 est bien $\frac{1}{3}$, celle qu'elle se fixe sur 0 est bien $\frac{2}{3}$.

1.2.2. Tirage de plusieurs événements. — Considérons maintenant le cas plus compliqué où l'on demande à l'appareil de faire le choix entre n éventualités incompatibles de probabilités respectives :

$$p_1, \quad p_2, \quad \ldots, \quad p_n. \tag{8}$$

En appelant toujours

$$X_1, \quad X_2, \quad \ldots, \quad X_m, \quad \ldots \qquad (X_m = 0 \text{ ou } 1),$$

la suite des résultats de tirage fournie par l'appareil, nous serons amenés à former n fonctions

$$F_k(X_1, X_2, \ldots, X_m, \ldots) \qquad (k = 1, 2, \ldots, n) \tag{9}$$

susceptibles de prendre les valeurs 0 et 1. Nous conviendrons que lorsque la suite des X_m sera telle que l'une des fonctions, soit F_k, prendra la valeur 1, cela signifiera que la $k^{\text{ième}}$ éventualité s'est produite. Lorsque l'une des fonctions, soit celle de rang k' prendra la valeur zéro, cela signifiera que l'éventualité de rang k' est exclue. En astreignant les fonctions F_k à l'identité

$$\sum_k F_k = 1, \tag{10}$$

nous serons certains que deux fonctions F_k d'indices différents ne peuvent prendre en même temps la valeur 1, et que si $n-1$ d'entre elles prennent la valeur zéro, la $n^{\text{ième}}$ prendra forcément la valeur 1. Le tirage des valeurs X_m devra se poursuivre jusqu'à ce qu'une des fonctions F_k prenne la valeur 1.

Pour assurer aux différentes éventualités, telles qu'elles seront indiquées par l'appareil, par l'intermédiaire des fonctions F_k, les probabi-

lités assignées en (8), nous devrons donner aux fonctions F_k une structure telle que

$$\mathfrak{M} F_k = p_k. \tag{11}$$

1.2.2.1. *Cas particulier* $\left(p_1 = p_2 = p_3 = \frac{1}{3}\right)$. — Traitons d'abord un cas particulier et cherchons à déterminer F_1, F_2, F_3 de manière à obtenir les probabilités

$$p_1 = p_2 = p_3 = \frac{1}{3}. \tag{12}$$

Nous prendrons pour F_1 la fonction déjà utilisée

$$F_1 = X_1 - X_1X_2 + X_1X_2X_3 - X_1X_2X_3X_4 + \dots. \tag{13}$$

Si $X_1 = 0$, nous savons que la première éventualité est exclue. Nous imposerons alors aux deux autres éventualités de se produire avec les probabilités $\frac{1}{3}$ et $\frac{2}{3}$ respectivement, ce qui nous conduit à poser

$$\begin{cases} F_2 = (1 - X_1)[X_2 - X_2X_3 + X_2X_3X_4 - \dots] + X_1G_2, \\ F_3 = (1 - X_1)[1 - X_2 + X_2X_3 - X_2X_3X_4 + \dots] + X_1G_3; \end{cases} \tag{14}$$

G_2, G_3 étant deux fonctions qui restent à déterminer. L'équation (10) sera satisfaite si

$$1 = F_1 + F_2 + F_3 = X_1(G_2 + G_3) + 1 - X_1(X_2 - X_2X_3 + X_2X_3X_4 - \dots),$$

ce qui nous conduit à

$$G_2 + G_3 = X_2 - X_2X_3 + X_2X_3X_4 - \dots.$$

Nous devons avoir de plus

$$\frac{1}{3} = \mathfrak{M} F_2 = \frac{1}{6} + \frac{1}{2}\mathfrak{M} G_2,$$
$$\frac{1}{3} = \mathfrak{M} F_3 = \frac{1}{3} + \frac{1}{2}\mathfrak{M} G_3.$$

Nous ferons $G_3 = 0$, ce qui conduit à

$$\begin{cases} F_1 = X_1 - X_1X_2 + X_1X_2X_3 - \dots = X_1(1 - F_2), \\ F_2 = X_2 - X_2X_3 + X_2X_3X_4 - \dots, \\ F_3 = (1 - X_1)(1 - F_2). \end{cases} \tag{15}$$

Les fonctions F_1, F_2, F_3 ne sont déterminées, c'est-à-dire l'une d'entre elles ne prend la valeur 1, que lorsque la fonction F_2 est elle-

même déterminée. Le premier coup sert à déterminer X_1 ; il faut ensuite en moyenne deux coups pour déterminer F_2. Il faut donc compter en moyenne *trois coups* pour déterminer quelle éventualité doit apparaître.

1.2.2.2. *Cas général* (*Intervention des probabilités* a priori *et* a posteriori). — Passons maintenant à quelques considérations sur le cas général. Il faut remarquer que le problème que nous cherchons à traiter n'est autre qu'un problème de probabilités des causes, au sens de Bayes. Appelons en effet

$$E_1, \quad E_2, \quad \ldots, \quad E_n$$

les éventualités que nous cherchons à faire apparaître, à partir des résultats du tirage, par l'intermédiaire des fonctions $F_1, F_2, \ldots, F_n$. L'événement E_k peut s'écrire

$$E_k = \{ F_k = 1 \}, \tag{16}$$

avec

$$\text{Prob} \{ F_k = 1 \} = p_k. \tag{17}$$

L'événement E_k est donc *lié*, au sens le plus classique, aux variables X_1, $X_2, \ldots, X_m$. Supposons les F_k choisies par un procédé quelconque et voyons quelles sont les conséquences de l'apparition de la valeur X_1. Si $X_1 = 0$ les événements E_k, qui avaient les probabilités p_k *a priori*, prennent les probabilités *a posteriori* p'_k. Si $X_1 = 1$, ils prennent les probabilités *a posteriori* p''_k. Les relations entre probabilités *a priori* et probabilités *a posteriori* sont

$$p_1 = \frac{p'_1 + p''_1}{2}, \qquad p_2 = \frac{p'_2 + p''_2}{2}, \qquad \ldots, \qquad p_n = \frac{p'_n + p''_n}{2}. \tag{18}$$

Supposons maintenant que, parmi tous les choix possibles de systèmes de fonctions F, le choix qui conduise au nombre moyen minimum de coups nécessaires corresponde à un nombre moyen

$$K(p_1, p_2, \ldots, p_n).$$

Ce nombre moyen minimum est bien une fonction des p_k (nous laissons de côté pour l'instant les difficultés relatives à l'existence des fonctions F permettant effectivement d'atteindre ce nombre moyen

minimum). D'après ce que nous avons vu sur l'interprétatiou de la valeur de X_1, la fonction K doit satisfaire à la relation fonctionnelle

$$(19)\quad K(p_1, p_2, \ldots, p_n) = 1 + \frac{1}{2}\,\mathrm{Min}[K(p'_1, p'_2, \ldots, p'_n) + K(p''_1, p''_2, \ldots, p''_n)].$$

Dans cette équation, il faut entendre que le minimum, dans le second membre, est pris parmi tous les systèmes de p'_k et p''_k satisfaisant aux équations (18).

1.3. **Détermination du nombre moyen de tirage à probabilité $\frac{1}{2}$ pour obtenir un tirage avec probabilités $\{p_k\}$.** — Pour éviter la difficulté relative au cas où l'on n'est pas sûr que la borne inférieure des nombres moyens de coups nécessaires est atteinte pour un choix convenable des F_k, nous allons supposer que les p_k sont de la forme

$$p_k = \frac{r_k}{2^N},$$

les r_k étant des entiers (ainsi que N). Nous supposerons de plus que, pour la formation des F_k, on s'en tient au procédé suivant :

La suite des N premiers tirages peut prendre 2^N formes différentes. Parmi ces 2^N formes, nous en affecterons r_1 à E_1, r_2 à E_2, ..., r_n à E_n. Cela revient à dire que F_1 est égale à 1 sur r_1 suites, et à zéro sur les $2^N - r_1$ autres, que F_2 est égale à 1 sur r_2 suites, et à zéro sur les $2^N - r_2$ autres, etc. Nous sommes ainsi certains qu'au bout de N coups tout est terminé, et que le rang d'un des événements E_k est déterminé. Il se peut naturellement que la détermination ait lieu avant que les N coups aient été tirés. Comme nous raisonnons sur un nombre fini de cas, et que les manières de choisir des groupes de $r_1, r_2, \ldots$ suites sont elles-mêmes en nombre fini, le nombre moyen minimum de coups à tirer est un nombre bien défini, que nous appellerons

$$K_N(p_1, p_2, \ldots, p_n).$$

Examinons maintenant l'influence de l'apparition de X_1. La connaissance de la valeur de X_1 revient à isoler 2^{N-1} suites parmi les 2^N possibles. Les probabilités p'_k, p''_k sont donc de la forme

$$p'_k = \frac{r'_k}{2^{N-1}}, \qquad p''_k = \frac{r''_k}{2^{N-1}}$$

et nous avons

$$r_k = r'_k + r''_k.$$

L'équation (19) prend alors la forme

$$(20)\quad K_N(p_1, p_2, \ldots, p_n) = 1 + \frac{1}{2}\,\mathrm{Min}\,\{K_{N-1}(p'_1, p'_2, \ldots, p'_n) + K_{N-1}(p''_1, p''_2, \ldots, p''_n)\}.$$

L'équation (20) montre comment la fonction K_N peut se calculer par récurrence. Nous allons voir qu'elle permet d'encadrer K_N entre des valeurs limites l'exprimant simplement en fonction des p_k.

Nous allons tout d'abord indiquer une borne inférieure de K_N. Nous utiliserons pour cela l'inégalité

$$(21)\quad \sum p'_k \log \frac{1}{p'_k} + \sum p''_k \log \frac{1}{p''_k} \geqq 2 \sum p_k \log \frac{1}{p_k} - 2 \log 2$$

qui est une conséquence immédiate des relations (18).

Nous allons montrer que

$$(22)\quad K_N(p_1, p_2, \ldots, p_n) \geqq \frac{1}{\log 2} \sum p_k \log \frac{1}{p_k}.$$

Pour $N = 1$, l'inégalité (22) (qui peut se réduire à l'égalité) est évidente. Supposons-la démontrée pour toutes les valeurs de l'indice inférieures ou égales à $N - 1$.

Nous déduisons de (20)

$$K_N(p_1, p_2, \ldots, p_n) \geqq 1 + \frac{1}{2 \log 2}\left[\sum p'_k \log \frac{1}{p_k} + \sum p''_k \log \frac{1}{p''_k}\right],$$

ce qui, en tenant compte de (21), conduit immédiatement à (22). Nous allons maintenant chercher une borne supérieure de K_N. Nous allons d'abord particulariser davantage la forme des p_k, en supposant que les r_k sont des puissances de 2.

1.3.1. Cas particulier ou les p_k sont des puissances négatives de 2. — Si les r_k sont des puissances de 2, les p_k sont de la forme

$$(23)\quad p_k = \frac{1}{2^m},$$

l'exposant m pouvant varier d'un p_k à un autre. Les F_k peuvent alors être pris de la forme

$$(24)\quad F_k = Y_1 Y_2 \ldots Y_m \qquad (Y_i = X_i \text{ ou } 1 - X_i).$$

Les fonctions F_k sont des produits de m facteurs, le $i^{\text{ième}}$ facteur étant égal à X_i ou $1 - X_i$. Nous remarquons tout d'abord qu'une telle forme assure

$$\mathfrak{M} F_k = p_k. \tag{25}$$

Il nous faut vérifier que l'on peut astreindre ces fonctions à la condition

$$\sum F_k = 1.$$

Nous allons pour cela procéder par récurrence. Soit M le plus grand des m. Certains des p_k ont pour valeur 2^{-M}. Ceux des p_k qui ont cette valeur sont forcément en nombre pair. Si en effet ces p_k étaient en nombre impair, l'égalité

$$\sum p_k = 1$$

multipliée par 2^M, conduirait à une égalité entre un nombre impair à gauche et un nombre pair à droite. Groupons deux par deux les p_k égaux à 2^{-M}, et convenons d'affecter à deux p_k d'une même paire des fonctions F de la forme

$$Y_1 Y_2 \dots Y_{M-1} X_M,$$
$$Y_1 Y_2 \dots Y_{M-1} (1 - X_M),$$

c'est-à-dire ne différant que par le dernier facteur. Nous voyons que si nous savons former les fonctions F, de la forme assignée, pour un système de p_k de la forme 2^{-m} avec $m \leqq M - 1$, nous saurons en déduire des fonctions F, également de la forme assignée, pour des p_k de la forme 2^{-m} avec $m \leqq M$. Soit en effet un tel système de probabilités, ordonnées suivant un ordre de grandeur non croissant :

$$p_1 \geqq p_2 \geqq \dots \geqq p_n = 2^{-M}.$$

Conservons les $p_k \geqq 2^{1-M}$ et groupons deux par deux les p_k égaux à 2^{-M}, lesquels sont forcément en nombre pair comme nous l'avons vu. Nous obtenons un système de probabilités

$$p'_1 \geqq p'_2 \geqq \dots \geqq p'_s = 2^{1-M}.$$

En appelant

$$F'_1, \quad F'_2, \quad \dots, \quad F'_1$$

les fonctions F associées aux probabités p'_k, fonctions que nous savons former par hypothèse, si nous posons

$$\begin{array}{lll} F_k = F'_k & \text{pour} & k = 1,\ 2,\ \ldots,\ 2s-n; \\ F_k = X_M F'_t & \text{»} & k = n-2h-1,\quad t = s-h \quad (h \leq n-s-1); \\ F_k = (1-X_M) F'_t & \text{»} & k = n-2h,\qquad t = s-h \end{array}$$

nous obtiendrons un système de fonction F, associées aux p_k, et de la forme assignée.

Comme nous savons former ces fonctions F pour $M = 1$, nous avons dans le procédé indiqué une méthode permettant de les former dans tous les cas où les p_k sont de la forme 2^{-m}.

Étant donné une telle fonction F, on n'est certain qu'elle est égale à 1 qu'au bout de m tirages, m étant le nombre des facteurs intervenant dans la fonction.

Le nombre moyen de coups nécessaires pour déterminer une éventualité se trouve donc être égal, pour ce choix des fonctions F, à

$$\sum m.2^{-m},$$

c'est-à-dire à

$$\frac{1}{\log 2} \sum p_k \log \frac{1}{p_k}.$$

Nous reconnaissons la valeur minimum de K. Nous en concluons que pour les p_k de la forme envisagée, nous avons exactement

$$K(p_1, p_2, \ldots, p_n) = \frac{1}{\log 2} \sum p_k \log \frac{1}{p_k}. \tag{26}$$

1.3.2. Cas ou les p_k sont des multiples entiers de puissances négatives de 2. — Revenons maintenant au cas où les p_k sont de la forme

$$p_k = r_k.2^{-N} \qquad (r_k \text{ entier}).$$

On peut mettre ces p_k sous la forme

$$p_k = \sum_h q_{kh},$$

les q_{kh} étant de la forme 2^{-m}, et les q_{kh} affectés à un même p_k étant *tous différents entre eux.*

Substituons au système des p_k le système des q_{kh}. Si nous savons former les fonctions F relatives au système des p_{kh}, nous saurons former

un système de fonction F relatives aux p_k. En appelant F_{kh} la fonction affectée à q_{kh}, il nous suffira de poser

$$F_k = \sum_h F_{kh}.$$

Ce système de F_k n'est peut être pas le plus avantageux en ce qui concerne la réduction maximum du nombre moyen de coups nécessaires. Mais son existence nous permet d'écrire à coup sûr :

$$K_N(\{p_k\}) \leqq K_N(\{q_{kh}\}),$$

en représentant par $\{p_k\}$ l'ensemble des p_k et par $\{q_{kh}\}$ l'ensemble des q_{kh}. Explicitons cette inégalité, en tenant compte des résultats précédemment acquis; il vient

$$K_N(\{p_k\}) \leqq \frac{1}{\log 2} \sum_{kh} q_{kh} \log \frac{1}{q_{kh}}. \tag{27}$$

Séparons dans le second membre l'expression

$$\varphi(p_k) = \sum_h q_{kh} \log \frac{1}{q_{kh}}. \tag{28}$$

La valeur numérique de cette expression dépend uniquement de la valeur numérique de p_k, c'est pourquoi nous la considérons comme la valeur prise pour $t = p_k$ par une certaine fonction que nous allons étudier.

Supposons $p_k < \frac{1}{2}$. Si nous doublons la valeur de p_k tous les q_{kh} se trouvent doublés, d'où

$$\varphi(2p_k) = \sum 2q_{kh} \log \frac{1}{2q_{kh}} = 2\varphi(p_k) - 2p_k \log 2.$$

La fonction $\varphi(t)$ satisfait donc à l'équation fonctionnelle

$$\varphi(2t) = 2\varphi(t) - 2t \log 2 \qquad \left(t < \frac{1}{2}\right), \tag{29}$$

étant entendu que t ne prend que des valeurs de la forme $r \cdot 2^{-m}$ (r entier). Faisons le changement de variable et de fonction

$$t = e^{-u \log 2}, \qquad \psi(u) = \frac{\varphi(t)}{t}.$$

L'équation (29) devient

$$\psi(u-1)=\psi(u)-\log 2$$

ou, en introduisant une nouvelle variable v :

$$v=u-1,$$
$$\psi(v+1)=\psi(v)+\log 2.$$

La fonction $\psi(v)$, dans son domaine d'existence, est donc de la forme

$$\psi(v)=v\log 2+\rho(v), \tag{30}$$

$\rho(v)$ étant une fonction périodique de v, de période 1.

L'équation (29) montre que $\varphi(t)$ est entièrement déterminée par les valeurs qu'elle prend dans l'intervalle

$$\frac{1}{2}<t<1. \tag{31}$$

L'équation (30) montre que $\psi(v)$ est également déterminée par les valeurs qu'elle prend dans l'intervalle correspondant

$$-1<v<0.$$

Dans l'intervalle (31), le développement binaire de t est constitué par des termes de la forme 2^{-m}, *tous différents*, et $\varphi(t)$ est de la forme

$$\sum 2^{-m}\log 2^m.$$

$\varphi(t)$ est donc bornée supérieurement par l'expression obtenue en prenant *tous* les termes possibles de la forme 2^{-m}; nous pouvons donc écrire

$$\varphi(t)<\sum_{m=1}^{\infty}2^{-m}\log 2^m=\log 2\sum_{m=1}^{\infty}\frac{m}{2^m}=2\log 2.$$

Nous en déduisons pour $\psi(v)$ l'inégalité

$$\psi(v)=v\log 2+\rho(v)<\frac{2\log 2}{t}\qquad(-1<v<0)$$

et pour $\rho(v)$ l'inégalité

$$\rho(v)<\left(\frac{2}{t}-v\right)\log 2=(2^{v+2}-v)\log 2.$$

Dans l'intervalle considéré, le second membre de l'inégalité ci-dessus

est croissant; donc il est au plus égal à la valeur qu'il prend pour $\varphi = 0$, et nous avons ainsi

$$\rho(\varphi) < 4 \log 2;$$

ρ étant périodique, l'inégalité que nous venons de trouver est valable quel que soit φ. Nous pouvons donc écrire

$$\psi(\varphi) < \varphi \log 2 + 4 \log 2$$

et également

$$\psi(u) < (u + 4) \log 2,$$
$$\varphi(p) < p \log 2 \left[4 - \frac{\log p}{\log 2} \right] = 4p \log 2 + p \log \frac{1}{p}.$$

Revenons à (27); nous déduisons de l'inégalité précédente :

$$K_N(\{p_k\}) \leq 4 + \frac{1}{\log 2} \sum p_k \log \frac{1}{p_k}. \tag{32}$$

1.3.3. Cas des p_k quelconques. — Dans le cas où les p_k sont quelconques, on peut tout d'abord montrer sans peine que

$$K(p_1, p_2, \ldots, p_n) \leq n.$$

Pour cela, on peut procéder par récurrence. Supposons l'inégalité démontrée pour tous les entiers inférieurs à n. Nous pouvons toujours substituer, aux n probabilités p_k, $n + 1$ probabilités p'_k telles que

$$p_1 = p'_1, \quad \ldots, \quad p_h = p'_h, \quad p_{h+1} = p'_{h+1} + p'_{h+2},$$
$$p_{h+2} = p'_{h+3}, \quad \ldots, \quad p_n = p'_{n+1};$$
$$p'_1 + \ldots + p'_{h+1} = p'_{h+2} + \ldots + p'_{n+1} = \frac{1}{2}.$$

Il en résulte évidemment que

$$K(p_1, p_2, \ldots, p_n) \leq K(p'_1, p'_2, \ldots, p'_{n+1}).$$

Par ailleurs, un premier tirage nous permet de choisir entre les deux groupes de p'_k, d'où la relation

$$K(p'_1, p'_2, \ldots, p'_{n+1}) \leq 1 + \frac{1}{2} [K(2p'_1, \ldots, 2p'_{h+1}) + K(2p'_{h+2}, \ldots, 2p'_{n+1})].$$

Sauf dans le cas où $n = 2$, on peut, en modifiant au besoin la numérotation des p_k, faire en sorte que

$$h + 1 < n \quad \text{et} \quad n - h < n.$$

Comme pour $n = 2$, on sait déjà que $K = 2$; les inégalités ci-dessus montrent bien que $K \leq n$ dans tous les cas.

Ceci étant établi, revenons aux p_k, et approchons-les inférieurement par des p'_k de la forme envisagée en 1.3.2.

Posons

$$\sum_{k=1}^{n} p'_k = P < 1, \qquad Q = 1 - P;$$

$$p'_k = P\bar{p}_k \leq p_k.$$

Les $\bar{p}_k$ sont également de la forme envisagée en 1.3.2.

Nous avons alors

$$p_k = P\bar{p}_k + Q\bar{q}_k.$$

Procédons d'abord au tirage des probabilités

$$P\bar{p}_1, \quad P\bar{p}_2, \quad \ldots, \quad P\bar{p}_n, \quad Q.$$

Si l'événement de probabilité P se produit, c'est-à-dire une des n premières éventualités, nous pourrons conclure à l'apparition d'une des éventualités primitivement considérées. Si c'est l'événement de probabilité Q, il faudra entreprendre un nouveau tirage. Ceci montre que

$$\begin{aligned} K(p_1, p_2, \ldots, p_n) &\leq PK(P\bar{p}_1, P\bar{p}_2, \ldots, P\bar{p}_n, Q) \\ &\quad + QK(\bar{q}_1, \bar{q}_2, \ldots, \bar{q}_n) \\ &\leq \frac{P}{\log 2}\left(P\bar{p}_1 \log \frac{1}{P\bar{p}_1} + \ldots + Q \log \frac{1}{Q}\right) + 4P + nQ. \end{aligned}$$

Cette dernière inégalité étant vraie quel que soit l'ordre de l'approximation faite, il en résulte à la limite

$$K(p_1, p_2, \ldots, p_n) \leq \frac{1}{\log 2} \sum p_k \log \frac{1}{p_k} + 4.$$

Cette inégalité peut sans doute être resserrée, mais elle suffit pour l'établissement des résultats que nous avons en vue.

L'inégalité (32) montre que l'expression

$$\mathcal{H}(\{p_k\}) = \frac{1}{\log 2} \sum p_k \log \frac{1}{p_k}$$

représente une valeur approchée du nombre moyen de coups nécessaires pour obtenir, à partir d'un étalon de probabilité $\frac{1}{2}$, un tirage avec les probabilités $(\{p_k\})$.

Cette valeur approchée peut être considérée comme une valeur limite; nous allons montrer dans quel sens.

Précisons le résultat déjà obtenu :

L'étalon de probabilité nous fournit une suite des valeurs :

$$X_1, \quad X_2, \quad \ldots, \quad X_m, \quad \ldots \qquad (X_m = 0 \quad \text{ou} \quad 1).$$

Nous en déduisons une suite de valeurs :

$$Y_1, \quad Y_2, \quad \ldots, \quad Y_m, \quad \ldots \qquad (Y_m = 1, 2, \ldots, n);$$

de manière que

$$\text{Prob}\{Y_m = k\} = p_k.$$

Y_1 sera déterminé par une suite

$$X_1, \quad X_2, \quad \ldots, \quad X_{\mu_1};$$

Y_2 par une suite

$$X_{\mu_1+1}, \quad X_{\mu_1+2}, \quad \ldots, \quad X_{\mu_1+\mu_2};$$

et ainsi de suite. Nous avons montré que l'on peut déterminer la loi de formation des Y_i de manière que

$$\mathfrak{M}\{\mu_1\} = \mathfrak{M}\{\mu_2\} = \ldots \leqq 4 + \mathfrak{H}(\{p_k\}).$$

La correspondance entre les X et les Y est telle que l'on détermine les Y un par un. Mais nous pourrions déterminer les Y par couples. Cela revient à substituer aux p_i (en nombre n) les n^2 probabilités de la forme $p_k p_h$.

Pour déterminer le premier couple Y_1, Y_2, il nous faudrait alors procéder aux tirages de

$$X_1, \quad X_2, \quad \ldots, \quad X_{\nu_1};$$

Pour déterminer le deuxième couple Y_3, Y, il nous faudrait procéder aux tirages de

$$X_{\nu_1+1}, \quad \ldots \quad X_{\nu_1+\nu_2}$$

et ainsi de suite. Nous aurions alors

$$\mathfrak{M}\{\nu_1\} = \mathfrak{M}\{\nu_2\} = \ldots \leqq 4 + \mathfrak{H}[\{p_k p_h\}],$$

de sorte que le nombre moyen de coups nécessaires pour déterminer un Y serait au plus

$$2 + \frac{1}{2}\mathfrak{H}[\{p_k p_h\}].$$

Nous représentons conventionnellement par $\{p_k p_h\}$ l'ensemble des n^2 probabilités associés à deux Y consécutifs. On montre sans peine que

$$\mathcal{H}[\{p_k p_h\}] = 2\mathcal{H}[\{p_k\}],$$

de sorte que si nous admettons que l'on puisse déterminer les Y par groupes de deux, on pourra opérer de manière que le nombre moyen de tirages sur les X nécessaires pour détenir *un* Y soit au plus de

$$2 + \mathcal{H}[\{p_k\}].$$

Nous pouvons aller plus loin dans la voie amorcée au paragraphe précédent. Nous pouvons supposer que l'on détermine les événements de probabilité p_k par groupes de ν événements consécutifs et indépendants. Le nombre minimum de coups nécessaire en moyenne pour déterminer un événement sera alors

$$\frac{1}{\nu}\mathrm{K}[\{p_k\}^\nu],$$

en désignant par $\{p_k\}^\nu$ l'ensemble des n^ν probabilités associées à une suite de ν tirages indépendants, les probabilités associées à chaque tirage étant les p_k. On vérifie sans peine que

$$\mathcal{H}[\{p_k\}^\nu] = \nu\mathcal{H}[\{p_k\}],$$

de sorte que le nombre moyen de coups nécessaires pour déterminer un événement pourra être amené, par un choix convenable de l'interprétation à donner aux résultats du tirage émanant de l'étalon considéré tout à fait au début, entre les limites

$$\mathcal{H}[\{p_k\}], \qquad \mathcal{H}[\{p_k\}] + \frac{4}{\nu}.$$

Si l'on suppose maintenant que l'on peut prendre ν suffisamment grand, on voit que l'expression $\mathcal{H}[\{p_k\}]$ peut être considéré comme le nombre moyen de coups nécessaires ponr déterminer un événement parmi n événements de probabilités $\{p_k\}$.

Exemple. — Pour montrer la nature du résultat qui vient d'être établi, nous allons considérer un exemple : soient les 10 probabilités

$$p_1, \quad p_2, \quad \ldots, \quad p_{10},$$

associées à 10 événements incompatibles. Procéder à une suite de

tirages indépendants, avec ces probabilités, revient à former une suite de variables indépendantes :

$$X_1, \quad X_2, \quad \ldots, \quad X_m, \quad \ldots \tag{33}$$

chacune de ces variables pouvant prendre les valeurs de 0 à 9; d'une manière précise, nous devons avoir quel que soit m :

$$\text{Prob}\{X_m = 0\} = p_1, \quad \text{Prob}\{X_m = 1\} = p_2, \quad \ldots, \quad \text{Prob}\{X_m = 9\} = p_{10}. \tag{34}$$

La suite des X_m peut être considérée comme la suite des décimales du développement décimal d'un nombre réel X compris entre 0 et 1. Nous faisons abstraction des difficultés qui peuvent apparaître si X est un nombre rationnel à dénominateur de la forme d'une puissance de 10, auquel cas deux suites de X_m peuvent correspondre à un même X. Dans ces conditions, tirer au hasard les m premiers termes de la suite (33) revient à tirer au hasard X, sur le segment (0, 1), avec une distribution de probabilités convenablement choisie, et à déterminer, par une loi précise, les valeurs des m premières décimales du développement de X.

Supposons maintenant que nous ne disposions d'aucun appareil pour procéder au tirage de X, mais que nous possédions l'étalon de probabilités auquel il a déjà été fait allusion; cet étalon peut nous fournir la suite de variables indépendantes :

$$Y_1, \quad Y_2, \quad \ldots, \quad Y_m, \quad \ldots,$$

avec

$$\text{Prob}\{Y_m = 0\} = \text{Prob}\{Y_m = 1\} = \frac{1}{2}.$$

Les Y_m déterminent un nombre

$$Y = \sum Y_m \cdot 2^{-m}$$

et, réciproquement, sauf en ce qui concerne les Y rationnels à dénominateur puissance de 2, la donnée de Y détermine les Y_m.

Pour déterminer les X_m à partir des Y_m, nous pouvons lier X et Y par une relation fonctionnelle

$$X = f(Y) \tag{35}$$

respectant la loi de probabilité à laquelle satisfont les X_m. Nous entendons par là que si l'on prend Y au hasard dans l'intervalle (0, 1), avec

densité de probabilité uniforme, que l'on calcule $f(Y)$ et qu'on détermine ces décimales, cette suite de décimales sera constituée par des variables indépendantes satisfaisant à (34).

Attachons-nous maintenant aux m premières décimales :

$$X_1, \quad X_2, \quad \ldots, \quad X_m \tag{36}$$

et procédons au tirage des Y_i :

$$Y_1, \quad Y_2, \quad \ldots, \quad Y_n. \tag{37}$$

Les n premiers termes de (37) ne permettent de calculer Y qu'avec une certaine approximation, à savoir à 2^{-n} près. De cette valeur approchée de Y, nous ne pouvons calculer X, d'après (35), qu'avec une certaine approximation ; et nous ne pourrons par conséquent calculer qu'un certain nombre des premières décimales (35). Ce nombre peut être éventuellement nul. Soit

$$N[Y_1, Y_2, \ldots, Y_n] \tag{38}$$

le nombre des décimales X_i que l'on peut calculer à partir de $Y_1, Y_2, \ldots, Y_n$. Ce nombre dépend non seulement de n, mais encore des valeurs prises par $Y_1, Y_2, \ldots, Y_n$. Nous devrons poursuivre le tirage jusqu'à ce que $N \geqq m$, pour que la suite (36) soit définie.

Les raisonnements faits plus haut nous montrent que si l'on pose

$$H = \frac{1}{\log 2} \sum_{k=1}^{10} p_k \log \frac{1}{p_k} \tag{39}$$

et si l'on donne un nombre ε arbitrairement petit, on pourra choisir m et la fonction $f(Y)$ de manière que le nombre moyen de Y_i nécessaires pour déterminer les X_i de la suite (36) soit compris entre

$$mH \quad \text{et} \quad m(H+\varepsilon).$$

Considérons le graphique dressé de la manière suivante :

On procède au tirage des Y_i, et, pour toute valeur entière i, on marque le point ayant pour abscisse i et pour ordonnée le nombre des décimales de X déterminées par $Y_1, Y_2, \ldots, Y_i$; en joignant tous ces points, on obtient une ligne brisée. Il résulte des considérations précédentes que l'on peut choisir la relation entre X et Y de manière que cette ligne brisée ait, avec probabilité 1, la direction de pente $\frac{1}{H}$ comme direction asymptotique.

Pour un très grand nomdre de tirages, nous voyons que N valeurs des Y_i détermineront des valeurs des X_i dont le nombre sera de l'ordre de $\frac{N}{H}$. La suite des $\frac{N}{H}$ premières valeurs des Y_i peut prendre 2^N formes différentes; la suite des $\frac{N}{H}$ premières valeurs des X_i peut prendre $10^{\frac{H}{N}}$ formes différentes. Comme les formes que peut prendre la suite des X_i se déduisent des formes prises par la suite des Y_i, et par conséquent ne peuvent être plus nombreuses, nous sommes amenés à penser que parmi les $10^{\frac{N}{H}}$ formes *théoriquement possibles* de la suite des premiers X_i, il n'y en a eu en réalité que 2^N qui sont pratiquement possibles. Nous allons montrer dans ce qui suit ce qu'il faut entendre par là.

1.4. **Notion de diversité et de nombre d'événements pratiquement possibles.** — 1.4.1. — DIVERSITÉ D'UN SYSTÈME D'ÉVÉNEMENTS. — Soit un système de n événements incompatibles de probabilités $p_1, p_2, \ldots, p_n$. Posons

$$(40)\qquad \begin{cases} \log\mu = -\sum p_i \log p_i, \\ (\log\sigma)^2 = \sum p_i (\log \mu p_i)^2 \qquad (\sigma > 1) \end{cases}$$

et soit ε un nombre positif arbitraire. Considérons les p_i tels que

$$(\log \mu p_i)^2 > k^2 (\log\sigma)^2 \qquad (k > 1).$$

La somme $\sum' p_i$ de ces p_i est au plus égale à $\frac{1}{k^2}$. Si nous prenons

$$k = \frac{1}{\sqrt{\varepsilon}}, \qquad \sum' p_i < \varepsilon,$$

nous voyons que dans un tirage au sort des événements considérés, il y aura une probabilité au moins égale à $1 - \varepsilon$ pour que l'événement qui apparaît, soit un événement dont la probabilité p_i d'apparition satisfasse à l'inégalité

$$(\log \mu p_i)^2 < \frac{(\log\sigma)^2}{\varepsilon}.$$

Cette égalité est équivalente à la double inégalité

$$(41)\qquad \frac{1}{\tau\mu} < p_i < \frac{\tau}{\mu} \qquad \left[\log^2\tau = \frac{1}{\varepsilon}\log^2\sigma,\ \tau > 1\right].$$

Il existe donc, parmi les p_i donnés, un certain groupe de p_i dont la somme est au moins égale à $1-\varepsilon$, et qui satisfait à (41). Il est facile de limiter le nombre de ces p_i. Ils sont en nombre au plus égal à $\tau\mu$. S'ils étaient plus nombreux, d'après (41), leur somme dépasserait l'unité.

Nous pouvons encore dire : μ et σ étant définies par les égalités (40), et τ étant un nombre positif plus grand que 1, si l'on considère, parmi les n événements possibles, les $\mu\tau$ plus probables, la somme des probabilités affectées à ces événements sera au moins égale à

$$1-\left(\frac{\log\sigma}{\log\tau}\right)^2. \tag{42}$$

Si τ est voisin de 1 et si la quantité (42) est également voisine de 1, nous voyons que μ peut être considéré comme le nombre d'événements pratiquement possibles parmi les événements considérés, parce que la probabilité correspondant aux $n-\mu$ événements les moins probables est pratiquement négligeable.

Le nombre μ sera appelé la « diversité » du système d'événements considéré.

Nous allons voir que dans le cas d'épreuves répétées, les conditions envisagées ci-dessus sont respectées.

Reprenons les n événements considérés. Supposons que l'on procède à N épreuves indépendantes; nous avons maintenant affaire à n^N événements. Si la diversité du système des n événements était μ, la diversité du système des n^N événements sera μ^N. On vérifie en effet que

$$-\sum p_{i_1}p_{i_2}p_{i_3}\dots p_{i_N}\log p_{i_1}p_{i_2}\dots p_{i_N}=N\log\mu.$$

Si nous appelons σ_N la quantité, analogue à σ, relative aux nouveaux événements, nous voyons que

$$(\log\sigma_N)^2=N(\log\sigma)^2.$$

Donnons-nous une valeur de τ, aussi voisine de 1 que nous voudrons; nous allons montrer que dans le décompte des événements pratiquement possibles, parmi les n^N événements théoriquement possibles, on peut admettre, pour N suffisamment grand, que chaque épreuve ne donne lieu qu'à $\mu\tau$ éventualités différentes. Cela revient en effet à dire que les N épreuves donnent lieu à $(\mu\tau)^N$ éventualités différentes. Or, si nous évaluons la somme des probabilités relatives aux $(\mu\tau)^N$ événements

les plus probables parmi les n^N événements possibles nous obtenons une quantité bornée inférieurement par une expression de la forme (42), avec la condition que $(\log\sigma)^2$ doit être multiplié par N et $\log\tau$ également par N. La probabilité en question est donc au moins égale à

$$1 - \frac{1}{N}\left(\frac{\log\sigma}{\log\tau}\right). \tag{43}$$

Elle tend vers 1 lorsque N tend vers l'infini.

La diversité μ d'un système pouvant prendre n états de probabilités $p_1, p_2, \ldots, p_n$, étant définie par la première des équations (40), il est facile de vérifier que $\mu \leq n$, l'égalité $\mu = n$ n'étant réalisée que si toutes les probabilités p_i sont égales entre elles.

μ est une évaluation du nombre d'états que peut pratiquement prendre le système, au sens que nous avons précisé plus haut. Cela tient essentiellement au fait que $\frac{1}{\mu}$ est la moyenne géométrique pondérée des probabilités p_i; $\frac{1}{\mu}$ est donc une probabilité moyenne. On conçoit que *l'inverse* d'une probabilité moyenne puisse être considérée comme un nombre d'états conventionnel. Nous allons voir pourquoi l'inverse de la probabilité moyenne géométrique est une notion plus utile que l'inverse de probabilités moyennes d'une autre nature.

Examinons par exemple ce qui se passe quand on considère la *moyenne harmonique* des p_i, moyenne pondérée avec les poids p_i. Cette moyenne λ est définie par

$$\frac{1}{\lambda} = \sum p_i \frac{1}{p_i} = n.$$

L'inverse de λ donne le nombre n lui-même, ce qui n'apprend rien de nouveau. Passons à la moyenne arithmétique, toujours pondérée; son inverse est défini par

$$\frac{1}{\nu} = \sum p_i p_i = \sum p_i^2 = \sum \left(p_i - \frac{1}{n}\right)^2 + \frac{1}{n}.$$

Nous constatons que $\nu < n$. Donc ν peut être considéré comme un nombre d'états possibles. Reste à calculer la probabilité attachée aux ν états les plus probables du système. Il nous faut considérer pour cela la quantité s définie par

$$s^2 = \sum p_i \left(p_i - \frac{1}{\nu}\right)^2,$$

caractérisant la dispersion des probabilités autour de la valeur moyenne $\frac{1}{\nu}$.

Les probabilités telles que

$$\left(p_i - \frac{1}{\nu}\right)^2 < \varepsilon^2 \qquad (\varepsilon > 0)$$

ont une somme au plus égale à $\frac{s^2}{\varepsilon^2}$.

Les probabilités au moins égales à $\frac{1}{\nu} - \varepsilon$ ont donc une somme au moins égale à

$$1 - \frac{s^2}{\varepsilon^2},$$

et les états admettant de telles probabilités sont en nombre égal au plus à $\frac{\nu}{1 - \nu\varepsilon}$.

Nous pouvons donc dire que si nous considérons les $\nu\tau$ états les plus probables ($\tau > 1$), la somme des probabilités correspondantes est au moins égale à

$$P = 1 - \frac{s^2\nu^2\tau^2}{(\tau - 1)^2}. \tag{44}$$

Cette probabilité est très voisine de 1 si s est petit, et dans ce cas on peut considérer $\nu\tau$ comme une borne du nombre d'états pratiquement possibles.

ν, τ, s étant donnés, nous allons voir comment varie P lorsque l'on considère au lieu d'un seul système à n états, N systèmes semblables, indépendants; le système total a n^N états possibles, avec un nombre égal de probabilités de la forme

$$p_i, \quad p_{i_2}, \quad \cdots, \quad p_{i_N}.$$

Dans le cas où $N = 2$, si nous appelons ν_2 la nouvelle valeur de ν, nous voyons sans peine que

$$\frac{1}{\nu_2} = \sum p_i^2 p_j^2 = \frac{1}{\nu^2}.$$

En appelant s_2 la nouvelle valeur de s, nous voyons que

$$s_2^2 = \sum p_i p_j \left(p_i p_j - \frac{1}{\nu^2}\right) = \left(s^2 + \frac{1}{\nu^2}\right)^2 - \frac{1}{\nu^4}, \qquad (s_2\nu_2)^2 = (s^2\nu^2 - 1)^2 - 1.$$

Si nous appelons ν_N et s_N les grandeurs relatives aux N systèmes, nous aurons ainsi

$$(s_N \nu_N)^2 = (s^2 \nu^2 + 1)^N - 1,$$

et l'expression (44) sera remplacée par

$$1 - \left(\frac{\tau}{\tau - 1}\right)^2 [(s^2 \nu^2 + 1)^N - 1].$$

Nous ne pouvons plus conclure, comme nous l'avions fait plus haut, que la probabilité en question tend vers 1 lorsque N augmente indéfiniment.

1.4.2. Diversité de plusieurs systèmes indépendants. — Pour montrer que la notion de diversité d'un système que nous avons adoptée correspond bien à la notion de nombre d'états pratiquement possibles pour le système, nous poserons les définitions suivantes :

Étant donné N systèmes indépendants $S_1, S_2, \ldots, S_N$ pouvant prendre chacun un certain nombre d'états, nous appellerons produit de ces systèmes, et nous représenterons par

$$\sum = \{ S_1 S_2 \ldots . S_N \},$$

le système dont chaque état est caractérisé par le fait que S_1 est dans un état déterminé, S_2 dans un état déterminé, etc., et ainsi de suite. Si $S_1, S_2, \ldots, S_N$ sont susceptibles de $n_1, n_2, \ldots, n_N$ états respectivement, $\sum$ sera susceptible de $n_1, n_2, \ldots, n_N$ états.

Si les systèmes S_i sont tous de même structure et qu'on puisse par conséquent poser $S_i = S$, nous appellerons $\sum$ la puissance $N^{\text{ième}}$ de S, et nous écrirons

$$\sum = \{ S \}^N.$$

Étant donné un système S quelconque, nous appellerons

$$n_\varepsilon(S) \qquad (\varepsilon > 0),$$

le nombre minimum des états, que peut prendre S, tel que la somme des probabilités affectées à ces états, soit au moins égale à $1 - \varepsilon$.

Considérons maintenant l'expression

$$[n_\varepsilon[\{ S \}^N]]^{\frac{1}{N}}.$$

Lorsque N tend vers l'infini, cette expression tend vers une limite, indépendante de ε, qui n'est autre que la diversité du système S telle qu'elle a été définie plus haut.

1.4.3. Diversité d'une suite de systèmes liés en chaine. — Soit une suite de systèmes

$$S_1, \quad S_2, \quad \ldots, \quad S_n$$

pouvant chacun prendre deux états, que nous représenterons par les symboles o et 1. Nous admettons que ces systèmes sont liés en chaîne simple de Markov, c'est-à-dire que les probabilités p_{ij}, probabilité que S_{n+1} soit dans l'état j quand le système S_n est dans l'état i, sont bien définies, et indépendantes des états pris par les systèmes S_{n-1}, S_{n-2}, ... Nous supposerons de plus que ces probabilités sont indépendantes du rang n considéré, ou encore que la chaîne est à liaisons constantes.

Nous supposerons que les probabilités que S_1 soit dans les états o et 1 sont des nombres p_0 et p_1 tels que

$$\begin{aligned} p_0 &= p_{00}p_0 + p_{10}p_1 \\ p_1 &= p_{01}p_0 + p_{11}p_1 \end{aligned} \qquad (p_{00} + p_{01} = p_{10} + p_{11} = 1).$$

Dans cette hypothèse, on vérifie que les probabilités pour qu'un système S_i de rang quelconque soit dans les états o ou 1 sont également p_0 et p_1.

Le système $\sum_n$ constitué par le produit des n premiers systèmes S_i est un système susceptible de 2^n états. Sa diversité est μ_n, dont le logarithme est de la forme

$$H_n = \log \mu_n = -\sum P_n \log P_n,$$

en appelant P_n une des 2^n probabilités attachées aux 2^n états possibles. En séparant parmi les 2^n suites possibles celles qui se terminent par un zéro, et celles qui se terminent par un 1, nous décomposons H_n en une somme de deux termes :

$$H_n = H_n^{(0)} + H_n^{(1)} = -\sum P_n^{(0)} \log P_n^{(0)} - \sum P_n^{(1)} \log P_n^{(1)}.$$

Nous appelons $P_n^{(0)}$ la probabilité d'une suite quelconque de n états

dont le dernier est zéro, $P_n^{(1)}$ se définit d'une manière analogue; ces probabilités satisfont, quel que soit n, à

$$\sum P_n^{(0)} = p_0, \qquad \sum P_n^{(1)} = p_1.$$

Les probabilités $P_n^{(0)}$ ont forcément une des deux formes

$$P_{n-1}^{(0)} p_{00} \quad \text{ou} \quad P_{n-1}^{(1)} p_{10}$$

tandis que les probabilités $P_n^{(1)}$ ont une des deux formes

$$P_{n-1}^{(0)} p_{01} \quad \text{ou} \quad P_{n-1}^{(1)} p_{11}.$$

Nous en déduisons

$$\begin{aligned}
H_n^{(0)} &= p_{00} H_{n-1}^{(0)} + p_{10} H_{n-1}^{(1)} - p_0 p_{00} \log p_{00} - p_1 p_{10} \log p_{10}, \\
H_n^{(1)} &= p_{11} H_{n-1}^{(1)} + p_{01} H_{n-1}^{(0)} - p_1 p_{11} \log p_{11} - p_0 p_{01} \log p_{01}, \\
H_n &= H_{n-1} - p_0 (p_{00} \log p_{00} + p_{01} \log p_{01}) - p_1 (p_{10} \log p_{10} + p_{11} \log p_{11}),
\end{aligned}$$

et, par conséquent, l'expression générale de H_n, qui peut se mettre facilement sous la forme

$$H_n = (n-1) H_2 - (n-2) H_1.$$

Exemple. — Prenons un exemple numérique. Supposons que

$$p_{01} = p_{10} = \frac{2}{3}, \qquad p_{00} = p_{11} = \frac{1}{3};$$

d'où l'on déduit sans peine

$$p_0 = p_1 = \frac{1}{2}.$$

Nous obtenons

$$H_n = H_{n-1} - \left(\frac{1}{3} \log \frac{1}{3} + \frac{2}{3} \log \frac{2}{3} \right), \qquad H_1 = \log 2;$$

$$\mu_n = 2.3^{n-1}.2^{-\frac{2(n-1)}{3}} \sim (1,89)^n.$$

Les systèmes liés suivant la loi considérée ont donc une diversité moyenne de 1,89. Pour n très grand, on pourra considérer que le nombre d'états pratiquement possibles est non pas 2^n, mais par exemple $(1,9)^n$.

1.4.4. Relation entre la diversité et la fréquence d'apparition d'un événement dans une suite de tirages. — Nous allons voir quelle relation existe entre la diversité d'une suite de tirages et la fréquence d'appa-

rition d'un résultat dans cette suite. Examinons d'abord une suite de tirages indépendants, pouvant donner lieu chacun à deux résultats, avec probabilités p et q. Nous représenterons symboliquement par 1 et 0 les résultats de chaque tirage.

La diversité d'une suite de n tirages est donnée par la formule

$$\log\mu = -n(p\log p + q\log q).$$

Les suites contenant np fois le résultat 1 et nq fois le résultat zéro sont en nombre ν tel que (1)

$$\nu = \frac{n!}{np!\,nq!}.$$

On vérifie facilement, par application de la formule de Stirling, que

$$\lim_{n=\infty} \frac{1}{n}\log\nu = \log\mu,$$

de sorte que pour évaluer l'ordre de grandeur du nombre des suites pratiquement possibles, on peut se contenter de considérer les suites contenant exactement np fois un des résultats et nq fois l'autre (1).

Pour étendre la portée du résultat que nous venons de trouver, considérons le cas de la chaîne de Markov envisagée plus haut. Les probabilités des groupements

$$00 \quad 01 \quad 10 \quad 11$$

sont respectivement

$$(45) \qquad P_{00} = p_0 p_{00}, \qquad P_{01} = P_{10} = p_0 p_{01} = p_1 p_{10}, \qquad P_{11} = p_1 p_{11}.$$

Cherchons d'abord le nombre $N(\alpha, \beta, \gamma, \delta)$ de suites dans lesquelles les groupements 00, 01, 10, 11 se présentent exactement α, β, γ, δ fois respectivement. De telles suites contiennent $\alpha+\beta+\gamma+\delta+1$ termes; nous appellerons

$$N_{00}, \quad N_{01}, \quad N_{10}, \quad N_{11}$$

les nombres respectifs de ces suites dont les premiers termes sont 0, 0, 1, 1 et les derniers termes 0, 1, 0, 1 respectivement.

Nous avons évidemment

$$\begin{aligned}
N_{00}(\alpha,\beta,\gamma,\delta) &= N_{00}(\alpha-1,\beta,\gamma,\delta) + N_{01}(\alpha,\beta,\gamma-1,\delta),\\
N_{01}(\alpha,\beta,\gamma,\delta) &= N_{00}(\alpha,\beta-1,\gamma,\delta) + N_{01}(\alpha,\beta,\gamma,\delta-1),\\
N_{10}(\alpha,\beta,\gamma,\delta) &= N_{10}(\alpha-1,\beta,\gamma,\delta) + N_{11}(\alpha,\beta,\gamma-1,\delta),\\
N_{11}(\alpha,\beta,\gamma,\delta) &= N_{10}(\alpha,\beta-1,\gamma,\delta) + N_{11}(\alpha,\beta,\gamma,\delta-1).
\end{aligned}$$

(1) Si np et nq ne sont pas entiers, on leur substituera les entiers les plus voisins.

Si nous ne nous attachons qu'au dernier terme, et que nous désignions par N_0, N_1 les nombres de suites se terminant par o ou 1, nous obtenons

$$N_0(\alpha, \beta, \gamma, \delta) = N_0(\alpha - 1, \beta, \gamma, \delta) + N_1(\alpha, \beta, \gamma - 1, \delta),$$
$$N_1(\alpha, \beta, \gamma, \delta) = N_0(\alpha, \beta - 1, \gamma, \delta) + N_1(\alpha, \beta, \gamma, \delta - 1).$$

On déduit de ces deux dernières relations

$$N(\alpha, \beta, \gamma, \delta) = N(\alpha - 1, \beta, \gamma, \delta) + N(\alpha, \beta, \gamma, \delta - 1) + N(\alpha, \beta - 1, \gamma - 1, \delta) - N(\alpha - 1, \beta, \gamma, \delta - 1).$$

Une solution de cette équation est

$$N(\alpha, \beta, \gamma, \delta) = \frac{(\alpha + \beta)!\,(\gamma + \delta)!}{\alpha!\,\beta!\,\gamma!\,\delta!}.$$

Ce n'est pas là la valeur exacte de $N(\alpha, \beta, \gamma, \delta)$; mais on peut montrer que c'est une valeur asymptotique suffisamment approchée pour que le calcul de

$$\lim_{n=\infty} \frac{1}{n} \log N$$

pour

$$\alpha + \beta + \gamma + \delta = n - 1,$$
$$(\alpha \sim n P_{00},\ \gamma \sim n P_{10},\ \beta \sim n P_{01},\ \delta \sim n P_{11})$$

soit correct. La limite que l'on trouve est

$$H_2 - H_1,$$

H_2 et H_1 étant les valeurs définies au paragraphe 1.4.3. Nous voyons donc que, pour une chaîne de Markov comme pour le cas de Bernoulli, la diversité du système étudié peut se calculer en décomptant les suites des résultats pour lesquelles les répétitions des groupes de deux résultats consécutifs sont, à une unité près, le produit de la probabilité par le nombre d'épreuves.

Examinons le cas d'un tirage de Poisson; les épreuves sont indépendantes comme dans le cas de Bernoulli, mais la probabilité d'un résultat donné peut varier d'une épreuve à une autre. Soient

$$p_1,\ p_2,\ \ldots,\ p_n$$

la suite des probabilités associées; aux différents tirages symbolisant toujours par 1 et o les deux résultats possibles, nous savons que dans n épreuves la répétition du résultat 1 sera de la forme

$$\sum_{k=1}^{n} p_k + K\sqrt{n},$$

K étant un nombre d'ordre zéro par rapport à n considéré comme infiniment grand principal. Posons

$$r = \sum_{k=1}^{n} p_k.$$

Le nombre des suites de résultats possibles ne sera pas de l'ordre de $\frac{n!}{r!(n-r)!}$.

On montre en effet que la diversité du système à 2^n états possibles, correspondant aux tirages considérés est donnée, sous des conditions très générales, par la formule

$$\log \mu = -\sum_{k=1}^{n} (p_k \log p_k + q_k \log q_k),$$

c'est-à-dire que la probabilité accumulée sur les $\mu(1+\varepsilon)^n$ suites les plus probables tend vers 1 lorsque n tend vers l'infini. D'autre part, quand n tend vers l'infini, le rapport $\frac{\log \mu}{\log \nu}$ tend en général vers une limite inférieure à 1.

1.4.5. Utilité de la notion de probabilité moyenne. — Nous avons vu dans ce qui précède que pour décompter le nombre d'états pratiquement possibles pour un système, il était utile de recourir à la notion de probabilité moyenne, au sens de moyenne géométrique. Nous allons maintenant voir si cette notion de probabilité moyenne peut servir à d'autres buts.

Soit un système pouvant prendre n états, avec probabilités

$$p_1, \quad p_2, \quad \ldots, \quad p_n.$$

Deux des problèmes fondamentaux qui se posent à propos de ce système sont les suivants :

1° Avant tirage, quels sont les états que l'on peut exclure comme étant de probabilité négligeable ?

2° Après tirage, dans quelle mesure peut-on considérer que le résultat du tirage est satisfaisant, c'est-à-dire compatible avec l'hypothèse faite sur les valeurs numériques des p_i ?

La diversité μ du système joue évidemment un rôle dans ces deux questions; les états non exclus seront en nombre de l'ordre de μ, dans

la réponse à 1°; de même, les états dont l'apparition sera considérée comme satisfaisante, en réponse à 2°, seront en nombre égal à μ. Il semble naturel, si l'on fait une théorie unitaire, de penser à prendre les mêmes états pour former les deux ensembles :

1° États prévus.

2° États dont l'apparition confirme l'hypothèse faite sur les probabilités.

Si l'on se donne la probabilité $1-\varepsilon$ accumulée sur ces états, on peut, entre autres conventions, choisir pour ces états :

a. soit les états les plus probables;

b. soit les états dont la probabilité est la plus voisine de la probabilité moyenne $\frac{1}{\mu}$.

Chacune des deux conventions ci-dessus conduit à un nombre d'états, choisis, de l'ordre de μ. Pour le premier problème (prévision), il semble normal de choisir les états les plus probables; mais pour le deuxième problème, il apparaît préférable de choisir les états dont la probabilité est la plus voisine de $\frac{1}{\mu}$. Un exemple simple le prouve :

Soit un tirage à pile ou face, de 1000 coups, avec une pièce pour laquelle les probabilités de pile et face sont respectivement 0,499 et 0,501. Le cas le plus probable correspond à l'apparition de 1000 fois face; cela n'empêche que si ce cas se produit, l'observateur aura des doutes très forts sur l'hypothèse relative aux valeurs 0,499 et 0,501 pour les probabilités de pile et face. Cette situation paradoxale, de refuser vertu confirmatoire à un résultat qui est pourtant le plus probable, tient à ce que l'on ne peut s'empêcher de chercher à vérifier en *même temps* la valeur de la probabilité et l'indépendance des coups, et le fait de voir apparaître 1000 fois le même côté de la pièce semble un argument très fort *contre* l'indépendance. Un logicien pourrait soutenir que les 1000 tirages, dans le cas d'indépendance, peuvent prendre 2^{1000} formes différentes, et que puisque la plus probable est apparue, il y a lieu d'être satisfait. Cet argument est peut-être bon, mais il est certain qu'il ne convaincra personne. Il y a là un fait psychologique qu'il est bon de fouiller. Il est incontestable que ce que l'on reproche à ce résultat comportant 1000 faces, c'est d'être trop probable pour être bon. L'hypothèse est trop vérifiée pour que cette survérification n'ait pas

une cause plus forte que le simple excès de 0,501 sur 0,499, par exemple une liaison entre les événements, ou un truquage. Il est non moins incontestable que reprocher à un résultat d'être trop probable, si on ne lui reproche *que cela*, est d'une logique douteuse. Si nous réfléchissons plus avant, nous voyons que nous reprochons au résultat des 1000 faces, non seulement d'être le plus probable, mais encore d'être *le seul* à jouir de cette propriété, ce qui lui confère un caractère très exceptionnel. Une fois ce point établi, nous pouvons passer à un autre exemple moins direct, mais plus simple parce que ne faisant pas intervenir de tirages successifs ni par conséquent de considérations d'indépendance.

Considérons 91 événements incompatibles, dont un a la probabilité 0,1, les 90 autres ayant chacun pour probabilité 0,01. Nous appellerons E l'événement de probabilité 0,1 et F un quelconque des 90 autres événements. Supposons que nous fassions une expérience. L'hypothèse relative aux valeurs des probabilités sera-t-elle mieux vérifiée si c'est E qui apparaît, ou un des F ? Si l'on s'attache à la probabilité maximum, c'est évidemment E qui apporte meilleure confirmation. Mais si nous groupons tous les F en un événement $\mathcal{F}$, il est évident que c'est l'apparition de $\mathcal{F}$ qui apporte une confirmation supérieure à celle apportée par l'apparition de E. L'apparition d'*un* F, étant donnée la structure du système des 91 probabilités supposées, doit jouir du même privilège que l'apparition de $\mathcal{F}$; en d'autres termes : on ne peut pas reprocher à un F de n'avoir que 0,01 comme probabilité, parce qu'ils sont 90 à avoir cette même probabilité. Si l'on adopte cette seconde attitude, on est amené à considérer comme confirmant l'hypothèse l'apparition d'un événement ayant une probabilité voisine de la probabilité moyenne, laquelle est ici très voisine de 0,01.

Dans le cas général, on est conduit à appliquer la règle suivante :

Si une hypothèse attribue à n éventualités incompatibles les probabilités

$$p_1, \quad p_2, \quad \cdots, \quad p_n,$$

on considèrera que l'apparition de l'éventualité de rang i (et par conséquent de probabilité p_i) apporte à cette hypothèse une confirmation dont l'évaluation numérique est

$$k_i = \sum p_i (\log \mu p_i)^2 - (\log \mu p_i)^2 \qquad \left(\log \mu = -\sum p_i \log p_i\right).$$

La valeur moyenne de la confirmation à attendre est

$$\sum p_i k_i = 0.$$

Avec cette règle, on voit que ce sont les éventualités de probabilité la plus voisine de $\frac{1}{\mu}$ qui apportent la meilleure vérification.

CHAPITRE II.

THÉORIE DES JEUX.

2.1. Habitude et habileté d'un joueur, d'après E. Borel. Égalité de von Neumann.

2.1.1. Description probabiliste d'une habileté. — La théorie des jeux s'est longtemps limitée aux jeux de hasard pur, tels le jeu de dés. A cette théorie des jeux de hasard pur se rattachent les calculs conduisant à l'évaluation des probabilités afférentes aux diverses compositions d'une donne, à la répartition des cartes entre les différentes mains. De cette théorie relèvent également les calculs des partis (au sens introduit par Pascal), les considérations sur la ruine des joueurs, etc. Ce sont là des questions classiques sur lesquelles nous ne reviendrons pas ici. Ces questions sont justiciables le plus souvent de l'analyse combinatoire.

La théorie des jeux de hasard pur est évidemment un simple chapitre de la théorie générale des jeux, puisque dans la plupart des jeux interviennent des *tactiques* variant avec les habitudes et l'habileté des joueurs. On peut se demander si les éléments « habitude » et « habileté » sont descriptibles en termes de probabilité. En ce qui concerne l'habitude, nous avons là une notion qui se prête très bien au calcul des probabilités; en ce qui concerne l'habileté, cela est moins évident. Le problème a été clairement posé, pour la première fois, par M. Émile Borel.

L'argumentation qui rattache l'habileté à des notions de probabilité est la suivante :

Supposons que deux joueurs A et B jouent à un jeu de cartes quelconque, et que ce soit à A à prendre la première initiative (jouer une carte, faire une déclaration, miser, etc.). Avant que A prenne cette

initiative, B peut se faire une idée des différentes probabilités attachées aux jeux que A peut avoir en mains. C'est une question de probabilités classiques. Après que A a pris une initiative, B, s'il connaissait les habitudes de A, pourrait modifier en conséquence ses estimations de probabilités. En effet, si A est bon joueur, il n'a pas pris cette initiative gratuitement; s'il pose une carte, ce n'est pas n'importe quelle carte de son jeu. B se trouve donc en face d'un problème de probabilités *a posteriori*, qu'il saurait résoudre, toujours d'après des méthodes classiques, s'il connaissait les probabilités *a priori* attachées aux réactions de A, c'est-à-dire, plus simplement, les habitudes de A. Les habitudes de A peuvent se représenter par un certain nombre de probabilités $p'_A, p''_A, \ldots$; celles de B par des probabilités $p'_B, p''_B, \ldots$. Pour un spectateur qui connaîtrait ces probabilités, le comportement de A et B serait assimilable au comportement de deux robots équipés chacun d'un dispositif interne de tirage au sort. Ce tirage au sort est celui qui est apparent lorsqu'un joueur hésite avant de prendre une décision. Pour ce spectateur, l'espérance mathématique de A serait une fonction déterminée des p_A, p_B, et des probabilités P attachées aux distributions de cartes : soit

$$G_A = F(p'_A, p''_A, \ldots; p'_B, p''_B, \ldots; P', P'', \ldots). \tag{1}$$

L'*habileté* de A consistera à choisir des p_A rendant G_A maximum; celle de B consistera à choisir les p_B de manière à rendre G_A minimum.

M. Émile Borel a précisé davantage encore cette notion d'habileté. Supposons que A connaisse les p_B; il pourra donner à G_A la valeur

$$\operatorname*{Max}_{p_A} F = I(p_B) \tag{2}$$

en choisissant les p_A de manière à rendre F maximum, les p_B restant constants. Ce maximum I est fonction des p_B. Pour se défendre contre cette maximisation de G_A, B, à supposer qu'il ait pu calculer I, pourra rendre I minimum. Soit I_0 ce minimum :

$$I_0 = \operatorname*{Min}_{p_B} I(p_B). \tag{3}$$

Nous voyons que A, s'il est en face d'un joueur B infiniment habile, pourra être mis par ce joueur dans une position telle que, *quelle que soit son attitude à lui*, A, il ait son espérance bornée par l'inégalité

$$G_A \leqq I_0. \tag{4}$$

L'argumentation peut se renverser. Si B connaît les p_A, il pourra donner à G_A la valeur

$$\operatorname*{Min}_{p_B} F = J(p_A).$$

A, à supposer qu'il connaisse $J(p_A)$, pourra choisir les p_A tels que $J(p_A)$ soit maximum : soit

$$J_0 = \operatorname*{Max}_{p_A} J(p_A)$$

ce maximum. A pourra donc, et cela *quoi que fasse* B, s'assurer une espérance minimum J_0, et par conséquent assurer

$$G_A \geqq J_0. \tag{5}$$

Si A agit de manière à assurer (5), et B de manière à assurer (4) le gain de A sera compris entre I_0 et J_0 :

$$J_0 \leqq G_A \leqq I_0. \tag{6}$$

L'intervalle (J_0, I_0) est donc un intervalle où l'habileté des joueurs *peut* amener G_A. Si $G_A < J_0$, A est un maladroit. Si $G_A > I_0$, c'est B qui est maladroit.

Il restait, dans la théorie de M. Borel, une incertitude sur l'intervalle (J_0, I_0), et sur la qualification des joueurs selon l'endroit où G_A se plaçait dans cet intervalle. Les exemples schématiques traités par M. Borel montraient toujours que $J_0 = I_0$, mais cela n'était pas suffisant pour en conclure que $I_0 = J_0$. La démonstration de cette égalité est due à M. von Neumann.

Cette démonstration a été simplifiée par la suite. Nous reviendrons plus loin sur la portée de la démonstration, et allons appliquer ce théorème à un jeu de tactique pure.

2.1.2. Vérification de l'égalité de von Neumann sur un jeu de tactique pure. — Considérons deux joueurs d'échecs, A et B. Nous supposons que A a les blancs, et joue le premier. A a à sa disposition un certain nombre, très grand, de tactiques :

$$a_1, \quad a_2, \quad \ldots, \quad a_N;$$

B a à sa disposition des tactiques :

$$b_1, \quad b_2, \quad \ldots, \quad b_M.$$

Nous entendons par tactique une règle de jeu très stricte, qui ne laisse aucune initiative au joueur qui l'emploie, et qui lui dicte sa conduite dans toutes les circonstances créées par l'adversaire. La seule initiative laissée à A et B est donc de choisir, avant l'engagement de la partie, une tactique. Ensuite, ils sont censés s'y tenir. Théoriquement, ce choix préliminaire et unique est équivalent à la suite des choix tels qu'ils se pratiquent en réalité en cours de partie. Supposons que le nombre de coups à jouer soit limité; si la tactique a_i de A s'oppose à la tactique b_j de B, le résultat final sera soit le gain de A, le gain de B, soit une partie nulle. Nous représenterons ce résultat par α_{ij}, avec la convention que $\alpha_{ij} = 1, -1$ ou 0 représentent respectivement les trois issues envisagées.

Si A choisit les tactiques a_i avec les probabilités p_i, et B les tactiques b_j avec les probabilités q_j, la probabilité, pour A, de gagner la partie, de perdre, ou d'aboutir à une partie nulle, sera, selon le cas :

$$\sum \alpha_{ij}^2 \frac{1+\alpha_{ij}}{2} p_i q_j, \quad \sum \alpha_{ij}^2 \frac{1-\alpha_{ij}}{2} p_i q_j, \quad \sum (1-\alpha_{ij}^2) p_i q_j.$$

Si A cherche à maximiser la différence entre la probabilité de gagner et la probabilité de perdre, il cherchera à rendre maximum

$$G_A = \sum \alpha_{ij} p_i q_j,$$

B de son côté cherchera à rendre cette même expression minimum. Nous nous trouvons formellement dans un cas particulier du jeu envisagé à propos de la théorie de M. Émile Borel. Nous allons vérifier sur ce cas particulier que

$$\operatorname*{Max}_{p} \operatorname*{Min}_{q} G_A = \operatorname*{Min}_{q} \operatorname*{Max}_{p} G_A.$$

Nous constaterons que pour ce cas particulier, la valeur commune des expressions ci-dessus ne peut être que $1, -1$ ou 0, c'est-à-dire que théoriquement, entre joueurs suffisamment habiles la partie est jouée d'avance. A est forcément gagnant s'il est assez adroit, ou forcément perdant, si B est suffisamment adroit, à moins que A ou B ne puissent se réfugier dans la partie nulle. Nous nous hâtons de dire que l'on a pu démontrer que les seules valeurs possibles étaient $1, -1$ ou 0, mais que la détermination de celle de ces valeurs qui apparaît effectivement est encore au-dessus des possibilités actuelles du calcul.

Pour conduire la démonstration, nous emploierons les notations suivantes :

X_{2i} représentera le coup joué par A au $2^{ième}$ coup. Y_{2i+1} représentera le coup joué par B au $2i+1^{ième}$ coup. Les X et les Y peuvent être des nombres entiers, si l'on convient d'un code représentant en nombre les différents coups possibles. Un exemple de chiffrement est le suivant :

X est un nombre pouvant aller de o à $2^{12}-1$,

X peut se mettre sous la forme

$$X = 2^6 x' + x'',$$

x' et x'' étant deux nombres pouvant aller de zéro à 2^6-1. On interprètera X comme le coup consistant à prendre la pièce située dans la case numérotée x' pour la mettre dans la case numérotée x''.

Une partie de 2M coups sera représentée par une suite

$$(7) \qquad X_0,\ Y_1,\ X_2,\ Y_3,\ \ldots,\ X_{2M-2},\ Y_{2M-1}.$$

Inversement, toute suite de la forme (7) représentera une partie, si nous convenons que le premier nombre de cette suite correspondant à un coup impossible (parce que la première case est vide, ou que le coup contredit les règles) a pour effet, s'il existe de faire perdre la partie à celui qui s'est rendu coupable de cette impossibilité, et que les coups suivants ne sont pas joués.

Dans ces conditions, à toute suite (7) nous pouvons associer un nombre égal à 1, -1 ou o selon que la partie est acquise à A, à B, ou est nulle. Nous définissons ainsi une fonction à 2M arguments entiers

$$(8) \qquad F_{2M}(X_0, Y_1, \ldots, Y_{2M-1})$$

qui ne peut prendre que les valeurs 1, -1 ou o.

Considérons maintenant des fonctions à un nombre moindre d'arguments, de la forme $F_n(X_0, Y_1, \ldots)$ ne dépendant, en ce qui concerne les fonctions d'indice n, que des n premiers termes de la suite (7). Astreignons ces fonctions aux relations de récurrence :

$$(9) \qquad \begin{cases} F_{2n}\ (X_0, Y_1, \ldots, Y_{2n-1}) = \operatorname{Max}_{X_{2n}} F_{2n+1}(X_0, Y_1, \ldots, Y_{2n-1}, X_{2n}), \\ F_{2n+1}(X_0, Y_1, \ldots, X_{2n}) = \operatorname{Min}_{Y_{2n+1}} F_{2n+2}(X_0, Y_1, \ldots, Y_{2n+1}). \end{cases}$$

Pour obtenir la valeur de F_{2n}, on considère la fonction F_{2n+1} sup-

posée déjà formée. On donne aux $2n$ premiers arguments de F_{2n+1} les valeurs pour lesquelles on veut calculer F_{2n}. F_{2n+1} dépend encore d'un $(2n+1)^{\text{ième}}$ argument X_{2n}. On choisit pour valeur de F_{2n} la plus grande des valeurs que prend F_{2n+1}, quand on fait varier le $(2n+1)^{\text{ième}}$ argument, les $2n$ premiers restant fixes.

Les fonctions F_{2n+1} sont déterminées d'une manière analogue. Nous poserons de plus

$$F_0 = \operatorname*{Max}_{X_0} F_1(X_0). \tag{10}$$

F_{2M} étant une fonction bien déterminée, nous pouvons en déduire successivement F_{2M-1}, F_{2M-2} et nous parviendrons enfin jusqu'à F_0 qui est un nombre. F_{2M} ne pouvant prendre que les valeurs 1, —1 ou 0, il en sera de même de toutes les fonctions de la suite

$$F_0, \quad F_1(X_0), \quad F_2(X_0, Y_1), \quad \ldots, \quad F_{2M}(X_0, \ldots, Y_{2M-1}). \tag{11}$$

Ces fonctions jouissent de la propriété suivante :

A *peut* agir de manière que la suite (11) ne soit pas décroissante.
B *peut* agir de manière que la suite (11) ne soit pas croissante.

En effet, A a la disposition des X, et B celle des Y. Si dans la suite (11) on est arrivé à la fonction F_{2n}, la première des égalités (9) montre que A, dont c'est le tour de jouer, peut choisir X_{2n} tel que

$$F_{2n+1} = F_{2n}.$$

Donc A peut maintenir la valeur de F_{2n}, mais ne peut l'augmenter. Si A se trompe, la valeur de F_{2n+1}, devient inférieure à celle de F_{2n}. On voit de même que dans le passage de F_{2n+1} à F_{2n+2}, B peut maintenir la valeur de F_{2n+1}, sans pouvoir la diminuer.

Ceci étant montré, il en résulte que si $F_0 = 1$, A peut jouer de manière à gagner la partie, en empêchant la suite (11) de décroître. Si $F_0 = -1$, B peut gagner la partie, en empêchant la suite de croître. Enfin, si $F_0 = 0$, A et B peuvent jouer de manière que la partie soit nulle. D'une manière plus précise : A peut jouer de manière que la partie soit nulle ou gagnée par lui, et B peut jouer de manière que la partie soit nulle ou gagnée par lui. Si A et B sont infiniment habiles, la partie sera donc finalement nulle. En revenant au théorème général,

nous voyons qu'il se vérifie bien pour le cas du jeu d'échecs que l'on a

$$\underset{p}{\text{Max}}\ \underset{q}{\text{Min}}\, G_A = \underset{q}{\text{Min}}\ \underset{p}{\text{Max}}\, G_A$$

et que l'on peut obtenir cette égalité sans procéder à un tirage au sort, c'est-à-dire avec des p tous nuls sauf un, et des q tous nuls sauf un.

2.1.3. Objection a la définition du joueur infiniment habile. Intervention du nombre de coups de la partie. — Nous allons voir maintenant que le résultat précis établi ci-dessus ne suffit pas à faire une théorie exhaustive du jeu d'échecs. Supposons en effet que nous ayons construit une machine qui puisse calculer toutes les fonctions F, et qui permette par conséquent de construire un automate joueur d'échecs, en théorie infiniment habile. Supposons de plus que la partie soit théoriquement gagnée d'avance par les blancs. Comme notre automate est infiniment habile, il doit pouvoir jouer correctement en prenant la position défavorable des noirs, et, bien qu'il soit théoriquement perdant, l'énorme avantage qu'il a de connaître entièrement la théorie du jeu doit lui permettre de vaincre tous les joueurs « ordinaires » qu'on peut lui opposer. Ceci étant posé, nous voyons tout de suite que la théorie que nous avons étudiée ne nous permet pas de spécifier sans ambiguïté la conduite de cet automate. En effet, la règle de jeu pour l'automate consiste, puisqu'il a la position B, à ne jamais faire d'opération qui augmente un terme de la suite (11). Si donc, tout au moins dans les premiers coups, A est parvenu à se maintenir sur la valeur 1, valeur maximum, B n'a absolument aucune règle à suivre. Il peut jouer n'importe quoi. Il n'y a rien de paradoxal là dedans. En effet, les fonctions F correspondent à l'hypothèse que A a en face de lui un joueur infiniment habile. Si donc dès le premier coup, B est déjà perdant, il ne se défend même pas. Cela est absurde, puisque tout joueur réel finira par faire une faute ; B doit jouer de manière à ne pas compromettre ses chances lorsque cette faute se produira. Il existe un moyen simple, mais un peu artificiel, de rétablir la situation. On observe d'abord que la définition des fonctions F subsiste même si F_{2M} a un champ de variation plus étendu que $+1$, -1 ou 0. De même subsiste la propriété fondamentale pour la suite (11) de pouvoir être maintenue non décroissante par A, non croissante par B. La seule modification qui intervient est que les fonctions F ont alors un champ de variation plus grand.

Au lieu d'attribuer à F_{2M} les valeurs $+1$, -1 ou 0, nous attribuerons à F_{2M} la valeur $2M-n$ si A a gagné au bout de n coups, et $n-2M-1$ si B a gagné au bout de n coups. Si A a gagné au $(2M-1)^{ième}$ coup, F_{2M} est bien égal à 1 ; si B a gagné au $2M^{ième}$ coup, F_{2M} est égal à -1. Si la partie est nulle, F_{2M} reste égale à zéro. Nous établissons ainsi une hiérarchie entre les différentes valeurs de F_{2M}, en attribuant une valeur plus grande aux parties gagnées rapidement.

Si donc $F_0 = 2M-\mu$, ce qui signifie que A peut gagner en μ coups au plus, et que $F_1 = 2M-\mu$, ce qui signifie que A n'a pas fait de faute au premier coup, B pourra maintenir F_2 à la valeur $2M-\mu$, ce qui entraîne que la défense de B est correcte, puisqu'il recule sa défaite théorique jusqu'à la dernière limite. Au cours du jeu, A fera certainement des fautes ayant pour effet d'augmenter μ, c'est-à-dire de diminuer les F, et B aura toujours, en suivant la règle dictée par les égalités (9), une réaction « habile ».

Nous voyons que l'étude d'un jeu, même théoriquement complètement déterminé comme le jeu d'échecs, laisse un certain arbitraire, qui s'est manifesté ici par la pondération choisie pour les valeurs de F_{2M}.

2.1.4. Démonstration de l'égalité de von Neumann. — Nous allons maintenant revenir au théorème de von Neumann. Si deux joueurs A et B peuvent choisir des tactiques :

$$\begin{array}{llll} t_1, & t_2, & \ldots, & t_n \quad \text{pour A ;} \\ t'_1, & t'_2, & \ldots, & t'_m \quad \text{» B ;} \end{array}$$

et si la tactique t_i opposée à la tactique t'_j asssure à A l'espérance α_{ij}, l'espérance de A, si les tactiques t_i sont choisies par lui avec les probabilités p_i et les tactiques t'_j par B avec les probabilités q_j, sera

$$K(p, q) = \sum \alpha_{ij} p_i q_j. \tag{12}$$

B peut astreindre $K(p, q)$ à l'inégalité

$$K(p, q) \leq \operatorname*{Max}_{p} \operatorname*{Min}_{q} K(p, q) = I_0.$$

Cela signifie qu'il existe un système de nombres $\bar{q}_j$ tels que

$$\sum \alpha_{ij} p_i \bar{q}_j \leq I_0, \tag{13}$$

quels que soient les p_i. Posons

$$\beta_i = \sum \alpha_{ij} q_j. \tag{14}$$

Les m formes β_i, aux variables q, sont telles que, quelles que soient les valeurs attribuées aux q_j, il existe toujours au moins une des formes qui satisfait l'inégalité

$$\beta_i \geqq I_0. \tag{15}$$

S'il n'en était pas ainsi, on pourrait déterminer un nombre ε et un système de valeurs des q_j, soient $\overline{\overline{q}}_j$, tels que, pour ces valeurs des q_j on ait, quel que soit i :

$$\overline{\overline{\beta}}_i \leqq I_0 - \varepsilon.$$

Si B adopte ce système de valeurs des q_j, il pourra astreindre K à l'inégalité

$$K(p, q) \leqq I_0 - \varepsilon,$$

ce qui est en contradiction avec la définition de I_0.

Revenons maintenant aux inégalités (15). On peut montrer que si, quels que soient les q_j, une au moins des formes B satisfait à (15), il existe une combinaison linéaire des B_i :

$$H = \sum \lambda_i \beta_i \qquad \left(\lambda_i \geqq 0, \sum \lambda_i = 1\right),$$

telle que, quels que soient les q_i, on ait

$$H \geqq I_0. \tag{16}$$

Supposons que A adopte des valeurs des λ_i pour les p_i. Nous voyons que A pourra astreindre K à l'inégalité

$$K \geqq I_0$$

et cela quoi que fasse B. Il en résulte évidemment que

$$\operatorname*{Max}_p \operatorname*{Min}_q K(p, q) = \operatorname*{Min}_p \operatorname*{Max}_q K(p, q) = I_0. \tag{17}$$

2.2. **Calcul des probabilités à attribuer aux tactiques.** — Appelons $\overline{p}_i$ un système des valeurs des p_i permettant à A d'assurer l'inégalité

$$K(p, q) \geqq I_0.$$

Un tel système de valeurs est tel que la forme linéaire, en q_j, $K(\bar{p}, q)$, satisfait à

$$K(\bar{p}, q) \geqq I_0. \tag{18}$$

Nous savons par ailleurs que B peut choisir les $\bar{q}_j$ de manière que la forme $K(p, \bar{q})$ aux variables p_i satisfasse à

$$K(p, \bar{q}) \leqq I_0.$$

Nous pouvons donc compléter l'inégalité (18) par l'égalité

$$K(\bar{p}, \bar{q}) = I_0, \tag{19}$$

I_0 est donc le minimum, *effectivement atteint*, de la forme $K(\bar{p}, q)$ aux variables q. Or, une forme linéaire, quand les variables qui y figurent varient dans un domaine, ne peut atteindre son minimum que dans des conditions très particulières. Considérons en effet le système des $\bar{q}_j$. Parmi ces $\bar{q}_j$, certains peuvent être nuls, d'autres sont différents de zéro. Donnons à ceux des $\bar{q}_j$ qui sont différents de zéro des accroissements infiniment petits, de somme nulle. Nous aurons

$$\delta K(\bar{p}, \bar{q}) = \sum_{ij} \alpha_{ij} \bar{p}_i \delta \bar{q}_j \geqq 0 \qquad \left(\sum \delta \bar{q}_j = 0\right) \tag{20}$$

comme il résulte de (18) et (19). Les $\delta\bar{q}_j$ sont complètement arbitraires à part la condition d'être de somme nulle, et d'être infiniment petits. Nous pouvons les prendre par exemple de la forme

$$\delta \bar{q}_j = \bar{q}_j \left[\sum_i \alpha_{ij} \bar{p}_i - K(\bar{p}, \bar{q}) \right] \delta t. \tag{21}$$

Cette forme assure que seuls les $\bar{q}_j$ positifs ont un accroissement, et que la somme des accroissements est nulle.

Nous déduisons de (20) et (21) :

$$\delta K = \frac{1}{\delta t} \sum \frac{(\delta \bar{q}_j)^2}{\bar{q}_j} \geqq 0. \tag{22}$$

Nous en déduisons, puisque le signe de δt est arbitraire, que les $\delta\bar{q}_j$ définis par (21) sont tous nuls, c'est-à-dire que, quel que soit j, nous avons

$$\bar{q}_j \left[\sum_i \alpha_{ij} \bar{p}_i - K(\bar{p}, \bar{q}) \right] = 0. \tag{23}$$

2.2.1. Élimination des tactiques erronnées. — La forme des équations (23) montre qu'il est important de déterminer les $\bar{q}_j$ nuls. Il y a en général plusieurs systèmes de $\bar{q}_j$ possibles. Nous appellerons tactique erronnée pour B toute tactique t'_j telle que le $\bar{q}_j$ correspondant soit nul quelle que soit la manière dont les $\bar{q}_j$ sont déterminés de manière à assurer

$$K(p, \bar{q}) \leqq I_0.$$

Remarquons que s'il existe deux systèmes distincts de $\bar{q}_j$, $\bar{q}'$ et $\bar{q}''$, le système $\lambda\bar{q}' + \mu\bar{q}''$ constitué par les valeurs $\lambda\bar{q}'_j + \mu\bar{q}''_j$, avec

$$\lambda \geqq 0, \qquad \mu \geqq 0, \qquad \lambda + \mu = 1$$

est un système possible de $\bar{q}_j$, puisque

$$K(p, \lambda q' + \mu\bar{q}'') = \lambda K(p, \bar{q}') + \mu K(\bar{p}, q'') \leqq I_0.$$

Si λ et μ sont positifs, seuls sont nuls dans $\lambda\bar{q}' + \mu\bar{q}''$ les $\bar{q}_j$ qui sont nuls à la fois dans $\bar{q}'$ et $\bar{q}''$. Nous voyons ainsi que si l'on sait déterminer les tactiques erronnées de B, on sait qu'il existe un système de $\bar{q}_j$ tel que tous les $\bar{q}_j$ correspondant à des tactiques non erronnées soient différents de zéro. Ce système $\bar{q}_j$ n'est pas forcément unique, mais il est le plus étendu possible, c'est-à-dire que le nombre des $\bar{q}_j$ différents de zéro y est maximum. Nous voyons également que si deux systèmes sont le plus étendus possibles, ce sont les $\bar{q}_j$ de mêmes indices qui sont nuls dons les deux systèmes, à savoir ceux correspondant aux tactiques erronnées. Revenons à (23). Nous voyons que ces équations peuvent être relatives au système de $\bar{q}_j$ le plus étendu possible, c'est-à-dire qu'il peut s'écrire

$$\sum_i \alpha_{ij} p_i = H \tag{24}$$

pour tous les j correspondant à des tactiques non erronées (de B) ; pour les tactiques erronnées nous aurons

$$\sum_i \alpha_{ij} q_i > H. \tag{25}$$

Nous déduisons immédiatement de (24) que

$$H = K(\bar{p}, \bar{q}) = I_0.$$

Nous constatons que la recherche des $\bar{p}_i$ et des $\bar{q}_j$ peut se décomposer

en deux stades. Le premier stade consiste à déterminer les fausses tactiques, qui sont entachées de fautes de jeu proprement dites. Ceci revient à déterminer les $\bar{p}_i$ et les $\bar{q}_i$ qui sont nécessairement nuls. Le deuxième stade consiste à déterminer les valeurs de $\bar{p}_i$ et $\bar{q}_i$. Ceci constitue la partie psychologique de l'étude. Supposons maintenant que nous ayons affaire à des joueurs connaissant parfaitement la technique du jeu, c'est-à-dire qui n'emploient aucune fausse tactique. La lutte entre les deux joueurs se place sur le plan psychologique. Les équations déterminant les $\bar{p}_i$ et $\bar{q}_j$ sont alors, puisque les fausses tactiques sont éliminées :

$$(26)\qquad \begin{cases} \sum_i \alpha_{ij} p_i = \mathrm{H}' & \text{quel que soit } j, \\ \sum \alpha_{ij} q_j = \mathrm{H}'' & \text{» » » } i. \end{cases}$$

On constate facilement que les équations (26) ont pour conséquence

$$(27)\qquad \mathrm{H}' = \mathrm{H}'' = \mathrm{I}_0,$$

ce qui entraîne que ces équations, considérées comme des équations p_i, q_j, H′, H″, ont une solution unique en H′ et H″, avec de plus l'égalité (27). Les équations de la première ligne dans (26) ont donc, à elles seules, une solution unique en H′. Ceci montre que dans l'hypothèse où les fausses tactiques sont éliminées, le joueur A, pour s'assurer le gain I_0, doit résoudre le système constitué par la première ligne (26). Supposons qu'il l'ait fait, et ait déterminé les $\bar{p}_i$ correspondants. Nous voyons alors que

$$\mathrm{K}(\bar{p}, q) = \mathrm{I}_0$$

quelles que soient les probabilités q_i. Nous constatons qu'une fois le jeu ramené à une lutte purement psychologique, le fait pour un joueur de se prémunir contre une perte par un choix convenable de son attitude psychologique lui interdit par cela même de tirer parti des fautes psychologiques possibles de son adversaire. Il ne pourra tirer parti que de ses fautes tactiques.

2.2.2. Détermination des probabilités. — 2.2.2.1. *Méthode discontinue.* — Le problème de la détermination des $\bar{p}_i$ se ramène à celui de la résolution des inégalités

$$(28)\qquad h_j = \sum_i \alpha_{ij} p_i \geqq \mathrm{H},$$

H étant le plus grand possible. Si les α_{ij} sont donnés par leurs valeurs numériques, on peut résoudre le système (28) de la manière suivante. On commence par l'écrire

$$h_j - \mathrm{H} \geqq 0, \qquad p_i \geqq 0. \tag{29}$$

Si dans les inégalités (29) on remplace p_n par $1 - p_1 - \ldots - p_{n-1}$, on obtient de nouvelles inégalités; si l'on met en évidence dans ces inégalités qui contiennent p_1 cette valeur p_1, on obtient des inégalités ayant une des trois formes

$$p_1 \geqq k, \qquad p_1 \leqq l, \qquad m \geqq 0\,; \tag{30}$$

k, l, m ne dépendant pas de p_1. Du système (30), on passe aux inégalités de la forme

$$l \geqq k, \qquad m \geqq 0, \tag{31}$$

étant entendu qu'il y a autant d'inégalités $l \geqq k$ dans (31) qu'il y a des couples l, k dans (30). Dans (31) l'inconnue p_1 a été éliminée. On élimine ensuite $p_2, p_3, \ldots, p_n$; il reste en fin de compte une condition pour H, de la forme

$$\mathrm{A} < \mathrm{H} \leqq \mathrm{B},$$

A et B ne dépendant que des α_{ij}. Nous savons *a priori* que $\mathrm{A} = -\infty$; en effet, en prenant H suffisamment grand négatif, on est certain que les inégalités (28) ont une solution. Nous arrivons enfin à une condition de la forme

$$\mathrm{H} \leqq \mathrm{B}(\alpha).$$

Comme le joueur A cherche à rendre H maximum, il adoptera

$$\mathrm{H} = \mathrm{B}(\alpha). \tag{32}$$

H étant ainsi déterminé, les inégalités obtenues au cours de l'élimination de $p_1, p_2, \ldots, p_n$ donnent des limites possibles pour les p_i.

Ce processus n'est évidemment applicable qu'en théorie, parce qu'il fait intervenir un nombre énorme d'inégalités.

2.2.2.2. — *Méthode utilisant un système d'équations différentielles.* — M. von Neumann a imaginé un procédé de détermination des p_i faisant appel à des équations différentielles; ce procédé suppose le jeu symétrique, ce qui se traduit par

$$\alpha_{ij} + \alpha_{ji} = 0. \tag{33}$$

Posons

$$\varphi_j = \frac{1}{2}\{|h_j| - h_j\} \tag{34}$$

et faisons varier les p_i, en fonction d'un paramètre t, de manière à satisfaire à

$$\frac{dp_i}{dt} = \varphi_i - p_i\varphi, \qquad \left\{\varphi = \sum \varphi_i\right\}. \tag{35}$$

Les équations (35) ont pour conséquence

$$\frac{d}{dt}\sum p_i = 0.$$

Si nous partons de conditions initiales telles que

$$p_i \geqq 0, \qquad \sum p_i = 1,$$

ces conditions seront respectées au cours de la variation. En ce qui concerne la somme des p_i, cela est évident. En ce qui concerne le signe des p_i, on voit que si $p_i = 0$ l'équation (35) donne

$$\frac{dp_i}{dt} \geqq 0,$$

ce qui montre que p_i ne peut traverser la valeur zéro.

Posons maintenant

$$\Phi = \frac{1}{2}\sum \varphi_i^2. \tag{36}$$

Un calcul simple montre que, tenu compte de (33) nous avons

$$\frac{d\Phi}{dt} = -\sum_{ij} \alpha_{ij}\varphi_j\left(\frac{dp_i}{dt} - \varphi_i\right) = +\varphi\sum \alpha_{ij}p_i\varphi_j,$$

$$\frac{d\Phi}{dt} = \varphi\sum h_j\varphi_j = -\varphi\sum |h_j|\varphi_j \leqq 0. \tag{37}$$

Pour établir les relations (37), nous avons tenu compte du fait que

$$\varphi_j \geqq 0, \qquad [h_j + |h_j|]\varphi_j = 0.$$

Nous pouvons déduire de (37)

$$\frac{d\Phi}{dt} = \frac{1}{2}\varphi\sum{}'\varphi_j(h_j - |h_j|) = -\varphi\Phi, \tag{38}$$

d'où il résulte que Φ tend vers zéro quand t tend vers l'infini. Tous les φ_j

tendent vers zéro, ce qui n'est possible que si toutes les formes h_j tendent vers des limites non négatives. On peut déduire du comportement, à la limite, des p_i, un système de $\bar{p}_i$ tels que

$$\sum \alpha_{ij}\bar{p}_i \geqq 0 \qquad \text{quel que soit } j.$$

Comme par raison de symétrie nous avons ici $I_0 = 0$, nous avons bien résolu le problème de la recherche des $\bar{p}_i$.

2.2.2.3. *Passage d'un jeu asymétrique à un jeu symétrique.* — Considérons d'abord le jeu dit de la pierre, du papier et du ciseau (nous dirons brièvement jeu P. P. C.) qui peut se ramener au schéma suivant :

Le joueur A peut prendre une des attitudes α, β, γ. Le joueur B peut prendre une des attitudes α, β, γ (les mêmes). Si A et B prennent la même attitude, la partie est nulle. Sinon, les règles sont :

α gagne β et rapporte c ;
β » γ » a ;
γ » α » b.

Dans le jeu classique P. P. C. α est le ciseau, β le papier, et γ la pierre. Le ciseau coupe le papier, qui couvre la pierre, qui casse le ciseau. a, b, c ont même valeur. Dans le cas général, si A choisit α, β, γ avec les probabilités p, q, r, ses espérances mathématiques si B choisit α, β, γ sont respectivement

$$rb - qc, \qquad pc - ra, \qquad qa - pb.$$

Les inégalités

$$rb - qc \geqq H, \qquad pc - ra \geqq H, \qquad qa - pb \geqq H$$

ont pour conséquence

$$H(p + q + r) = H \leqq 0.$$

La valeur maximum de H est nulle, ce qui conduit à

$$p = \frac{a}{a+b+c}, \qquad q = \frac{b}{a+b+c}, \qquad r = \frac{c}{a+b+c}.$$

On peut, en faisant appel à la théorie du jeu P. P. C. ramener tout jeu asymétrique à un jeu symétrique. Soit en effet un jeu J asymétrique où le gain d'un joueur est

$$K(p, q) = \sum \alpha_{ij} p_i q_j$$

s'il adopte les probabilités p_i, son adversaire adoptant les probabilités q_j. Nous dirons que le premier joueur a la position P, le second joueur la position Q. Soit maintenant une somme S plus grande que la valeur absolue d'un quelconque des α_{ij}. Supposons que deux joueurs A et B jouent au jeu suivant : A et B choisissent les positions α, β, γ. Les conventions sont les suivantes :

β gagne α et rapporte S ;
γ » α » S ;
α » β » S.

De plus, dans ce dernier cas, celui des joueurs A ou B qui prend la position α joue le jeu J avec la position P, l'autre joue avec la position Q.

L'attitude de A sera définie par les probabilités P_1, P_2, P_3 de choisir α, β, γ, par les probabilités p'_i, p''_i de choisir la tactique n° i s'il est en position P ou Q. L'attitude de B sera définie par les probabilités analogues Q_1, Q_2, Q_3, q'_j, q''_i.

Le gain moyen de A sera ainsi :

$$(39)\qquad \begin{aligned}\mathcal{H} = {} & P_1 Q_2 [S + K(p', q')] - P_1 Q_3 S \\ & + P_2 Q_1 [- S - K(q'', p'')] + P_2 Q_2 S + P_3 (Q_1 - Q_2) S.\end{aligned}$$

Le nouveau jeu ainsi défini est évidemment symétrique. Les probabilités choisies par B sont

$$Q_2 q'_j, \quad Q_3, \quad Q_1 q''_i.$$

Le jeu étant symétrique, le choix optimum des probabilités pour A sera tel que les coefficients de ces probabilités dans $\mathcal{H}$ soient non négatifs. En ce qui concerne le coefficient de Q_3, cela nous donne

$$(40)\qquad P_2 - P_1 \geqq 0.$$

P_1 et P_2 ne peuvent s'annuler en même temps ; (39) montre en effet que dans ce cas

$$\mathcal{H} = P_3 (Q_1 - Q_2) S,$$

et que par conséquent B peut gagner. L'inégalité (40) montre que P_2 est alors forcément positif.

P_3 ne peut être nul, puisque si $P_3 = 0$ on obtient avec $Q_1 = 1$:

$$\mathcal{H} = P_2 [- S - K(q'', p'')] < 0 \qquad (S > |\alpha_{ij}|).$$

Enfin, P_1 ne peut être nul, puisque pour $Q_2 = 1$, nous obtiendrions alors

$$\mathcal{H} = - P_3 S < 0.$$

Le joueur A doit donc choisir trois probabilités P_1, P_2, P_3 positives. Posons maintenant :

$$\sum \alpha_{ij} p'_i = h_j, \qquad \sum \alpha_{ij} p''_j = k_i;$$

d'où

$$K(p', q') = \sum_j h_j q'_j, \qquad K(q'', p'') = \sum_i k_i q''_i.$$

Les conditions à réaliser par A sont

$$\begin{aligned} &P_3 S - P_2[S + k_i] \geqq 0, \\ &P_1[S + h_j] - P_3 S \geqq 0, \\ &(P_2 - P_1) S \geqq 0 \end{aligned}$$

et nous avons vu que P_1, P_2, P_3 sont positifs. Il en résulte que

$$h_j \geqq k_i \qquad \text{quels que soient } i \text{ et } j.$$

Nous voyons donc que si l'on sait calculer les $\bar{p}_i$, $\bar{q}_j$, dans un jeu symétrique, on saura, en se ramenant à un tel jeu, déterminer pour un jeu asymétrique des $\bar{p}_i$ et des $\bar{q}_j$ tels que

$$\sum_i \alpha_{ij} \bar{p}_i \geqq \sum_j \alpha_{ij} \bar{q}_j \qquad \text{quels que soient } i \text{ et } j. \tag{41}$$

Posons

$$\lambda = \operatorname*{Min}_j \sum \alpha_{ij} \bar{p}_i,$$

$$\mu = \operatorname*{Max}_i \sum \alpha_{ij} \bar{p}_j.$$

D'après (41), nous avons $\lambda \geqq \mu$. Si le joueur en position P choisit les probabilités $\bar{p}_i$, il est certain de gagner au moins λ. Nous avons ainsi :

$$\operatorname*{Max}_p \operatorname*{Min}_q K(p, q) \geqq \lambda.$$

Si le joueur en position Q choisit les probabilités $\bar{q}_i$, il peut astreindre K à être au plus égal à μ ; d'où

$$\operatorname*{Min}_p \operatorname*{Max}_q K(p, q) \leqq \mu.$$

Nous voyons ainsi que

$$\operatorname*{Max}_p \operatorname*{Min}_q K(p, q) \geqq \operatorname*{Min}_p \operatorname*{Max}_q K(p, q).$$

Comme on sait *a priori* que

$$\operatorname*{Max}_p \operatorname*{Min}_q K(p, q) \leqq \operatorname*{Min}_p \operatorname*{Max}_q K(p, q),$$

il en résulte forcément l'égalité, ce qui d'une part démontre le théorème fondamental, et d'autre part montre que les $\bar{p}_i$ et $\bar{q}_i$ calculés sont bien les $\bar{p}_i$ et $\bar{q}_i$ cherchés.

2.3. **Jeu à trois partenaires.** — Le résultat fondamental de la théorie des jeux à deux partenaires est que lorsque la règle d'un jeu est définie, ainsi que les probabilités des différentes circonstances pouvant dépendre du hasard, la « valeur » d'une position, supposée occupée par un joueur infiniment habile, est bien déterminée. En effet, pour ce joueur, cette position est définie par les α_{ij} envisagés plus haut, c'est-à-dire que les coefficients de la forme $\mathrm{K}(p, q)$. Pour son adversaire, la forme correspondante est $-\mathrm{K}(p, q)$ la valeur des positions des deux joueurs est, respectivement

$$\operatorname*{Max}_{p}\operatorname*{Min}_{q}\mathrm{K}(p, q),$$
$$\operatorname*{Max}_{p}\operatorname*{Min}_{q}[-\mathrm{K}(p, q)].$$

La somme de ces valeurs est bien nulle.

Considérons maintenant un jeu à trois joueurs. Les probabilités que peut choisir A seront symbolisées par p, celles choisies par B seront appelées q, celles choisies par C appelées r.

Les espérances de A, B, C seront de la forme

$$(42)\qquad \begin{cases} \mathrm{A}(p, q, r) = \sum \alpha_{ijk} p_i q_j r_k, \\ \mathrm{B}(p, q, r) = \sum \beta_{ijk} p_i q_j r_k, \\ \mathrm{C}(p, p, r) = \sum \gamma_{ijk} p_i q_j r_k. \end{cases}$$

Nous sommes tentés de définir une valeur minimum du jeu de A de la manière suivante. Si les p sont donnés, il existe une valeur minimum de A :

$$(43)\qquad \operatorname*{Min}_{q, r}\mathrm{A}(p, q, r)$$

qui dépend des p. Si le joueur A rend cette quantité maximum, il obtiendra une valeur minimum de son espérance :

$$(44)\qquad \underline{\mathrm{A}} = \operatorname*{Max}_{p}\operatorname*{Min}_{q,r}\mathrm{A}(p, q, r).$$

A peut choisir les p de manière que, quoi que fassent B et C, A réalise l'inégalité

$$A(p, q, r) \geqq \underline{A}. \tag{45}$$

Supposons maintenant que B et C *se concertent* de manière à faire perdre à A le maximum. Ils peuvent choisir q et r de manière à atteindre le minimum

$$\overline{A} = \operatorname*{Min}_{q,r} \operatorname*{Max}_{p} A(p, q, r). \tag{46}$$

Nous ne sommes plus assurés que $\overline{A} = \underline{A}$. Il n'en serait immédiatement ainsi que si les joueurs B et C pouvaient se comporter comme un seul joueur, c'est-à-dire choisir non les produits $q_j r_k$, mais des probabilités liées P_{jk}. Or, choisir de telles probabilités revient évidemment à se communiquer leurs intentions au cours du jeu ; c'est une tricherie manifeste. Si les joueurs B et C, sans se concerter autrement que se liguer contre A, ne se mettent pas d'accord à l'avance sur la conduite à tenir, il peut leur être difficile d'astreindre A à l'inégalité

$$A < \overline{A}.$$

2.3.1. Étude d'un exemple simple. — Un exemple très simple montrera la nature de cette difficulté. Il existe un jeu se jouant à trois où chacun des trois partenaires pose une pièce de monnaie. Celui dont la pièce présente un côté différent des deux autres ramasse le tout. Si les pièees sont posées sur le même côté, la partie est nulle.

On voit facilement que A (p, q, r) est de la forme

$$A(p, q, r) = 2(p_1 q_2 r_2 + p_2 q_1 r_1) - (q_1 r_2 + q_2 r_1). \tag{47}$$

Cherchons le maximum de A, q et r étant donnés. Quand p_1 varie de 0 à 1, A varie entre

$$2 q_2 r_2 - (q_1 r_2 + q_2 r_1)$$
$$2 q_1 r_1 - (q_1 r_2 + q_2 r_1),$$

Nous avons donc

$$\operatorname*{Max}_{p} A = 2 \operatorname{Max} \{ q_2 r_2, q_1 r_1 \} - q_1 r_2 - q_2 r_1.$$

Nous pouvons écrire cette expression sous la forme équivalente :

$$\operatorname*{Max}_{p} A = q_2 r_2 + q_1 r_1 + | q_2 r_2 - q_1 r_1 | - q_1 r_2 - q_2 r_1. \tag{48}$$

Si B et C sont décidés à jouer contre A, et cherchent chacun de leur côté à minimiser cette expression, supposée calculée par chacun d'eux, ils sont assez embarrassés. Si B *savait* que C a choisi $r_2 = 0$, il lui suffirait de prendre $q_1 = 0$ pour que $A = -1$, c'est-à-dire pour donner à (48) sa valeur minimum. Mais si B ignore la valeur r, choisie par C, il ne peut rendre (48) minimum. En effet, si B considère r_1 comme un paramètre inconnu, il constate que A est une fonction de q_1 prenant les valeurs suivantes :

$$(49) \qquad \begin{cases} \text{pour } q_1 = 0 : & A = r_2 - r_1 ; \\ \text{« } \quad q_1 = r_2 : & A = -(r_1 - r_2)^2 ; \\ \text{« } \quad q_1 = 1 : & A = r_1 - r_2. \end{cases}$$

Pour les valeurs intermédiaires de q_1, A est une fonction linéaire de q_1. Le choix le plus défavorable pour A est ainsi

$$q_1 = \frac{1}{2} - \frac{1}{2}\frac{r_1 - r_2}{|r_1 - r_2|},$$

ce qui peut s'écrire d'une façon symétrique :

$$(50) \qquad q_1 - q_2 = \frac{r_2 - r_1}{|r_2 - r_1|}.$$

Supposons que C admette que B fasse ce raisonnement. L'espérance de A ne dépend plus que du choix de C. C calcule alors le résultat de la substitution de (50) dans (48), ce qui conduit à

$$(51) \qquad A = -\frac{1 + |r_1 - r_2|}{2}.$$

Il peut donc être admis que C, sachant que B est disposé à l'aider, choisisse r_1 et r_2 tels que

$$(52) \qquad |r_1 - r_2| = 1.$$

B peut donc compter sur l'égalité (52). Mais cette égalité ne lui permet pas, *à elle seule*, de déterminer q_1. Dans le cas particulier qui nous occupe, B pourra dès le premier coup connaître le choix de C, et par conséquent agir. Mais nous voyons que nous sortons du cadre de la théorie générale d'après laquelle on cherche à déterminer les probabilités *avant* le début de la partie.

2.3.2. Exploitation d'une situation par deux individus. — Les difficultés apparues dans le cas de trois joueurs apparaissent presque au

même degré quand deux individus ont à exploiter une situation. Soient A et B pouvant prendre différentes attitudes devant une situation. Si A prend l'attitude i et B l'attitude j, leurs gains seront respectivement α_{ji} et β_{ij}. L'« entente », pour A et B, sera de choisir i et j de manière à rendre maximum

$$\alpha_{ij} + \beta_{ij}$$

et ensuite, de se répartir au mieux cette somme. La difficulté vient de ce que les règles de répartition dépendent essentiellement *de ce qui se passerait si* A *et* B *ne s'entendaient pas*. La structure des deux tableaux $[\alpha_{ij}]$ et $[\beta_{ij}]$ décrit la force économique respective de A et B, c'est-à-dire la manière dont ils sont armés pour exploiter la situation, non seulement vis-à-vis du milieu extérieur, mais encore vis-à-vis l'un l'autre. Si A veut fixer son attitude vis-à-vis de B, il peut lui dire :

Si vous choisissez la position j, je choisirai la position i avec la probabilité p_{ji}. B peut répondre de la même manière, en proposant des q_{ij}. On voit que cette manière d'aborder la question pose de graves difficultés, parce que les p_{ji} et les q_{ij} ne peuvent conduire à un choix des p_i et des q_j que si le rapport $\frac{p_{ji}}{q_{ij}}$ est le quotient d'une fonction de i par une fonction de j.

On doit se contenter, même dans une question aussi simple, de poser des règles conventionnelles. Si au lieu de chercher à maximiser $A(p, q)$, le partenaire A cherche à maximiser la différence

$$A(p, q) - B(p, q),$$

c'est-à-dire s'il cherche à gagner plus que B, et si B de son côté en fait autant, c'est-à-dire s'il cherche à maximiser

$$B(p, q) - A(p, q),$$

nous nous retrouvons dans les conditions du théorème fondamental. Il existe une valeur bien déterminée de

$$\Delta = \operatorname*{Max}_{p} \operatorname*{Min}_{q} [A(p, q) - B(p, q)]$$

dont la définition est symétrique par rapport à A et B. Ce nombre peut être défini comme mesurant la *supériorité économique* de A sur B (s'il est positif).

Soit $\bar{p}, \bar{q}$ le système des probabilités (ou un des systèmes de proba-

bilités) attachés à la réalisation de Δ. Parmi tous les systèmes possibles, il se peut qu'il y en ait certains qui donnent des valeurs différentes à $A(\bar{p}, \bar{q})$ et, par conséquent à $B(\bar{p}, \bar{q})$, puisque la différence est constante. Un premier pas de collaboration entre A et B sera de choisir le système $(\bar{p}, \bar{q})$ maximisant $B(\bar{p}, \bar{q})$. Dans cette recherche, les intérêts de A et B sont concordants. Soient $\overline{A}$ et $\overline{B}$ les valeurs ainsi obtenues. Nous avons certainement

$$\overline{A} + \overline{B} \leqq \alpha = \operatorname*{Max}_{ij}(\alpha_{ij} + \beta_{ij}).$$

Un second pas vers l'entente serait de choisir la position i et la position j conduisant à α, et à se répartir α en donnant

$$\frac{\alpha}{2} + \frac{\Delta}{2} \quad \text{à A},$$
$$\frac{\alpha}{2} - \frac{\Delta}{2} \quad \text{à B}.$$

La situation commune est ainsi utilisée au mieux, et l'on respecte l'inégalité des positions économiques de A et B. Si B refuse cette convention, A peut user de représailles en adoptant les probabilités $\bar{p}$, ce qui a pour effet d'entraîner

$$A - B \geqq \Delta,$$
$$A + B \leqq \alpha,$$
$$B \leqq \frac{\alpha}{2} - \frac{\Delta}{2}.$$

De même B peut user de représailles envers A, si A refuse la transaction. La transaction proposée semble donc raisonnable; mais la théorie que nous venons d'esquisser reste muette sur le prix dont se payent les représailles, parce que les quantités $\overline{A}$ et $\overline{B}$ ne sont définies que dans le cas où une première collaboration est supposée entre A et B. Si l'on admet que la victime d'une mesure de représailles réagit en minimisant l'effet des représailles sur lui-même, cet effet, pour A, sera évidemment :

$$\frac{\alpha}{2} + \frac{\Delta}{2} - \overline{A}.$$

Il en serait autrement si A réagissait à une mesure de représailles en maximisant l'effet de cette mesure *sur* B.

CHAPITRE III.

THÉORIE DES SIGNAUX.

Nous définirons un signal comme une énergie qui se propage, depuis un émetteur jusqu'à un récepteur, et est susceptible de se présenter sous un grand nombre de formes différentes. Chacune de ces formes constitue un *message*. Par exemple, un signal sonore est constitué par une onde sonore, qui peut varier de forme, de durée, d'intensité; l'émetteur peut être un organe vocal, un haut parleur; le récepteur peut être une oreille, un microphone.

3.1. **Système de transmission dont l'émetteur et le récepteur comportent chacun un nombre fini de positions.** — Dans ce qui suit, nous ne traiterons que de signaux électriques, la plupart des problèmes de transmissions pouvant se ramener à l'étude de tels signaux. Le schéma de transmission le plus élémentaire est alors le suivant :

Un transmetteur, ou émetteur, est manœuvré d'une manière quelconque; nous assimilons cette manœuvre au positionnement d'une clé à n positions. Le récepteur est lié à l'émetteur par une ligne de transmission. Nous assimilons sa réaction, aux messages reçus, au mouvement d'une aiguille pouvant se placer devant n graduations. Le système de transmission est en principe spécifié de manière que si la clé du transmetteur est dans la position i, l'aiguille du récepteur se mette sur la position i. Le signal lui-même est le courant passant dans la ligne. La position de la clé du transmetteur est le *message transmis*, la position de l'aiguille est le *message reçu*.

Si l'on suppose les appareils parfaits, le nombre de signaux distincts possibles est bien égal à n. Si les appareils sont imparfaits, le nombre des signaux distincts est réduit. Nous allons examiner dans quelle proportion.

3.1.1. Cas où l'émetteur et le récepteur sont tous deux à n positions. — Supposons que, le transmetteur étant à n positions, le récepteur soit à n positions; supposons de plus que les défauts du systèmes soient tels que si l'on place le transmetteur sur la position i, la probabilité pour que le récepteur se place sur la position j soit p_{ij}. Le tableau des p_{ij}

caractérise entièrement la qualité du système. Ce tableau, à première vue, ne peut pas être quelconque. Si l'agent récepteur interprète le message reçu j comme étant effectivement le message transmis j, cette interprétation ne se justifie que si

$$p_{jj} \geqq p_{ij} \qquad (i \neq j),$$

c'est-à-dire si, parmi les différentes positions du récepteur correspondant à une position donnée du transmetteur, la plus probable est la position correcte.

3.1.2. Cas ou l'émetteur est a n positions et de le récepteur a m positions. — Plaçons-nous dans le cas plus général où il n'y a pas correspondance entre le nombre n des positions du transmetteur et le nombre m des positions du récepteur. Le tableau des p_{ij} est alors un tableau rectangulaire. Supposons que l'on tire au hasard une position du transmetteur. La probabilité pour que le transmetteur soit sur la position i et le récepteur sur la position j est alors

$$\frac{1}{n} p_{ij} = \bar{\omega}_{ij}. \tag{1}$$

La probabilité *a posteriori* pour que, si le récepteur est sur la position j, le message envoyé ait été le message i est par conséquent

$$q_{ij} = \frac{\bar{\omega}_{ij}}{q_j}, \qquad q_j = \sum_i \bar{\omega}_{ij} = \frac{1}{n} \sum_i p_{ij}. \tag{2}$$

La transmission parfaite correspond au cas où tous les q_{ij} associés à un même indice j sont nuls *sauf un*. Cela exige que tous les $\bar{\omega}_{ij}$ associés à un même indice j soient nuls sauf un, et qu'il en soit de même des p_{ij}. Le nombre des messages que l'on peut transmettre est alors le nombre des indices i pour lesquels $p_{ij} > 0$. Ce nombre est égal à n. En effet, chaque position du transmetteur correspond à au moins une position du récepteur, et chaque position du récepteur donne une indication sans ambiguïté. On peut exprimer ce nombre n par une expression analytique fonction des $\bar{\omega}_{ij}$.

Cette expression analytique nous sera utile par la suite. Remarquons que pour j constant, q_{ij} ne peut être égal par hypothèse qu'à 0 ou 1 ; donc

$$\sum_i q_{ij} \log q_{ij} = 0,$$

et, par conséquent :

$$\sum_{ij} q_{ij} \log q_{ij} = 0 \tag{3}$$

Considérons maintenant l'expression

$$-\mathrm{H}(\bar{\omega}) = \sum_{ij} \bar{\omega}_{ij} \log \bar{\omega}_{ij}.$$

Elle peut s'écrire

$$\sum_{ij} \bar{\omega}_{ij} (\log q_j + \log q_{ij})$$

ou encore

$$\sum_{ij} q_{ij} q_j (\log q_j + \log q_{ij}).$$

D'après ce qui a été remarqué, la sommation par rapport à i nous conduit à

$$\mathrm{H}(\bar{\omega}) = -\sum_{j} q_j \log q_j. \tag{4}$$

Posons

$$p_i = \frac{1}{n}$$

et

$$\mathrm{H}(p) = -\sum p_i \log p_i, \qquad \mathrm{H}(q) = -\sum q_j \log q_j.$$

Nous obtenons l'expression de $\log n$:

$$\log n = \mathrm{H}(p) + \mathrm{H}(q) - \mathrm{H}(\bar{\omega}). \tag{5}$$

3.1.3. Définition du nombre de messages transmis et du nombre de messages a l'entrée. — Généralisons encore les considérations du paragraphe précédent. Supposons les p_{ij} quelconques, et associons des probabilités p_i aux messages à l'entrée.

Supposons pour un instant que l'on convienne d'appeler nombre de messages *transmis* à travers un système caractérisé par le tableau p_{ij}, quand les probabilités des positions du manipulateur sont p_i et que par conséquent les positions du récepteur ont les probabilités

$$q_j = \sum p_i p_{ij},$$

le nombre dont le logarithme est

$$\mathrm{I} = \mathrm{H}(p) + \mathrm{H}(q) - \mathrm{H}(\bar{\omega}). \tag{6}$$

Nous reconnaissons en I la quantité qui était égale à $\log n$ dans le paragraphe précédent.

Convenons, de plus, d'appeler nombre des messages à *l'entrée* le nombre dont le logarithme est $H(p)$ et nombre de messages à la sortie le nombre dont le logarithme est $H(q)$. Nous allons voir que ces conventions sont cohérentes avec les résultats relatifs à la transmission parfaite.

Nous avons en effet les résultats suivants :

1. *Le nombre des messages à l'entrée est au plus égal au nombre de positions du transmetteur.* Il n'est égal à ce nombre que si tous les p_i sont égaux.

Ce résultat est une conséquence de l'inégalité

$$-\sum_{i=1}^{n} p_i \log p_i \leqq \log n.$$

2. *Le nombre des messages transmis est au plus égal au nombre des messages à l'entrée.*

Cette proposition est une conséquence de l'inégalité

$$H(\bar{\omega}) \geqq H(q)$$

qui peut s'écrire

$$-\sum_{ij} \bar{\omega}_{ij} \log \frac{\bar{\omega}_{ij}}{q_j} \geqq 0.$$

Cette inégalité est évidente, puisque $\bar{\omega}_{ij} \leqq q_j$.

3. *Le nombre des messages transmis est au plus égal au nombre des messages à la sortie.*

Cette proposition n'est pas distincte de la précédente, étant donnée la symétrie de I.

4. *Le nombre des messages transmis est toujours positif (ou nul).*

Cette proposition est une conséquence de l'inégalité

$$\sum_{ij} \bar{\omega}_{ij} \log \frac{\bar{\omega}_{ij}}{p_i q_j} \geqq 0.$$

Les quatre résultats qui précèdent sont vrais quels que soient les p_i

et les p_{ij}. Nous allons maintenant en venir à la condition de transmission parfaite.

5. *La condition nécessaire et suffisante pour que la transmission soit parfaite est que le nombre des messages transmis soit égal au nombre de messages à l'entrée.*

Cette proposition est une conséquence immédiate de la proposition 2. L'égalité entre $H(\bar{\omega})$ et $H(q)$ ne peut avoir lieu que si $\bar{\omega}_{ij}$ ne peut prendre que les valeurs zéro ou q_j. Il en résulte bien que lorsque $q_j \neq 0$, le message reçu de rang j a une interprétation unique. Tout message présenté au transmetteur est donc transmis correctement. Le nombre maximum des messages que l'on peut transmettre s'obtient en faisant tous les p_i égaux. On retombe sur les résultats notés plus haut.

3.2. **Conséquences de la définition du nombre de messages transmis.** — 3.2.1. NOMBRE MAXIMUM DE MESSAGES TRANSMISSIBLES. — Nous allons voir maintenant où conduit la définition (6). Le nombre I est fonction des probabilités p_i; nous supposerons que l'on peut choisir ces probabilités. Les p_{ij} sont supposés fixes. Il se peut que la transmission soit parfaite pour certaines valeurs des p_i et ne le soit plus pour d'autres. Cherchons à déterminer les p_i de manière à rendre I maximum; puisque les p_{ij} sont supposés fixes, I ne dépend que des p_i.

Supposons que nous ayons choisi les p_i de manière que I soit maximum. Donnons aux p_i des accroissements de la forme

$$\delta p_i = k_i p_i \, \delta t,$$

les k_i étant des constantes finies. Les q_j subissent des accroissements de la forme

$$\delta q_j = h_j q_j \, \delta t,$$

les h_j étant des constantes finies. Pour qu'il n'en soit pas ainsi, il faudrait que nous ayons à la fois

$$q_j = \sum_i p_{ij} p_i = 0, \qquad \delta q_j = \sum_i p_{ij} \, \delta p_i \neq 0.$$

Cela est impossible parce que les p_{ij} et les p_i étant non négatifs, l'égalité $q_j = 0$ n'est possible que si les n égalités

$$p_{ij} p_i = 0 \qquad (i = 1, 2, \ldots, n)$$

ont lieu; comme seuls subissent des accroissements ceux des p_i qui ne sont pas nuls, il est impossible que l'on ait

$$q_j = 0, \qquad \delta q_j \neq 0.$$

Nous en déduisons que pour les systèmes d'accroissements considérés, I est une fonction différentiable. Nous obtenons alors :

$$\delta \mathrm{I} = \sum_i \left[\sum_j p_{ij} \log \frac{p_{ij}}{q_j} \right] \delta p_i. \tag{7}$$

Dans l'équation (7) interviennent seuls les crochets pous lesquels $p_i > 0$. On voit alors qu'ils ont tous une valeur bien déterminée. Supposons en effet que pour un de ces crochets on ait

$$q_j = 0.$$

Il en résulte que $p_{ij} = 0$, puisque

$$q_j = \sum_i p_{ij} p_i > p_i p_{ij}$$

et nous pouvons donc écrire

$$p_{ij} \log \frac{p_{ij}}{q_j} = 0,$$

en convenant que

$$0 \log 0 = 0.$$

Dans l'équation (7), les coefficients des δp_i sont positifs ou nuls. Nous pouvons écrire

$$\delta \mathrm{I} = \sum \mathrm{J}_i \, \delta p_i, \qquad \mathrm{J}_i \geqq 0. \tag{8}$$

Nous remarquons par ailleurs que

$$\sum_i p_i \mathrm{J}_i = \mathrm{I}. \tag{9}$$

Si par conséquent nous choisissons les δp_i de la forme

$$\delta p_i = p_i (\mathrm{J}_i - \mathrm{I}) \, \delta t \qquad (\delta t > 0), \tag{10}$$

nous respectons la condition

$$\sum p_i = 1$$

et nous avons

$$\delta \mathrm{I} = \frac{1}{\delta t} \sum \frac{(\delta p_i)^2}{p_i} > 0.$$

Le maximum est ainsi forcément atteint pour des p_i satisfaisant aux équations

$$p_i(J_i - I) = 0 \qquad (i = 1, 2, \ldots, n). \tag{11}$$

Supposons que les équations (11) admettent une solution $\{\bar{p}_i\}$ constituée par des probabilités toutes positives. Nous allons voir que nous avons effectivement atteint le maximum de I. Appelons $\bar{I}$, $\bar{J}_i$, $\bar{q}_j$, les grandeurs relatives à la solution du système

$$\bar{J}_i = \bar{I}$$

et soient p_i, q_j, I, J_i les grandeurs analogues relatives à un système quelconque. Nous aurons

$$\begin{aligned} \bar{I} - I &= \sum (\bar{p}_i \bar{J}_i - p_i J_i) = \sum_i p_i (\bar{J}_i - J_i) \\ &= \sum_{ij} p_i p_{ij} \log \frac{q_j}{\bar{q}_j} = \sum_j q_j \log \frac{q_j}{\bar{q}_j} \geq 0. \end{aligned} \tag{12}$$

Les équations (11) sont de la forme

$$\sum_j p_{ij} \log q_j = \sum_j p_{ij} \log p_{ij} - I \qquad (p_i > 0).$$

Ce sont des équations linéaires par rapport aux inconnues $\log q_j$. Dans le second membre figure une quantité I inconnue. Posons

$$\log I = \lambda, \qquad \lambda q_j = z_j.$$

Nous sommes ramenés au système

$$\sum_j p_{ij} \log z_j = \sum_j p_{ij} \log p_{ij} \tag{13}$$

aux inconnues $\log z_j$. Seules seront admissibles les solutions telles que si l'on pose

$$q_j = \frac{z_j}{\sum_j z_j}$$

on obtienne un système de valeurs des q_j tel que le système

$$\sum_i p_i p_{ij} = q_j$$

aux inconnues p_i admette une solution positive. Comme nous savons que de toute manière I a un maximum que l'on peut atteindre, le système (12) a toujours des solutions admissibles, après suppression éventuelle de certaines équations (correspondant à des p_i nuls).

3.2.2. NOMBRE MAXIMUM DE MESSAGES TRANSMISSIBLES POUR UN NOMBRE DONNÉ DE MESSAGES A L'ENTRÉE. — Il peut être intéressant de chercher le maximum de I quand on s'impose une valeur de $H(p)$. Revenons à l'expression (7). Donnons aux p_i des accroissements de la forme

$$\delta p_i = p_i[J_i - I + \mu(\log p_i + H)]\,\delta t. \tag{14}$$

La condition $\sum p_i = 1$ est satisfaite. La condition $H(p) = \text{const.}$ est vérifiée si

$$\sum \log p_i\,\delta p_i = \sum p_i J_i \log p_i + HI + \mu \sum p_i (\log p_i)^2 - H^2\mu = 0$$

est satisfaite, ce qni nous fixe la valeur de μ :

$$\mu = \frac{-\sum p_i \log p_i (J_i - I)}{\sum p_i (\log p_i)^2 - H^2}. \tag{15}$$

Avec ces accroissements, nous voyons que

$$\delta I = \sum \frac{1}{p_i} \frac{(\delta p_i)^2}{\delta t}.$$

Donc le maximum de I, à $H(p)$ constant, n'est obtenu que si les δp_i sont nuls, c'est-à-dire si

$$J_i = I - \mu(\log p_i + H) \qquad (\text{pour } p_i > 0). \tag{16}$$

Les équation (16) et (15) définissent les p_i de manière à maximiser I, à H constant. Supposons que les p_i soient toujours ainsi définis, et *faisons varier* $H(p)$.

Nous obtiendrons en différentiant (16), les relations auxquelles satisfont les δp_i :

$$\delta J_i = \delta I - \delta\mu(\log p_i + H) - \mu \frac{\delta p_i}{p_i} - \mu\delta H. \tag{17}$$

Par ailleurs, nous avons toujours, quelle que soit la manière dont varient les δp_i (pourvu que les p_i nuls le restent) :

$$\delta I = \delta \sum J_i p_i = \sum J_i\,\delta p_i,$$

d'où

$$\sum p_i \delta J_i = 0.$$

Cette dernière équation permet de tirer de (17) la relation

$$0 = \delta I - \mu \delta H. \tag{18}$$

La variation de I avec H, étant entendu que pour une valeur H de $H(p)$, I est rendu maximum, est donc régie par l'équation

$$\frac{\delta I}{\delta H} = \mu. \tag{19}$$

Il existe une valeur de $H(p)$ pour laquelle I est maximum; c'est celle qui correspond à la recherche du maximum abolu de I. Soit $\overline{H}$ cette valeur. Pour $H < \overline{H}$, nous avons, au moins dans un certain domaine

$$0 < \frac{\delta I}{\delta H} < 1,$$

μ est donc positif et nous avons, dans ce domaine

$$\frac{\delta}{\delta H}(H - I) = 1 - \mu > 0,$$

ce qui montre que pour $H < \overline{H}$, l'excès des messages à l'entrée sur les messages transmis (excès exprimé par $H - I$) est une fonction croissante de H. En d'autres termes, si l'on diminue H, on rapproche I de H. Il y a donc intérêt à faire un sacrifice sur H, parce que l'on réduit moins le nombre des messages transmis que le nombre des messages présentés. On améliore ainsi la transmission, puisque l'incertitude sur la signification des messages reçus provient en partie de l'excès du nombre des messages à l'entrée sur le nombre des messages transmis. Les règles à suivre pour améliorer la transmission apparaissent sur la formule (12), d'où il résulte que

$$I = \overline{I} - \sum q_j \log \frac{q_j}{\overline{q_j}}. \tag{20}$$

Pour diminuer $H(p)$ sans beaucoup diminuer I, il faut modifier les p_i de manière à avoir une forte variation de $H(p)$ et une *faible* variation des q_j.

Supposons par exemple que le système de transmission soit symétrique par rapport aux différentes positions du transmetteur, c'est-à-dire,

lorsqu'on passe d'une valeur de i à une autre, les m nombres p_{ij} ne changent pas de valeur mais s'échangent simplement.

Si de plus le tableau des p_{ij} est symétrique, et que par conséquent $m = n$, le système (13) admet la solution

$$p_i = q_j = \frac{1}{n}.$$

Le nombre positif

$$h = -\sum_j p_{ij} \log p_{ij}$$

est bien défini, puisque la somme qui le définit est indépendante de i. La valeur maximum de I est dans ces conditions

$$\bar{\mathrm{I}} = \log n - h.$$

Pour toute autre valeur des p_i, nous avons

$$\mathrm{I} = -\sum_j q_j \log q_j - h$$

et par conséquent

$$\mathrm{H}(p) - \mathrm{I} = -\sum_i p_i \log p_i - \sum_j q_j \log q_j - h.$$

Si nous pouvons modifier les p_i de manière à obtenir l'égalité

$$\mathrm{H}(p) = \log n - h$$

et cela sans beaucoup modifier les q_j, de manière que

$$-\sum q_j \log q_j = \log n - \varepsilon,$$

nous aurons obtenu

$$\mathrm{I} = \mathrm{H}(p) - \varepsilon,$$

c'est-à-dire que nous aurons une transmission presque parfaite.

Dans la technique des transmissions, on constate effectivement que si un transmetteur à n positions, et qu'un récepteur à m positions est actionné par ce transmetteur à travers un système pouvant introduire des erreurs, si l'on calcule, d'après les règles indiquées ci-dessus, la valeur maxima de $\bar{\mathrm{I}}$, et qu'ensuite on réduise le nombre des positions du transmetteur à un nombre inférieur à $\bar{\mathrm{I}}$, ces positions étant naturellement choisies au mieux, le système de transmission devient tel que les erreurs sont considérablement diminuées, c'est-à-dire que l'on peut

associer avec une grande certitude à chaque position du récepteur une position du transmetteur. L'inverse n'a naturellement pas lieu ; à chaque position du transmetteur correspondent plusieurs positions du récepteur. Le choix effectif des positions du transmetteur à conserver pose de difficiles problèmes d'analyse combinatoire que nous ne pouvons traiter ici.

3.3. — **Application de la définition du nombre de messages transmis à la théorie de l'observation indirecte d'une variable aléatoire.** — 3.3.1. OBSERVATION INDIRECTE D'UNE VARIABLE ALÉATOIRE A L'AIDE D'UNE AUTRE VARIABLE ALÉATOIRE OBSERVABLE DIRECTEMENT. — Nous allons appliquer la notion de nombre de messages transmis au problème de l'observation indirecte d'une variable aléatoire. Soit X une variable aléatoire, de densité de distribution $p(x)$, qui n'est pas observable directement. Nous supposons qu'un observateur ne peut mesurer que la valeur z d'une variable aléatoire Z liée à X. Cette liaison est caractérisée par la distribution conditionnelle de Z lorsque X est connue. Nous supposerons que cette distribution peut se représenter par la densité $p(x\,;\,z)$.

La distribution marginale de Z est alors (en densité)

$$q(z) = \int p(x)p(x\,;\,z)\,dx.$$

Assimilons les différentes valeurs de X aux différentes positions d'un transmetteur, les différentes valeurs de Z aux différentes positions d'un récepteur. Nous obtenons en

$$\mathrm{I} = \int p(x)p(x\,;\,z)\log\frac{p(x\,;\,z)}{q(z)}\,dx\,dy, \tag{21}$$

le logarithme du nombre que nous avons appelé nombre de messages transmis, et que nous appellerons ici « nombre des valeurs de X discernables par observation de Z ».

Une propriété importante de I est la suivante :

I reste invariant par tous changements de variable (monotones) effectué séparément sur X et Z.

En d'autres termes, si l'on pose

$$\mathrm{X}' = f(\mathrm{X}), \qquad \mathrm{Z}' = g(\mathrm{Z}),$$

f et g étant deux fonctions à dérivée de signe constant, le nombre I

relatif à X′ observé à travers Z′ est le même que le nombre I relatif à X et Z. Cette propriété de I se démontre en faisant les changements de variables classiques dans l'intégrale définissant I.

Si l'on fait intervenir la distribution conditionnelle de X connaissant Z, soit

$$q(z;\,x)=\frac{p(x)p(x;\,z)}{q(z)},$$

on peut écrire I sous la forme

$$\mathrm{I}=\int q(z)q(z;\,x)\log\frac{q(z;\,x)}{p(x)}\,dx\,dz. \tag{22}$$

3.3.2. Cas ou la variable observée directement ne peut prendre que n valeurs distinctes. — Si Z ne peut prendre que n valeurs distinctes, la probabilité de la $j^{\text{ième}}$ valeur quand x est connue étant

$$\bar{\omega}_j(x);\qquad \left[\sum\bar{\omega}_j(x)=1\right]. \tag{23}$$

nous avons, pour la probabilité de cette $j^{\text{ième}}$ valeur quand x est inconnue (probabilité marginale) :

$$q_j=\int p(x)\bar{\omega}_j(x)\,dx. \tag{24}$$

La distribution conditionnelle de x quand z est connue devient alors

$$p_j(x)=\frac{p(x)\,\bar{\omega}_j(x)}{q_j} \tag{25}$$

et nous pouvons mettre $p(x)$ sous la forme

$$p(x)=\sum_j q_jp_j(x)\qquad\left[\int p_j(x)\,dx=1\right]. \tag{26}$$

L'expression I s'écrit alors

$$\mathrm{I}=\sum_j\int dx\left[q_jp_j(x)\log\frac{p_j(x)}{p(x)}\right]. \tag{27}$$

Si l'on forme la différence

$$-\sum q_j\log q_j-\mathrm{I},$$

on constate qu'on peut l'écrire

$$\sum_j\int dx\,p_j(x)\,q_j\log\frac{p(x)}{q_jp_j(x)}. \tag{28}$$

Il apparaît sous cette forme qu'il s'agit d'une expression essentiellement positive (ou nulle). Nous avons donc

$$\mathrm{I} \leqq -\sum q_j \log q_j. \tag{29}$$

Le signe « égal » n'a lieu que si

$$p_j(x)[p(x) - q_j p_j(x)] = 0,$$

c'est-à-dire si en tous les points où $p_j(x) > 0$, $p_j(x)$ est proportionnelle à $p(x)$. Il en résulte que

$$\varpi_j(x) = 0 \quad \text{ou} \quad 1,$$

c'est-à-dire qu'il y a correspondance univoque de X vers Z, ou encore que l'on a une relation fonctionnelle entre Z et X :

$$\mathrm{Z} = f(\mathrm{X}). \tag{30}$$

Dans ces conditions, I est maximum lorsque les p_j sont tous égaux. Les indices j ne sont alors autres que les indices de n ensembles équiprobables, disjoints, découpés dans le champ de variation de X.

3.3.3. Observation d'une variable aléatoire X par l'intermédiaire d'une variable Z vue elle-même a travers une variable T observable directement. — Nous allons voir maintenant comment se modifie I lorsque l'on suppose connue la valeur t d'une variable T liée à X par l'intermédiaire de Z. Nous entendons par là que la densité de probabilité liée de X, Z, T est de la forme

$$r(t)\, r(t;\, z)\, q(z;\, x). \tag{31}$$

Nous avons alors

$$q(z) = \int r(t) r(t;\, z)\, dt. \tag{32}$$

La loi liant x et z reste

$$q(z)\, q(z;\, x) = p(x)\, p(x;\, z). \tag{33}$$

La distribution (31) peut encore s'écrire

$$p(x)\, p(x;\, z)\, \rho(z;\, t) \tag{34}$$

En intégrant (34) et (31) par rapport à x, nous obtenons

$$q(z)\, \rho(z;\, t) = r(t)\, r(t;\, z). \tag{35}$$

Nous pouvons donc, pour définir la liaison de T avec X et Z, nous donner arbitrairement $\rho(z;\ t)$. Supposons cette fonction choisie. Si la valeur t de T est connue, la fonction de distribution de X n'est plus $p(x)$ mais

$$p_t(x) = \frac{P(x,\ t)}{r(t)}, \tag{36}$$

$P(x,\ t)$ étant la densité de probabilité liée de X et T :

$$\begin{aligned} P(x,\ t) &= r(t) \int r(t;\ z)\, q(z;\ x)\, dz \\ &= p(x) \int p(x;\ z)\, \rho(z;\ t)\, dz. \end{aligned}$$

Reportons-nous à la formule (22) définissant I quand X est vue à travers Z. La quantité analogue J relative à X vue à travers T sera

$$J = \int r(t)\, p_t(x) \log \frac{p_t(x)}{p(x)}\, dx\, dt. \tag{37}$$

Cette quantité J est au plus égale à I. Nous avons en effet

$$I - J = \int q(z)\, \rho(z;\ t)\, dz\, dt \int q(z;\ x) \log \frac{q(z;\ x)}{p_t(x)}\, dx \geqq 0. \tag{38}$$

Supposons que T ne puisse prendre que n valeurs entières, et que $T = k$ ait lieu si Z se trouve dans un domaine Ω_k. Nous supposons les domaines Ω_k disjoints. Dans ce cas, la probabilité que $T = k$ et que X, Z soient dans des intervalles dx, dz est de la forme

$$r(k)\, r(k;\ z)\, q(z;\ x)\, dx\, dz. \tag{39}$$

Ceci peut encore s'écrire

$$p(x)\, p(x;\ z)\, \rho(z;\ k)\, dx\, dz, \tag{40}$$

avec

$$\begin{cases} q(z)\, \rho(z;\ k) = r(k)\, r(k;\ z), \\ q(z) = \sum_k r(k)\, r(k;\ z). \end{cases} \tag{41}$$

La probabilité $r(k)$ s'obtient en intégrant $q(z)$ dans le domaine Ω_k. En effet $\rho(z;\ k)$ est égal à 1 dans ce domaine, à zéro en dehors. Nous avons ainsi

$$\begin{cases} r(k;\ z) = \dfrac{q(z)}{r(k)} & \text{pour } z \text{ dans } \Omega_k; \\ r(k;\ z) = 0 & \text{pour } z \text{ en dehors de } \Omega_k. \end{cases} \tag{42}$$

Il en résulte évidemment que

$$r(k) = \int_{\Omega_k} q(z)\, dz.$$

Nous obtenons pour la loi liée de X et T :

$$P(x, k) = r(k) \int r(k; z)\, q(z; x)\, dz = \int q(z)\, q(z; x)\, dz$$

et par conséquent, pour la distribution de X quand T est connue

$$p_k(x) = \frac{\int_{\Omega_k} q(z)\, q(z; x)\, dz}{\int_{\Omega_k} q(z)\, dz}. \tag{43}$$

Nous vérifions que

$$p(x) = \sum_k r(k) p_k(x).$$

Nous pouvons écrire encore

$$p_k(x) = \frac{p(x)}{r(k)} \int_{\Omega_k} p(x; z)\, dz = \frac{p(x)}{r(k)} \alpha_k(x),$$

de sorte que J prend la forme

$$J = -\sum r(k) \log r(k) + \sum_k \int p(x)\, \alpha_k(x) \log \alpha_k(x)\, dx. \tag{44}$$

Rendre J maximum pose des problèmes difficiles. On peut voir sur (44) que puisque $\alpha_k(x) < 1$, J sera toujours au plus égal à $-\sum r(k) \log r(k)$. Nous pouvons nous imposer comme condition que tous les $r(k)$ soient égaux. Dans ce cas nous aurons

$$J = \log n + \sum_k \int p(x)\, \alpha_k(x) \log \alpha_k(x)\, dx \tag{45}$$

et nous aurons à déterminer les domaines Ω_k, tels que

$$\int_{\Omega k} q(z)\, dz = \frac{1}{n}$$

et que l'intégrale figurant au second membre de (45) soit maximum (c'est-àdire que sa valeur absolue soit minimum).

Exemple. — Pour voir à quoi correspond cette recherche du maximum de J, considérons un exemple.

Soit X une variable normale, et U une autre variable normale indépendante de X. Supposons que nous ayons

$$Z = X + U,$$

que Z, X, U soient centrées, que X et U aient pour variances s^2 et σ^2. La variance de Z est alors $s^2 + \sigma^2$.

Nous avons

$$p(x;\ z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2\sigma^2}},$$

d'où

$$\log p(x;\ z) = -\frac{u^2}{2\sigma^2} - \log\sigma\sqrt{2\pi}.$$

De même

$$\log q(z) = -\frac{z^2}{2(s^2+\sigma^2)} - \log\sqrt{(s^2+\sigma^2)\,2\pi}.$$

Nous en déduisons, d'après la formule (2) définissant I, que

$$I = \log\sqrt{\frac{s^2+\sigma^2}{\sigma^2}} = \log\frac{s}{\sigma} + \log\sqrt{1+\frac{\sigma^2}{s^2}},$$

de sorte que si σ est petit devant s, le nombre de valeurs de X discernables à travers Z est de $\frac{s}{\sigma}$; ce nombre est le quotient de la dispersion de X par l'erreur quadratique moyenne σ de l'observation.

Supposons maintenant que nous voulions déterminer J, c'est-à-dire déterminer les valeurs de X discernables à travers une variable T ne pouvant prendre que n valeurs, et elle-même liée à X par l'intermédiaire de Z. Lier à Z une variable T ne pouvant prendre que n valeurs distinctes revient à mettre $q(z)$ sous la forme

$$q(z) = \sum_{k=1}^{n} r_k q_k(z) \qquad \left[\int q_k(z)\,dz = 1\right].$$

Supposons que $n = 2$. Nous pouvons définir la liaison entre Z et T par deux fonctions $\rho_1(z)$, $\rho_2(z)$ telles que

$$\rho_1(z) + \rho_2(z) = 1, \qquad \rho_1(z) \geqq 0, \qquad \rho_2(z) \geqq 0.$$

Nous aurons alors la décomposition

$$q(z) = r_1 q_1(z) + r_2 q_2(z),$$

avec

$$r_1 = \int q(z)\,\rho_1(z)\,dz, \qquad r_1 q_1(z) = q(z)\,\rho_1(z);$$

$$r_2 = \int q(z)\,\rho_2(z)\,dz, \qquad r_2 q_2(z) = q(z)\,\rho_2(z).$$

La fonction de distribution de X quand T = 1 sera

$$p_1(x) = \int q_1(z)\, q(z;\, x)\, dz = \frac{1}{r_1} \int \rho_1(z)\, \mathrm{P}(x;\, z)\, dz,$$

$\mathrm{P}(x;\, z)$ étant la loi liant X et Z. De même, si T = 2, nous aurons

$$p_2(x) = \frac{1}{r_2} \int \rho_2(z)\, \mathrm{P}(x;\, z)\, dz.$$

Nous avons naturellement

$$r_1 p_1(x) + r_2 p_2(x) = p(x).$$

La quantité J a alors pour expression :

$$\mathrm{J} = r_1 \int p_1(x) \log \frac{p_1(x)}{p(x)}\, dx + r_2 \int p_2(x) \log \frac{p_2(x)}{p(x)}\, dx.$$

Modifions légèrement $\rho_1(z)$ en lui ajoutant, dans un intervalle infiniment petit Δ, situé autour du point ξ, une quantité ε telle que

$$\rho_1(\xi) + \varepsilon > 0, \qquad \rho_2(\xi) - \varepsilon > 0.$$

Nous aurons alors

$$\delta r_1 = \varepsilon \Delta q(\xi), \qquad \delta r_2 = -\delta r_1,$$

$$\delta p_1(x) = \frac{\varepsilon \Delta}{r_1} [\mathrm{P}(x,\, \xi) - p_1(x)\, q(\xi)],$$

$$\delta \int p_1(x) \log \frac{p_1(x)}{p(x)}\, dx = \frac{\varepsilon \Delta}{r_1} \int [\mathrm{P}(x,\, \xi) - p_1(x)\, q(\xi)] \log \frac{p_1(x)}{p(x)}\, dx,$$

$$\delta \mathrm{J} = \varepsilon \Delta q(\xi) \int \left[p_1(x) \log \frac{p_1(x)}{p(x)} - p_2(x) \log \frac{p_2(x)}{p(x)} \right] dx$$

$$+ \varepsilon \Delta \int [\mathrm{P}(x,\, \xi) - p_1(x)\, q(\xi)] \log \frac{p_2(x)}{p(x)}\, dx$$

$$- \varepsilon \Delta \int [\mathrm{P}(x,\, \xi) - p_1(x)\, q(\xi)] \log \frac{p_2(x)}{p(x)}\, dx$$

$$= \varepsilon \Delta \int \mathrm{P}(x,\, \xi) \log \frac{p_1(x)}{p_2(x)}\, dx.$$

Nous voyons que si $p_1(x) = p_2(x)$, nous ne pouvons pas faire varier J d'une quantité de l'ordre de $\varepsilon\Delta$ par le procédé indiqué. Mais nous savons alors que J = 0 et a atteint son minimum. Le maximum sera donc atteint en général lorsque l'opération indiquée sera impossible, c'est-à-dire lorsque l'on aura toujours soit $\rho_1(z) = 0$, soit $\rho_1(z) = 1$ et que *de plus* la quantité

$$\int \mathrm{P}(x,\, \xi) \log \frac{p_1(x)}{p_2(x)}\, dx$$

sera du signe de $\rho_1(\xi) - \rho_2(\xi)$, c'est-à-dire positive ou nulle si $\rho_1(\xi) = 1$, négative ou nulle si $\rho_1(\xi) = 0$.

Si dans l'exemple considéré, où il s'agit de variables normales, nous prenons

$$\begin{aligned} \rho_1(z) &= 1 \quad \text{pour} \quad z < 0, \\ \rho_1(z) &= 0 \quad \text{»} \quad z > 0, \end{aligned}$$

nous aurons évidemment pour $p_1(x)$ et $p_2(x)$ deux fonctions de x symétriques, telles que

$$p_1(x) = p_2(-x).$$

Comme $\mathrm{P}(x, \xi)$ est symétrique en x, nous aurons par conséquent

$$\int \mathrm{P}(x, \xi) \log p_1(x)\, dx = \int \mathrm{P}(x, \xi) \log p_1(x)\, dx.$$

Nous sommes dans les conditions du maximum.

Si σ est petit, nous avons à peu près

$$\begin{aligned} p_1(x) &= 2p(x) \quad \text{pour} \quad x < 0, \\ p_1(x) &= 0 \quad \text{»} \quad x > 0, \end{aligned}$$

de sorte que J atteint presque sa valeur maximum $\log 2$.

La recherche du maximum de J présente un grand intérêt théorique. Nous avons vu en effet que lorsqu'on cherchait à lier X et Z le plus fortement possible, Z étant astreint à ne prendre que n valeurs, on était amené à associer ces n valeurs à n régions, équiprobables, de domaine où variait X. Mais rien ne nous indiquait comment découper ces régions : rien ne nous imposait de prendre n intervalles consécutifs plutôt que n ensembles tout à fait quelconques $\left(\text{à part d'être de mesure } \frac{1}{n} \text{ dans la densité de probabilité considérée}\right)$. Nous établissions ainsi une relation fonctionnelle entre X et Z, de la forme

$$\mathrm{Z} = f(\mathrm{X}),$$

de manière que

$$\mathrm{Prob}\,\{\mathrm{Z} = k\} = \frac{1}{n},$$

mais, cette relation fonctionnelle étant choisie d'après une règle ne faisant intervenir que des densités de probabilité, nous n'avions aucun moyen de lui associer une continuité quelconque.

Si au contraire nous ne découpons pas en domaines équiprobables le domaine de X, mais le domaine de Z, supposé liée à X, avec la condition que la fonction

$$T = \varphi(Z)$$

satisfaisant à

$$\text{Prob}\{T = k\} = \frac{1}{n}$$

soit le plus fortement liée à X (et non plus à Z), nous sommes amenés à choisir parmi tous les découpages possibles. A une valeur de T correspondent des valeurs de Z et de X qui sont fortement liées, et par conséquent nous respectons une certaine continuité. On peut donner une figuration géométrique aux considérations qui précèdent.

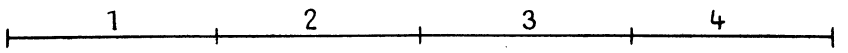

Fig. 1.

Soit X une variable à densité de probabilité uniforme dans le segment (0, 1). Si nous voulons « représenter » cette variable par un paramètre prenant quatre valeurs, nous pouvons partager le segment en quatre parties égales, et les numéroter de 1 à 4 (*fig.* 1). Si l'on donne la valeur du paramètre, il en résultera que nous connaîtrons la position de X avec une précision quatre fois plus grande puisque X ne sera inconnue que sur une longueur de 0,25 au lieu de l'être sur un segment 1. Si nous partageons le segment en quatre ensembles numérotés de 1 à 4, par exemple selon la figure 2 :

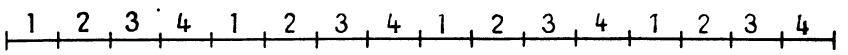

Fig. 2.

la donnée du paramètre nous donnera un renseignement de même valeur, du point de vue auquel nous nous sommes placés.

Supposons maintenant que nous étudions X par l'intermédiaire d'une autre variable Z, et que le point X, Z puisse occuper au hasard une position quelconque dans la bande diagonale d'un carré, comme il est représenté sur la figure 3. Cherchons maintenant à définir

une fonction à valeurs entières de Z, prenant seulement quatre valeurs distinctes, de manière que la donnée de cette fonction réduise au maximum le champ de variation de X. Nous voyons qu'une subdivision du type indiqué sur la figure 1 donne une réduction plus grande qu'une subdivision du type indiqué sur la figure 2. Donc, le fait d'observer une variable X à travers une variable Z permet de distinguer, parmi les différentes manières de partager en cellules équiprobables le champ de variation de Z certains modes optimum quant à l'observation de X. Il y a donc là ouverte une voie pour donner une approximation de X par une variable aléatoire susceptible de valeurs discrètes. La définition de cette approximation, il faut le noter, reste invariante par transformation biunivoque de X et Z (considérés séparément)

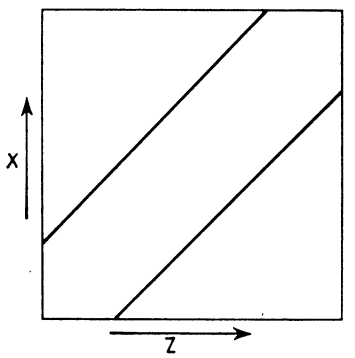

Fig. 3.

ne touchant pas à la liaison entre X et Z. En effet, les calculs qui mènent à cette approximation ne font intervenir que des rapports de densité de probabilité qui restent eux-mêmes invariants.

3.3.4. Approximation d'une variable aléatoire. Limite de la précision nécessaire a la réception d'une information. — Si la voie que nous avons esquissée dans ce qui précède pour approcher une variable par une série de valeurs discrètes n'était applicable qu'à des variables aléatoires, on pourrait penser, avec raison, qu'il y a là beaucoup d'efforts dépensés pour un problème qui admet d'autres solutions plus simples et aussi satisfaisantes. Par exemple on peut approcher une variable aléatoire en numérotant ses centiles. Le numéro du centile est une variable aléatoire

liée à la variable considérée, et qui, dans la pratique, jouerait à peu près le même rôle que toute variable définie par le procédé ci-dessus exposé, pourvu que la liaison entre X et Z ne soit pas choisie trop compliquée. Mais la voie que nous avons indiquée s'applique à un élément abstrait quelconque, pour lequel on peut définir une densité de probabilité.

Reprenons en effet les considérations faites plus haut, en désignant par X, Z deux points dans des espaces à un nombre quelconque de dimensions, et par dx, dz les mesures des éléments de volume dans ces espaces. Nous pourrons, dans des conditions assez générales, définir des densités de probabilités liées :

$$p(x)\,p(x;\,z) = q(z)\,q(z;\,x)$$

et définir le nombre de valeurs de X discernables à travers Z comme le nombre dont le logarithme est

$$\mathrm{I}(\mathrm{X},\,\mathrm{Z}) = \int p(x)\,p(x;\,z)\,\log\frac{p(x;\,z)}{q(z)}\,dx\,dz.$$

Si la valeur de Z est connue, la fonction de distribution de X a pour densité $q(z;\,x)$. Si la valeur de Z n'est pas connue exactement, on sait seulement que Z est dans un domaine ω, la densité de X devient

$$\mathrm{P}_\omega(x) = \frac{\int_\omega q(z)\,q(z;\,x)\,dz}{\int_\omega q(z)\,dz}.$$

La substitution de $q(z;\,x)$ à $p(x)$ constitue une information. Celle de $\mathrm{P}_\omega(x)$ à $p(x)$ constitue une information de qualité moindre; cette dernière information est d'autant plus précise que ω est plus petit. On comprend facilement qu'il existe un degré, dans la diminution de ω, où la valeur de l'information n'augmente plus beaucoup quand on fait tendre ω vers zéro. En effet, on a pour limite de $\mathrm{P}_\omega(x)$ une distribution $q(z;\,x)$ qui reste une distribution aléatoire. Il est intéressant de savoir jusqu'où il faut aller dans le partage du domaine de variation de Z, en cellules ω, pour arriver à un point ou un partage plus fin n'est plus « payant ». On peut reconnaitre là la généralisation immédiate du problème de la recherche de la variable T étudié plus haut. Comme des considérations trop générales dans ce domaine nous entraîneraient trop loin, nous allons traiter un exemple concret.

3.4. Application de la théorie de signaux a l'étude de la signification des spectres de signal et de bruit. — Remarquons que si une variable Z, à travers laquelle on observe X, est de la forme

$$Z = X + \Delta,$$

Δ étant indépendante de X, on peut mettre I sous la forme

$$I = \int p(x)\, p(x; z) \log \frac{p(x; z)}{q(z)}\, dz,$$

avec

$$p(x; z) = r(z - x) = r(u);$$

d'où on tire sans peine

$$I = -\int q(z) \log q(z)\, dz + \int r(u) \log r(u)\, du.$$

Cette remarque s'applique, *mutatis mutandis*, lorsque X et Z sont des vecteurs aléatoires à un nombre quelconque de coordonnées.

Soit maintenant $X(t)$ une fonction aléatoire, observable à travers une fonction aléatoire $Z(t)$ de la forme

$$Z(t) = X(t) + \Delta(t),$$

$\Delta(t)$ étant une fonction aléatoire indépendante de $X(t)$. Nous sommes dans le champ d'application de la remarque précédente.

Supposons $X(t)$ et $\Delta(t)$ stationnaires, et considérons une suite d'instants $t_1, t_2, \ldots, t_n$. Nous appellerons $\vec{X}$ le vecteur de coordonnées $X(t_1), X(t_2), \ldots, X(t_n)$. Ce vecteur est aléatoire à n dimensions. Nous définirons également $\vec{Z}$ et $\vec{\Delta}$ de la même manière. Appelant $d\bar{\omega}$ l'élément de volume dans l'espace décrit par $\vec{\Delta}$, cherchons à évaluer

$$J = -\int r\left(\vec{\Delta}\right) \log r\left(\vec{\Delta}\right) d\bar{\omega}.$$

Supposons que $\Delta(t)$ soit une fonction aléatoire gaussienne : $r\left(\vec{\Delta}\right)$ est la densité de probabilité d'une distribution gaussienne à n dimensions, laquelle est de la forme

$$A\, |\rho_{ij}|^{-\frac{1}{2}} e^{-\frac{Q}{2}},$$

en appelant $[\rho_{ij}]$ la matrice formée avec les valeurs moyennes

$$\rho_{ij} = \mathfrak{M}\, \Delta(t_i)\, \Delta(t_j).$$

A est une constante ne dépendant que du nombre de dimensions, Q est une forme quadratique par rapport aux variables $\Delta(t_i)$. Cette forme quadratique est de la forme

$$Q(\Delta_i \Delta_j) = \sum \rho_{ij}^{*} \Delta_i \Delta_j,$$

la matrice ρ_{ij}^{*} étant inverse de la matrice des ρ_{ij}.

L'expression J peut donc s'écrire :

$$J = -\log A + \frac{1}{2} \log |\rho_{ij}| + \mathfrak{M} \frac{Q}{2},$$

le signe $\mathfrak{M}$ désignant la valeur moyenne. Nous avons évidemment

$$\mathfrak{M} Q = \sum_{ij} \rho_{ij}^{*} \rho_{ij} = \sum_{ij} \rho_{ij}^{*} \rho_{ji} = n,$$

d'où

$$J = -\log A + \frac{1}{2} \log |\rho_{ij}| + \frac{n}{2}.$$

Si nous faisons intervenir maintenant

$$r_{ij} = \mathfrak{M} Z_i Z_j = \mathfrak{M} Z(t_i) Z(t_j)$$

nous obtiendrons pour I l'expression

$$I = \frac{1}{2} \log |r_{ij}| - \frac{1}{2} \log |\rho_{ij}|.$$

Les termes en A et n ont disparu. Il nous reste à évaluer les déterminants intervenant dans l'expression de J. Supposons les instants t_i en progression arithmétique. Comme les ρ_{ij} (et les r_{ij}) ne dépendent que de la différence $t_j - t_i$, nous pouvons écrire

$$|\rho_{ij}| = \begin{vmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots \\ \rho_1 & \rho_0 & \rho_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{vmatrix},$$

avec

$$\rho_k = \mathfrak{M} \Delta(t) \Delta(t + k\theta),$$

θ étant l'écartement de deux t_i consécutifs.

On obtient une valeur approchée du déterminant ρ_{ij} en lui substituant un déterminant circulant. Nous supposerons pour cela n impair.

$$n = 2m + 1$$

et substituerons à $|\rho_{ij}|$ le déterminant circulant défini par les nombres

$$\rho_m,\quad \rho_{m-1},\quad \ldots,\quad \rho_1,\quad \rho_0,\quad \ldots,\quad \rho_m.$$

Il est connu que la valeur d'un tel déterminant est

$$\prod_{k=0}^{n-1} F(\omega^k), \qquad \omega = e^{-\frac{2\pi i}{n}},$$

avec

$$F(\omega) = \rho_m + \rho_{m-1}\omega + \ldots + \rho_0\omega^m + \ldots + \rho_m\omega^{2m+1}.$$

Considérons l'expression

$$\varphi(f) = \theta \sum_{k=0}^{n-1} \rho_{k-m} \exp\{-2\pi i f(k-m)\theta\}.$$

Nous pouvons exprimer $F(\omega)$ sous la forme

$$F(\omega) = \omega^m \frac{1}{\theta}\varphi\left(\frac{1}{n\theta}\right),$$

$$F(\omega^k) = \omega^{mk} \frac{1}{\theta}\varphi\left(\frac{k}{n\theta}\right),$$

ce qui nous conduit à l'expression approchée suivante pour $\log[\rho_{ij}]$:

$$\log[\rho_{ij}] = n\log\frac{1}{\theta} + \sum_{k=0}^{n-1} \log\varphi\left(\frac{k}{n\theta}\right).$$

En ce qui concerne la variable $\vec{Z}$, nous aurons un calcul analogue à faire. Posant

$$\psi(f) = \theta \sum_{k=0}^{n-1} r_{k-m} \exp\{-2\pi i f(k-m)\theta\},$$

nous obtiendrons

$$\log[r_{ij}] = n\log\frac{1}{\theta} + \sum_{k=0}^{n-1} \log\psi\left(\frac{k}{n\theta}\right)$$

et par conséquent

$$I = \frac{1}{2}\sum_{k=0}^{n-1} \log\psi\left(\frac{k}{n\theta}\right) - \frac{1}{2}\sum_{k=0}^{n-1} \log\varphi\left(\frac{k}{n\theta}\right).$$

Laissant inchangées les valeurs t_1 et t_n, faisons tendre n vers l'infini.

Nous supposerons par exemple que $t_1 = 0$, $t_n = T$; nous voyons alors que si nous posons

$$r(u) = \mathfrak{M}\, Z(t)\, Z(t+u),$$
$$\rho(u) = \mathfrak{M}\, \Delta(t)\, \Delta(t+u),$$

$\varphi(f)$ et $\psi(f)$ ont les limites :

$$\varphi(f) = \int_{-\frac{T}{2}}^{+\frac{T}{2}} \rho(u)\, e^{-2\pi i f u}\, du,$$

$$\psi(f) = \int_{-\frac{T}{2}}^{+\frac{T}{2}} r(u)\, e^{-2\pi i f u}\, du$$

et que

$$\frac{1}{T} I = \frac{1}{2T} \sum_{k=0}^{n-1} \log \left\{ \frac{\psi\left(\frac{k}{T}\right)}{\varphi\left(\frac{k}{T}\right)} \right\}.$$

Si maintenant nous faisons tendre T vers l'infini, nous voyons que

$$\lim_{T=\infty} \frac{I}{T} = \frac{1}{2} \int_0^\infty \log \frac{\psi(f)}{\varphi(f)}\, df.$$

Nous pouvons encore transformer cette formule en tenant compte de la forme de Z. Faisons intervenir

$$R(u) = \mathfrak{M}\, X(t)\, X(t+u) :$$

Nous aurons évidemment

$$r(u) = \rho(u) + R(u)$$

et, en posant

$$\Phi(f) = \int e^{-2\pi i f u} R(u)\, du,$$

il nous vient

$$\lim_{T=\infty} \frac{I}{T} = \frac{1}{2} \int_0^\infty \log \left[1 + \frac{\Phi(f)}{\varphi(f)} \right] df = \mu.$$

Nous pouvons ainsi, si nous considérons des observations durant des temps assez longs, dire que la fonction $X(t)$, observée à travers $Z(t)$, si elle est observée pendant T unités de temps, est susceptible de prendre un nombre de formes discernables dont le logarithme est μT. Si donc cette fonction $X(t)$ est chargée de porter un message, et qu'elle soit couverte par un bruit $\Delta(t)$ on pourra transmettre, pendant T secondes,

un nombre de messages de l'ordre de $e^{\mu T}$. Pour voir à quoi correspond ce chiffre, comparons un système de transmission utilisant la fonction $X(t)$ [avec le bruit $\Delta(t)$] à un système de transmission télégraphique envoyant à cadence de ν par seconde, des signaux susceptibles de deux polarités.

Dans le système télégraphique pris comme étalon, un message de durée T est susceptible de $2^{\nu T}$ formes différentes. Donc la fonction $X(t)$ est analogue à un pareil système si

$$2^{\nu T} = e^{\mu T},$$

c'est-à-dire si

$$\nu = \frac{\mu}{\log 2}.$$

Nous pouvons conclure en disant que si une fonction $X(t)$ a une fonction de corrélation $r(t)$ dont le spectre est $\Phi(f)$, et si elle est couverte par un bruit dont la fonction de corrélation a le spectre $\varphi(f)$, cette fonction peut être assimilée, du point de vue du nombre de formes discernables à travers le bruit, à un système télégraphique envoyant des signaux bivalents à raison de

$$\nu = \frac{1}{2}\int_0^\infty \log_2 \left| 1 + \frac{\Phi(f)}{\varphi(f)} \right| df$$

signaux pendant l'unité de temps.

BIBLIOGRAPHIE.

BERGE (C.) :

[1] *Sur une théorie ensembliste des jeux alternatifs* (*Thèse*) (*J. Math. pures et appl.*, t. 32, 1953, p. 129-184).

BOREL (E.) :

[1] *La théorie du jeu et les équations intégrales à noyau symétrique gauche* (*C. R. Acad. Sc.*, t. 173, 1921, p. 1304-1308).

[2] *Sur les jeux où interviennent le hasard et l'habileté des joueurs* (*Association française pour l'avancement des sciences*, 1923, p. 79-85).

[3] *Éléments de la théorie des probabilités*, 3e édit., Note IV, p. 204-221 ; *Sur les jeux où interviennent le hasard et l'habileté des joueurs*, Hermann, Paris, 1924.

[4] *Traité du Calcul des probabilités et de ses applications*, t. IV, fasc. II (*Applications aux jeux de hasard*, rédigé par J. VILLE, Gauthier-Villars, Paris, 1938). [1] et [3] traduits en Anglais par L. J. SAVAGE, *Econometrica*, vol. 21, n° 1, 1953, p. 97-115.

FORTET (R.) :

[1] *Processus stationnaire et de Markoff et entropie* (*La Cybernétique*, Paris, 1951, p. 9).

GUILBAUD (G. TH.) :

[1] *La théorie des jeux. Contributions critiques à la théorie de la valeur. Econométrie appliquée*, vol. 2, 1949. p. 275-319.

KHINTCHINE (A.) :

[1] *Notion d'entropie dans la théorie des probabilités* (*U.M. Nauk, S.S.S.R.*, t. 8, 1953, p. 3).

NEUMANN (J. VON) :

[1] *Zur Theorie der Gesellschaftsspiele* (*Math. Ann.*, vol. 100, 1928, p. 295-320).

NEUMANN (J. VON) et MORGENSTERN (O.) :

[1] *Theory of games and economic behavior*, Princeton University Press, Princeton, 1943 (2e édit., 1947).

POSSEL (R. DE) :

[1] *Sur la théorie mathématique des jeux de hasard et de réflexion* (*Actualités scientifiques et industrielles*, Hermann, Paris, 1936).

SHANNON (C. E.) :

[1] *A mathematical theory of communications* (*Bell Syst. Thec. J.*, vol., 27, 1948, n° 3, p. 399-429 et n° 4, p. 623-656).

VILLE (J.) :

[1] *Sur la théorie générale des jeux où intervient l'habileté du joueur* (*Traité du Calcul des probabilités et ses applications*, t. IV, fasc. II, Note, Gauthier-Villars, Paris, 1938).