

J. BOUZITAT, D. DUGUÉ, R. FORTET,  
A. FUCHS, F. GENUYS, G.-Th. GUILBAUD,  
A. HUISMAN, E. MOURIER, J. VILLE

---

# Le calcul des probabilités et l'enseignement

**Les brochures de l'A.P.M.**

**4**

### EXTRAIT DES STATUTS

Article II. — L'Association a pour but l'étude des questions intéressant l'enseignement des Mathématiques et la défense des intérêts professionnels de ses membres. Elle institue ou encourage des réunions, des discussions, des enquêtes sur l'enseignement des Mathématiques en France ou à l'étranger...

L'A.P.M. est ouverte à tous les Collègues enseignant dans les Facultés, les Grandes Ecoles, les Lycées, les Collèges Classiques, Modernes ou Techniques, les Ecoles Nationales Professionnelles, les Cours Complémentaires ou les Centres d'Apprentissage.

COTISATION. — Elle comprend l'abonnement au Bulletin, ainsi que les fascicules d'énoncés.

Cotisation normale ..... 10 NF

Cotisation réduite (stagiaires C.P.R., élèves des E.N.S. et des I.P.E.S., jeunes gens accomplissant leur service militaire, retraités) ..... 5 NF

ABONNEMENT (personnes n'appartenant pas à l'Enseignement Public, bibliothèques, etc...) :

France et Communauté : 12 NF - Autres pays : 15 NF

Le numéro : 3 NF

MODE DE PAIEMENT : Virement postal (adressé au centre de chèques du tireur) ou mandat-carte à l'adresse :

A.P.M., 29, rue d'Ulm - PARIS, 5<sup>e</sup> - C.C.P. Paris 5708-21

RECOMMANDATIONS DU TRESORIER. — Indications à porter sur le talon du chèque : 1<sup>o</sup> Nom (en majuscules) et prénom. — 2<sup>o</sup> Adresse où doit être envoyé le Bulletin. — 3<sup>o</sup> Ancienne adresse en cas de changement. — 4<sup>o</sup> Nom de l'établissement où l'on exerce. — 5<sup>o</sup> Nom de l'établissement précédent en cas de mutation en fin d'année scolaire.

**N. B.** — Toute nouvelle adhésion demandée en cours d'année scolaire compte à partir du 1<sup>er</sup> octobre précédent. Elle donne droit à tous les bulletins déjà parus au cours de l'année scolaire, sous réserve qu'ils ne soient pas épuisés.

J. BOUZITAT, D. DUGUÉ, R. FORTET,  
A. FUCHS, F. GENUYS, G.-Th. GUILBAUD,  
A. HUISMAN, E. MOURIER, J. VILLE

---

*[Bareil Henri  
-  
déc. 1961  
-*

# Le calcul des probabilités et l'enseignement

Association des Professeurs  
de Mathématiques de l'Enseignement Public  
PARIS — 1961

« Afin que, ... aux dépens d'autrui  
Sage, je m'enseignasse. »

REGNIER.

« C'est assez désagréable... de ne pouvoir plus rien  
apprendre pour toute la vie ! Nos aïeux s'en tenaient aux  
enseignements qu'ils avaient reçus dans leur jeunesse :  
mais, nous, il nous faut recommencer tous les cinq ans, si  
nous ne voulons pas être complètement démodés. »

GÆTHER (Les affinités électives).

**Les brochures de l'A.P.M.** mettent à la disposition des professeurs des textes utiles à l'enseignement.

Ou bien ces textes sont inédits, ou bien ils ont déjà paru, soit dans le Bulletin de l'A.P.M., soit ailleurs. Dans tous les cas, il a paru intéressant de regrouper des écrits sous une forme commode pour les maîtres qui auront à s'en servir.

#### **Brochures parues :**

1. Le langage simple et précis des mathématiques modernes, par A. REVUZ et L. LESIEUR, Professeurs à la Faculté des Sciences de Poitiers (avril 1960), (épuisé).
2. Congruences Paratactiques de cycles, par Paul ROBERT, Inspecteur général honoraire de l'Instruction Publique (avril 1960).
3. Recherche d'une axiomatique commode pour le premier enseignement de la géométrie élémentaire, par Gustave CHOQUET, Professeur à la Sorbonne (février 1961).
4. Le calcul des probabilités et l'enseignement, par A. HUISMAN, R. FORTET, E. MOURIER, A. FUCHS, D. DUGUE, G.-T. GUILBAUD, J. BOUZITAT, J. VILLE et F. GENUYS (novembre 1961).
5. L'enseignement de la mécanique, par P. GERMAIN, J. KAMPE DE FERIET et R. MAZET (novembre 1961).

#### **Brochures en préparation :**

6. L'algèbre et l'enseignement. Tome I : groupes, anneaux et corps, par André et Germaine REVUZ.
7. L'enseignement de l'astronomie.
8. Le cinéma dans l'enseignement des mathématiques.

## AVERTISSEMENT

« Nous errons dans les ténèbres, ou nous marchons avec perplexité entre des préjugés et des probabilités. »

BUFFON.

*Lorsque la Société Mathématique de France et l'Association des Professeurs de Mathématiques de l'Enseignement Public conjuguèrent leur action pour organiser des conférences à l'intention des professeurs, elles pensèrent d'abord aux sujets qui font l'objet de l'enseignement élémentaire : algèbre, topologie, mesure des grandeurs. Traiter, ensuite, du calcul des probabilités, ce n'est plus voir l'enseignement tel qu'il est, mais, peut-être, tel qu'il devrait être. Ce n'est plus parfaire la formation des professeurs, mais, pour beaucoup d'auditeurs, entreprendre leur initiation, c'est ouvrir, pour la plupart, de vastes aperçus sur les domaines si variés de l'activité humaine, des jeux à la prévision économique, où l'on entreprend d'analyser l'aléatoire.*

*Dans ces conditions, on pouvait craindre que le sujet fût trop éloigné de ce qui s'enseigne dans les classes, que l'attention des auditeurs fût donc moins grande, leur assiduité plus relâchée. Il n'en fut rien, bien au contraire. La curiosité était plus vive. Par le fait de la diversité des sujets traités, l'intérêt d'une conférence à l'autre était renouvelé. D'octobre 1958 à juin 1959, au cours d'un cycle de plus de dix conférences, l'auditoire montra la même ardeur attentive qui fut, pour les organisateurs, la meilleure des satisfactions.*

*Il est évident que le mérite en revient d'abord aux conférenciers. Ce qui ne s'explique pas seulement par leur évidente qualification. Il faut aussi tenir compte, disons, de leur éloquence, cette flamme qui témoignait de leur amour de la science et de leur talent pédagogique. Il nous semble que cela se sent encore dans la rédaction...*

*Pour donner à la réunion de ces textes toute sa portée, nous avons demandé à notre Collègue Huisman, qui avait pris la plus grande part à l'organisation du cycle de conférences, de préciser quelle pouvait être la place des probabilités dans l'enseignement du Second Degré. Le sujet n'est pas nouveau, mais le problème n'a pas encore trouvé sa solution. On peut penser, maintenant, que le premier programme réalisable en a été rédigé.*

*Puisse cette réalisation ne plus tarder ! Les auteurs de cette brochure y trouveront la récompense de leurs efforts.*

La Rédaction de l'A.P.M.

## TABLE DES MATIÈRES

André HUISMAN :	
« Le bon sens réduit au calcul » .....	5
Exercices .....	26
Bibliographie .....	28
Robert FORTET :	
Le calcul des probabilités, les grandes lignes de son développement et ses principaux champs d'application	31
Edith MOURIER :	
Quelques grandes théories mathématiques intervenant dans le calcul des probabilités .....	43
Aimé FUCHS :	
Introduction à l'étude des processus stochastiques ....	49
Daniel DUGUE :	
Statistique mathématique, plans d'expériences .....	59
Georges-Th. GUILBAUD :	
La Mathématique des programmes économiques .....	67
Jean BOUZITAT :	
Quelques aspects théoriques et pratiques des jeux de stratégie .....	85
J. VILLE :	
La structure mathématique simple des grandes machines à calculer .....	123
François GENUYS :	
Les calculateurs électroniques et le développement des Mathématiques numériques .....	129
N.D.L.R. — Les textes réunis dans cette brochure ont fait l'objet de publications antérieures, soit, dans l'ordre de la table des matières :	
1) <i>Bull. de l'A.P.M.</i> , n° 215 (juin 1961). — 2) <i>Bull. A.P.M.</i> , n° 198 (mars 1959). — 3) <i>Bull. A.P.M.</i> , n° 199 (juin 1959). — 4) <i>Bull. A.P.M.</i> , n° 205 (janv.-février 1960). — 5) <i>Bull. A.P.M.</i> , n° 206 (mars 1960). — 6) <i>Science et action économique</i> (B.U.R.O.). — 7) <i>Science et action économique</i> (B.U.R.O.) et <i>Bull. A.P.M.</i> , n° 201 (oct.-novembre 1959). — 8) et 9) <i>Bull. A.P.M.</i> , n° 213 (décembre 1960).	

## « LE BON SENS REDUIT AU CALCUL »

par André HUISMAN, *professeur au Lycée Montaigne (Paris)*

« On peut même dire, à parler en rigueur, que presque toutes nos connaissances ne sont que probables ; et, dans le petit nombre des choses que nous pouvons savoir avec certitude, dans les sciences mathématiques elles-mêmes, les principaux moyens de parvenir à la vérité, l'induction et l'analogie, se fondent sur les probabilités ; en sorte que le système entier des connaissances humaines se rattache à la théorie exposée dans cet essai. »

(LAPLACE, *Essai philosophique sur les probabilités*).

\*  
\*\*

En 1783, C.-F. de Biquilley, Garde du Corps du Roi, publie une étude (rééditée en 1805), intitulée « Du calcul des probabilités ». Dans sa préface, l'auteur observe : « La théorie des probabilités, ébauchée par des géomètres célèbres, m'a paru susceptible d'être approfondie et de faire partie de l'enseignement élémentaire. » (\*).

Voilà qui témoigne d'un enthousiasme bien sympathique ; il rappelle celui de La Fontaine pour Baruch. Convenons cependant qu'il était quelque peu prématuré.

Trente ans plus tard, Laplace termine son *Essai* par une remarque analogue, mais avec plus d'autorité :

« On voit par cet essai que la théorie des probabilités n'est au fond que le bon sens réduit au calcul ; elle fait apprécier avec certitude ce que les esprits justes savent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte. Elle ne laisse rien d'arbitraire dans le choix des opinions et des partis à prendre, toutes les fois que l'on peut, à son moyen, déterminer le choix le plus avantageux. Par là, elle devient le supplément le plus heureux à l'ignorance et à la faiblesse de l'esprit humain. Si l'on considère les méthodes analytiques auxquelles cette théorie a donné naissance, la vérité des principes qui lui servent de base, la logique fine et délicate qu'exige leur emploi dans la solution des problèmes, les établissements d'utilité publique qui s'appuient sur elle, et l'extension qu'elle a reçue et qu'elle peut recevoir encore par son application aux questions les plus importantes de la philosophie naturelle et des sciences morales ; si l'on observe ensuite que, dans les choses mêmes qui ne peuvent être soumises au calcul, elle donne les aperçus les plus sûrs qui puissent nous guider dans nos jugements, et qu'elle apprend à

---

(\*) Avoient traité le même sujet, avant le Garde du Corps, entre autres : Pascal, Fermat, Huygens, Jacques Bernoulli, de Moivre, de Montmort, Bayes.

se garantir des illusions qui souvent nous égarent, on verra qu'il n'est point de science plus digne de nos méditations, et qu'il soit plus utile de faire entrer dans le système de l'instruction publique. » (\*).

Plus près de nous, en 1923, Emile Borel reprenait, sans se faire beaucoup d'illusions, cette idée vieille de près d'un siècle et demi :

« Le calcul des probabilités est une des branches les plus attrayantes et les moins ardues de la Mathématique. C'est simplement pour des raisons de tradition, l'on n'ose écrire de routine, que les éléments de ce calcul ne figurent pas aux programmes de l'enseignement secondaire, où ils remplaceraient avantageusement bien des matières qui y subsistent, pour le seul motif que personne ne se donne la peine de les supprimer. »

Enfin, plus récemment encore, en 1937, dans leur « Préparation à l'étude des probabilités », MM. Leconte et Deltheil, tout en se défendant de vouloir alourdir des programmes déjà fort encombrés, insistaient sur l'intérêt qu'il y aurait à utiliser, au moins à titre d'exercices, les notions de probabilité :

« A un point de vue précisément pédagogique, les exercices que fournit le calcul des probabilités sont en nombre illimité. Ils contribuent à renouveler l'intérêt des calculs en arithmétique et en algèbre. Ils sont faciles à graduer. Même très simples, ils ne se résolvent le plus souvent qu'à condition de ne pas cesser d'employer les nombres avec attention et réflexion, de sorte qu'ils apportent de nouveaux moyens de lutte contre l'appel à la mémoire et l'automatisme dont le rôle est si fâcheux dans l'étude des sciences exactes. »

A part une timide apparition des probabilités dans la classe de Sciences Expérimentales (et, bien entendu, les sections techniques spécialisées), on ne peut pas dire que ces divers plaidoyers aient été écoutés. Il faut reconnaître d'ailleurs que, s'il est né de questions très accessibles relatives aux jeux de hasard, le calcul des probabilités a tout d'abord trouvé son champ d'activité dans des questions très techniques, notamment en physique, dans la théorie cinétique des gaz et la mécanique statistique, ce qui ne semblait pas justifier son introduction dans un enseignement de culture générale.

Participant ensuite au mouvement général des Mathématiques vers la rigueur axiomatique, bénéficiant de l'introduction de la notion de mesure, des progrès de la théorie des ensembles et de celle des espaces abstraits, la théorie des probabilités, selon la remarque de Bourbaki, « autrefois prétexte à devinettes et à paradoxes, est devenue une branche de la théorie de l'intégration depuis son axiomatisation par Kolmogoroff ».

Il est cependant permis de penser que ce jugement est un peu brutal. D'une part, les « récréations mathématiques » ne sont pas méprisables si l'on songe à un enseignement d'initiation. D'autre part, l'évolution à laquelle nous assistons depuis le début de ce siècle montre le rôle sans

---

(\*) L'Essai a été publié en 1814, mais il provient de leçons données par Laplace aux Ecoles Normales en 1795.

cesse accru des probabilités dans l'étude des phénomènes économiques, biologiques, sociologiques. On assiste, en somme, à un retour du calcul des probabilités vers ses origines, selon le programme magistralement esquissé par Jacques Bernoulli dans son « *Ars coniectandi* » (1713) :

« Pour ce qui est sûr et hors de doute, nous parlons de connaissance et de compréhension ; pour tout le reste, nous disons seulement conjecture et opinion. Conjecturer quelque chose, c'est mesurer son degré de probabilité. Ainsi, le Savoir Conjecturer, ou Stochastique, se définit pour nous comme savoir mesurer, le plus exactement possible, les degrés de probabilités, afin que, dans nos décisions et nos actions, nous puissions toujours choisir ou accepter ce qui nous aura paru meilleur, plus satisfaisant, plus sûr, plus prudent, seul objet à quoi s'applique toute la sagesse du Philosophe, toute la prévoyance du Politique. » (\*).

Nous étant ainsi abrités derrière des autorités incontestées, concluons qu'on peut estimer que l'introduction des probabilités dans l'enseignement du second degré est, plus que jamais, souhaitable. Préparés dès le début des études par des exercices gradués de dénombrement (qui aboutiraient progressivement aux éléments de l'analyse combinatoire), et par des notions sur les ensembles et la logique, les élèves pourraient, sans grand effort semble-t-il, aborder au terme du second cycle une étude des principales méthodes, tout au moins en se bornant à des ensembles finis d'événements.

C'est une telle étude que nous essayons de résumer dans l'exposé fort classique qui suit, à titre d'introduction aux savantes conférences organisées, en 1958-59, par la Société Mathématique de France et l'A.P.M.

## I. — Généralités.

On procède à une expérience, qui donne lieu à un nombre fini d'événements formant un ensemble  $E$ . Par exemple, si l'on jette une pièce, on a :  $E = (\text{pile, face})$ . Si l'on jette un dé cubique, on a :  $E = (1, 2, 3, 4, 5, 6)$ .

A chaque événement de  $E$  doit être attachée une mesure appelée probabilité, et c'est sur le choix de cette mesure que reposeront tous les calculs ultérieurs. Pour un enseignement d'initiation, on peut sans doute s'en tenir à la définition classique de Laplace, malgré ses défauts :

*La probabilité d'un événement est le rapport du nombre de cas favorables (à l'événement) au nombre total des cas possibles, supposés également possibles.*

Une première critique qu'on a faite à cette définition, c'est de ne pas s'appliquer au cas où l'ensemble des événements est infini. Elle ne nous arrête pas, puisque nous n'avons à étudier que des ensembles finis.

La deuxième critique concerne l'égalité des probabilités des événements élémentaires, égalité qui est utilisée sans être définie.

Si l'expérience envisagée consiste à lancer un dé cubique, et si nous savons que ce dé a été fabriqué correctement, de manière à n'être pas

---

(\*) Voir G.-Th. Guilbaud : *Leçons sur les éléments principaux de la théorie mathématique des Jeux.*

déséquilibré (notamment par le creusement des points sur les faces), il nous paraît justifié d'admettre que chaque face a la même probabilité de se montrer après le lancement du dé, soit  $1/6$ .

De toute façon, il nous faut une hypothèse pour bâtir nos calculs, et cette hypothèse pourra être soumise au contrôle de l'expérience qui décidera finalement si l'hypothèse peut, ou non, être conservée. Ce contrôle, c'est la loi des grands nombres qui nous en donnera les moyens.

En définitive, on voit que notre attitude n'a rien d'insolite ; elle est celle qu'on adopte en toute science, lorsqu'on veut soumettre une hypothèse au contrôle expérimental.

En résumé, jusqu'à plus ample informé, nous admettrons que, dans le cas d'une pièce, pile et face ont chacun la probabilité  $1/2$  de se réaliser, et que, dans le cas d'un dé, chacun des six points a la probabilité  $1/6$  d'être obtenu. Ce sont là, si l'on veut, les définitions d'une pièce ou d'un dé « parfaits » et lancés « loyalement ».

Si l'on jette un dé deux fois de suite (ou deux dés parfaits en une seule fois), l'ensemble des événements sera le produit cartésien  $E \times E$ ,  $E$  étant l'ensemble  $\{1, 2, 3, 4, 5, 6\}$ , obtenu avec un seul dé (fig. 1). D'où 36 événements de même probabilité  $1/36$ .

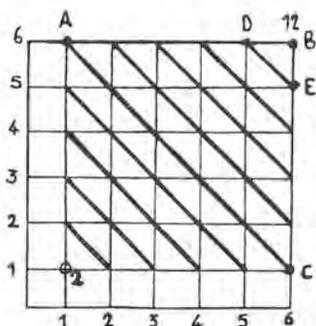


Fig. 1

On peut partager  $E \times E$  en sous-ensembles, en plaçant dans un même sous-ensemble tous les événements qui correspondent à une même somme des points montrés par les deux dés (somme qui varie de 2 à 12). Quant à la mesure attachée à chacun des sous-ensembles, elle résulte de la définition de Laplace, qui nous donne le tableau suivant :

Somme des points :	2	3	4	5	6	7	8	9	10	11	12
Nombre des cas :	1	2	3	4	5	6	5	4	3	2	1
Probabilités :	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

De la définition précédente de la probabilité résultent les remarques suivantes :

1) Si un phénomène ne se produit dans aucun des événements de l'ensemble (phénomène impossible), sa probabilité est nulle.

2) Si un phénomène se produit dans chacun des événements de l'ensemble (phénomène certain), sa probabilité est égale à 1.

Autrement dit, si l'on considère le sous-ensemble des événements dans lesquels se produit le phénomène étudié, ce sous-ensemble est l'ensemble vide dans le cas du phénomène impossible, la partie pleine dans le cas du phénomène certain.

3) Quel que soit le phénomène étudié, sa probabilité est un nombre rationnel compris entre 0 et 1, lorsqu'on suppose, comme nous l'avons fait, que le nombre des événements possibles est fini.

4) Supposons que, dans l'ensemble E des événements possibles, on considère un sous-ensemble  $E_1$ . Par exemple, dans le cas du lancement de deux dés,  $E_1$  est le sous-ensemble des coups pour lesquels la somme S des points vérifie la condition :

$$5 \leq S \leq 8.$$

La réalisation de  $E_1$  a une probabilité  $\text{Pr}(E_1)$  qu'on déduit du tableau donné plus haut.

Supposons un autre partage de E, dans lequel on envisage le sous-ensemble  $E_2$  défini par :

$$6 \leq S \leq 8,$$

auquel correspond une probabilité  $\text{Pr}(E_2)$  ; on vérifie que l'on a :

$$\text{Pr}(E_2) < \text{Pr}(E_1)$$

et, d'une manière générale, on peut écrire :

$$E_2 \subset E_1 \Rightarrow \text{Pr}(E_2) < \text{Pr}(E_1).$$

## II. — Partition de E ; système complet d'événements.

Dans l'exemple des deux dés, le partage de E en sous-ensembles correspondant aux diverses valeurs de la somme des points réalise une partition de E : les sous-ensembles ne sont pas vides, ils sont deux à deux disjoints, et leur réunion constitue E.

D'une manière générale, quand on a réalisé une partition de l'ensemble E des événements en sous-ensembles  $E_1, E_2, \dots, E_n$ , on dit que les  $E_i$  constituent un système complet d'événements. Les propriétés d'un tel système sont donc :

1) les événements  $E_i$  sont deux à deux incompatibles, c'est-à-dire que si l'un d'eux se réalise, aucun des autres ne peut se réaliser en même temps ;

2) l'un des événements du système se réalise certainement, ce qu'on exprime en disant que l'événement : «  $E_1$  ou  $E_2$  ou ... ou  $E_n$  » est certain.

Un dénombrement facile montre que l'on a :

$$\text{Pr}(E_1 \text{ ou } E_2 \text{ ou } \dots \text{ ou } E_n) = \sum_{i=1}^n \text{Pr}(E_i) = 1.$$

D'une manière plus générale, si A est une partie de E, et si  $A_1, A_2, \dots, A_p$  constituent une partition de A, on a :

$$\text{Pr}(A) = \sum_{i=1}^p \text{Pr}(A_i).$$

*Cas particulier : événements contraires.*

Supposons  $E$  partagé en deux sous-ensembles disjoints, autrement dit en un ensemble  $E_1$  et son complémentaire par rapport à  $E$ , noté  $\bar{E}_1$ . Ces deux événements constituent un système complet. Si l'on note  $E_1 \cup \bar{E}_1$  l'événement certain «  $E_1$  ou  $\bar{E}_1$  » et  $E_1 \cap \bar{E}_1$  l'événement impossible «  $E_1$  et  $\bar{E}_1$  », on peut écrire :

$$\Pr(E_1 \cup \bar{E}_1) = 1, \quad \Pr(E_1 \cap \bar{E}_1) = 0$$

(à rapprocher du principe du tiers exclu et du principe de contradiction, dans la logique des propositions).

Le système  $(E_1, \bar{E}_1)$  étant complet, on a :

$$\Pr(E_1) + \Pr(\bar{E}_1) = 1.$$

*Exemple.* — Avec deux dés, on a une telle partition avec :

$E_1$  : ensemble des coups pour lesquels la somme des points est strictement inférieure à 7 ;

$\bar{E}_1$  : ensemble des coups pour lesquels la somme des points est supérieure ou égale à 7.

*Exemple.* — Avec deux dés, quelle est la probabilité d'obtenir au moins un 6 ?

Si  $E_1$  est l'événement : « on a au moins un 6 », l'événement contraire  $\bar{E}_1$  est : « on n'obtient aucun 6 ». Or,  $\Pr(\bar{E}_1)$  s'obtient aisément : c'est le nombre d'éléments de l'ensemble produit de  $(1, 2, 3, 4, 5)$  par lui-même. D'où :

$$\Pr(\bar{E}_1) = 25/36, \text{ et, par suite : } \Pr(E_1) = 11/36.$$

### III. — Événement $E_1 \cap E_2$ ; probabilités liées.

Si les événements  $E_1$  et  $E_2$  ne sont pas incompatibles, ils peuvent se produire simultanément : on a alors l'événement  $E_1 \cap E_2$ .

*Exemple.* — Avec deux dés, l'événement  $E_1$  est : « la somme des points est paire », et  $E_2$  est : « la somme des points est multiple de 3 ». L'événement  $E_1 \cap E_2$  est alors : « la somme des points est multiple de 6 ».

Le tableau déjà utilisé montre que :

$$\Pr(E_1) = 18/36, \quad \Pr(E_2) = 12/36, \quad \Pr(E_1 \cap E_2) = 6/36.$$

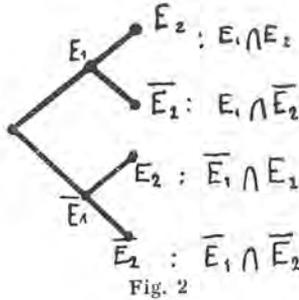
$E_1$  et  $E_2$  ne forment pas un système complet, mais, de ces deux événements, on peut déduire aisément un tel système ; il suffit pour cela de considérer les événements :

$$E_1 \cap E_2, \quad E_1 \cap \bar{E}_2, \quad \bar{E}_1 \cap E_2, \quad \bar{E}_1 \cap \bar{E}_2.$$

Ils sont manifestement incompatibles deux à deux ; de plus, leur réunion donne  $E$ . On peut le vérifier par un raisonnement « dichotomique », analogue à celui qui permet de trouver le nombre  $2^n$  des parties

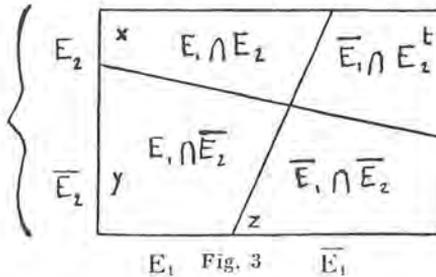
d'un ensemble formé de  $n$  éléments (fig. 2). On peut aussi utiliser le schéma classique dans l'étude des ensembles (fig. 3). Enfin, à titre d'exercice, on peut démontrer formellement la formule :

$$(E_1 \cap E_2) \cup (E_1 \cap \bar{E}_2) \cup (\bar{E}_1 \cap E_2) \cup (\bar{E}_1 \cap \bar{E}_2) = E_1 \cup \bar{E}_1 = E \quad (*)$$



Désignons par  $x, y, z, t$  le nombre des éléments contenus dans chacun des ensembles précédents, conformément aux indications de la figure 3. Nous obtenons immédiatement :

$$\Pr(E_1 \cap E_2) = \frac{x}{x+y+z+t}, \quad \Pr(E_1) = \frac{x+y}{x+y+z+t}$$



Désignons par  $\Pr(E_2/E_1)$  la probabilité de l'événement  $E_2$  lorsqu'on sait que l'événement  $E_1$  est réalisé, c'est-à-dire :

$$\Pr(E_2/E_1) = \frac{x}{x+y}$$

et de même :  $\Pr(E_1/E_2) = \frac{x}{x+t}$

d'où résultent les formules :

$$\Pr(E_1 \cap E_2) = \Pr(E_1) \cdot \Pr(E_2/E_1) = \Pr(E_2) \cdot \Pr(E_1/E_2).$$

*Exemple.* — Avec deux dés jetés une seule fois, quelle est la probabilité d'obtenir un total supérieur à 10, sachant qu'au moins un dé a donné un 6 ; ou bien sachant que le premier dé a donné un 6 ?

Les réponses se lisent sur la figure 1 : si l'un des dés a donné 6, les

(\*) A rapprocher de la forme normale disjonctive de la tautologie dans la logique des propositions.

événements sont les 11 points du contour ABC ; si, en outre, le total a dépassé 10, les événements favorables sont les points D, B, E, d'où la probabilité  $3/11$  d'obtenir plus de 10, lorsqu'on sait qu'un dé a donné 6.

La probabilité d'avoir au moins un 6, lorsqu'on n'a aucune information, est  $11/36$ , et la probabilité d'avoir plus de 10 avec au moins un 6 est  $3/36$  ; on vérifie bien que l'on a :

$$3/36 = (11/36) \times (3/11).$$

Dans le deuxième cas, où l'on sait que le premier dé a donné 6, les événements sont les 6 points du segment BC, et les événements favorables à une somme supérieure à 10 sont les points B et E, d'où la probabilité « liée »  $1/3$ .

La probabilité pour que le premier dé donne 6 est  $6/36$  ; la probabilité d'avoir plus de 10 avec 6 au premier dé est  $2/36$  ; on vérifie encore la formule :

$$2/36 = 6/36 \times 1/3.$$

*Cas particulier.* — Si l'on a :  $\Pr(E_1/E_2) = \Pr(E_1)$ , on dit que  $E_1$  est indépendant de  $E_2$ . Dans ce cas, les formules données plus haut montrent que l'on a aussi :

$$\Pr(E_2/E_1) = \Pr(E_2).$$

L'indépendance, au sens des probabilités, signifie donc : le fait que l'un des événements est réalisé ou non n'influe pas sur la probabilité de l'autre. C'est une relation symétrique.

Lorsque  $E_1$  et  $E_2$  sont indépendants, on a :

$$\Pr(E_1 \cap E_2) = \Pr(E_1) \Pr(E_2) \quad (1).$$

Par exemple, si l'on a un sac dans lequel on a placé  $n$  boules noires et  $b$  boules blanches, si l'on tire une boule du sac, puis une deuxième, *sans remettre la première dans le sac*, les deux tirages ne sont pas indépendants. Ils le deviennent si l'on remet la première boule dans le sac avant de tirer la deuxième (et si l'on agit avant de faire le second tirage !).

*Exercice.* — Soient  $n$  événements indépendants  $E_1, E_2, \dots, E_n$ , c'est-à-dire que la probabilité de chacun d'eux n'est pas modifiée par la réalisation des autres. Généraliser la formule (1) par récurrence :

$$\Pr(E_1 \cap E_2 \cap E_3 \dots \cap E_n) = \Pr(E_1) \cdot \Pr(E_2) \dots \Pr(E_n) \quad (2) \quad (*).$$

(\*) Dans la pratique, il faut bien s'assurer, avant d'appliquer la formule, que les événements sont indépendants. Dans « Le Mystère de Marie Roget », Edgar Poe a fait une curieuse application du principe des probabilités « composées ». Marie Roget a disparu, on a trouvé un cadavre non identifiable, et certains indices donnent à croire que ce pourrait être celui de la jeune fille. Les journaux contestent ces indices, prétextant que chacun d'eux peut convenir à un grand nombre de personnes. Mais, remarque Edgar Poe, « l'important n'est pas que le cadavre ait les jarrettières de la jeune fille perdue, ou ses souliers, ou son chapeau, ou les fleurs de son chapeau, ou ses pieds, ou son aspect et ses proportions générales ; l'important est que le cadavre a chacune de ces choses, et les a toutes collectivement ».

Cependant, plusieurs des indices sont relatifs à des détails de toilette. Ils ne peuvent donc pas être, de façon certaine, considérés comme indépendants, si l'on suppose que Marie Roget était sensible à l'influence de la mode, supposition qui n'a rien d'in vraisemblable, mais ne diminue en rien la valeur de l'argument d'Edgar Poe.

Un autre exemple célèbre d'erreur due à la méconnaissance de la dépendance des événements est celui de Condorcet, quand il calculait le nombre de juges d'un tribunal pour qu'il soit pratiquement à l'abri de l'erreur judiciaire.

**IV. — Événement  $E_1 \cup E_2$ .**

L'événement «  $E_1$  ou  $E_2$  » se note  $E_1 \cup E_2$ . Nous avons vu que, dans le cas d'un système complet, nous avons :

$$\Pr(E_1 \text{ ou } E_2 \text{ ou } \dots \text{ ou } E_n) = \Pr(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_1^n \Pr(E_i) = 1.$$

et :  $\Pr(A) = \sum \Pr(A_i)$ ,

si les  $A_i$  réalisent une partition de  $A$ .

Soient  $E_1$  et  $E_2$  deux sous-ensembles quelconques de  $E$ . On peut décomposer  $E_1$  en un système complet en utilisant les deux événements  $E_1 \cap E_2$  et  $E_1 \cap \bar{E}_2$  ; ce qui permet d'écrire :

$$\Pr(E_1) = \Pr(E_1 \cap E_2) + \Pr(E_1 \cap \bar{E}_2) \quad (3)$$

et de même :

$$\Pr(E_2) = \Pr(E_1 \cap E_2) + \Pr(\bar{E}_1 \cap E_2).$$

Un dénombrement direct facile conduit à la formule :

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2) \quad (4).$$

On peut aussi en donner une démonstration formelle intéressante comme suit :

Appliquant (3), on peut écrire :

$$\Pr(\bar{E}_1 \cap \bar{E}_2) = \Pr(\bar{E}_1) - \Pr(\bar{E}_1 \cap E_2).$$

Mais les événements  $\bar{E}_1 \cap \bar{E}_2$  et  $E_1 \cup E_2$ ,  $E_1$  et  $\bar{E}_1$  sont contraires ; l'égalité précédente peut donc s'écrire :

$$1 - \Pr(E_1 \cup E_2) = 1 - \Pr(E_1) - \Pr(\bar{E}_1 \cap E_2),$$

ou :

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(\bar{E}_1 \cap E_2),$$

et comme, en utilisant de nouveau (3) :

$$\Pr(\bar{E}_1 \cap E_2) = \Pr(E_2) - \Pr(E_1 \cap E_2),$$

on obtient finalement la formule (4).

En général, on a donc :

$$\Pr(E_1 \cup E_2) \leq \Pr(E_1) + \Pr(E_2),$$

l'égalité n'ayant lieu que si :  $\Pr(E_1 \cap E_2) = 0$ , c'est-à-dire si les deux événements sont incompatibles.

*Exemple.* — D'après l'horaire, deux trains doivent arriver en gare à la même heure. L'un a une probabilité 8/10 de n'avoir pas de retard, l'autre une probabilité 7/10. Quelle est la probabilité pour que l'un au moins des trains n'ait pas de retard ?

Soient les événements :

$E_1$  : « le premier train n'a pas de retard » ;

$E_2$  : « le deuxième train n'a pas de retard ».

Nous aurons :

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2).$$

Admettons que les arrivées des deux trains sont des événements indépendants ; alors :

$$\Pr(E_1 \cap E_2) = 8/10 \times 7/10 = 56/100,$$

et, par suite :

$$\Pr(E_1 \cup E_2) = 8/10 + 7/10 - 56/100 = 94/100.$$

*Exemple.* — Les statistiques démographiques donnent une probabilité  $p$  pour qu'une naissance soit masculine, et la probabilité  $q = 1 - p$  pour qu'elle soit féminine. Une famille a six enfants (sans jumeaux). Quelle est la probabilité pour qu'il y ait deux filles au plus ?

Soient  $P$  la probabilité cherchée,  $P_0, P_1, P_2$  les probabilités pour que la famille ait exactement 0, 1 ou 2 filles. Comme ces trois derniers événements sont incompatibles, nous avons :  $P = P_0 + P_1 + P_2$ .

Calculons  $P_0$ . Si nous admettons que chaque naissance n'a aucune influence sur les suivantes, les six naissances sont des événements indépendants, d'où :  $P_0 = p^6$ .

Calculons  $P_1$ . L'existence d'une seule fille est possible de six manières, selon le rang de la naissance féminine dans les six naissances. Chacun de ces événements est le concours de six événements indépendants, de probabilité  $p$  pour un garçon,  $q$  pour une fille, d'où la probabilité  $p^5q$  pour l'un quelconque des six événements. Enfin, ces événements étant incompatibles, nous avons :

$$P_1 = 6p^5q.$$

Pour le calcul de  $P_2$ , le raisonnement est le même. Par dénombrement direct ou par utilisation des combinaisons, on voit que :

$$P_2 = 15p^4q^2.$$

D'où finalement :

$$P = p^6 + 6p^5q + 15p^4q^2.$$

Rappelons qu'on pourrait obtenir ce résultat en utilisant la formule du binôme pour développer  $(p + q)^6$ .

En remplaçant  $q$  par  $1 - p$ , on trouve :

$$P = 10p^6 - 24p^5 + 15p^4.$$

## V. — Probabilités complètes ; formule de Bayes.

La représentation graphique (fig. 4) permet de résoudre aisément le problème suivant :

Pour étudier un phénomène  $A$ , on procède à une expérience  $E$  qui donne lieu à un système complet d'événements  $E_1, E_2, \dots, E_n$ . La réalisation d'un  $E_i$  entraîne celle de  $A$ , avec une certaine probabilité. Quelle est la probabilité de réalisation de  $A$  ?

Si N est le nombre total d'événements, la figure 4 montre que :

$$\Pr(A) = \sum \frac{x_i}{N}, \quad \Pr(E_i) = \frac{x_i + y_i}{N}, \quad \Pr(A/E_i) = \frac{x_i}{x_i + y_i}$$

d'où :

$$\Pr(E_i) \cdot \Pr(A/E_i) = \frac{x_i}{N}$$

d'où enfin :

$$\Pr(A) = \sum \frac{x_i}{N} = \sum \Pr(E_i) \cdot \Pr(A/E_i) \quad (5).$$

C'est la formule de la probabilité complète.

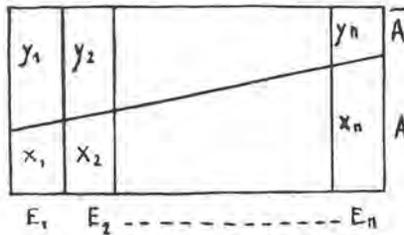


Fig. 4

Lorsqu'on sait que A s'est produit, quelle est la probabilité pour que ce soit par l'événement \$E\_i\$ ? Nous avons :

$$\Pr(E_i \cap A) = \Pr(E_i) \cdot \Pr(A/E_i) = \Pr(A) \cdot \Pr(E_i/A),$$

d'où la formule de Bayes :

$$\Pr(E_i/A) = \frac{\Pr(E_i) \cdot \Pr(A/E_i)}{\Pr(A)}$$

c'est-à-dire, en utilisant (5) :

$$\Pr(E_i/A) = \frac{\Pr(E_i) \cdot \Pr(A/E_i)}{\sum \Pr(E_k) \cdot \Pr(A/E_k)} \quad (6).$$

Cette formule permet de résoudre le problème de la « probabilité des hypothèses ». Supposons, par exemple, une usine qui fabrique des ampoules électriques. La probabilité pour qu'une ampoule soit « bonne » est 0,96. Pour contrôler la fabrication, on adopte une méthode rapide qui déclare bonne une ampoule, avec la probabilité 0,97 lorsqu'elle est vraiment bonne, et la probabilité 0,04 lorsqu'elle est mauvaise. Le contrôle a déclaré bonne une certaine ampoule ; quelle est la probabilité pour qu'elle soit vraiment bonne ?

Nous avons ici :

\$E\_1\$ : l'ampoule est bonne, \$E\_2 = \bar{E}\_1\$ : l'ampoule est mauvaise ;

$$\Pr(E_1) = 0,96, \quad \Pr(\bar{E}_1) = 0,04.$$

Le phénomène A qui est réalisé est : « le contrôle a déclaré bonne l'ampoule examinée », et on cherche \$\Pr(E\_1/A)\$. La formule de Bayes donne :

$$\Pr(E_1/A) = \frac{0,96 \times 0,97}{0,96 \times 0,97 + 0,04 \times 0,04} \approx 0,998.$$

On peut donc espérer que, sur 1 000 ampoules examinées et acceptées par le contrôle, il y en aura environ 2 de mauvaises, alors que, sans le contrôle, on pourrait craindre qu'il y en ait environ 40.

*Exercice.* — Dans un examen, on propose aux candidats un choix de 10 questions, et on en tire une au sort pour chaque candidat. De deux candidats, l'un est bon et connaît 9 des 10 questions, l'autre est moins bon et ne connaît qu'une question sur les dix. L'un des deux tire une question, et il se trouve qu'il la connaît. Quelle est la probabilité pour que ce soit le bon candidat ? (Rép. 9/10).

Dans la pratique, il est souvent difficile de fixer les probabilités *a priori*, c'est-à-dire celles des événements complets  $E_i$ . Un exemple classique de cette difficulté est le problème de Poincaré :

A joue à l'écarté avec un inconnu B qui, dès la première partie, retourne un roi. Quelle est la probabilité pour que B soit un tricheur ?

Soient les probabilités *a priori* :  $p$  pour que B soit tricheur ;  $1 - p$  pour qu'il soit non-tricheur.

Soit de même  $p'$  la probabilité pour qu'un tricheur retourne le roi. Quant à la probabilité pour qu'un non-tricheur retourne le roi, elle est bien connue : c'est  $1/8$ , puisqu'il y a 4 rois dans un jeu de 32 cartes.

La formule de Bayes (ou une figure analogue aux précédentes) donne :

$$P = \frac{pp'}{pp' + \frac{1}{8}(1-p)}$$

P étant la probabilité pour que B soit un tricheur, sachant qu'il a retourné un roi au premier coup.

Poincaré admettait :  $p = 1/2$ ,  $p' = 1$ , d'où une probabilité P égale à  $8/9$ .

Emile Borel remarquait avec raison qu'il n'est pas raisonnable de jouer avec quelqu'un qui a une chance sur deux d'être tricheur, et que, de plus, un tricheur ne doit pas être assez maladroit pour retourner le roi à chaque coup. Il proposait :  $p = 1/10$ ,  $p' = 1/4$ , c'est-à-dire que le tricheur retourne le roi deux fois plus souvent que le joueur honnête. Dans ces conditions, on trouve :  $P = 2/11$ .

## VI. — Espérance mathématique ; écart-type.

Jetant un coup d'œil d'ensemble sur ce qui précède, nous voyons l'analogie entre le calcul des probabilités (pour un ensemble fini d'événements) et l'algèbre des ensembles et celle des propositions. Le seul élément nouveau est qu'à chaque événement  $E_i$  (sous-ensemble de l'ensemble référentiel E), est attachée une mesure  $\text{Pr}(E_i)$  qui vérifie les propriétés :

$$\left\{ \begin{array}{l} 0 \leq \text{Pr}(E_i) \leq 1 \quad \text{pour tout } E_i \subset E, \\ \text{Pr}(E_i) = 0 \text{ si et seulement si } E_i = \emptyset, \\ \text{Pr}(E) = 1, \\ E_1 \subset E_2 \Rightarrow \text{Pr}(E_1) < \text{Pr}(E_2), \\ E_1 \cap E_2 = \emptyset \Rightarrow \text{Pr}(E_1 \cup E_2) = \text{Pr}(E_1) + \text{Pr}(E_2). \end{array} \right.$$

Une telle présentation ne paraît pas inaccessible à nos grands élèves si, comme il est souhaitable, ils ont été, dès le début de leurs études, familiarisés avec le langage des ensembles et des propositions.

Néanmoins, se pose la question : que valent ces considérations, si l'on veut les appliquer à des situations concrètes ? Pour répondre à cette question, et donner une valeur pratique aux éléments qui précèdent, il faut établir la « loi des grands nombres » et tout d'abord définir les deux moyennes appelées « *espérance mathématique* » et « *écart-type* ».

*L'espérance mathématique.*

Une « variable aléatoire »  $X$  (la somme des points donnés par deux dés par exemple) peut prendre des valeurs  $x_1, x_2, \dots, x_n$ , avec les probabilités  $p_1, p_2, \dots, p_n$ . On appelle espérance mathématique l'expression :

$$E(X) = \sum_1^n p_i x_i.$$

Par exemple, si  $a$  est un nombre certain, on a :  $E(a) = a$ .

*Exemple.* — On jette un dé,  $X$  est le point obtenu ; on trouve immédiatement que :  $E(X) = 7/2$ .

On jette deux dés ; les résultats connus nous montrent que si  $X$  est la somme des points, nous avons :  $E(X) = 7$ . C'est la somme des espérances relatives à chaque dé.

D'une manière générale, si  $X$  et  $Y$  sont deux variables aléatoires quelconques, on a :

$$E(X + Y) = E(X) + E(Y).$$

Soient  $X$  et  $Y$  deux variables aléatoires définies par leurs lois de répartition, c'est-à-dire les tableaux des valeurs qu'elles prennent et les probabilités correspondantes :

$$\begin{array}{l} X \} \begin{array}{l} x_1 x_2 \dots x_n, \\ p_1 p_2 \dots p_n \end{array} \\ Y \} \begin{array}{l} y_1 y_2 \dots y_m, \\ q_1 q_2 \dots q_m. \end{array} \end{array}$$

La variable  $X + Y$  prendra les valeurs :  $x_i + y_k$  avec les probabilités  $p_{ik}$ ,  $i$  parcourant l'intervalle  $(1, n)$  et  $k$  l'intervalle  $(1, m)$ . Nous avons d'ailleurs :

$p_{ik} = \Pr(x_i) \cdot \Pr(y_k/x_i) = \Pr(y_k) \cdot \Pr(x_i/y_k)$ ,  
en désignant par  $\Pr(x_i)$  la probabilité de l'événement «  $X = x_i$  », et par  $\Pr(y_k/x_i)$  la probabilité de l'événement «  $Y = y_k$ , sachant que  $X = x_i$  est réalisé ». Nous aurons :

$$E(X + Y) = \sum (x_i + y_k) p_{ik} = \sum x_i p_{ik} + \sum y_k p_{ik}$$

les sommations devant se faire par rapport aux deux indices  $i$  et  $k$ .

Pour calculer  $\sum x_i p_{ik}$ , laissons d'abord  $i$  constant et faisons varier  $k$ . Si nous marquons les points  $(x_i, y_k)$ , cela revient à faire la somme pour les points d'une parallèle à  $Oy$  (fig. 5). Nous avons alors :

$\sum x_i p_{ik} = \sum x_i \Pr(x_i) \cdot \Pr(y_k/x_i) = x_i \Pr(x_i) \cdot \sum \Pr(y_k/x_i)$ ,  
 $x_i$  étant fixé, les événements :

$$Y = y_1, Y = y_2, \dots, Y = y_m$$

forment un système complet, et nous avons :

$$\sum_1^m \Pr(y_k/x_i) = 1,$$

d'où :

$$\sum x_i p_{ik} = x_i \cdot \Pr(x_i).$$

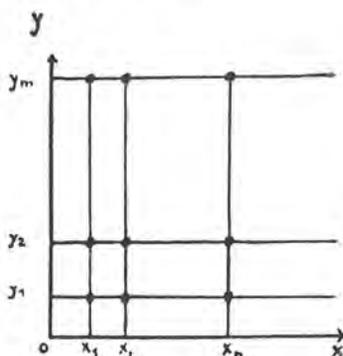


Fig. 5

Faisons maintenant la somme des expressions précédentes pour toutes les parallèles à  $Oy$  ; nous aurons :

$$\sum x_i \cdot \Pr(x_i) = E(X).$$

On procède de même pour calculer le deuxième terme  $\sum y_k p_{ik}$ , en faisant d'abord la somme sur une parallèle à  $Ox$ , et on obtient finalement la formule annoncée.

Le théorème se généralise par récurrence à la somme d'un nombre quelconque de variables aléatoires :

$$E(\sum X_i) = \sum E(X_i).$$

En particulier, si  $a$  est un nombre certain, on a :

$$E(X + a) = E(X) + a,$$

et, comme il est clair que :

$$E(aX) = aE(X),$$

on peut écrire la formule générale, dans laquelle les  $a_i$  sont des nombres certains :

$$E(\sum a_i X_i) = \sum a_i E(X_i).$$

*Exemple.* — On jette un dé ; soit  $X$  une variable aléatoire égale à 0 si le point obtenu est différent de 6, égale à 1 si le point est 6. Les probabilités associées sont évidemment  $5/6$  et  $1/6$ . On a :

$$E(X) = 0.5/6 + 1.1/6 = 1/6.$$

Si l'on jette le dé  $n$  fois, la variable  $Z$  égale à la somme des  $X$  donne le nombre de 6 obtenus dans les  $n$  coups, et l'on a :

$$E(Z) = E(nX) = nE(X) = n/6.$$

Remarquons que l'analyse combinatoire fournit un autre calcul de  $E(Z)$ , plus pénible il est vrai. Résumons-le :

Le nombre de manières d'obtenir  $k$  fois 6 en  $n$  coups est :  $C_n^k$  ; à chacune de ces manières correspond la probabilité :

$$\left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

d'où la probabilité d'obtenir  $k$  fois 6 en  $n$  coups :

$$C_n^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

et  $E(Z)$  sera la somme :

$$\sum k C_n^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

où  $k$  prend les valeurs entières de 1 à  $n$ . Resterait à vérifier que cette somme est égale à  $n/6$ .

*Espérance d'un produit.*

Par analogie avec l'expression de  $E(X + Y)$ , on peut se demander si l'on a :

$$E(XY) = E(X) \cdot E(Y) ;$$

il n'en est rien, comme le montre l'exemple suivant :

On jette un dé, et la variable aléatoire  $X$  est égale à 1 si le point obtenu est pair, égale à 0 si le point est impair. On trouve aisément que  $E(X)$  est égal à  $1/2$ .

Soit de même  $Y$  une variable aléatoire égale à 1 si le point est 4, égale à 0 dans les autres cas, d'où  $E(Y) = 1/6$ .

Considérons d'autre part la variable  $XY$  ; elle vaut 1 lorsque le point est 4, et 0 dans les autres cas ; d'où  $E(XY) = 1/6$ , tandis que

$$E(X) \cdot E(Y) = 1/12.$$

Remarquons que  $X$  et  $Y$  ne sont pas indépendantes : si  $Y = 1$ , le point est 4, donc  $X = 1$ .

Par contre, si les variables aléatoires  $X$  et  $Y$  sont indépendantes (au sens des probabilités), on a :

$$E(XY) = E(X) \cdot E(Y),$$

comme on le voit facilement, en remarquant que la probabilité de l'événement : «  $X = x_i$  et  $Y = y_k$  » est :  $\text{Pr}(x_i) \cdot \text{Pr}(y_k)$ , avec les mêmes notations que précédemment.

*Exercice.* — Démontrer que :  $E(X - Y) = E(X) - E(Y)$ .

L'espérance mathématique a son origine dans l'étude des jeux ; Jacques Bernoulli l'a définie comme « l'espérance que nous avons d'obtenir le meilleur, tempérée par la crainte du pire, de sorte que la valeur de notre attente est quelque chose d'intermédiaire entre le meilleur que nous espérons et le pire que nous craignons ».



Pour le joueur A, la probabilité de gagner la partie s'obtient en faisant la somme des probabilités qui correspondent aux divers chemins qui, partant de la position considérée, aboutissent à une position finale où A est gagnant :

On voit ainsi qu'en  $A^2B$ , il y a deux chemins qui aboutissent aux positions (gagnantes pour A)  $A^3B$ ,  $A^3B^2$ , dont les probabilités respectives sont  $1/2$ ,  $1/4$ . Il n'y a qu'un chemin gagnant pour B, celui qui aboutit en  $A^2B^3$ , de probabilité  $1/4$ . En résumé, les probabilités de gain sont  $3/4$  pour A et  $1/4$  pour B, d'où les espérances mathématiques :

$$64 \times 3/4 = 48, \quad 64 \times 1/4 = 16.$$

Dans la position A, les probabilités de gain sont  $11/16$  pour A et  $5/16$  pour B, et les espérances 44 et 20. Enfin, dans la position  $A^2$ , les probabilités sont  $7/8$  et  $1/8$  et les espérances 56 et 8, et ce sont bien les solutions de Pascal.

Dans une certaine mesure, l'espérance mathématique permet de comparer deux situations aléatoires ; la situation préférable est celle qui correspond à une plus grande espérance mathématique, et deux situations sont équivalentes si elles correspondent à une même espérance. Ainsi peut être définie une relation d'ordre dans les situations aléatoires, ce qui est un problème essentiel dans les applications économiques du calcul des probabilités.

Mais cette relation d'ordre est un peu grossière, en ce sens qu'elle néglige les considérations psychologiques qui interviennent, pour chaque individu, dans le choix qu'il a à faire entre deux situations (notamment, elle néglige le risque, notion importante quand il s'agit des assurances, des loteries, des placements spéculatifs ou « de père de famille ») (\*).

### Variance et écart-type.

Même si l'on écarte les applications économiques, l'espérance mathématique ne renseigne pas suffisamment sur le comportement d'une variable aléatoire. Imaginons, par exemple, les deux lois de répartition suivantes :

Valeurs . . . . .	— 0,02	+ 0,02	Valeurs . . . . .	— 50	+ 50
Probabilités ..	0,5	0,5	Probabilités . . . . .	0,5	0,5

(\*) Pour toute fortune, un clochard possède un billet de loterie qui lui donne une chance sur deux de gagner 100 000 F et une chance sur deux de ne rien gagner, ce qui lui fait une espérance de 50 000 F. Personne ne trouverait singulier que le clochard accepte de vendre son billet 20 000 F, alors qu'une personne riche préférerait sans doute le garder et tenter sa chance.

C'est ce qui a incité Daniel Bernoulli à définir l'« espérance morale ». Puisque l'attrait d'un gain est plus ou moins grand selon ce qu'on possède déjà, on peut admettre que la satisfaction procurée par l'accroissement  $dx$  apporté à une fortune  $x$  est mesuré par  $kdx/x$ . Donc, si  $S$  est une fonction de  $x$  qui mesure cette satisfaction, on peut poser :  $dS = kdx/x$ , ce qui conduit à :

$$S(b) - S(a) = k \text{ Log } b/a$$

Naturellement, cela suppose que la fortune initiale  $a$  n'est pas nulle. Mais, observe Laplace, « celui même qui ne possède rien donne toujours au produit de son travail et à ses espérances une valeur au moins égale à ce qui lui est rigoureusement nécessaire pour vivre ».

Dans les deux cas, l'espérance est zéro, mais les écarts par rapport à zéro sont très différents. Cette situation est celle de deux correcteurs d'examen ; pour chacun d'eux, on connaît la loi de répartition des notes attribuées de 0 à 20. Il peut arriver que les deux correcteurs donnent une même espérance mathématique, mais que l'un utilise toute la gamme des notes, tandis que l'autre se cantonne dans un petit intervalle autour de la moyenne commune.

Il est donc nécessaire d'avoir des renseignements sur la dispersion par rapport à l'espérance mathématique ; on y parvient avec la *variance* et l'*écart-type*.

Soit la variable aléatoire  $X$ , d'espérance  $E(X)$ , que nous désignerons par  $m$ . La variable  $Z = X - m$  est l'« écart », et nous avons :

$$E(Z) = E(X - m) = E(X) - E(m) = m - m = 0.$$

La « variance » est  $E(Z^2)$  :

$$\begin{aligned} V = E(Z^2) &= E[(X - m)^2] = \sum p_i(x_i - m)^2 = \sum p_i x_i^2 + \sum p_i m^2 - 2 \sum p_i x_i m \\ &= \sum p_i x_i^2 + m^2 \sum p_i - 2m \sum p_i x_i = \sum p_i x_i^2 + m^2 - 2m^2 = \sum p_i x_i^2 - m^2 \end{aligned}$$

ou : 
$$E(X^2) = [E(X)]^2 + V.$$

L'« écart-type » est : 
$$\sigma = \sqrt{V}.$$

La variance est évidemment strictement positive ; elle ne peut être nulle que si l'on a, quel que soit  $i$  :  $x_i = m$ . Donc, en général :

$$E(X^2) \geq [E(X)]^2,$$

l'égalité n'ayant lieu que si la variable est constante.

*Exemple.* — Dans l'ensemble des entiers de 1 à  $n$ , on en choisit un au hasard, avec la même probabilité pour chacun d'eux. Calculer l'espérance et l'écart-type.

Si  $X$  est l'entier, nous avons : 
$$E(X) = \sum_{k=1}^n \frac{k}{n} = \frac{1}{n} \sum k = \frac{n+1}{2}$$

et :

$$V = E\left[\left(k - \frac{n+1}{2}\right)^2\right] = \sum \frac{k^2}{n} + \frac{(n+1)^2}{4} - \sum \frac{(n+1)k}{n} = \frac{n^2-1}{12}, \quad \sigma = \sqrt{\frac{n^2-1}{12}}$$

*Théorème d'addition des variances.*

Soient  $n$  variables aléatoires  $x_i$ , indépendantes deux à deux, et la variable  $Z = \sum a_k X_k$ , les  $a_k$  étant des nombres certains. Nous avons :

$$E(Z) = \sum a_k E(X_k), \quad \sigma_z^2 = E[(Z - E(Z))^2].$$

Posons  $E(X_k) = m_k$ , donc  $E(Z) = \sum a_k m_k$ . D'où :

$$\begin{aligned} \sigma_z^2 &= E\left[\left(\sum a_k (X_k - m_k)\right)^2\right] = E\left[\sum a_k^2 (X_k - m_k)^2 + 2 \sum a_i a_k (X_i - m_i)(X_k - m_k)\right] \\ &= \sum a_k^2 E[(X_k - m_k)^2] + 2 \sum a_i a_k E[(X_i - m_i)(X_k - m_k)] \end{aligned}$$

comme les  $X_i$  sont indépendantes deux à deux, nous avons :

$E[(X_i - m_i)(X_k - m_k)] = E(X_i - m_i) \cdot E(X_k - m_k) = 0$ ,  
d'où enfin :

$$\sigma_z^2 = \sum a_k^2 \sigma_k^2.$$

*Exemple.* — On répète  $n$  fois une même expérience ; à chaque fois, il y a une probabilité  $p$  en faveur d'un certain événement,  $1 - p$  pour l'événement contraire. Soit  $X$  une variable aléatoire, égale à 1 si l'événement a lieu, à 0 dans le cas contraire. Pour une expérience, nous avons donc :  $E(X) = p$ . Pour  $n$  expériences, si  $Z$  est la somme des  $X$ , nous aurons, en vertu de la propriété « linéaire » de l'espérance :  $E(Z) = np$ .

Calculons l'écart-type pour une expérience. La loi de répartition de l'écart  $X - p$  est :

$$\begin{array}{cc} 1 - p & - p \\ p & 1 - p \end{array}$$

d'où :

$$\sigma_x^2 = (1 - p)^2 p + p^2 (1 - p) = p(1 - p).$$

Les variables  $X$  étant indépendantes, nous pouvons appliquer à  $Z$  le théorème d'addition des variances, d'où,  $\sigma_z$  étant l'écart-type de  $Z$  :

$$\sigma_z^2 = np(1 - p).$$

Soit par exemple une partie de pile ou face, donc  $p = 1/2$ , où l'on fait 900 coups. L'espérance, pour le nombre d'apparitions de pile, est donc 450, et l'écart-type de  $Z$  est égal à 15.  $\sigma$  donne une indication sur la dispersion du nombre des apparitions de pile autour de 450, indication que nous précisons dans ce qui suit.

### VII. — Inégalité de Tchebicheff ; loi des grands nombres.

Soit une variable aléatoire  $X$ , d'espérance  $E(X) = m$  et d'écart-type  $\sigma$ . Etant donné un nombre positif  $t$ , cherchons :

$$\Pr (|X - m| \geq t\sigma).$$

Dans  $\sigma^2 = \sum p_i (x_i - m)^2$ , les termes tels que  $|x_i - m| \geq t\sigma$  ont une probabilité totale inférieure ou égale à  $1/t^2$ . En effet, si  $\Sigma'$  est la somme de ces termes, nous avons :

$$\Sigma' p_i (x_i - m)^2 \geq \Sigma' p_i t^2 \sigma^2 = (\Sigma' p_i) t^2 \sigma^2,$$

or,  $\Sigma$  ne s'étend qu'à une partie des termes de  $\sigma^2$  ; donc :

$$(\Sigma' p_i) t^2 \sigma^2 \leq \sigma^2 \Rightarrow \Sigma' p_i \leq \frac{1}{t^2}.$$

En résumé :

$$\Pr (|X - m| \geq t\sigma) \leq \frac{1}{t^2}, \quad \Pr (|X - m| \leq t\sigma) \geq 1 - \frac{1}{t^2}.$$

Inégalités établies par Tchebicheff en 1846, mais que l'Inspecteur Général des Finances Bienaymé, ami de Cournot, connaissait aussi.

Supposons par exemple  $t = 5$  ; nous aurons :

$$\Pr (|X - m| \leq 5\sigma) \geq 0,96.$$

C'est la probabilité pour que toutes les valeurs de  $X$  soient contenues dans un intervalle de longueur  $10\sigma$  centré sur l'espérance  $m$ .

Appliquons les inégalités à la variable aléatoire  $F = Z/n$ ,  $Z$  étant la variable étudiée dans l'exemple ci-dessus.  $Z$  est le nombre des succès dans  $n$  expériences, donc  $F$  est la « fréquence » de ces succès. La loi de répartition de  $F$  est celle de  $Z$ , avec la seule modification que les valeurs de  $Z$  sont toutes divisées par  $n$  ; d'où il résulte que :

$$E(F) = p ; \sigma_F = \frac{1}{n} \sigma_Z = \sqrt{\frac{p(1-p)}{n}}$$

et :

$$\Pr (|F - p| \leq t\sigma_F) \geq 1 - 1/t^2.$$

Nous pouvons choisir  $t$  de manière que cette probabilité soit aussi voisine que l'on peut de 1.  $t$  étant ainsi choisi, on peut prendre  $n$  assez grand pour que  $t\sigma_F$  soit aussi petit que l'on veut (car  $\sigma_F$  tend vers zéro quand  $n$  croît).

*Exemple.* — Soit une suite de coups « pile ou face » ;  $p = 1/2$ ,  $\sigma_F = \frac{1}{2\sqrt{n}}$ . Imposons à la probabilité précédente d'être supérieure à 0,99,

ce qui exige  $t = 10$ . Imposons maintenant à  $|F - p|$  d'être inférieur à 0,02. Nous trouvons que cela a lieu pour :  $n > 62\,500$ .

Rappelons que la méthode de Tchebicheff, si elle est simple et élégante, conduit à une majoration beaucoup plus grossière que la méthode « des épreuves répétées » et la loi de Laplace-Gauss.

Les résultats précédents constituent le théorème de J. Bernoulli, ou « loi des grands nombres » ; on peut encore l'énoncer :

Etant donné un nombre positif  $\varepsilon$  aussi petit que l'on veut, si l'on répète  $n$  fois une expérience pouvant donner lieu à un certain événement avec la probabilité  $p$ , la probabilité pour que l'écart  $|F - p|$  entre la fréquence de l'événement et sa probabilité, soit en valeur absolue supérieure à  $\varepsilon$ , tend vers 0 lorsque  $n$  augmente indéfiniment.

La loi des grands nombres donne au calcul des probabilités une valeur concrète. Si l'on a fixé *a priori* la probabilité d'un événement, on pourra justifier ce choix par l'expérience, en procédant à un grand nombre d'épreuves répétées. La fréquence observée devra être voisine de la probabilité, et cela d'autant plus que le nombre d'expériences sera plus élevé.

Inversement, si l'on ne sait pas fixer *a priori* de valeur à la probabilité d'un événement, l'expérience donne un moyen d'obtenir une valeur approchée de cette probabilité (\*).

*Exemple.* — Le biologiste Weldon, collaborateur de Fr. Galton, a lancé 12 dés 26 306 fois, soit en somme un dé 315 672 fois, pour détermi-

(\*) En supposant que la fréquence observée reste stable autour d'une valeur fixe, ce qui permet d'admettre que l'événement est « probabilisable ».

ner la probabilité d'obtenir « 5 ou 6 ». Avec des dés parfaits, cette probabilité serait  $1/3$ . Weldon a trouvé 106 602 cas favorables, ce qui donne une fréquence de 0,33777 environ et un écart de 0,00437. Le calcul donne pour écart-type 0,00084, ce qui fait que l'écart observé est plus de cinq fois l'écart-type. D'où la conclusion que les dés employés n'étaient pas parfaits ; l'indication des points sur les faces avait entraîné une perte de matière variable avec la face, d'où vraisemblablement un léger déséquilibre des dés.

### *La justification de Pascal.*

Dès les débuts du calcul des probabilités, l'étude que fit Pascal des questions posées par son ami de Méré constituait déjà une justification de la loi des grands nombres. L'expérience de Méré comme joueur de dés lui avait fait observer des particularités que Pascal expose à Fermat dans sa lettre du 29 juillet 1654 :

« Si on entreprend de faire un six avec un dé, il y a avantage de l'entreprendre en quatre coups ; si on entreprend de faire "sonnez" (c'est-à-dire 6 et 6) avec deux dés, il y a désavantage de l'entreprendre en 24. Et, néanmoins, 24 est à 36 comme 4 est à 6. Voilà quel était son grand scandale qui lui faisait dire hautement que les propositions n'étaient pas constantes et que l'arithmétique se démentait. »

Si  $E$  est l'événement, « on obtient 6 avec un dé » ;  $\bar{E}$  est l'événement, « on n'obtient pas de 6 ». On a évidemment :  $\Pr(\bar{E}) = 5/6$ . Si on lance le dé  $n$  fois, les événements successifs sont indépendants ; la probabilité de n'avoir aucun 6 est donc  $(5/6)^n$ , et la probabilité d'avoir au moins un 6 est :

$$1 - \left(\frac{5}{6}\right)^n.$$

On vérifie qu'elle dépasse  $1/2$  à partir de  $n = 4$ . Pour  $n = 4$ , il y a 671 cas favorables et 625 défavorables, ce qui confirme l'expérience de Méré en ce qui concerne le jeu à un seul dé.

Le même raisonnement montre qu'avec deux dés, la probabilité d'obtenir « 6 et 6 » au moins une fois en  $n$  coups est :

$$1 - \left(\frac{35}{36}\right)^n.$$

Elle dépasse  $1/2$  à partir de  $n = 25$ . L'observation de Méré était donc juste, et elle témoignait d'une expérience enviable du jeu de dés. Mais son indignation ne s'explique que par une erreur familière aux élèves étourdis : voir de la proportionnalité là où il n'y en a pas.

### *Les probabilités négligeables.*

D'après le théorème de Bernoulli, on peut obtenir une probabilité aussi voisine que l'on veut de 1, pour qu'une variable aléatoire soit aussi voisine que l'on veut de sa valeur moyenne  $E(X)$ . Mais cette probabilité n'est jamais égale à 1, et la probabilité de l'événement contraire n'est jamais nulle. Aussi est-on amené à se demander : à partir de quelle valeur une probabilité doit être considérée comme donnant une « quasi-

certitude » ? A partir de quelle valeur une probabilité doit-elle être considérée comme négligeable, et correspondant à un événement « quasi-impossible » ?

Emile Borel a longuement étudié cette question, et le mieux est de se reporter à son livre intitulé « Valeur pratique et Philosophie des probabilités ». Rappelons seulement qu'il considère que tout homme, dans la conduite ordinaire de la vie, néglige les probabilités de l'ordre de  $10^{-6}$ , et qu'« un homme qui voudrait tenir compte de possibilités aussi peu probables deviendrait rapidement un maniaque ou même un fou ».

### Conclusion

Les notions très classiques qui précèdent ne mettent pas en jeu des Mathématiques très savantes, et nous pensons qu'elles seraient fort accessibles à nos élèves du Second Cycle. On pourrait par exemple imaginer une répartition du genre de la suivante :

Seconde : Eléments d'analyse combinatoire ; calcul de probabilités simples, fondé seulement sur l'analyse combinatoire.

Première : Règles générales (probabilités totales, probabilités composées).

Classe terminale : L'espérance mathématique ; l'écart-type ; le théorème de Bernoulli, à partir des inégalités de Tchebicheff.

Terminons par quelques exercices, et par une bibliographie sommaire relative aux sujets traités dans les conférences que le bulletin a publiées et que l'A.P.M. va réunir en volume.

\*

\*\*

### EXERCICES

1. — Dans un examen, on pose aux candidats dix questions ; à chacune desquelles il faut répondre par « vrai » ou « faux ». Si un candidat choisit au hasard chaque réponse, quelle est la probabilité pour qu'il ait au moins 7 réponses justes ?

2. — On a écrit 4 lettres différentes, et d'autre part les noms et adresses des destinataires sur 4 enveloppes. On met au hasard une lettre dans chaque enveloppe. Quelles sont les probabilités pour qu'il y ait exactement 4, 3, 2, 1, 0 lettres dans les enveloppes qui leur étaient destinées ?

3. — On suppose un dé tel que la probabilité d'amener un point soit proportionnelle à ce point. Calculer les diverses probabilités. On jette ce dé 2 fois. Calculer les probabilités des divers événements possibles, et trouver la loi de répartition pour la somme des points donnés par les 2 dés.

4. — Un dé parfait est lancé 2 fois. Soient les variables aléatoires :  $X_1$  : point donné par le 1<sup>er</sup> lancé,  $X_2$  : point donné par le 2<sup>e</sup> lancé. Calculer les probabilités des événements suivants :

$$X_1 = 3 \text{ et } X_2 < 4 ; X_1 \neq 6 \text{ et } X_2 > 3 ; X_1 \neq 7 \text{ ou } X_2 < 5 ;$$

$$X_1 = 7 \text{ ou } X_2 \neq 3.$$

Mêmes questions pour le dé de l'exercice 3.

5. — Démontrer :

$$\begin{aligned} \Pr(E_1 \cap E_2) &\leq \Pr(E_1) ; \Pr(E_1 \cap E_2) \leq \Pr(E_2) ; \\ \Pr(E_1 \cup E_2) &\geq \Pr(E_1) ; \Pr(E_1 \cup E_2) \geq \Pr(E_2). \end{aligned}$$

6. — Démontrer :

$$\begin{aligned} E_1 \subset E_2 &\Rightarrow \Pr(E_1/A) < \Pr(E_2/A), \\ \Pr(E_1/A) + \Pr(\bar{E}_1/A) &= 1. \end{aligned}$$

7. — Quand a-t-on :

$$\Pr(B/A) = 1 ; \Pr(B/A) = 0 ?$$

8. — Si  $E_1$  et  $E_2$  sont indépendants, il en est de même pour les événements  $E_1$  et  $\bar{E}_2$  ; pour  $\bar{E}_1$  et  $\bar{E}_2$ .

9. — Démontrer :

$$\Pr(E_1 \cap E_2 \cap E_3) = \Pr(E_1) \cdot \Pr(E_2/E_1) \cdot \Pr(E_3/E_1 \cap E_2).$$

Généraliser.

10. —  $E_3$  peut être indépendant de  $E_1$  et  $E_2$  sans l'être de  $E_1 \cup E_2$  et de  $E_1 \cap E_2$ . Soit par exemple l'expérience suivante :

On jette une pièce (parfaite) 2 fois, et on suppose :  $E_1$  : le 1<sup>er</sup> coup donne pile ;  $E_2$  : le 2<sup>e</sup> coup donne pile ;  $E_3$  : les deux coups donnent le même résultat. Calculer :

$$\Pr(E_1) ; \Pr(E_3/E_1) ; \Pr(\bar{E}_3/E_2) ; \Pr(E_3/E_1 \cap E_2) ; \Pr(E_3/E_1 \cup E_2).$$

11. — Dans un examen, chaque question a 4 réponses possibles. Un bon candidat connaît 90 % des questions, un candidat médiocre en connaît 50 %. En admettant que tout candidat a la probabilité 1 de donner la bonne réponse à une question qu'il connaît, et 1/4 à une question qu'il ne connaît pas : si un bon candidat a donné la bonne réponse à une certaine question, quelle est la probabilité pour qu'il ait répondu au hasard ? Même question pour un candidat médiocre.

12. — Si les événements  $E_i$  sont indépendants deux à deux, démontrer :

$$\begin{aligned} \Pr(\bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n) &= \Pr(\bar{E}_1) \cdot \Pr(\bar{E}_2) \dots \Pr(\bar{E}_n) ; \\ \Pr(E_1 \cup E_2 \dots \cup E_n) &= 1 - [1 - \Pr(E_1)] \dots [1 - \Pr(E_n)]. \end{aligned}$$

13. — On jette 4 fois successives une pièce (parfaite). Si la variable aléatoire  $X$  est le nombre de piles obtenu dans la partie, trouver la loi de répartition de  $X$ , calculer l'espérance et l'écart-type.

14. — Dans les mêmes conditions que dans l'exercice 13, la variable  $X$  est le nombre de fois où le joueur « pile » est gagnant au cours de la partie (c'est-à-dire le nombre de cas où, depuis le début de la partie, il est sorti plus de piles que de faces). Trouver la loi de répartition, l'espérance et l'écart-type.

Dans les deux exercices précédents, on pourra faire le graphe de la partie, comme il a été fait dans le texte à plusieurs reprises.

15. —  $n$  hommes entrent au café, accrochent leur chapeau au porte-

manteau. En partant, chacun d'eux reprend un chapeau au hasard. Si  $x_i$  est une variable aléatoire égale à 1 si l'homme  $i$  a repris son chapeau, égale à 0 s'il s'est trompé, calculer  $E(x_i)$  et  $E(\Sigma x_i)$ , et les écarts-type de  $x_i$  et de  $\Sigma x_i$ .

16. — Calculer l'espérance et l'écart-type pour la variable aléatoire qui prend les valeurs : (— 10 ; 0 ; 5 ; 10 ; 100), avec des probabilités égales.

17. — La variable aléatoire  $X$  prend les valeurs (— 1, 0, 1) avec des probabilités égales. Calculer l'espérance et l'écart-type, la variance de  $(X + X^2)$  et la somme : variance de  $X +$  variance de  $X^2$ .

18. — La variable aléatoire  $X$  prend les valeurs 0 et 1 avec des probabilités inconnues. Montrer que : variance de  $X \leq 1/4$ . Donner un exemple où la variance de  $X$  est égale à  $1/4$ .

19. — On écrit un développement décimal dans lequel les chiffres décimaux sont successivement pris au hasard parmi les chiffres 0, 1, 2, ..., 9. Evaluer la probabilité pour que la somme des 100 000 premiers chiffres s'écarte de sa valeur moyenne de moins de 3 000.

20. — Montrer que l'inégalité de Tchebicheff peut se mettre sous la forme suivante :

$$\Pr (|X - m| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

étant un nombre positif donné et  $\sigma$  l'écart-type.

Dans la démonstration du théorème de Bernoulli, nous avons supposé la répétition d'expériences identiques, de même écart-type. Plus généralement, soient  $n$  variables aléatoires  $x_i$ , de valeurs moyennes  $a_i$  et d'écart-type  $\sigma_i$ , les variables étant de plus indépendantes deux à deux. Soit  $X$  la variable aléatoire  $\Sigma x_i$ . En supposant que tous les écarts-type  $\sigma_i$  sont majorés par un nombre positif donné  $b$ , montrer que l'on a :

$$\Pr (|X - E(X)| > \varepsilon) \leq \frac{b^2}{n\varepsilon^2}.$$

## BIBLIOGRAPHIE

1. E. BOREL. — Les probabilités et la vie, P.U.F., 1946.
2. BOREL. — Probabilité et certitude, P.U.F., 1950.
3. BOREL, DELTHEIL et HURON. — Probabilités, Erreurs, A. Colin, 1958.
4. R. FORTET. — Eléments de calcul des probabilités, C.N.R.S., 1950.
5. FRÉCHET. — Généralités sur les probabilités, Eléments aléatoires, Gauthier-Villars, 1950 (Traité de Borel).
6. FRÉCHET. — Méthode des fonctions arbitraires ; théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles, G.V., 1952 (Traité de Borel).
7. LECONTE et DELTHEIL. — Préparation à l'étude des probabilités, Vuibert, 1937.

8. D. DUGUÉ. — *Traité de Statistique théorique et appliquée (analyse aléatoire, algèbre aléatoire)*, Masson, 1958.
9. D. DUGUÉ. — *Arithmétique des lois de probabilité*, G.V., 1957.
10. D. DUGUÉ. — *Ensembles mesurables et probabilisables*, Dunod, 1958.
11. D. DUGUÉ et M. GIRAULT. — *Analyse de variance et plans d'expérience*, Dunod, 1959.
12. GIRAULT. — *Initiation aux processus aléatoires ; processus de Poisson ; files d'attente, pannes de machines*, Dunod, 1959.
13. DELTHEIL et HURON. — *Statistique mathématique*, A. Colin, 1959.
14. G. DARMOIS. — *Calcul des probabilités*, C.D.U., 1955.
15. GNÉDENKO et KHINTCHINE. — *Introduction à la théorie des probabilités*, Dunod, 1960.
16. YAGLOM. — *Probabilités et Information*, Dunod, 1959.
17. BLANC-LAPIERRE et FORTET. — *Théorie des fonctions aléatoires*, Masson, 1953.
18. G.-Th. GUILBAUD. — *Leçons sur les éléments principaux de la théorie mathématique des jeux*, C.N.R.S., 1954.
19. J. VILLE. — *Leçons sur quelques aspects nouveaux de la théorie des probabilités*, Institut H.-Poincaré, 1954.
20. KEMENY, SNELL et THOMPSON. — *Finite Mathematical Structures*, Prentice Hall, 1959.
21. KEMENY, SNELL et THOMPSON. — *Introduction to finite Mathematics*, Prentice Hall, 1959.
22. S. VAJDA. — *Théorie des jeux et programmation linéaire*, traduction Bouzitat, Dunod, 1959.
23. Mc KINSEY. — *Introduction to the Theory of Games*, Mc Graw Hill, 1952.
24. R. DUNCAN LUCE et H. RAÏFFA. — *Games and decisions*, J. Wiley, 1957.
25. J. D. WILLIAMS. — *The Compleat Strategyst*, Mac Graw Hill, 1954 (traduction française sous le titre « La stratégie dans les actions humaines », Dunod, 1956).
26. C. BERGE. — *Théorie générale des jeux à  $n$  personnes*, Gauthier-Villars, 1957.
27. C. BERGE. — *Théorie des graphes et applications*, Dunod, 1958.
28. Mc CLOSKEY et TREFETHEN. — *Introduction à la recherche opérationnelle*, Dunod, 1957.
29. Mc CLOSKEY et COPPINGER. — *Recherche opérationnelle ; cas pratiques et méthodes*, Dunod, 1959.
30. S. VAJDA. — *Readings in linear programming*, Pitman, 1958.
31. J. LESOURNE. — *Technique économique et gestion industrielle*, Dunod, 1958.
32. CHURCHMANN, ACKOFF et ARNOFF. — *Introduction to Operations Research*, J. Wiley, 1958.
33. A. TORTRAT. — *Principes de statistique mathématique*, Dunod, 1961.



# LE CALCUL DES PROBABILITES, LES GRANDES LIGNES DE SON DEVELOPPEMENT ET SES PRINCIPAUX CHAMPS D'APPLICATION

par R. FORTET  
*Professeur à la Sorbonne*

## A. — AXIOMATIQUE ET GÉNÉRALITÉS

1° La notion de probabilité est intuitive ; pourtant, elle n'est devenue une notion scientifique, plus précisément mathématique, que tardivement ; cette élaboration s'est opérée principalement par deux voies à certains égards bien différentes : les jeux de hasard et les dénombrements statistiques. Dans les jeux de hasard, les événements intéressant le joueur sont combinaisons d'un nombre fini d'événements fondamentaux ; les dispositions sont prises pour que ces événements fondamentaux soient équivalents du point de vue de la question de savoir lequel d'entre eux se produira, ce qu'on peut traduire en disant qu'ils ont la même probabilité : c'est ce qu'on a appelé le « principe de symétrie » ; la probabilité  $\text{Pr}(A)$  d'un événement  $A$  quelconque est alors repérée par le rapport  $n/N$  du nombre  $n$  de ceux des événements fondamentaux qui réalisent  $A$  au nombre total  $N$  des événements fondamentaux ; et rien n'empêche de poser conventionnellement  $\text{Pr}(A) = \frac{n}{N}$ . La théorie des « événements équivalents » de de Finetti peut être considérée comme une forme moderne et généralisée du « principe de symétrie ».

Les premiers dénombrements statistiques portant sur des populations assez larges et prolongés assez longtemps concernaient surtout des caractères sociaux ou économiques ; un tel dénombrement fait apparaître, par exemple, que, sur les  $N$  individus recensés,  $n$  ont le caractère  $C$  étudié (d'être célibataire, chômeur, etc...) ; il est arrivé que, pour une population donnée (mais qui évolue nécessairement au cours du temps), la fréquence empirique  $\frac{n}{N}$  garde une stabilité frappante ; on peut traduire cette observation concrète en disant qu'il y a une probabilité  $\text{Pr}(C)$  qu'un individu de la population en question ait le caractère  $C$  ; la fréquence empirique  $\frac{n}{N}$  diffère peu de  $\text{Pr}(C)$ , si  $N$  est assez grand, et, intuitivement, d'autant moins que  $N$  est plus grand :  $\frac{n}{N}$  fournit donc une estimation plus ou moins précise de la valeur inconnue

de  $\text{Pr}(C)$  : ce qui conduit naturellement à attribuer aux probabilités  $\text{Pr}(C)$  les propriétés mathématiques des rapports  $\frac{n}{N}$ .

Les probabilités auxquelles on accède par l'un ou l'autre des deux procédés ci-dessus sont-elles bien de même nature ? En tout cas, elles ont les mêmes propriétés formelles ; ce sont ces propriétés qui, exprimées sous une forme mathématique rigoureuse et imposées par axiomes aux probabilités, constituent l'axiomatique du calcul des probabilités ; je vais indiquer cette axiomatique, sous la forme aujourd'hui classique dite de « Kolmogorov ».

2° *Axiomatique de Kolmogorov* : Un problème de calcul des probabilités concerne un système  $\sum$  quelconque, mais déterminé, qui a plusieurs configurations distinctes possibles, parmi lesquelles il en adopte une déterminée sous l'influence du hasard. Soit  $\mathcal{U}$  l'ensemble de toutes les configurations possibles,  $u$  un élément quelconque de  $\mathcal{U}$  ; dans la terminologie Fréchet,  $\mathcal{U}$  est la catégorie d'épreuves,  $u$  une épreuve ; dans la terminologie de Kolmogorov,  $u$  est un « événement élémentaire »,  $\mathcal{U}$  l'ensemble des « événements élémentaires ». Un événement  $A$  (concernant  $\sum$ ) est toujours définissable sous la forme suivante :  $A$  est réalisé si l'épreuve  $u$  réalisée possède un certain caractère ou propriété  $C$  ; si alors on appelle  $A$  le sous-ensemble de  $\mathcal{U}$  constituée par les  $u$  qui ont le caractère  $C$ , on peut identifier l'événement  $A$  et le sous-ensemble  $A$  : les événements sont les sous-ensembles de  $\mathcal{U}$ . Les éléments généraux de la théorie des ensembles sont donc applicables aux événements : notion de réunion, intersection, limite, etc... ; des sous-ensembles disjoints de  $\mathcal{U}$  sont des événements incompatibles, etc...

La probabilité  $p(A)$  d'un événement  $A \subset \mathcal{U}$  est une fonction d'ensemble  $A$  ; l'un des axiomes comporte que c'est une fonction complètement additive ; comme le domaine de définition d'une telle fonction est nécessairement une  $\sigma$ -algèbre (corps de Borel), nous pouvons dire finalement que :

AXIOME 1 : Tout problème de calcul des probabilités porte :

- a) sur un ensemble arbitraire  $\mathcal{U}$  (catégorie d'épreuves) d'éléments  $u$  de nature quelconque (épreuves) ; tout sous-ensemble  $A$  de  $\mathcal{U}$  est un événement ;
- b) sur une  $\sigma$ -algèbre (corps de Borel)  $\mathcal{B}$  de sous-ensembles  $A \subset \mathcal{U}$ , contenant  $\mathcal{U}$  lui-même, mais pouvant être quelconque par ailleurs ;
- c) sur une mesure finie, ou probabilité,  $p$  définie sur  $\mathcal{B}$ , c'est-à-dire sur une fonction d'ensemble  $p(A)$  réelle non négative, complètement additive ( $\sigma$ -additive), définie sur  $\mathcal{B}$ , telle que  $p(\mathcal{U}) = 1$ , mais pouvant être quelconque par ailleurs ;  $p$  est la loi de probabilité de la catégorie  $\mathcal{U}$ .

Les jeux de hasard et les relevés statistiques dont j'ai parlé au § 1 invitent évidemment à imposer à  $p$  d'être additive : aucune évidence

expérimentale n'oblige à imposer à  $p$  la restriction supplémentaire d'être *complètement* additive, mais sans elle trop de questions devraient rester sans réponse.

On peut imposer à  $p$  la condition supplémentaire d'être une mesure *complète*, mais ce n'est pas à proprement parler une restriction, puisque si  $p$  n'est pas complète, il est toujours possible de la compléter, par un élargissement convenable du  $\sigma$ -corps  $\mathcal{B}$ .

On notera qu'en général il existe des sous-ensembles  $A$  de  $\mathcal{U}$  qui ne font pas partie de  $\mathcal{B}$ , donc il existe des événements  $A$  qui n'ont pas de probabilité : non-probabilisés.

*Probabilités conditionnelles* : Un point très important est que la probabilité d'un événement n'est pas une propriété intrinsèque de cet événement ; elle dépend de lui, mais aussi de la catégorie d'épreuves  $\mathcal{U}$  à laquelle on réfère l'événement.

Soit  $\mathcal{U}'$  une partie quelconque de  $\mathcal{U}$  : c'est aussi une catégorie d'épreuves, plus restreinte que  $\mathcal{U}$  ; comme sous-ensemble de  $\mathcal{U}$ ,  $\mathcal{U}'$  représente un événement  $B$ , et  $\mathcal{U}'$  est la catégorie des épreuves qui réalisent  $B$ .

Supposons un événement  $\varepsilon$  consistant en ce que l'épreuve réalisée  $u$  possède un certain caractère  $C$  ; par rapport à la catégorie  $\mathcal{U}$ ,  $\varepsilon$  est représenté par le sous-ensemble  $A \subset \mathcal{U}$  constitué par les  $u \in \mathcal{U}$  qui possèdent le caractère  $C$  ; par rapport à la catégorie  $\mathcal{U}'$ ,  $\varepsilon$  est représenté par le sous-ensemble  $A'$  des  $u$  de  $\mathcal{U}'$  qui ont le caractère  $C$  ; évidemment :

$$A' = A \cap B ;$$

supposons  $\varepsilon$  probabilisé dans  $\mathcal{U}$  [ $\varepsilon \in \mathcal{B}$  ;  $\Pr(\varepsilon)$  ou  $\Pr(A) = p(A)$ ] ; pour que  $\varepsilon$  soit aussi probabilisé dans  $\mathcal{U}'$ , il faut d'abord que  $\mathcal{U}'$  soit munie d'une loi de probabilité ; donc :

- qu'il existe un  $\sigma$ -corps  $\mathcal{B}'$  de sous-ensembles de  $\mathcal{U}'$ , contenant  $\mathcal{U}'$  ; avec, définie sur  $\mathcal{B}'$ , une mesure de probabilité  $p'$  [ $p'(\mathcal{U}') = 1$ ] ;
- ensuite, il faut que  $A' = A \cap B \in \mathcal{B}'$  ; alors la probabilité de  $\varepsilon$  par rapport à  $\mathcal{U}'$  est  $p'(A \cap B)$  ; on l'appelle la probabilité conditionnelle de  $\varepsilon$  (ou de  $A$ ) par rapport à  $B$  [quand  $B$  est réalisé] et on la note souvent :  $\Pr(A/B)$ .

Les jeux de hasard et les dénombrements statistiques suggèrent que  $\Pr(A/B)$  est normalement différente de la probabilité  $\Pr(A)$  de  $\varepsilon$  dans  $\mathcal{U}$  ; et plus précisément que :

$$(2,1) \quad \Pr(B) \Pr(A/B) = \Pr(A \cap B) \quad [p(B) p'(A \cap B) = p(A \cap B)] ;$$

on est donc amené à imposer la propriété (2,1) par axiome ; mais on s'aperçoit rapidement que cet axiome est insuffisant. Le problème typique que l'on rencontre est le suivant : en considérant le triplet  $(\mathcal{U}, \mathcal{B}, p)$  comme donné, soit donnée également une famille  $\mathcal{F}$  de sous-ensembles  $e \in \mathcal{B}$ , deux à deux disjoints et dont la réunion est  $\mathcal{U}$  ;  $A$  désignant un événement quelconque de  $\mathcal{B}$ , la probabilité conditionnelle  $q(e; A) = \Pr(A/e)$ , si elle existe pour tout  $e \in \mathcal{F}$  et tout  $A \in \mathcal{B}$ , doit

être, pour tout  $e$  fixe, une fonction complètement additive de  $A \in \mathcal{B}$ , avec  $q(e; \mathcal{U}) = 1$ . Complété par (2,1), cela ne suffit pas à déterminer  $q$ ; il est naturel de substituer à (2,1) la propriété plus stricte suivante; soit  $\mathcal{B}_{(\mathcal{F})}$  le plus petit  $\sigma$ -corps contenant tous les  $e$ ; pour tout événement  $E \in \mathcal{B}_{(\mathcal{F})}$ , on doit avoir :

$$(2,2) \quad \int_E q(e; A) d\mu = p(e \cap A),$$

où  $\mu$  est la mesure induite par  $p$  sur  $\mathcal{B}_{(\mathcal{F})}$ ; nous adopterons (2,2) comme second axiome (axiome des probabilités composées).

Alors  $q(e; A)$  est déterminée par le théorème de Radon-Nicodym; mais elle n'est pas ainsi déterminée de façon unique; et, en général, parmi les solutions ainsi offertes par le théorème de Radon-Nicodym, il n'y en a aucune qui, comme fonction de  $A$ , soit complètement additive.

On s'est alors demandé quelles conditions, au moins suffisantes, il convient d'imposer à  $(\mathcal{U}, \mathcal{B}, p)$ , — et à la famille  $\mathcal{F}$  —, pour qu'il y ait au moins une solution complètement additive; les résultats dans ce sens les plus intéressants semblent ceux obtenus par Irgina, qui a utilisé le concept de « mesure parfaite » introduit par Gnedenko et Kolmogorov. Il serait utile d'avoir des conditions assurant l'existence, mais aussi l'unicité d'une solution complètement additive; pour le moment, on n'a de telles conditions que dans des cas assez particuliers.

J'ai, ici, considéré le problème de la probabilité conditionnelle en partant de  $p$  (ou probabilité *a priori*) comme donnée, la probabilité conditionnelle étant à déduire de  $p$ ; c'est ainsi que le problème se présente naturellement en statistique. Mais, dans d'autres domaines (processus stochastiques), ce qui est naturellement donné, c'est un jeu de probabilités conditionnelles, à partir duquel il faut construire  $p$ ; dans les cas simples, le théorème de Fubini règle la question; dans des cas compliqués, on se trouve en présence d'un problème non encore résolu.

3° *Commentaires sur l'axiomatique de Kolmogorov*: On a critiqué (et d'ailleurs Kolmogorov lui-même) l'axiomatique précédente, en particulier sur les points suivants :

1) elle prend comme base les notions d'épreuve et de catégorie d'épreuves; or ces notions sont très abstraites;

2) un événement de probabilité 1 n'est pas forcément certain, mais seulement presque-certain; cette distinction entre événements certains et événements presque-certains n'est pas immédiatement intuitive, ou suggérée par l'expérience;

3) l'additivité complète, adoptée pour les raisons dites plus haut, garde cependant un certain caractère d'arbitraire.

On peut éviter ces objections, en prenant comme base les événements probabilisés eux-mêmes; ils forment une algèbre de Boole, et finalement tout problème de probabilité concerne alors une algèbre de Boole métrique, qu'on peut toujours supposer complète; divers auteurs ont étudié cette autre forme d'axiomatique, montrant qu'elle échappe effectivement aux objections 1), 2), 3) ci-dessus, tout en étant en fait

strictement équivalente à l'axiomatique de Kolmogorov (les objections 1), 2), 3) sont donc de forme plutôt qu'essentiellles) ; jusqu'à présent, on utilise peu cette axiomatique booléenne.

4° *Indépendance* : Revenant au triplet  $(\mathcal{U}, \mathcal{B}, p)$ , deux événements A et B sont indépendants si :

$$\left. \begin{aligned} \Pr(A \cap B) &= \Pr(A) \times \Pr(B) \\ \Pr(\bar{A} \cap B) &= \Pr(\bar{A}) \times \Pr(B) \\ \Pr(A \cap \bar{B}) &= \Pr(A) \times \Pr(\bar{B}) \\ \Pr(\bar{A} \cap \bar{B}) &= \Pr(\bar{A}) \times \Pr(\bar{B}) \end{aligned} \right\}$$

en désignant par  $\bar{E}$  le contraire (complémentaire) d'un événement E ; ces quatre conditions se réduisent à une seule si ni  $\Pr(A)$ , ni  $\Pr(B)$  ne valent 0 ou 1 ; on définit de façon analogue l'indépendance « mutuelle » d'un nombre fini quelconque d'événements ; noter que trois événements A, B, C peuvent être indépendants deux à deux sans l'être mutuellement.

5° *Notion générale d'élément aléatoire* : Soit  $\mathcal{X}$  un espace quelconque et  $x = x(u)$  une application quelconque de  $\mathcal{U}$  dans  $\mathcal{X}$  ; c'est un élément aléatoire, à valeurs dans  $\mathcal{X}$ , qu'on désignera souvent par une majuscule X. En général, on se limite aux éléments aléatoires X, ou applications  $x = x(u)$ , tels que les événements intéressants concernant X soient probabilisés. Un événement concernant X est toujours du type :  $X \in e$ , où  $e$  est un certain sous-ensemble de  $\mathcal{X}$ . L'ensemble des  $\Pr(X \in e)$  pour tous les  $e$  tels que l'événement  $X \in e$  soit probabilisé constitue la loi de probabilité de X.

## B. — VARIABLES ALÉATOIRES

6° *Définition et description d'une variable aléatoire* : Naturellement, la première catégorie d'éléments aléatoires qu'on a considérés sont ceux pour lesquels  $\mathcal{X}$  est l'ensemble R des nombres réels ; il s'agit alors des nombres (réels) ou variables aléatoires ; pour une variable aléatoire X, les événements naturellement intéressants sont ceux du type  $X < x$ ,  $x$  nombre réel donné quelconque ; on supposera donc que ces événements sont probabilisés [ $x(u)$  est une fonction mesurable- $p$  de  $u \in \mathcal{U}$ ], et on posera :

$$\Pr(X < x) = F(x) :$$

$F(x)$  est la fonction de répartition de X ; c'est une fonction monotone non-décroissante, continue à gauche. On écarte en général le cas où il y aurait une probabilité positive que X soit infini ; on a alors :

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0 ;$$

$$F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1.$$

$F(x)$  détermine  $\Pr(X \in e)$ , du moins pour tout sous-ensemble  $e \subset R$  d'intérêt pratique, en particulier si  $e$  est un ensemble de Borel.

Lorsque  $F(x)$  est dérivable, sa dérivée  $f(x)$  est la densité de probabilité de  $X$ .

*Espérance mathématique et moments* : L'espérance mathématique  $E(X)$  de  $X$  est définie par :

$$E(X) = \int_{\mathcal{U}} x(u) f(d:u) = \int_{-\infty}^{+\infty} x dF(x),$$

où, par définition, l'intégrale existe si, et seulement si, elle est absolument convergente.

Le moment algébrique d'ordre  $K$  ( $K$  entier  $\geq 0$ ), rapporté à l'abscisse  $a$ ,  $a^m K$  est défini par :

$$a^m K = E[(X - a)^K] = \int_{-\infty}^{+\infty} (x - a)^K dF(x);$$

le moment absolu d'ordre  $K$  ( $K$  réel  $\geq 0$ ) rapporté à l'abscisse  $a$ ,  $a^{m*} K$ , est défini par :

$$a^{m*} K = E(|X - a|^K) = \int_{-\infty}^{+\infty} |x - a|^K dF(x).$$

Quels que soient  $K > 0$  et  $h > 0$ , on a :

$$(6,1) \quad \Pr (|X - a| \geq h) \leq \frac{a^{m*} K}{h^K}$$

(inégalités de Bienaymé-Tschebichev).

*Extensions* :

a) *Variables aléatoires complexes* : Si  $\mathcal{X}$ , au lieu d'être l'espace  $R$  des nombres réels, est l'espace  $C$  des nombres complexes,  $X$  est une variable aléatoire complexe.

b) *Variables aléatoires réelles à plusieurs dimensions* : Si  $\mathcal{X}$  est un espace  $E_n$  affine proprement euclidien,  $X$  est une variable aléatoire réelle à  $n$  dimensions ; si on rapporte  $E_n$  à un repère cartésien, si  $X_1, X_2, \dots, X_n$  sont les coordonnées de  $X$  par rapport à ce repère, il apparaît qu'une variable à  $n$  dimensions n'est autre qu'un système de  $n$  variables aléatoires (en général non indépendantes). Une variable complexe peut être considérée comme une variable réelle à deux dimensions.

Si  $X = (X_1, X_2, \dots, X_n)$  est une variable réelle à  $n$  dimensions,  $F(x_1, x_2, \dots, x_n) = \Pr (X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$  est sa fonction de répartition ; c'est une fonction de  $n$  variables.

*Fonction caractéristique* : Revenons au cas d'une variable réelle  $X$  à une dimension. Sa fonction caractéristique  $\varphi(u)$  est la fonction de  $u$  réel définie par :

$$\varphi(u) = E[e^{iux}] = \int_{-\infty}^{+\infty} e^{iux} dF(x);$$

autrement dit,  $\varphi(u)$  est la transformée de Fourier de  $F(x)$ .  $\varphi(0) = 1$ ,  $\varphi(u)$  est uniformément continue sur  $(-\infty, +\infty)$  et est définie positive ; toute fonction  $\varphi(u)$  continue et définie positive avec  $\varphi(0) = 1$  est la transformée de Fourier d'une fonction de répartition et d'une seule ; la

correspondance biunivoque entre  $F(x)$  et  $\varphi(u)$  est bicontinue (avec des topologies convenables).

Avec des adaptations convenables, les notions d'espérance mathématique, de moments, de fonction caractéristique s'étendent aux variables complexes, ou aux variables à plusieurs dimensions.

Si  $X$  et  $Y$  sont deux variables réelles (à 1 seule dimension chacune), on a toujours :

$$(6,1) \quad E(X + Y) = E(X) + E(Y).$$

Si on pose :

$$\sigma_X^2 = E(|X - E(X)|^2)$$

$$\sigma_Y^2 = E(|Y - E(Y)|^2)$$

$$\sigma_X \sigma_Y r = E\{[X - E(X)][Y - E(Y)]\},$$

on a :  $0 \leq |r| \leq 1$  d'après l'inégalité de Schwarz ;  $|r| = 1$  si, et seulement si, il y a entre  $X$  et  $Y$  une relation linéaire à coefficients non aléatoires :  $r$  s'appelle la *coefficient de corrélation* de  $X$  et  $Y$ .

$X$  et  $Y$  sont indépendants si, quels que soient les ensembles  $e_1, e_2$  de nombres réels, les deux événements  $X \in e_1, Y \in e_2$  sont indépendants. Si  $X$  et  $Y$  sont indépendants, on a :

$$(6,2) \quad E(XY) = E(X) \times E(Y),$$

et par suite le coefficient de corrélation de  $X$  et de  $Y$  est nul.

7° *Addition des variables aléatoires indépendantes* : Si  $X$  et  $Y$  sont indépendantes, et si  $S = X + Y$  est leur somme, on a :

$$(7,1) \quad E(S) = E(X) + E(Y)$$

d'après (6,2) ; si  $E(X) = E(Y) = E(S) = 0$ , on voit avec (6,2) que :

$$(7,2) \quad E(S^2) = E(X^2) + E(Y^2) ;$$

d'une façon générale, il n'est pas difficile de calculer un moment algébrique quelconque  $E(S^K)$  de  $S$ , en fonction des moments, d'ordre  $\leq K$ , de  $X$  et de  $Y$  ; les formules obtenues sont de plus en plus compliquées quand  $K$  augmente.

Si  $F(x), G(y), H(s)$  sont les fonctions de répartition de  $X, Y, S$  respectivement, on a :

$$(7,3) \quad H(s) = \int_{-\infty}^{+\infty} G(s-x) dF(x) = \int_{-\infty}^{+\infty} F(s-y) dG(y),$$

autrement dit,  $H$  est le produit de convolution de  $F$  et  $G$  ; par suite, si  $\varphi_1(u), \varphi_2(u), \varphi_3(u)$  sont les caractéristiques de  $X, Y, S$ , on a :

$$(7,4) \quad \varphi_3(u) = \varphi_1(u) \times \varphi_2(u).$$

8° *Convergence en probabilité, presque-sûre, en moyenne* : Les convergences en mesure, presque-partout, en moyenne d'ordre  $\alpha$ , usuelles dans la théorie de la mesure et des fonctions mesurables, se transforment immédiatement en calcul des probabilités, où elles prennent les noms, respectivement, de convergence en probabilité, presque-sûre, en moyenne d'ordre  $\alpha$  ; par exemple, une variable aléatoire  $X_K$  ( $K = 1, 2, 3$ ) tend, lorsque  $K \rightarrow +\infty$ , vers une variable aléatoire limite  $X$  en moyenne d'ordre 2 (en moyenne quadratique) si  $E[(X_K - X)^2] \rightarrow 0$  ; alors, d'après (6,1),  $X_K$  tend vers  $X$  en probabilité. Si  $X_K$  tend vers  $X$  en pro-

tabilité, cela ne signifie pas que la différence  $X_K - X$  est petite, mais seulement qu'il y a, pour  $K$  assez grand, une probabilité arbitrairement voisine de 1 pour que  $X_K - X$  soit arbitrairement petite.

9° *Lois asymptotiques de l'addition des variables aléatoires indépendantes* : Le chapitre le plus anciennement développé du calcul des probabilités consiste à étudier la somme  $S_n$  d'un grand nombre  $n$  de variables aléatoires indépendantes  $X_1, X_2, \dots, X_n$  :

$$S_n = X_1 + X_2 + \dots + X_n ;$$

on dégage ainsi des lois asymptotiques (pour  $n \rightarrow +\infty$ ) qui appartiennent à deux types principaux :

1° *Lois des grands nombres* : Les lois des grands nombres proprement dites sont des théorèmes établissant, sous des conditions convenables, que lorsque  $n \rightarrow +\infty$  la moyenne arithmétique  $S_n/n$  tend (par exemple en probabilité) vers une limite  $L$ .

Exemple : Soient une succession de parties indépendantes de pile ou face, numérotées 1, 2, ...,  $n$  ; à chaque partie, les probabilités de pile et face ont les mêmes valeurs respectivement  $p$  et  $q = 1 - p$  (on suppose  $0 < p < 1$ ). Si  $X_j$  vaut 1 ou 0 suivant que la partie N°  $j$  donne pile ou face,  $S_n/n$  est la fréquence relative de pile sur les  $n$  parties ; on prouve facilement que  $S_n/n$  tend en moyenne quadratique (donc en probabilité), et aussi presque-sûrement vers  $p$  ; car

$$(9,1) \quad E(|S_n/n - p|^2) \sim \frac{1}{n} ;$$

ce qui rejoint l'idée intuitive que la fréquence empirique, sur un grand nombre  $n$  d'expériences, de l'événement pile doit être en un certain sens une valeur approchée de sa probabilité  $p$ .

2° *Lois de probabilités limites* : Reprenons l'exemple de pile ou face ; pour  $n$  grand,  $S_n$  est probablement grand ; sa fonction de répartition  $F_n(x)$  tend donc vers 0 pour tout  $x$  fini ; mais cette constatation n'apporte rien de nouveau ; mais normons  $S_n$  convenablement, après l'avoir rapporté à une origine également convenable ; en fait, considérons, au lieu de  $S_n$ ,

$$U_n = (S_n - np) / \sqrt{npq},$$

qui pour tout  $n$  a une espérance nulle et un moment d'ordre 2 constamment égal à 1. On démontre que la fonction de répartition  $G_n(u)$  de  $U_n$  tend pour tout  $u$  vers :

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2}} dx,$$

fonction de répartition connue sous le nom de loi de Laplace (ou de Gauss).

De telles lois-limites peuvent être mises en évidence dans des cas beaucoup plus généraux que celui de pile ou face ; l'usage des caractéristiques et de (7,4) est fondamental pour cette étude, aujourd'hui à peu près terminée ; elle a mis en évidence les notions de lois stables, quasi-stables, indéfiniment divisibles (la loi de Laplace est stable, et par suite indéfiniment divisible) ; les efforts les plus récents ont visé à préciser la rapidité des convergences [de  $G_n(u)$  vers  $\Phi(u)$ , par exemple], ou à

savoir si la convergence des fonctions de répartition s'accompagne de la convergence des densités de probabilités (en supposant que celles-ci existent).

### C. — LES GRANDS CHAPITRES DU CALCUL DES PROBABILITÉS

10° *Les processus stochastiques* : En dehors de l'addition des variables aléatoires indépendantes, le calcul des probabilités, dans son état actuel, est principalement consacré à l'étude des processus stochastiques, ou évolutions aléatoires au cours du temps d'un système quelconque. On peut concevoir une infinie variété de processus stochastiques ; pour des raisons théoriques et, plus souvent, pratiques, on a étudié principalement :

a) les processus à accroissements indépendants, dont la théorie se fonde sur l'addition des variables aléatoires indépendantes ;

b) les processus de Markov, dont l'étude peut être envisagée comme une application de la théorie des semi-groupes d'opérations linéaires dans un espace de Banach ; les processus de Wiener-Lévy, qui sont à la fois de Markov et à accroissements indépendants, présentent d'intéressantes connexions avec la théorie du potentiel ;

c) les processus stationnaires du second ordre, dont l'analyse harmonique est un chapitre de l'analyse harmonique générale.

Je ne donne pas d'autres indications sur ces questions qui seront reprises dans des conférences suivantes.

Je note seulement que les processus stochastiques sont des éléments aléatoires d'un certain type, évidemment beaucoup plus complexes que les variables aléatoires, et qu'une façon indirecte de les étudier est de développer la théorie des éléments aléatoires les plus généraux possibles [c'est-à-dire celle des applications  $x(u)$  dans des espaces  $\mathcal{X}$  aussi généraux que possible] ; cette théorie, encore à son début, comporte déjà des résultats prometteurs.

### D. — LES APPLICATIONS DU CALCUL DES PROBABILITÉS

11° *Les applications du calcul des probabilités* : Les premières applications, historiquement, du calcul des probabilités concernèrent la démographie, les assurances, divers phénomènes économiques et sociaux. Ce domaine reste important avec une branche nouvelle : la psychométrie ; toutefois, l'introduction des processus stochastiques en économétrie, bien qu'apparemment souhaitable, reste très timide.

Mais le calcul des probabilités a trouvé, plus récemment, un immense champ d'application en physique ou dans des techniques dérivées de la physique, dans tous les cas où intervient un ensemble considérable de particules : diffusion et mouvement brownien, mécanique statistique (classique ou quantique), phénomènes d'émission (bruit de fond dans les tubes électroniques), semi-conducteurs... ; je citerai aussi, d'un type différent et également très importantes, la théorie de la transmission de l'information et celle de la détection. En retour, ces applications ont

notablement influé sur le cours du développement moderne du calcul des probabilités.

Je ne puis donner aucun détail sur ces applications ; d'ailleurs, leur existence au moins est bien connue ; je préfère m'arrêter sur une autre application, très curieuse, qui peut éventuellement être utilisée dans n'importe quel domaine, et qui est généralement ignorée en dehors des spécialistes : les méthodes de Monte-Carlo.

12° *Les méthodes de Monte-Carlo* : Je veux d'ailleurs simplement faire comprendre le principe de ces méthodes, sans entrer dans la technique de leur utilisation.

Pour obtenir la valeur numérique  $\theta_0$ , supposée inconnue *a priori*, d'un paramètre  $\theta$  dont dépend la fonction de répartition d'une variable aléatoire  $X$ , il n'y a pas d'autre moyen que le procédé empirique : on procédera à  $n$  déterminations, par exemple indépendantes, de  $X$  et on déduira de leurs résultats une estimation de  $\theta_0$  qui sera adoptée comme valeur approchée de  $\theta_0$ , généralement en application d'une loi des grands nombres [comme, d'après (9,1),  $S_n/n$  peut être adopté comme estimation de  $p$ ] ; sous le nom de « théorie de l'estimation des paramètres », les statisticiens ont de longue date développé des méthodes perfectionnées (notion de résumé exhaustif, analyse séquentielle, etc...) pour réduire l'erreur  $e$  à craindre, et le temps nécessaire (c'est-à-dire  $n$ ) ; le calcul des probabilités montre [cf. (9,1)] que  $e$  est ordinairement de l'ordre de  $n^{-\frac{1}{2}}$  :

$$(12,1) \quad e \sim n^{-\frac{1}{2}} .$$

Soit  $\theta_0$  un nombre ; supposons que l'on possède de  $\theta_0$ , non sa valeur, mais une définition mathématique qui le détermine entièrement : par exemple,  $\theta_0$  est la plus petite racine d'une équation algébrique donnée à racines toutes positives ; de la définition mathématique de  $\theta_0$ , on peut déduire en général une méthode de calcul au moins approchée de sa valeur : de telles méthodes reposent souvent sur l'itération d'une opération plus ou moins complexe, de sorte que l'erreur  $e$  a en général un ordre de grandeur du type :

$$(12,2) \quad e \sim a^n \quad (a < 1),$$

où  $n$  est le nombre de fois qu'on a répété l'opération.

Mais soit une famille de variables aléatoires  $\{X_\theta\}$  dont la fonction de répartition  $F(\theta ; x)$  dépend d'un paramètre  $\theta$  ; et supposons qu'on sache procéder à des déterminations empiriques de  $X_{\theta_0}$  ; comme la valeur de  $\theta_0$  n'est pas connue, il faudra que la définition mathématique de  $\theta_0$  suffise à permettre de telles déterminations empiriques ; s'il en est ainsi, on voit qu'on pourra évaluer  $\theta_0$  par une méthode d'estimation de paramètre : on dira alors qu'on évalue  $\theta_0$  par une *méthode de Monte-Carlo*.

*Exemple 1* : Soit  $\theta_0 = \pi$  ;  $\pi$  possède des propriétés mathématiques qui le définissent absolument et dont on a depuis longtemps déduit sa valeur numérique avec une précision considérable ; mais imaginons qu'on n'ait pas encore effectué ce calcul ; on pourrait obtenir la valeur de  $\pi$  par une méthode de Monte-Carlo, par exemple de la façon suivante :

— On connaît le « jeu de l'aiguille » de Buffon ; soit  $X$  le nombre des intersections de l'aiguille et des parallèles tracées sur le plan ; si la longueur de l'aiguille est la moitié de l'écart des parallèles (de sorte que  $X$  ne peut valoir que 0 ou 1), on a :

$$\text{Pr}(X = 1) = \frac{1}{\pi} ;$$

c'est pile ou face, avec  $p = \frac{1}{\pi}$  ; la fréquence  $S_n/n$  correspondante (cf. § 9) sera donc une estimation de  $\pi$ .

*Exemple 2 :*  $u$  variant de 0 à 1, soit  $f(u)$  une fonction continue donnée, avec :  $0 \leq f(u) \leq 1$  ; et soit :

$$\theta_0 = \int_0^1 f(u) du.$$

Soient  $U$  et  $V$  deux variables aléatoires indépendantes de loi uniforme (densité constante) sur  $(0,1)$ , et soit  $X$  la variable aléatoire qui vaut 1 si  $V \leq f(U)$  et 0 si  $V > f(U)$  ; on a :

$$E(X) = \theta_0.$$

On peut procéder à des déterminations empiriques de  $U$  et de  $V$ , donc de  $X$  ; pour  $n$  déterminations indépendantes  $x_1, x_2, \dots, x_n$  de  $X$ ,  $\frac{1}{n} \sum_j x_j$  tend en moyenne quadratique vers  $E(X) = \theta_0$  (loi des grands nombres) ; on peut donc estimer  $\theta_0$  par  $\frac{1}{n} \sum_j x_j$ , et l'erreur à craindre est de l'ordre de :

$$\left\{ E \left( \frac{1}{n} \sum_j x_j - \theta_0 \right)^2 \right\}^{1/2} = \sqrt{\frac{\theta_0(1-\theta_0)}{n}}.$$

On a recouru (von Neumann, Ulam) aux méthodes de Monte-Carlo en 1942 pour des problèmes de diffusion de neutrons, difficiles à étudier par les procédés mathématiques ordinaires et se rattachant à la fabrication de la bombe atomique ; dans ce domaine, ce recours est demeuré usuel.

Naturellement, pour pouvoir évaluer un nombre  $\theta_0$  dont on possède une définition mathématique par une méthode de Monte-Carlo, il faut trouver un schéma stochastique auquel rattacher  $\theta_0$  et réalisable sur la base de la définition mathématique possédée de  $\theta_0$  ; cela peut être difficile ; mais, souvent aussi, un tel schéma est en évidence parce que  $\theta_0$  est précisément apparu à propos d'un phénomène aléatoire (cas de la diffusion des neutrons, du fonctionnement des centraux téléphoniques, etc...) ; il suffit alors de reproduire ce phénomène, naturellement, ou artificiellement avec une machine électronique.

L'usage de méthodes statistiques pour évaluer des quantités relevant normalement d'un calcul strictement mathématique semble cependant paradoxal ; la comparaison de (12,1) avec (12,2) met en évidence que, pour le même nombre  $n$  d'opérations, l'erreur est ordinairement beaucoup plus forte pour une méthode de Monte-Carlo que pour une méthode mathématique ; à vrai dire, les opérations à faire ne sont pas

les mêmes dans les deux cas, et peuvent être beaucoup plus simples et rapides pour la méthode de Monte-Carlo.

Il reste qu'avec une méthode statistique,  $n$  devra généralement être pris très grand, ce qui pose le problème pratique de fabriquer du hasard, en quelque sorte, en grande quantité ; plus correctement à opérer (assez rapidement), un grand nombre  $n$  de tirages au sort. On peut toujours se ramener à des tirages au sort (mutuellement indépendants) d'un nombre ou variable aléatoire compris entre 0 et 1, selon une loi uniforme sur  $(0,1)$  [densité de probabilité de la variable aléatoire constante et égale à 1, sur  $(0,1)$ ]. On a eu l'idée d'observer empiriquement de tels nombres produits par un phénomène naturel aléatoire (bruit de fond dans un tube électronique), et on a dressé et publié une table de  $1,2 \cdot 10^6$  tels nombres. Cette table est difficile à utiliser si on emploie par ailleurs une machine à calculer électronique, et on préfère généralement faire fabriquer les nombres aléatoires par la machine elle-même, selon un procédé arithmétique qui, naturellement, fournit des nombres qui sont en réalité des nombres déterminés, qu'on ne peut considérer comme aléatoires que par approximation.

*N.D.L.R.* — Nous croyons utile de signaler l'ouvrage récent de M. FORTET : « Some Aspects of Analysis and Probability », John Wiley and Sons, New-York (en collaboration avec I. KAPLANSKY, E. HEWITT et MARSHALL HALL), ainsi que les suivants, en préparation :

Collection universitaire de Mathématiques (Dunod) :

VIII. R. FORTET : « Matrices ».

X. R. FORTET : « Processus stochastiques ».

# QUELQUES GRANDES THEORIES MATHEMATIQUES INTERVENANT DANS LE CALCUL DES PROBABILITES

par E. MOURIER

*Maitre de Conférences à la Faculté des Sciences de Poitiers*

Le calcul des probabilités a souvent ses méthodes qui lui sont propres et ceci, en particulier, est dû au fait que les probabilités sont des nombres compris entre zéro et un ; toutefois, il fait également intervenir de grandes théories mathématiques, et des plus modernes. Tout d'abord, il résulte de l'axiomatique même du calcul des probabilités, développée dans la conférence précédente, que toute loi de probabilité est une mesure finie ; les liens entre le calcul des probabilités et la théorie de la mesure sont alors si clairement en évidence qu'il est inutile d'insister sur ce point. Soulignons seulement le fait que c'est en tant que fonction d'ensembles complètement additive, définie sur une  $\sigma$ -algèbre (ou corps de Borel), de sous-ensembles d'un ensemble E, qu'une mesure s'identifie à une loi de probabilité sur E ; que l'existence et la définition d'une loi de probabilité sur E sont indépendantes de la définition d'une topologie sur E.

Un des premiers, et des plus importants problèmes qui se soient posés en calcul des probabilités, est celui de l'étude de la fréquence de la réalisation d'un événement A au cours de  $n$  expériences indépendantes, le résultat de chaque expérience étant soit la réalisation de A, de probabilité  $p$ , soit sa non-réalisation, de probabilité  $q = 1 - p$  (cas de Bernoulli). Si on associe à chaque expérience la variable aléatoire X prenant les valeurs 1 ou 0 selon que A est réalisé ou non, la fréquence de A au cours des  $n$  expériences indépendantes est :

$$F = \frac{1}{n} [X_1 + X_2 \dots + X_n],$$

Or :

$$\begin{aligned} EX &= p \\ EX^2 &= p \end{aligned}$$

d'où :

$$E(F) = \sum_i \frac{EX_i}{n} = p.$$

$$\sigma^2(F) = E[F - E(F)]^2 = \frac{1}{n^2} \sum_i E(X_i - EX)^2 = \frac{p(1-p)}{n} = \frac{pq}{n}.$$

Il résulte donc de l'inégalité de Bienaymé :

$$\Pr [ |X - EX| \geq \tau ] \leq \frac{\sigma^2}{\tau^2}.$$

que  $\Pr [ |F - p| \geq \tau ] \leq \frac{n\tau^2}{pq}$  tend vers zéro lorsque  $n \rightarrow +\infty$ , c'est-à-dire que F converge en probabilité vers  $p$ .

Ce résultat constitue la première et la plus simple « loi des grands nombres ». Le même raisonnement s'étend immédiatement à une variable aléatoire  $X$  quelconque ayant des moments du 1<sup>er</sup> et du 2<sup>e</sup> ordre.

$EX = m$ ,  $E(X - EX)^2 = \sigma^2$ , si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes de même loi que  $X$ ,

$$Y_n = \frac{X_1 + \dots + X_n}{n} \text{ converge en probabilité vers } m, \text{ en effet}$$

$$EY_n = m \text{ et } \sigma^2 Y_n = E[Y_n - EY_n]^2 = E[Y_n - m]^2 = \frac{\sigma^2}{n}$$

et par conséquent :

$$\text{Pr } [|Y_n - m| \geq \tau^2] \leq \frac{\sigma^2}{n\tau^2}$$

donc  $Y_n$  converge en probabilité vers  $m$ .

Remarquons d'ailleurs que, pour démontrer cette convergence en probabilité, nous utilisons la propriété plus stricte :

$$\sigma^2 Y_n = E[Y_n - m]^2 = \frac{\sigma^2}{n} \text{ tend vers zéro lorsque } n \rightarrow +\infty,$$

qui, par définition même, exprime que la variable aléatoire  $Y_n$  converge en moyenne quadratique vers  $m$  ; ce résultat est la « loi des grands nombres en moyenne d'ordre 2 ».

En se rappelant qu'une variable aléatoire  $X$  est une application mesurable,  $x(u)$ , de l'espace probabilisé des épreuves  $(\mathcal{U}, \mathcal{S}, \mu)$  dans l'espace  $\mathbb{R}$  des nombres réels, la convergence en moyenne quadratique d'une suite de variables aléatoires est la convergence forte dans l'espace  $L^2(\mathcal{U}, \mu)$  des fonctions réelles  $x(u)$ ,  $u \in \mathcal{U}$ , de carré sommable par rapport à la mesure  $\mu$ , la norme d'un élément  $x \in L^2(\mathcal{U}, \mu)$  étant définie par :

$$\|x\| = \left[ \int_{\mathcal{U}} |x(u)|^2 d\mu \right]^{1/2}$$

L'énoncé et la démonstration des lois des grands nombres rappelées ci-dessus semblent propres au calcul des probabilités. Un des problèmes essentiels du calcul des probabilités classique, c'est-à-dire concernant les variables aléatoires numériques, a été de préciser de telles lois des grands nombres et de remplacer les conditions énoncées ci-dessus : variables aléatoires indépendantes, de même loi, ayant des moments du 1<sup>er</sup> et du 2<sup>e</sup> ordre, par des conditions plus faibles.

On dit que des éléments aléatoires  $X_n$  forment une suite strictement stationnaire si, quel que soit l'entier positif  $s$ , quels que soient les entiers  $n_1, n_2, \dots, n_s$  et quel que soit l'entier  $h$ , la loi de probabilité de l'élément complexe  $(X_{n_1+h}, X_{n_2+h}, \dots, X_{n_s+h})$  ne dépend pas de  $h$ . Kolmogorov a montré que si des variables aléatoires  $X_i$  forment une suite strictement stationnaire et si  $EX_i$  existe (et alors  $EX_i$  ne dépend pas de  $i$ , du fait de la stationnarité), la moyenne arithmétique  $Y_n = \frac{X_1 + \dots + X_n}{n}$  tend, lorsque  $n \rightarrow +\infty$ , presque-sûrement vers une variable aléatoire limite  $\bar{Y}$  telle que  $EY$  existe et que  $EY = EX_i$ .

Dans le cas particulier où les  $X_i$  sont des variables aléatoires indépendantes et de même loi,  $Y$  se réduit à l'élément certain  $EX_i$ .

Ce théorème constitue la « loi forte des grands nombres ».

On s'aperçoit alors que résultats et méthodes font intervenir la théorie ergodique, dont l'origine ne se trouve pas en calcul des probabilités, mais dans l'étude des systèmes dynamiques.

La première hypothèse ergodique est relative au comportement des trajectoires, dans l'espace des phases, d'un système dynamique ; elle fut énoncée par Maxwell en 1850 : pour les systèmes ayant un très grand nombre de degrés de liberté, par exemple gaz de la théorie cinétique, la moyenne statistique est égale à la moyenne temporelle. Les premiers théorèmes ergodiques furent démontrés par G.-D. Birkhoff, I. Carleman, D.-O. Koopman, J. von Neumann, etc..., au sujet du comportement de moyennes de transformations. Depuis, cette théorie ergodique a connu un développement considérable, qu'il est impossible de résumer ici ; j'indiquerai seulement deux théorèmes essentiels, à savoir :

**THÉORÈME ERGODIQUE DE BIRKHOFF :** Soit  $\Omega$  un ensemble d'éléments  $\omega$ , pourvu d'une mesure bornée  $\mu$ , on supposera  $\mu(\Omega) = 1$ . Soit  $\varphi$  une transformation (application biunivoque de  $\Omega$  sur lui-même) conservant la mesure  $\mu$ . Soit  $\mathcal{F}$  l'espace de Banach des fonctions numériques réelles,  $f(\omega)$ , sommables  $\mu$  sur  $\Omega$ , avec la norme :

$$\|f\| = \int_{\Omega} |f(\omega)| \mu(d\omega).$$

Soit  $T$  l'opération linéaire dans  $\mathcal{F}$  définie par :

$$Tf = f[\varphi(\omega)],$$

$T$  est une opération bornée,  $\|T\| = 1$ .

**THÉORÈME :** Pour presque tout  $\omega$  (au sens de  $\mu$ ),

$$\lim_{n \rightarrow +\infty} \frac{1}{n} (T^1 + T^2 + \dots + T^n) f(\omega) = \lambda(\omega)$$

existe, et  $\lambda \in \mathcal{F}$ .

Remarquons que les entiers  $1, 2, \dots, n, \dots$  peuvent s'interpréter comme des temps (temps discret), et alors la moyenne arithmétique est la moyenne temporelle.

**THÉORÈME ERGODIQUE DE YOSIDA ET KAKUTANI :** Soit  $T$  un opérateur linéaire borné qui transforme un espace de Banach  $\mathcal{X}$  en lui-même, et tel que  $T^n$  soit uniformément borné, c'est-à-dire qu'il existe un nombre  $C$ , indépendant de  $n$ , tel que  $\|T^n\| \leq C$ ,  $n = 1, 2, \dots$ . Si pour tout  $x \in \mathcal{X}$ , la suite

$$\{x_n\} \text{ où } x_n = \frac{1}{n} [T + T^2 + \dots + T^n]x, n = 1, 2, \dots,$$

contient une sous-suite qui converge faiblement vers un point  $\bar{x} \in \mathcal{X}$ , la suite  $\{x_n\}$  converge fortement vers  $\bar{x}$ .

Il faut noter que le théorème de Yosida et Kakutani est de nature beaucoup plus général que celui de Birkhoff, puisque les opérateurs linéaires  $T$  sont seulement astreints à ce que  $T^n$  est uniformément borné, alors que les opérateurs de Birkhoff sont de type très particulier et de

norme 1 ; d'autre part, le théorème de Yosida et Kakutani ne fait pas intervenir de mesure.

Kolmogorov eut le premier le mérite de montrer qu'il y a équivalence entre la loi forte des grands nombres pour des variables aléatoires et le théorème de Birkhoff.

Les recherches modernes en calcul des probabilités ont principalement pour objet de formuler des « lois des grands nombres », c'est-à-dire d'étudier le comportement asymptotique de

$$Y_n = \frac{1}{n} (X_1 + \dots + X_n),$$

non plus seulement pour des variables aléatoires, mais pour des éléments aléatoires  $X$  de nature quelconque et sous des hypothèses aussi générales que possible. La nature même du problème impose alors de considérer des éléments aléatoires à valeurs dans un espace vectoriel topologique ; d'autre part, une des généralisations les plus importantes de la notion de variable aléatoire est celle de fonction aléatoire qui intervient dans de nombreux problèmes tant pratiques que théoriques : problèmes posés par la théorie des télécommunications, de l'information, etc... Or, un moyen commode d'étudier une fonction aléatoire est de la considérer comme élément aléatoire à valeurs dans un espace approprié, dans de nombreux cas un espace  $L_p$  de Lebesgue. Par conséquent, la théorie des espaces vectoriels topologiques, et en particulier des espaces de Banach, intervient naturellement dans toute étude générale de lois des grands nombres. D'autre part, l'hypothèse de l'existence de  $EX_i$  intervenant dans la loi forte des grands nombres, pour des variables aléatoires, nécessite la définition de l'espérance mathématique d'un élément aléatoire à valeurs dans un espace vectoriel topologique  $\mathcal{X}$ , et ainsi intervient la théorie de l'intégration d'une fonction définie sur un ensemble  $\mathcal{U}$  quelconque, à valeurs dans un tel espace  $\mathcal{X}$ .

Lorsqu'on considère des éléments aléatoires à valeurs dans un espace de Banach  $\mathcal{X}$ , il convient de se limiter aux applications  $x(u)$  de  $\mathcal{U}$  dans  $\mathcal{X}$  telles que  $\langle x^*, x(u) \rangle$  soit mesurable pour tout  $x^* \in \mathcal{X}^*$  ( $\mathcal{X}^*$  dual fort de  $\mathcal{X}$ ) ; on dira alors que  $X = x(u)$  est un L-élément aléatoire. Alors, par définition, l'espérance mathématique de  $X$ ,  $EX$ , est l'élément de  $\mathcal{X}$ , s'il existe, tel que  $\langle x^*, EX \rangle = E \langle x^*, X \rangle$  pour tout  $x^* \in \mathcal{X}^*$ , c'est-à-dire que  $EX$  est l'intégrale de Pettis de  $x(u)$ .

Alors, sans grandes difficultés, le théorème ergodique de Birkhoff permet de démontrer la « loi forte des grands nombres dans un espace de Banach » :

Si  $\mathcal{X}$  est un espace de Banach, séparable, si  $\{X_i\}$  est une suite strictement stationnaire de L-éléments aléatoires à valeurs dans  $\mathcal{X}$  et si  $E(\|X_i\|) < +\infty$ , presque-sûrement, quand  $n \rightarrow +\infty$ , la moyenne temporelle

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i$$

tend fortement vers une limite  $Y$  qui est un

L-élément aléatoire à valeurs dans  $\mathcal{X}$  et tel que  $EY = EX_i$ .

Si les  $X_i$  sont des éléments aléatoires indépendants de même loi,  $Y$  se réduit encore à l'élément certain  $EX_i$ .

Le théorème ergodique de Yosida et Kakutani ne faisant pas intervenir de mesure ne peut évidemment pas permettre d'obtenir une loi forte des grands nombres, c'est-à-dire une convergence presque-sûre, mais il permet d'établir une « loi des grands nombres en moyenne d'ordre  $\alpha$  dans un espace de Banach ».

Si  $\mathcal{X}$  est un espace de Banach séparable, si  $\{X_i\}$  est une suite strictement stationnaire de L-éléments aléatoires à valeurs dans  $\mathcal{X}$  et tels que  $E(\|X\|^\alpha) < +\infty$ , avec  $1 \leq \alpha < +\infty$ ; il existe un L-élément aléatoire  $Y$  à valeurs dans  $\mathcal{X}$ , tel que  $E(\|Y\|^\alpha) < +\infty$ ,  $EY = EX_i$ , et que

$$\lim_{n \rightarrow +\infty} E\left(\left\|\frac{1}{n}(X_1 + \dots + X_n) - Y\right\|^\alpha\right) = 0.$$

Toutefois, ce théorème est loin de découler directement de celui de Yosida et Kakutani; sa démonstration nécessite en particulier l'étude d'un espace  $\overset{a}{\mathcal{X}}$  associé à l'espace de Banach, séparable,  $\mathcal{X}$ , et des fonctionnelles linéaires définies sur  $\overset{a}{\mathcal{X}}$ .

$\overset{a}{\mathcal{X}}$  est défini de la façon suivante: tout L-élément aléatoire  $X$  à valeurs dans  $\mathcal{X}$  et tel que  $E(\|X\|^\alpha) < +\infty$  est considéré comme un élément  $\overset{a}{X}$  d'un espace  $\overset{a}{\mathcal{X}}$  normé en posant:

$$\|\overset{a}{X}\| = [E(\|x\|^\alpha)]^{1/\alpha}$$

ou encore  $\overset{a}{\mathcal{X}}$  est l'espace constitué par les fonctions  $X = x(u)$  telles que

$$\int_{\mathcal{U}} \|x(u)\|^\alpha d\mu < +\infty \text{ avec } \|\overset{a}{X}\| = \left[ \int_{\mathcal{U}} \|x(u)\|^\alpha d\mu \right]^{1/\alpha}$$

ainsi défini,  $\overset{a}{\mathcal{X}}$  est un espace de Banach.



## INTRODUCTION

### A L'ETUDE DES PROCESSUS STOCHASTIQUES

par A. FUCHS, *Professeur à la Faculté des Sciences de Strasbourg*

#### 1. — Rappel de la notion de variable aléatoire.

Commençons par traiter le cas simple suivant. On considère l'épreuve consistant à jeter un dé parfait. Cette épreuve peut aboutir à l'un des six résultats  $\omega_1, \omega_2, \dots, \omega_6$  où  $\omega_k$  ( $k = 1, \dots, 6$ ) désigne l'événement consistant en le fait d'amener la face portant le numéro  $k$ . L'ensemble des résultats logiquement possibles, que nous désignerons par  $\Omega$ , sera formé des six événements  $\omega_1, \dots, \omega_6$ , appelés événements élémentaires :

$$\Omega = \{ \omega_1, \dots, \omega_6 \} .$$

Plus généralement, on appellera *événement* toute partie (ou tout sous-ensemble) de  $\Omega$ .

*Exemples.* —

$E = \Omega - \omega_6 = \omega_1 \cup \omega_2 \cup \dots \cup \omega_5$  : ne pas amener le 6,

$E = \omega_2 \cup \omega_4 \cup \omega_6$  : amener un numéro pair,

$E = \omega_3 \cup \omega_6$  : amener un numéro multiple de 3.

Parmi les événements, il y en a deux qui jouent un rôle d'événements extrêmes ; ce sont :

$\emptyset$  : événement logiquement *impossible*, par exemple amener le 7 ;

$\Omega$  : événement logiquement *certain* : amener au moins un numéro de 1 à 6.

On désignera l'ensemble de tous les événements par  $B_\Omega$  ; cet ensemble possède les propriétés suivantes :

$$a) E \in B_\Omega \Rightarrow \bar{E} \in B_\Omega .$$

En d'autres termes, si  $E$  est un événement, il en est de même de son contraire (événement contraire).

$$a) E, F \in B_\Omega \Rightarrow \left\{ \begin{array}{l} E \cup F \in B_\Omega \\ E \cap F \in B_\Omega \end{array} \right. .$$

En d'autres termes, si  $E$  et  $F$  sont des événements, il en est de même de leur réunion et de leur intersection.

On dit que l'ensemble  $B_\Omega$  forme une *tribu* (ou un anneau de Boole, ou un corps d'ensembles).

Revenons à notre jeu de dés.

À chaque événement on pourra associer un poids positif, appelé *probabilité*, de la manière suivante : à chaque événement élémentaire  $\omega_k$ ,  $k = 1, \dots, 6$ , on associera la probabilité  $\frac{1}{6}$  et si un événement  $E$  est constitué de  $l$  événements élémentaires disjoints, on lui associera la probabilité  $\frac{l}{6}$ . On postule en outre que la probabilité de l'événement  $\emptyset$  est nulle. Avec ces conventions, la probabilité de l'événement certain  $\Omega$  est égale à 1.

À chaque événement  $E \in B_\Omega$  nous avons ainsi pu associer une probabilité  $P(E)$ . La probabilité apparaît donc comme une fonction définie sur les éléments de  $B_\Omega$  et possédant les propriétés suivantes :

- a) pour tout  $E \in B_\Omega$ ,  $P(E) \geq 0$ ,
- b)  $P(\Omega) = 1$ ,
- c)  $E, F \in B_\Omega$ ,  $E \cap F = \emptyset \Rightarrow P(E \cup F) = P(E) + P(F)$ .

En d'autres termes, si  $E$  et  $F$  sont deux événements disjoints, la probabilité de leur réunion est égale à la somme de leurs probabilités.

Finalement, tout problème relatif au jeu de dé peut être résolu au moyen du triplet  $(\Omega, B_\Omega, P)$  où  $B_\Omega$  et  $P$  possèdent les propriétés énoncées ci-dessus.

Pour bâtir une théorie générale des probabilités, A. Kolmogoroff a érigé en axiomes les propriétés que nous avons mises en évidence sur le cas simple du jeu de dé. Il a en outre formulé ses axiomes de façon suffisamment générale pour pouvoir englober le cas où l'espace de base  $\Omega$  ne possède plus nécessairement un nombre fini d'éléments comme c'était le cas dans le jeu de dé.

**DÉFINITION 1.** — On appelle espace probabilisé tout triplet  $(\Omega, B_\Omega, P)$  formé des éléments suivants :

a)  $\Omega$  est un espace abstrait qui, dans les applications, peut être identifié avec l'ensemble des résultats logiquement possibles d'une épreuve bien définie.

b)  $B_\Omega$  est une famille de sous-ensembles de  $\Omega$ , appelés événements, qui possède les propriétés suivantes :

$$E \in B_\Omega \Rightarrow \bar{E} \in B_\Omega,$$

$$E_n \in B_\Omega; n = 1, 2, \dots \Rightarrow \left\{ \begin{array}{l} \bigcup_{n=1}^{\infty} E_n \in B_\Omega \\ \bigcap_{n=1}^{\infty} E_n \in B_\Omega \end{array} \right.$$

On dit que  $B_\Omega$  est une *tribu borélienne* (ou corps borélien d'ensembles).

c)  $P$  est une fonction définie sur  $B_\Omega$  qui possède les propriétés suivantes :

— pour tout  $E \in \mathcal{B}_\Omega$ ,  $P(E) \geq 0$  ;

—  $P(\Omega) = 1$  ;

— Si  $E_n \in \mathcal{B}_\Omega$ ,  $n = 1, 2, \dots$ , et  $E_m \cap E_n = \emptyset$ , quels que soient  $m$  et  $n$ , pourvu que  $m \neq n$ ,

alors :

$$P \left\{ \bigcup_{n=1}^{\infty} E_n \right\} = \sum_{n=1}^{\infty} P(E_n).$$

Cette dernière propriété est connue sous le nom d'axiome d'additivité complète.

DÉFINITION 2. — On appelle *variable aléatoire*, et on désigne par  $X$ , toute application mesurable de  $\Omega$  dans  $\mathbb{R}$  (espace des nombres réels) :

$$X : \Omega \rightarrow \mathbb{R}.$$

Postuler que l'application est *mesurable* revient, dans le cas d'une variable aléatoire réelle auquel nous nous limitons ici, à postuler que :

$$\{ \omega : X(\omega) < x \} = X^{-1}(] - \infty, x[) \in \mathcal{B}_\Omega.$$

Dans cette notation, le premier et le second membre désignent l'ensemble de tous les éléments  $\omega \in \Omega$  qui ont une image  $< x$ . Une variable aléatoire apparaît ainsi comme une fonction réelle d'une variable  $\omega \in \Omega$  que l'on pourra désigner par  $X(\omega)$ .

DÉFINITION 3. — On appelle *fonction de répartition* de la variable aléatoire  $X$ , et on désigne par  $F(x)$ , la fonction :

$$F(x) = P \{ \omega : X(\omega) < x \} = P \{ X^{-1}(] - \infty, x[) \}.$$

Pour éviter des difficultés d'écriture, on écrit souvent :

$$F(x) = P \{ X < x \}.$$

On démontre que  $F(x)$  est une fonction non décroissante et continue à gauche. On fait généralement l'hypothèse supplémentaire que :

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \text{et} \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

ce qui revient à postuler que la probabilité pour que  $|X|$  prenne la valeur  $\infty$  est nulle.

La fonction de répartition définit entièrement la loi de probabilité de  $X$  (plus précisément, elle permet de déterminer la probabilité de tout événement du type  $X \in E$ ,  $E$  étant un ensemble linéaire mesurable au sens de Borel).

Pour bien mettre en relief les relations entre l'espace  $\Omega$  et la variable aléatoire  $X$ , donnons quelques exemples simples tirés du jeu de dé.

*Exemple 1.* — Au jeu de dé, soit  $X$  la variable aléatoire égale au numéro de la face amenée :

$$X : \Omega \rightarrow \mathbb{R} \iff \omega_k \rightarrow k, \quad k = 1, 2, \dots, 6.$$

La loi de probabilité de  $X$  sera définie par :

$$p(k) = P \{ X^{-1}(k) \} = P \{ \omega_k \} = \frac{1}{6}, \quad k = 1, 2, \dots, 6.$$

*Exemple 2.* — Toujours au jeu de dé, convenons que je gagne 1 franc en amenant le 6, et 0 franc dans le cas contraire. Le gain  $G$ , en une épreuve, est une variable aléatoire définie par :

$$G : \Omega \rightarrow \mathbb{R} \iff \begin{cases} \omega_6 \rightarrow 1 \\ \omega_k \rightarrow 0 \end{cases} \quad k = 1, 2, \dots, 5$$

et la loi de probabilité de  $G$  sera définie par :

$$p(0) = P \{ G^{-1}(0) \} = P \{ \omega_1 \cup \dots \cup \omega_5 \} = \frac{5}{6},$$

$$p(1) = P \{ G^{-1}(1) \} = P \{ \omega_6 \} = \frac{1}{6}.$$

*Remarque.* — Nous avons vu l'importance que revêt l'ensemble  $\Omega$  des résultats d'une épreuve. Mais il ne faudrait pas croire que cet ensemble soit unique. Ainsi, dans le jeu de dé, on peut considérer que le résultat de l'épreuve est connu si l'on connaît la position et la vitesse initiale du dé au moment où je le lâche. En d'autres termes, si l'on connaît un point de  $\mathbb{R}^{12}$ . En toute rigueur on pourrait également prendre  $\mathbb{R}^{12}$  comme ensemble fondamental  $\Omega$  et pondérer  $\mathbb{R}^{12}$ .

Cette indétermination dans la définition de  $\Omega$  n'a pourtant pas d'influence sur l'élaboration de la théorie. Ainsi, dans le cas ci-dessus, on pourra se ramener à la pondération usuelle de la manière suivante : on partage  $\mathbb{R}^{12}$  en six sous-ensembles disjoints  $E_1, \dots, E_6$  tels que :

$$x \in E_k \iff \text{le dé tombe sur la face } n^\circ k, \quad k = 1, 2, \dots, 6.$$

Il suffit alors d'affecter chaque  $E_k$  du poids  $\frac{1}{6}$  pour avoir un modèle équivalent au modèle usuel.

## 2. — Fonction aléatoire sur un intervalle.

Soit  $T$  un intervalle de la droite réelle, pouvant d'ailleurs s'étendre jusqu'à l'infini.

*Définition :* On appelle fonction aléatoire définie sur  $T$  et l'on désigne par  $X$  toute application de  $\Omega \times T$  dans  $\mathbb{R}$ , qui est en outre mesurable pour tout  $t \in T$  :

$$X : \Omega \times T \rightarrow \mathbb{R}.$$

Une fonction aléatoire sur  $T$  apparaît ainsi comme une fonction réelle de deux variables  $t \in T, \omega \in \Omega$ , que l'on pourra désigner par

$$X(t, \omega).$$

Dire que cette fonction est mesurable pour tout  $t$  fixé  $\in T$ , c'est dire que pour tout  $t_0 \in T, X(t_0, \omega)$  est une variable aléatoire au sens où nous l'avons définie au premier paragraphe.

Dans les cas usuels, on identifie  $t$  avec le temps et on préfère alors parler de *processus stochastiques* plutôt que de fonction aléatoire.

Pour étudier les fonctions aléatoires, on peut adopter deux points de vue, d'ailleurs difficiles à concilier en toute rigueur, suivant que l'on attache plus d'importance à  $\Omega$  ou à  $T$ .

a) ÉCOLE DE J.-L. DOOB. — Pour  $\omega$  fixé  $\in \Omega$ , c'est-à-dire pour un

résultat d'épreuve donné,  $X(t, \omega)$  est une fonction de  $t$  seul que nous désignerons par  $X_\omega(t)$ . Ce sera une fonction non aléatoire de  $t$  (c'est-à-dire une fonction réelle au sens usuel) que l'on appelle *une réalisation* de la fonction aléatoire  $X$ . On pourra d'ailleurs sans inconvénient identifier  $\omega$  et  $X_\omega(t)$ .

b) ECOLE DE P. LÉVY. — Pour  $t$  fixé  $\in T$ ,  $X(t, \omega)$  est une fonction de  $\omega$  seul que nous désignerons par  $X_t(\omega)$ . D'après nos hypothèses, c'est une variable aléatoire.

La notion de fonction aléatoire apparaît ainsi sous un double aspect.

c) Elle traduit tout d'abord l'idée de choisir, suivant une certaine loi de probabilité à définir, une fonction dans une classe de fonctions  $\{X_\omega(t)\} = \Omega$ . Ceci justifie la notation  $X_\omega(t)$ .

Dans ce cas, l'élément aléatoire n'est plus un nombre, mais une fonction.

d) Dans le deuxième point de vue, la fonction aléatoire  $X(t, \omega)$  est considérée comme une famille de variables aléatoires dépendant d'un paramètre  $t \in T$ . Ceci justifie la notation  $X_t(\omega)$ .

Dans le cas où  $t$  est identifié avec le temps, cette dernière manière de voir semble plus conforme à l'intuition. Si l'on considère en effet  $X_t(\omega)$  comme une quantité caractérisant l'état aléatoire à l'instant  $t$  d'un certain système physique, elle traduit immédiatement l'évolution aléatoire, au cours du temps, de ce système.

*Exemple.* — Soit à étudier la pression atmosphérique dans un intervalle de temps  $T$  au moyen d'un baromètre enregistreur. La variation de cette pression dans cet intervalle de temps est une fonction aléatoire définie sur  $T$ .

Un relevé fourni par le baromètre enregistreur après écoulement de l'intervalle de temps  $T$  est une *réalisation* de cette fonction aléatoire.

Le point de vue de J.-L. Doob consiste à considérer comme élément aléatoire toute fonction réelle continue définie sur  $T$ ; celui de P. Lévy, au contraire, consiste à considérer comme élément aléatoire la valeur de la pression à un instant déterminé  $t \in T$ , et à suivre l'évolution au cours du temps de cet élément.

### 3. — Loi temporelle ; principaux types de processus.

a) LOI TEMPORELLE. — On dit que l'on connaît la *loi temporelle* de la fonction aléatoire  $X(t, \omega)$  définie sur  $T$  si, quels que soient l'entier  $n$ , les instants  $t_1, \dots, t_n \in T$  et les ensembles linéaires  $E_1, \dots, E_n$  mesurables au sens de Borel, on connaît la probabilité :

$$P \{ X(t_1, \omega) \in E_1, \dots, X(t_n, \omega) \in E_n \} .$$

Pour  $n$  et  $t_1, \dots, t_n$  fixés, l'ensemble de ces probabilités définit une loi de probabilité dans  $R^n$ . Un cas important est celui où cette loi de probabilité est une loi normale dans  $R^n$  (fonctions aléatoires gaussiennes ou laplaciennes).

Nous ne nous occuperons que des fonctions aléatoires dont les propriétés statistiques usuelles sont parfaitement définies par la donnée de

la loi temporelle. De telles fonctions sont appelées *séparables*. On démontre que ce sont les fonctions aléatoires telles que, quel que soit l'intervalle  $I \subset T$ , il existe une suite d'instants  $t_1, \dots, t_n \in I$  avec la propriété qu'avec une probabilité 1, on a :

$$\inf_{t \in I} X(t, \omega) = \inf_{t_i \in I} X(t_i, \omega),$$

$$\sup_{t \in I} X(t, \omega) = \sup_{t_i \in I} X(t_i, \omega).$$

b) FONCTIONS ALÉATOIRES STATIONNAIRES. — On dit que la fonction aléatoire  $X(t, \omega)$  définie sur  $T$  est (strictement) stationnaire si, quels que soient l'entier  $n$ , les instants  $t_1, \dots, t_n \in T$ , les ensembles linéaires  $E_1, \dots, E_n$  mesurables au sens de Borel et le nombre  $h$  tel que :  $t_1 + h, \dots, t_n + h \in T$ , on a :

$$P \{ X(t_1 + h, \omega) \in E_1, \dots, X(t_n + h, \omega) \in E_n \} = P \{ X(t_1, \omega) \in E_1, \dots, X(t_n, \omega) \in E_n \}.$$

On peut dire en gros que la fonction aléatoire  $X(t, \omega)$  est (strictement) stationnaire si la loi temporelle est invariante par translation sur l'axe des  $t$ .

*Remarque.* — Dans les applications, il est souvent suffisant de ne postuler la propriété ci-dessus que pour  $n = 1, 2$ . On dit alors que la fonction aléatoire  $X(t, \omega)$  est *stationnaire du second ordre*.

c) PROCESSUS DE MARKOFF. — Soit  $X(t, \omega)$  un processus stochastique défini pour  $t \geq 0$ . On dit que ce processus est *de Markoff* si, quels que soient  $t, \tau$ , avec  $0 \leq t < \tau$ , la loi de probabilité de  $X(\tau, \omega)$  dépend de la valeur prise par  $X(t, \omega)$  à l'instant  $t$ , mais, une fois cette valeur fixée, est indépendante de l'ensemble des variables aléatoires  $X(t', \omega)$ ,  $0 \leq t' < t$ .

La loi temporelle d'un processus de Markoff est déterminée entièrement par la donnée simultanée de la loi de probabilité initiale (si le processus a débuté à l'instant  $t = 0$ , ce que nous supposons ici) et de la loi de probabilité dite *de passage*, à savoir :

$$P \{ X(\tau, \omega) \in E \mid X(t, \omega) = x \} = F(t, x; \tau, E)$$

où  $E$  est un ensemble linéaire mesurable au sens de Borel.

La fonction  $F$  satisfait nécessairement à l'équation fonctionnelle suivante, dite de *Chapman-Kolmogoroff* :

Quels que soient  $0 \leq t < t' < \tau$ , on a :

$$(1) \quad F(t, x; \tau, E) = \int_{-\infty}^{+\infty} F(t, x; t', dy) F(t', y; \tau, E).$$

L'intégrale du second membre étant entendue au sens de Lebesgue-Stieltjes. Remarquons que l'équation (1) n'est pas caractéristique du processus de Markoff. P. Lévy a construit un processus qui satisfait à (1) mais qui n'est pas de Markoff.

*Remarque 1.* — Un processus de Markoff est dit *homogène dans le temps*, si la fonction  $F(t, x; \tau, E)$ ,  $t < \tau$ , ne dépend de  $t$  et  $\tau$  que par leur différence  $\tau - t$ .

*Remarque 2.* — Un processus de Markoff est dit à *accroissements indépendants* ou *additifs* si, quels que soient  $0 \leq t < t' < \tau$ , les variables aléatoires  $X(t', \omega) - X(t, \omega)$  et  $X(\tau, \omega) - X(t', \omega)$  sont *indépendantes*. Le deuxième membre de l'équation (1) de Chapman-Kolmogoroff se réduit alors à une convolution.

#### 4. — Continuité des fonctions aléatoires.

##### a) CONTINUITÉ LOCALE.

1) *En probabilité.* — Soit toujours  $X(t, \omega)$  une fonction aléatoire définie sur  $T$ . On dit que  $X(t, \omega)$  est continue en probabilité au point  $t_0 \in T$  si, à tout couple de nombres  $\varepsilon, \eta > 0$  on peut associer un voisinage  $\mathcal{O}(t_0, \varepsilon, \eta) \subset T$  du point  $t_0$  tel que :

$$t \in \mathcal{O} \Rightarrow P \{ |X(t, \omega) - X(t_0, \omega)| > \varepsilon \} < \eta.$$

Remarquons que la seule loi temporelle est suffisante pour définir la continuité locale en probabilité.

2) *Presque sûre.* — On dit que  $X(t, \omega)$  est continue presque sûrement au point  $t_0 \in T$ , si à tout couple  $\varepsilon, \eta > 0$  on peut associer un voisinage  $\mathcal{O}(t_0, \varepsilon, \eta)$  de  $t_0$  tel que :

$$P \left\{ \sup_{t \in \mathcal{O}} |X(t, \omega) - X(t_0, \omega)| > \varepsilon \right\} < \eta.$$

Remarquons que la seule loi temporelle ne suffit pas pour définir la continuité locale presque sûre. En effet, l'ensemble :

$$\left\{ \sup_{t \in \mathcal{O}} |X(t, \omega) - X(t_0, \omega)| > \varepsilon \right\} = \bigcup_{t \in \mathcal{O}} \{ |X(t, \omega) - X(t_0, \omega)| > \varepsilon \}.$$

est la réunion d'une infinité non dénombrable d'éléments de  $B_\Omega$ , et cette réunion n'est pas en général un élément de  $B_\Omega$ , donc n'est pas en général affecté d'une probabilité.

On démontre que si la fonction aléatoire est *séparable*, les notions de continuité locale en probabilité et presque sûre coïncident.

*Remarque.* — La continuité presque sûre au point  $t_0 \in T$  entraîne l'absence, avec probabilité 1, au point  $t_0$ , d'une discontinuité de  $X(t)$  (discontinuité fixe).

##### b) CONTINUITÉ GLOBALE.

Le fait qu'une fonction aléatoire soit localement presque sûrement continue en tout point de  $T$  entraîne l'absence, avec probabilité 1, de toute discontinuité fixe dans  $T$ . Mais il n'entraîne pas pour autant que cette fonction soit, avec probabilité 1, une fonction continue dans  $T$ . Il peut en effet exister des discontinuités dont la position ne peut être fixée *a priori* ; en d'autres termes, dont la position est aléatoire. De telles discontinuités s'appellent discontinuités *mobiles*. Exemple : Processus de Poisson.

Soit par exemple  $N(t)$ ,  $t > 0$ , le nombre de tops enregistrés par un compteur Geiger dans l'intervalle de temps  $[0, t[$ . On démontre que, moyennant des hypothèses extrêmement générales sur l'arrivée des tops,

$N(t)$  définit un processus à accroissements indépendants, ne pouvant varier que par sauts d'intensité 1 aux instants des tops et dont les accroissements obéissent à la loi de Poisson suivante :

$$h > 0 : P \{ X(t+h) - X(t) = k \} = e^{-\lambda h} \frac{(\lambda h)^k}{k!}, \quad k \text{ entier } \geq 0,$$

où  $\lambda$  est un paramètre numérique appelé *densité* du processus, et qui représente le nombre moyen de tops par unité de temps.

a)  $N(t)$  est localement continue en probabilité en tout point  $t > 0$ . En effet, quel que soit  $h > 0 : t-h > 0$ , on a (pour  $0 < \varepsilon < 1$ ) :

$$P \{ X(t+h) - X(t-h) > \varepsilon \} = \sum_{k=1}^{\infty} e^{-2\lambda h} \frac{(2\lambda h)^k}{k!} = 1 - e^{-2\lambda h},$$

qui tend vers 0 lorsque  $h \rightarrow 0$ .

b) Quel que soit  $\varepsilon > 0$ , il existe un intervalle de temps  $[T_0, T_1[$ ,  $T_0 < T_1$ , suffisamment grand, tel que  $X(t)$  soit discontinu dans cet intervalle avec probabilité  $> 1 - \varepsilon$ . En effet :

$$P \{ X(T_1) - X(T_0) > 0 \} = \sum_{k=1}^{\infty} e^{-\lambda(T_1-T_0)} \frac{[\lambda(T_1-T_0)]^k}{k!} = 1 - e^{-\lambda(T_1-T_0)}.$$

On peut naturellement choisir  $T_1 - T_0$  assez grand pour que :

$$1 - e^{-\lambda(T_1-T_0)} > 1 - \varepsilon.$$

Pour que la fonction aléatoire  $X(t, \omega)$  ne présente, dans un intervalle donné, ni discontinuités fixes, ni discontinuités mobiles, il faut donc lui imposer des conditions plus strictes que des conditions de continuité locale : de telles conditions sont dites de *continuité globale*. Nous ne les énoncerons pas, en raison de leur complexité. Nous nous bornerons à signaler que, dans le cas des processus de Markoff, la condition suivante, dite de W. Feller, entraîne la continuité globale :

Condition de Feller (pour les processus de Markoff) :

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|y-x| > \varepsilon} F(t, x; t + \Delta t, dy) = 0 \quad \text{quel que soit } \varepsilon > 0,$$

la limite étant atteinte uniformément par rapport à  $t$  et à  $x$ .

## 5. — Mouvement brownien linéaire.

Soit  $X(t, \omega)$  une fonction aléatoire réelle, définie sur toute la droite réelle, à *accroissements indépendants*, la loi de probabilité de l'accroissement  $\Delta X(t, \omega)$  sur un intervalle  $[t, t + \Delta t[$ ,  $\Delta t > 0$  étant une loi de Laplace-Gauss indépendante de  $t$ , avec :

$$\begin{aligned} E \{ \Delta X(t, \omega) \} &= 0, \\ \sigma^2 \{ \Delta X(t, \omega) \} &= E \{ [\Delta X(t, \omega)]^2 \} = \Delta t. \end{aligned}$$

Une telle fonction porte le nom de fonction aléatoire du *mouvement brownien*. Si  $t$  est le temps, elle définit un processus de Markoff à la fois homogène dans le temps et à accroissements indépendants.

Nous nous bornerons à établir deux propriétés importantes de cette fonction.

a) Nous allons voir que  $X(t, \omega)$  nous fournit un exemple de fonction aléatoire n'admettant, avec probabilité 1, ni discontinuités fixes, ni discontinuités mobiles. En d'autres termes, à l'exception d'un ensemble de valeurs de  $\omega$  de mesure nulle, toute réalisation de la fonction aléatoire est une fonction continue.

Pour le voir, il suffit de vérifier que la condition de Feller ci-dessus est satisfaite. En effet, on a : pour  $\Delta t > 0$  :

$$\frac{1}{\Delta t} \int_{|y-x|>\varepsilon} F(t, x; t + \Delta t, dy) = \frac{1}{\Delta t} P \{ |\Delta X(t, \omega)| > \varepsilon \} = \frac{1}{\Delta t} \frac{1}{\sqrt{2\pi\Delta t}} \int_{|u|>\varepsilon} e^{-\frac{u^2}{2\Delta t}} du,$$

d'où, en posant :  $\frac{u}{\sqrt{\Delta t}} = v$  :

$$\frac{1}{\Delta t} \cdot \frac{1}{\sqrt{2\pi}} \int_{|v|>\frac{\varepsilon}{\sqrt{\Delta t}}} e^{-\frac{v^2}{2}} dv = \frac{2}{\Delta t \sqrt{2\pi}} \int_{v>\frac{\varepsilon}{\sqrt{\Delta t}}} e^{-\frac{v^2}{2}} dv = \frac{2}{\Delta t} \left[ 1 - \Phi\left(\frac{\varepsilon}{\sqrt{\Delta t}}\right) \right],$$

où l'on a posé :

$$\Phi(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} e^{-\frac{v^2}{2}} dv,$$

or, si  $\xi > 0$ , on a :

$$1 - \Phi(\xi) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\xi} e^{-\frac{\xi^2}{2}}.$$

donc :

$$\begin{aligned} \frac{1}{\Delta t} \int_{|y-x|>\varepsilon} F(t, x; t + \Delta t, dy) &\leq \frac{\varepsilon}{\Delta t} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{\Delta t}}{\varepsilon} \cdot e^{-\frac{1}{2} \frac{\varepsilon^2}{\Delta t}} \\ &= \frac{\sqrt{2}}{\varepsilon \sqrt{2\pi\Delta t}} e^{-\frac{1}{2} \frac{\varepsilon^2}{\Delta t}}, \end{aligned}$$

quantité qui tend vers 0 lorsque  $\Delta t \rightarrow 0$ , et ceci quel que soit  $\varepsilon > 0$ , et uniformément en  $t$  et en  $x$ .

b) La fonction aléatoire  $X(t, \omega)$  n'admet de dérivée finie en aucun point, et ceci avec probabilité 1.

En effet, on sait que toute variable aléatoire est de l'ordre de grandeur de sa dispersion ; donc, pour  $\Delta t > 0$  :

$$X(t + \Delta t, \omega) - X(t, \omega) = O(\sqrt{\Delta t})$$

où  $O(x)$  désigne une quantité du même ordre de grandeur que  $x$  ou :

$$\frac{X(t + \Delta t, \omega) - X(t, \omega)}{\Delta t} = \frac{O(\sqrt{\Delta t})}{\Delta t} = O\left(\frac{1}{\sqrt{\Delta t}}\right).$$

Le second membre tend vers  $+\infty$  lorsque  $\Delta t \rightarrow 0$  ; il en est donc de même du premier.



# STATISTIQUE MATHÉMATIQUE

## PLANS D'EXPERIENCES

Daniel DUGUÉ

*Professeur à la Sorbonne*

### I. — DEFINITION DE LA STATISTIQUE MATHÉMATIQUE A L'INTERIEUR DU CALCUL DES PROBABILITES

Dans le calcul des probabilités, la notion de convergence se fractionne en deux notions différentes qui se réduisent à une seule dans le cas d'éléments certains :

- 1) la convergence en probabilité ou faible ;
- 2) la convergence presque certaine (ou presque sûre) ou forte, qui entraîne la précédente.

1° Une suite d'éléments aléatoires  $X_1, X_2, \dots, X_n$  tend en probabilité vers un élément  $X$  si :

$$\lim. \text{Prob} [|X_n - X| < \varepsilon] = 1,$$

quelque petit que soit  $\varepsilon$ .

2° Une suite d'éléments aléatoires  $X_1, X_2, \dots, X_n$  tend presque sûrement vers un élément  $X$  si :

$$\text{Prob} [\lim X_n = X] = 1.$$

A mon avis (cette opinion est contestée, et je citerai, parmi ceux qui ne la partagent pas, MM. Fréchet et Brard), la statistique mathématique est le domaine de la convergence en probabilité, le calcul des probabilités théoriques est celui de la convergence presque sûre.

Cela tient au fait que la statistique, science appliquée, ne peut s'intéresser à « une suite infinie de résultats », et que dire qu'une suite de variables aléatoires converge presque sûrement, c'est dire qu'il y a une probabilité unité pour qu'une suite infinie de résultats tende vers une limite.

Je rappelle, comme je l'ai dit au début de ce paragraphe, que si l'on a affaire à des quantités certaines, les deux manières de converger coïncident.

On peut présenter cette remarque en disant que le calcul des probabilités est une extension des mathématiques certaines au même titre que la géométrie non-euclidienne est une extension de la géométrie euclidienne (cette dernière s'obtenant à partir de la géométrie non-euclidienne par l'intervention de l'axiome d'Euclide : par tout point passe une seule parallèle à une droite).

Les mathématiques certaines sont le domaine du calcul des proba-

bilités dans lequel on ne considère que les éléments pour lesquels convergence en probabilité et convergence presque certaine coïncident.

En introduisant cette idée, je me permets d'insister d'une manière plus générale sur le caractère tout relatif d'un travail d'axiomatisation. Ce sont certes des recherches intéressantes, mais il ne faut pas perdre de vue que c'est une tâche qui ne sera jamais terminée. La rigueur définitive ne peut jamais être atteinte. Toute axiomatique sous-entend obligatoirement un certain nombre de conventions auxquelles on peut renoncer.

Un autre point qui montre le calcul des probabilités comme une extension des mathématiques certaines est le suivant, d'ailleurs lié au développement précédent.

Un ensemble  $X_1, X_2, X_3, \dots, X_n$  de quantités certaines est connu si l'on connaît chaque élément *isolément*. Il n'en est plus de même si  $X_1, \dots, X_n$  sont aléatoires. La connaissance de chacun des éléments  $X_1, \dots, X_n$ , ou connaissance marginale, ne renseigne pas complètement sur l'ensemble des  $n$  éléments. C'est le phénomène de la corrélation qui est typiquement aléatoire.

Ces quelques idées théoriques étant amorcées, je voudrais décrire un problème statistique particulier, celui des *plans d'expériences* (*design of experiments*, en anglais).

## II. — PLANS D'EXPERIENCES

La statistique, je l'ai dit, ne s'intéresse qu'à un nombre fini de résultats. Elle va avoir, dans ce problème des plans d'expériences, à utiliser des notions d'algèbre finie et en particulier celle de corps de Galois et d'espaces construits sur ces corps de Galois.

Soient  $n$  variables aléatoires ( $n = pq$ ), à deux indices,  $i$  prenant  $p$  valeurs et  $j$  en prenant  $q$  :

- 1) normales (obéissant à une loi de Laplace-Gauss) ;
- 2) indépendantes ;
- 3) ayant toutes le même écart-type  $\sigma$ .

Les  $p$  valeurs de l'indice  $i$  représentent l'influence d'un facteur, les  $q$  valeurs de l'indice  $j$  représentent l'influence d'un autre facteur ; répartissons les résultats dans les cases d'un tableau de  $p$  lignes et  $q$  colonnes. L'un de ces résultats sera  $x_{ij}$ , situé dans la  $i^{\text{e}}$  ligne et la  $j^{\text{e}}$  colonne.

				$x_{ij}$	

Posons-nous le problème suivant : les  $n = pq$  variables aléatoires ont-elles toutes la même valeur moyenne ; autrement dit,  $m_{ij}$ , valeur moyenne de la variable aléatoire  $X_{ij}$ , dont le résultat est  $x_{ij}$ , sera-t-elle indépendante de  $i$  et  $j$  ?

Pour arriver à résoudre « statistiquement » ce problème, la méthode va être de construire une ou plusieurs variables aléatoires, combinaisons linéaires des  $X_{ij}$ , et de calculer leur fonction de répartition dans l'hypothèse où  $m_{ij} = m$ , quels que soient  $i$  et  $j$ .

Supposons que l'une de ces variables aléatoires  $L$  ait une réalisation  $l$ . On rejetera l'hypothèse ( $m_{ij} = m$  pour tous les  $i$  et  $j$ ) si la probabilité de l'écart  $|l - E(L)|$  est inférieure à ce qu'on appelle le *seuil de signification* (1/20 ou 1/100 sont les seuils les plus usuels et pour lesquels les tables sont dressées).

Dans le cas actuel où les réalisations sont  $x_{ij}$ , on utilise l'égalité suivante :

$$\sum_{ij} (x_{ij} - m)^2 = \sum_{ij} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 + p \sum_j (x_{.j} - x_{..})^2 + q \sum_i (x_{i.} - x_{..})^2 + px(x_{..} - m)^2$$

où  $x_{i.}$  est la moyenne dans la ligne  $i$  :  $x_{i.} = \frac{1}{q} \sum_j x_{ij}$

$x_{.j}$  est la moyenne dans la colonne  $j$  :  $x_{.j} = \frac{1}{p} \sum_i x_{ij}$

$x_{..}$  est la moyenne dans l'ensemble du tableau :

$$x_{..} = \frac{1}{pq} \sum_i \sum_j x_{ij} = \frac{1}{p} \sum_i x_{i.} = \frac{1}{q} \sum_j x_{.j}$$

C'est ce qu'on appelle une *décomposition orthogonale* (car les termes rectangles disparaissent dans le développement) de la forme quadratique  $\Sigma(x_{ij} - m)^2$ .

La variable  $L$  définie tout à l'heure sera le quotient  $\frac{q_c^2}{Q^2}$  ou  $\frac{q_l^2}{Q^2}$ , où :

$$Q^2 = \sum_{ij} (x_{ij} - x_{i.} - x_{.j} + x_{..})^2 ; q_c^2 = p \sum_j (x_{.j} - x_{..})^2 ; q_l^2 = q \sum_i (x_{i.} - x_{..})^2$$

Ce sont des variables dont la loi peut être aisément calculée et qu'on appelle loi de Behrens-Fisher.

Si  $\frac{q_c^2}{Q^2} (p - 1)$  s'écarte significativement, au sens précisé, de sa valeur moyenne qui est l'unité,  $m_{ij}$  ne pourra pas être indépendante de  $j$ .

De même, si  $\frac{q_l^2}{Q^2} (q - 1)$  s'écarte significativement de l'unité,  $m_{ij}$  ne pourra être indépendante de  $i$ . Cela revient à rejeter le fait qu'un événement est dû au hasard s'il a une probabilité trop faible (« trop faible » étant fixé d'avance) de se réaliser.  $\sigma^2$  est « estimé », dans le cas où  $m_{ij}$

ne dépend pas de  $i$  et  $j$ , soit par  $\frac{q_c^2}{q - 1}$ , soit par  $\frac{q_l^2}{p - 1}$ .

La méthode du plan d'expérience que je viens de décrire, et qui est connue sous le nom d'*analyse de variance*, se généralise aisément à plus de deux indices.

Pour le cas de trois indices  $i, j, k$  prenant respectivement  $p, q, r$  valeurs, on a :

$$\begin{aligned} \sum_{ijk} (x_{ijk} - m)^2 &= \sum_{ijk} (x_{ijk} - x_{ij.} - x_{i.k} - x_{.jk} + x_{i..} + x_{.j.} + x_{..k} - x_{...})^2 \\ &+ r \sum_{ij} (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2 + q \sum_{ik} (x_{i.k} - x_{i..} - x_{..k} + x_{...})^2 \\ &+ p \sum_{jk} (x_{.jk} - x_{.j.} - x_{..k} + x_{...})^2 + pq \sum_k (x_{..k} - x_{...})^2 \\ &+ qr \sum_i (x_{i..} - x_{...})^2 + pr \sum_j (x_{.j.} - x_{...})^2 + pqr (x_{...} - m)^2, \end{aligned}$$

$x_{ij.}$ , par exemple, sera la moyenne des résultats dont les deux premiers indices sont  $i$  et  $j$ .

Chacune des sommes de carrés divisées par ce qu'on appelle le nombre de degrés de liberté, — pour  $r \sum_{ij} (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2$  ce sera  $(p-1)(q-1)$  —, fournit une estimation de  $\sigma^2$  dans le cas où  $m_{ijk}$  est indépendant de  $i, j, k$ .

Ces estimations sont indépendantes au sens du calcul des probabilités et leur quotient ne doit pas s'écarter significativement de l'unité. Les tables de Behrens-Fisher-Snedecor permettent encore de résoudre la question de la signification.

Dans le cas de trois indices, il faut donc  $pqr$  résultats, et si  $p = q = r$ ,  $p^3$  résultats pour appliquer cette méthode.

Nous allons étudier un procédé qui permet de n'en utiliser que  $p^2$ . Cette économie est un des buts du plan d'expérience, l'autre étant la simplification des calculs qui deviennent rapidement d'une complication monstrueuse à mesure qu'augmente le nombre des indices.

Supposons, par exemple, que  $p$  soit égal à 5 et examinons la figure suivante, que l'on appelle un *carré latin* :

A	B	C	D	E
C	D	E	A	B
E	A	B	C	D
B	C	D	E	A
D	E	A	B	C

Dans chaque ligne figure une fois et une seule chaque lettre. Il en est de même dans chaque colonne. Supposons que chaque case contienne un résultat aléatoire gaussien ; tous ces résultats étant indépendants les uns des autres et ayant même écart-type. Nous voulons encore tester l'hypothèse que la moyenne est la même pour toutes les variables.

Nous aurons ici une décomposition quadratique de la forme suivante :

$$\sum_{ij} (x_{ij} - m)^2 = \sum_{ij} (x_{ij} - x_i - x_j - x_t + 2x_{..})^2 + 5 \sum_i (x_i - x_{..})^2 + 5 \sum_j (x_j - x_{..})^2 + 5 \sum_t (x_t - x_{..})^2 + 25(x_{..} - m)^2,$$

$x_i$ , et  $x_j$  étant encore les moyennes des résultats dans les colonnes  $i$  et  $j$  et  $x_t$  la moyenne des résultats dans les cases portant la même lettre que la case considérée.

On aura encore, si la moyenne est la même pour toutes les cases (c'est-à-dire ne dépend ni de la ligne, ni de la colonne, ni de la lettre), des variables indépendantes dont les quotients obéiront aux lois de Behrens-Fisher-Snedecor, avec les degrés de liberté appropriés.

On dit ici que les lignes, colonnes et lettres sont orthogonales. Les termes rectangles disparaissent, car il y a une lettre donnée et une seule dans chaque ligne et dans chaque colonne.

De cette façon, avec  $5^2$  résultats, on obtient la même précision d'analyse de la variance qu'avec  $5^3$  dans le modèle factoriel à trois indices, et les calculs sont beaucoup plus simples.

On peut pousser plus loin l'orthogonalité en prenant deux carrés latins orthogonaux. Prenons un carré de 5 cases de côté, et considérons les deux carrés suivants :

A B C D E	$\alpha \beta \gamma \delta \varepsilon$
C D E A B	$\delta \varepsilon \alpha \beta \gamma$
E A B C D	$\beta \gamma \delta \varepsilon \alpha$
B C D E A	$\varepsilon \alpha \beta \gamma \delta$
D E A B C	$\gamma \delta \varepsilon \alpha \beta$

Nous pouvons constater que si on les superpose et si l'on considère les couples formés par la lettre latine et la lettre grecque qui se trouvent dans une case, tous les couples sont représentés :

A $\alpha$	B $\beta$	C $\gamma$	D $\delta$	E $\varepsilon$
C $\delta$	D $\varepsilon$	E $\alpha$	A $\beta$	B $\gamma$
E $\beta$	A $\gamma$	B $\delta$	C $\varepsilon$	D $\alpha$
B $\varepsilon$	C $\alpha$	D $\beta$	E $\gamma$	A $\delta$
D $\gamma$	E $\delta$	A $\varepsilon$	B $\alpha$	C $\beta$

A est ainsi associé à  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$ , une fois et une seule, et de même pour toutes les lettres latines.

Cette disposition permet la décomposition de la forme quadratique suivante :

$$\sum_{ij} (x_{ij} - m)^2 = \sum_{ij} (x_{ij} - x_i - x_j - x_t - x_\tau + 3x_{..})^2 + 5 \sum_i (x_i - x_{..})^2 + 5 \sum_j (x_j - x_{..})^2 + 5 \sum_t (x_t - x_{..})^2 + 5 \sum_\tau (x_\tau - x_{..})^2$$

à cause de l'orthogonalité des lignes et des colonnes, des lettres latines et des lettres grecques.

Ici, 5<sup>2</sup> résultats suffisent pour une analyse qui en aurait nécessité 5<sup>4</sup> avec le modèle factoriel ordinaire, puisqu'on a affaire à quatre indices (lignes, colonnes, lettres latines et lettres grecques).

Dans le cas d'un carré de 5 cases de côté, on peut construire quatre carrés latins orthogonaux deux à deux au sens que nous avons précisé :

A α a 0	B β b 1	C γ c 2	D δ d 3	E ε e 4
C δ b 4	D ε c 0	E α d 1	A β e 2	B γ a 3
E β c 3	A γ d 4	B δ e 0	C ε a 1	D α b 2
B ε d 2	C α e 3	D β a 4	E γ b 0	A δ c 1
D γ e 1	E δ a 2	A ε b 3	B α c 4	C β d 0

Dans ce cas, la forme quadratique  $\Sigma(x_{ij} - m)^2$  peut se décomposer de la façon suivante :

$$\begin{aligned} & \Sigma(x_{ij} - x_i - x_j - x_{\tau} - x_{\tau} - x_l - x_n + 5x_{..})^2 \\ & + 5 \Sigma(x_i - x_{..})^2 + 5 \Sigma(x_j - x_{..})^2 + 5 \Sigma(x_{\tau} - x_{..})^2 + 5 \Sigma(x_l - x_{..})^2 \\ & + 5 \Sigma(x_{\tau} - x_{..})^2 + 5 \Sigma(x_n - x_{..})^2 + 25(x_{..} - m)^2 \end{aligned}$$

$x_{\tau}$  étant la moyenne des résultats dans les cases contenant la même lettre majuscule que la case considérée et des définitions analogues pour les autres indices. Ici, le plan d'expérience permet, avec 5<sup>2</sup> résultats, une analyse de variance qui en aurait nécessité 5<sup>6</sup> avec le plan factoriel ordinaire.

On voit aisément qu'on ne peut avoir plus de  $p - 1$  carrés latins, de  $p$  cases de côté, orthogonaux deux à deux. Peut-on les avoir tous ? On conçoit que c'est là une question importante pour le plan d'expérience. On dit qu'on a alors orthogonalisation complète.

La solution de ce problème n'est pas obtenue à l'heure actuelle. On sait que pour les nombres  $p$  tels qu'il existe un corps de Galois ayant  $p$  éléments (c'est-à-dire tels que  $p = \omega^n$ ,  $\omega$  étant un nombre premier), il y a possibilité d'orthogonalisation complète. Mais on ne sait pas si la réciproque est exacte. Cela pose un problème d'algèbre finie très difficile.

On sait que pour tous les nombres  $p$  qui ne sont pas de la forme  $2(2n + 1)$ , il existe au moins deux carrés latins orthogonaux ayant  $p$  cases de côté. On sait qu'il n'y a pas deux carrés latins orthogonaux de 6 cases de côté. C'est le vieux problème des 36 officiers posé par Euler au milieu du XVIII<sup>e</sup> siècle et résolu par la négative par Tarry, en 1900. On ne sait rien pour le cas  $p = 10$ .

Je vais maintenant présenter un autre modèle que l'ensemble orthogonal de carrés latins, qui permet à la fois une simplification des calculs et une économie des résultats. Il s'agit de ce que Yates appelle le *bloc incomplet équilibré*.

Il s'agit dans ce problème de placer  $v$  lettres dans  $b$  lignes (blocs) de  $k$  cases chacune :

- 1) dans chaque ligne, chaque lettre figure 0 ou 1 fois ;
- 2) chaque lettre est répétée le même nombre de fois  $r$  dans tout le modèle (donc  $bk = rv$ ) ;

3) chaque combinaison de deux lettres figure le même nombre de fois  $\lambda$  dans tout le modèle. Cette condition entraîne que l'on doit avoir :

$$r(k-1) = \lambda(v-1).$$

La solution générale de ce problème reste encore à trouver. On a des solutions particulières qui s'obtiennent à partir des géométries projectives ou euclidiennes construites sur des corps de Galois, également à partir de modules, mais la solution générale n'est pas encore atteinte.

Donnons un exemple :

A B D H	B C F M	C H J K	G H I M
A C E I	B E G K	D E J M	
A F G J	B I J L	D F I K	
A K L M	C D G L	E F H L	

Ici :  $b = 13$  ;  $v = 13$  ;  $r = 4$  ;  $k = 4$  ;  $\lambda = 1$ .



## LA MATHÉMATIQUE DES PROGRAMMES ÉCONOMIQUES

par Georges-Th. GUILBAUD

Directeur d'études à l'École des Hautes Etudes

Ce n'est pas toute la recherche opérationnelle, loin de là. Peut-être même devrait-on dénoncer la faute de perspective qui, pour une partie de l'opinion, tendrait à réduire la nouveauté à quelque Mathématique savante. Mais ce n'est pas notre sujet.

Je me contenterai de dessiner, très sommairement et sans détails techniques, ce qu'on pourrait appeler la théorie mathématique des programmes économiques. Mais, comme on ne peut tout dire, il me faut écarter quelques thèmes importants que je dois, au moins, signaler avant de commencer.

Je suppose d'abord qu'on cherche une décision *rationnelle* : je laisse de côté — quitte à risquer les sarcasmes traditionnels : « Médecin, guériss-toi toi-même ! » — le *premier point* qui est de savoir quand une telle décision est souhaitable et quand elle est impossible. Evoquons simplement, pour fixer les idées, les cas dits d'urgence, où il faut choisir, et choisir vite, sans avoir le temps d'examiner le détail des dossiers, des plans, des devis. Écartons ces procédures accélérées et supposons, au contraire, qu'on ait le loisir de peser, de mesurer, de compter. Ce qui signifie que nous nous occuperons ici d'une *méthode* — que nous laisserons de côté la manière de s'en servir.

*Second point* : pour qu'une décision puisse passer pour rationnelle, elle doit fournir la liste complète des choix offerts et justifier le choix accompli par une comparaison avec tous les autres. Même, c'est le cas le plus simple, s'il n'y avait que deux issues, il faut dire pourquoi on a préféré A à B. Ne nous attardons pas sur le critère de préférence. Non que ce soit un petit détail : c'est même probablement l'essentiel. Mais une façon justement de s'en rendre compte est de voir à quoi sert ce critère et à quoi il conduit.

Pour faire bref, nous nous plaçons dans une situation idéale : il s'agit de choisir, mais : 1° on est capable d'énumérer tous les choix possibles A, B... ; 2° on est capable de donner une règle précise pour savoir si A est préférable à B, ou bien si c'est l'inverse, et ainsi pour toute comparaison par paire.

Le sens commun déclare alors volontiers se désintéresser de l'affaire, tout paraissant réglé d'avance. Mais c'est une illusion, car, même en de telles situations idéales, une difficulté menace, que nous nommerons la *complexité*.

### PROGRAMMES D'AFFECTATION

Prenons un exemple très simple, fort célèbre d'ailleurs et en passe de devenir scolaire. Il s'agit de l'attribution des tâches. Il ne sera peut-être pas inutile de signaler que ce modèle avait déjà servi à l'économiste Ricardo en 1817 pour faire comprendre la nature du « calcul économi-

que ». Deux ouvriers, disait Ricardo, savent faire l'un et l'autre des chapeaux et des souliers, mais ils sont inégaux en habileté : quelles sont les règles qui permettent une affectation de chacun et qui conduisent au plus grand avantage collectif ? Généralisons : un nombre quelconque d'ouvriers, un égal nombre de besognes, quelle est la meilleure distribution des tâches ? Comme nous ne nous intéressons qu'à la forme abstraite du problème, nous pouvons traduire le même énoncé de plusieurs façons. Ce peuvent être aussi bien N machines et N opérations à effectuer, chaque machine pouvant effectuer n'importe laquelle des opérations, mais avec des rendements variés. Ou bien encore N véhicules tous semblables, mais dispersés sur le terrain, et N lieux d'emploi, l'affectation de tel véhicule à tel emploi se traduisant par une dépense proportionnelle à la distance que le véhicule doit parcourir pour rejoindre le lieu d'emploi qui lui est attribué. En bref, on a deux ensembles comportant le même nombre d'éléments, et l'on doit opérer une affectation, établir une correspondance entre les deux ensembles. Les diverses possibilités d'affectation seront, d'autre part, comparées entre elles par leur rendement global, somme des rendements individuels. Le modèle est apparemment d'une grande simplicité — et il fut un temps où l'on aurait conclu qu'il n'y avait rien à en dire, sinon : il n'y a plus qu'à faire les calculs. Mais, aujourd'hui, on devient plus exigeant et l'on demandera : quels calculs ? Et l'on en vient à considérer qu'il n'est peut-être pas indigne d'un mathématicien de se préoccuper de certaines difficultés, naguère encore jugées très matérielles.

Prenons deux ensembles — origines et destination des véhicules, si l'on veut fixer les idées — réduits chacun à quatre éléments : A, B, C et D d'un côté ; P, Q, R et S de l'autre. On demande d'étudier les divers plans d'affectation. L'énumération est facile : il y a vingt-quatre possibilités, puisque pour A on a le choix entre P, Q, R et S — et, pour B, on n'aura plus que trois choix, etc. On présente le tableau des rendements de chaque affectation sous la forme suivante :

	A	B	C	D
P	1	2	2	5
Q	2	5	6	4
R	3	3	8	7
S	3	7	6	9

le nombre écrit dans chaque case indiquant la valeur de l'association correspondante.

Énumérer les vingt-quatre possibilités, calculer la valeur économique de chacune d'elles, ne sera pas très long. On aura donc vingt-quatre devis : choisir le meilleur est bien facile.

Mais, lorsque les ensembles qu'il faut affecter deviennent plus nombreux, le nombre des possibilités augmente : 24 possibilités pour 4 éléments, 120 pour 5, déjà plus de trois millions pour 10, treize cents milliards pour 15..., pour 30 objets, un nombre de trente-trois chiffres !

Les machines, dit-on, calculent vite et c'est vrai. Peuvent-elles nous aider ? Il s'agit d'établir en un temps raisonnable les devis des diverses

combinaisons possibles, de façon à choisir la meilleure. Supposons qu'on dispose d'un procédé électronique très perfectionné qui calcule un devis en un millième de seconde (c'est déjà optimiste). Cela ne fait pas encore 10 puissance 10 dans l'année — c'est-à-dire à peine le nombre des combinaisons pour treize objets à associer à treize autres. Si — ce qui est un rêve — notre machine faisait chaque calcul élémentaire en une micro-seconde, elle ne pourrait pas encore atteindre en un siècle la liste complète des affectations de vingt objets à vingt emplois !

Il faut donc voir la réalité telle qu'elle est : on ne peut espérer énumérer les cas possibles, même pour des problèmes d'affectation de taille fort raisonnable.

Alors intervient une analyse mathématique fort précieuse : dans l'état actuel de nos connaissances pour vingt machines à affecter à vingt tâches, une méthode de calcul permet, avec un papier et un crayon, de trouver l'optimum en quelques heures (et quelques minutes avec les machines électroniques qui, tout à l'heure, semblaient exiger plusieurs siècles).

Que s'est-il passé ? C'est justement ce qui nous intéresse ici.

Nous avons à choisir un programme d'affectation parmi un nombre considérable de programmes possibles. Ce qui nous manquait encore, ce qu'on nommerait probablement — mais la métaphore est encore vague — un « fil directeur », c'était une mise en ordre de cette foule, une idée de structure.

Donner une structure mathématique convenable à l'ensemble des programmes possibles, c'est là le principe fondamental. Mais qu'est-ce qu'une structure ? Comme il n'est pas question d'entrer trop avant dans la technicité du sujet, nous devons nous contenter de quelques vues rapides.

### LES APPROXIMATIONS SUCCESSIVES

Notons en premier lieu le caractère particulier du problème posé, qui évoque davantage la devinette que le problème du concours d'entrée de nos grandes écoles. Nous ne sommes plus, en effet, dans le domaine familier du continu ; s'il s'agissait de rendre maximum une fonction de variables numériques, le réflexe usuel jouerait, et sans doute efficacement : calculer la dérivée. Mais, ici, la variable n'est pas un nombre, c'est un ordre, un arrangement de  $n$  objets. Il vaudrait la peine d'étudier le rôle de certaines devinettes savantes, des « récréations mathématiques » comme on dit, qui ont parfois, au cours des âges, servi de refuge à certaines recherches, apparemment futiles, souvent profondes, mais situées hors du courant principal des Mathématiques « appliquées ».

Examinons en second lieu la signification des approximations successives. Une partie de l'enseignement mathématique traditionnel risquerait, si l'on n'y prenait garde, de donner le change. Il arrive qu'on oppose volontiers, à l'âge scolaire, les problèmes qui auraient une vraie solution, donnée par quelque formule, et les problèmes, inférieurs en dignité, justiciables seulement de l'approximation. Je ne parle pas seulement du goût de l'enfant pour les divisions qui « tombent juste », mais aussi de l'état d'esprit de l'étudiant devant une équation algébrique de degré élevé (il a entendu dire qu'il existe des formules jusqu'au quatrième degré, mais pas au-delà), ou bien devant les équations différentielles qu'il classe volontiers en deux familles, selon que la solution est ou non susceptible d'une expression en termes finis. Ce genre de classification a bien un fondement réel ; il est vrai que la recherche des solutions d'une équation

tion algébrique peut ou non se ramener à des extractions de radicaux, d'une équation différentielle à des quadratures, etc. Et il est utile de savoir, pour un problème donné, quels sont les instruments qui permettent de la résoudre. Mais il n'y a pas lieu de canoniser certains instruments : les anciens géomètres grecs voulaient se limiter à la règle et au compas ; leurs préjugés ont passé dans le langage courant et tel chroniqueur politique nous parle de la « quadrature du cercle », comme d'une difficulté insoluble.

Pour les problèmes qui nous intéressent ici, il faut faire table rase, il faut effacer les préjugés : un seul point est important, celui de savoir si le calcul est possible avec les moyens matériels dont on dispose, et, finalement, combien de temps il durera.

Quand on se place résolument à ce point de vue de réalisme extrême, l'opposition entre le continu et le discontinu s'estompe : toute formule mathématique, fût-ce une intégrale, renvoie finalement à un calcul numérique, et tout calcul, qu'il soit fait de tête, à la main ou à la machine, n'est en fin de compte qu'une manipulation effectuée sur des nombres entiers.

Revenons à la recherche d'un maximum (ou bien d'un minimum : la logique serait la même, il suffirait de changer quelques mots). Imaginons un explorateur, dans un pays inconnu, à la recherche du point culminant de la région. Puisqu'il cherche le point le plus élevé, il doit monter — mais non pas toujours, peut-être. Car, si le relief est assez compliqué, il peut arriver en un sommet — c'est-à-dire d'un point tel que tout déplacement à partir de ce point sera forcément une descente — et cependant apercevoir dans le lointain un *autre* sommet, encore plus élevé. Si notre voyageur est sûr de voir toujours assez loin, sa tactique est toute tracée. Mais si sa vue est limitée, tout devient plus difficile, puisque le fait d'être en un sommet ne prouve pas que l'on soit au sommet culminant. En langage technique : il faut distinguer un maximum local (tous les points *voisins* sont plus bas) du maximum absolu.

Si la vue est limitée, sans qu'on sache rien d'autre, l'exploration doit être complète avant qu'on puisse être sûr d'avoir découvert le point culminant. Mais on peut compenser les inconvénients d'une courte vue par quelque information sur la structure géographique de la région explorée. C'est ainsi que, pour un paysage déterminé, on comprend qu'il existe un horizon suffisant pour déjouer les pièges du maximum local : c'est-à-dire que si notre explorateur est toujours assuré de voir autour de lui dans un rayon  $R$  et s'il cherche constamment à atteindre le point le plus élevé situé dans ce cercle, on peut assurer que son chemin le conduira inéluctablement au point culminant. A condition, bien entendu, que ce rayon  $R$  soit suffisamment étendu. On peut dire la même chose d'une autre façon : il s'agit de définir convenablement les points voisins d'un point donné. Si cette relation de voisinage est prise en un sens trop étroit, il peut se faire qu'un point plus élevé que tous ses voisins ne soit cependant pas le point culminant ; mais, en élargissant suffisamment, les anomalies (les maximums locaux) disparaissent.

On pourrait développer diverses métaphores analogues : par exemple, dans un domaine à trois dimensions où chaque point est caractérisé par une grandeur (la température si l'on veut), la recherche du point le plus chaud (ou le plus froid) peut se faire, sans qu'il soit forcément nécessaire de comparer chaque point à tous les autres, mais seulement aux points d'un « voisinage » convenablement défini. Mais ce ne sont que des métaphores : car, pour le genre de problèmes auquel nous nous intéressons ici, les « points » sont des éléments discrets, les programmes

d'affectation par exemple. On comprend cependant qu'il soit judicieux de se demander comment définir deux programmes « voisins » l'un de l'autre (c'est-à-dire vraisemblablement peu « différents » l'un de l'autre) pour qu'on puisse atteindre le programme le meilleur par approximations successives. On partirait d'un programme quelconque, on le comparerait à tous ses « voisins » — si, parmi ces voisins, il s'en trouve un meilleur, on s'y transporte et l'on recommence la comparaison avec les voisins du nouveau programme (qui ne sont pas tous voisins de l'ancien) —, et ainsi de suite, jusqu'à aboutir à un programme meilleur que tous ses voisins.

Tout revient à savoir dans quelles conditions cette procédure d'amélioration de proche en proche conduira à un optimum. Dans le cas du problème d'affectation cité plus haut, la réponse est simple : on doit considérer comme « voisins » deux programmes qui se déduisent l'un de l'autre par un simple échange. Ainsi le programme d'affectation :

$$\begin{pmatrix} A & B & C & D \\ P & R & S & Q \end{pmatrix}$$

et le programme :

$$\begin{pmatrix} A & B & C & D \\ S & R & P & Q \end{pmatrix}$$

qui ne diffèrent que par les affectations de S et P.

Mais est-ce le seul cas de « voisinage » ? Pour le savoir, on devra d'abord établir que, lorsque deux programmes ne sont pas voisins au sens qu'on aura choisi, on peut trouver une chaîne de programmes dont les deux considérés sont les extrémités et tel que *tout* programme de la chaîne soit « voisin » de celui qui le précède et de celui qui le suit. Ici, on invoquera la proposition bien connue : toute permutation est un produit de cycles disjoints. Pour aller de (P R S Q) à (Q S R P), on passera par (Q R S P). Cette possibilité de « connexion » est tout à fait essentielle en ce genre d'analyse. On étudiera ensuite comment varie la valeur économique du programme quand on se déplace le long d'une pareille chaîne. Nous ne pouvons ici entrer dans le détail. Nous engageons simplement le lecteur à prendre un papier et un crayon et à examiner tous les programmes voisins du programme suivant :

$$\begin{pmatrix} A & B & C & D \\ S & R & P & Q \end{pmatrix}$$

et à calculer leurs valeurs économiques d'après le tableau donné plus haut, qui fournit les valeurs de chaque affectation. Ainsi, on a :

$$\begin{aligned} \text{Valeur} \begin{pmatrix} A & B & C & D \\ S & R & P & Q \end{pmatrix} &= \text{Val. (AS)} + \text{Val. (BR)} + \text{Val. (CP)} + \text{Val. (DQ)} \\ &= 3 + 3 + 2 + 4 = 12. \end{aligned}$$

Il découvrira sans doute pourquoi, si l'on connaît les variations de valeur pour tous les changements de voisinage, on en déduira, par simple addition, les variations de valeur pour les changements plus compliqués. *Et c'est la clef de la théorie.*

Tout n'est pas fini, bien sûr. Le fil directeur est trouvé : le cheminement de proche en proche. Il reste encore à perfectionner les méthodes de calcul, de façon à rendre praticable la détermination effective du programme le meilleur.

### PROGRAMMES DE TRANSPORT

Des problèmes d'affectation aux problèmes de transport, la transition est toute trouvée puisqu'un problème d'affectation est déjà, en quelque

manière, un problème de transport ; nous l'avons vu au moyen de l'allégorie des véhicules : on a des véhicules actuellement placés en A, B, C, D... et on veut obtenir une nouvelle disposition sur le terrain définie par les points P, Q, R, S..., en chacun desquels on exige que se transporte l'un des véhicules de notre parc. Les dits véhicules sont évidemment censés interchangeables, et le seul critère adopté est le coût de l'opération, somme des coûts individuels tels que (AP), (BR), (CQ)...

Supposons maintenant qu'au point A nous ayons, non pas un objet transportable, mais un stock d'un matériel quelconque que l'on peut fractionner *ad libitum* ; de même pour les autres points. Et que le nouveau dispositif soit lui aussi défini par des points P, Q..., et des quantités affectées à chaque point.

Il semble que le choix d'un programme de transport économique devrait être une question très classique. Les illustrations concrètes ne manqueraient pas, dans les travaux civils ou militaires. Mais on est bien forcé de constater que le sujet n'a guère été vulgarisé avant la dernière guerre mondiale. Ce sont les études de F. L. Hitchcock (U.S.A., 1941) et L. Kantorovich (U.R.S.S., 1942) qui ont ramené l'attention sur ce thème et ont suscité des améliorations des techniques de calcul très considérables et probablement non encore terminées.

Pourtant, déjà en 1776, alors qu'il venait d'être nommé professeur à l'École du génie de Mézières, l'illustre Monge avait commencé l'étude mathématique du problème des « remblais et déblais », c'est-à-dire du transport des terres (et, bien entendu, de matériaux quelconques). Mais il ne semble pas que les ingénieurs, civils ou militaires, aient beaucoup cherché à exploiter pratiquement les découvertes de Monge. Quelques travaux mathématiques théoriques en 1818 (Ch. Dupin) et 1884 (P. Appell) ne font que perfectionner et préciser les théorèmes de Monge, sans aucun effort ni vers le calcul numérique effectif, ni vers l'élargissement du modèle primitif.

Si le choix d'un programme de transport peut être présenté comme une généralisation naturelle de celui d'un programme d'affectation, on doit s'attendre à retrouver les mêmes idées directrices et à y découvrir quelques éléments nouveaux. L'élément nouveau le plus important peut être désigné par les concepts associés de « programme extrême » et de « programme admissible » que nous allons examiner brièvement.

Un programme de transport sera complètement défini lorsqu'on précisera quelles quantités doivent être transportées de tel endroit à tel autre. Dans le cas (auquel on se limitera ici) où les points de départ et les points d'arrivée sont en nombre fini, un tableau tel que le suivant dit ce qu'il faut dire :

		origines :			
		A	B	C	D
Destinations :	P	15	10	20	0
	Q	9	7	12	40
	R	10	12	8	5

les nombres de chaque case figurant les quantités à déplacer. En possession d'un programme, on va se demander s'il est possible de le modifier, et si certaines modifications sont, ou non, économiquement avantageuses.

Dans un tableau tel que le précédent, on voit bien une foule de modifications possibles qui n'altèrent pas les quantités totales, soit reçues, soit expédiées (c'est-à-dire les sommes en ligne ou en colonne). Ainsi, on peut amener non plus 15, mais 20 de A à P, à condition de diminuer de 5 unités une autre expédition issue de A, par exemple celle destinée à Q, qui deviendra 4, et, en même temps, de diminuer aussi de 5 unités un autre transport destiné à P, par exemple celui qui vient de B qui sera alors 5. Et, enfin, de compenser ces deux dernières diminutions par une augmentation du transport BQ.

On peut, d'autre part, remarquer qu'une telle modification que nous représenterons par le schéma suivant :

	A	B
P	+5	-5
Q	-5	+5

peut être variée en quantité (3 au lieu de 5) et en signe ( $-5$  au lieu de  $+5$  et inversement). C'est, en fait, toute une *famille* de modifications que désignera le schéma :

$$\begin{array}{cc} +x & -x \\ -x & +x \end{array}$$

et cette famille est bilatérale : le paramètre  $x$  peut prendre des valeurs positives aussi bien que négatives.

La première remarque est alors banale, mais grosse de conséquences : s'il existe, comme ci-dessus, des modifications bilatérales, le programme n'est certainement pas le meilleur du point de vue économique. En effet, si une valeur positive de  $x$  augmente le prix de revient, une valeur négative le diminuera. Ceci suppose évidemment que les frais sont toujours proportionnels aux quantités, ce que nous supposerons désormais.

La première chose à faire est donc d'éliminer d'abord tous les programmes qui, à l'instar de l'exemple ci-dessus, acceptent des modifications bilatérales. Après ces éliminations, il ne restera que les programmes dits « admissibles » (ce mot peut être entendu comme on le fait dans la pratique des examens en plusieurs degrés ; on pourrait dire aussi, comme chez les sportifs, les « finalistes »).

Il est facile de construire un programme admissible. En conservant les mêmes disponibilités en A, B, C, D et les mêmes besoins en P, Q, R, on peut écrire, par exemple :

34	11	0	0
0	18	40	10
0	0	0	35

On devra évidemment étudier les caractères d'un programme admissible au sens qui vient d'être défini ; ce sont ces caractères que l'on désigne par l'épithète « *extrême* ». Disons, sans insister, que le tableau doit comporter le plus grand nombre possible de zéros. D'une façon plus précise, il faut même, et il suffit, que la position de ces zéros permette de déterminer le programme tout entier.

Il faut ensuite organiser les épreuves éliminatoires. Et c'est ici qu'on va retrouver les techniques de cheminement, d'amélioration de proche en proche, qui ont si bien réussi dans les problèmes d'affectation.

On va, en effet, pouvoir définir une notion correcte de « voisinage » (officiellement, on préfère dire programmes « adjacents » plutôt que programmes « voisins »). Comme il est naturel, deux programmes extrêmes seront déclarés voisins si leurs tableaux diffèrent le moins possible, c'est-à-dire s'ils ont le maximum de zéros en commun. On montre alors — en introduisant les précisions nécessaires à la rigueur — que si un programme est extrême et si aucun adjacent n'est meilleur, alors ce programme ne peut absolument pas être amélioré, il est un optimum. D'où la règle de recherche de l'optimum, passer d'un programme extrême à un adjacent meilleur, et continuer ainsi tant qu'on le peut.

Toute la technique nécessaire sera donc rassemblée dès qu'on saura :

- 1° Définir les programmes extrêmes,
- 2° Reconnaître deux extrêmes adjacents,
- 3° Chercher un adjacent meilleur qu'un extrême donné.

### PROGRAMMES DE FABRICATION

Or, il est très remarquable que l'efficacité de cette organisation logique n'est pas limitée aux problèmes de transport, mais qu'au contraire son champ d'application est très vaste. Citons quelques exemples de problèmes où l'on cherche « le meilleur programme » et où une méthode de cheminement analogue à celle qui vient d'être décrite conduit sûrement à la solution.

On dispose de certaines matières premières A, B, C, D, etc., en quantités limitées ; mais on ne peut les vendre que sous forme de mélanges ou de combinaisons. Divers types ou mélanges M, N, P, Q, R, etc., sont commercialisables à des prix de vente qui varient selon les spécifications. Ces spécifications, de nature commerciale ou technologique, peuvent prendre des formes diverses. Le cas le plus simple est celui de proportions définies ; ainsi, le mélange M doit contenir 50 % de A, 25 % de B, 15 % de C, 10 % de D. Mais on peut rencontrer des formules moins simples : le mélange N doit contenir 40 % de A, *plus* de 40 % de B et *moins* de 20 % de C. Ou bien encore : le mélange est dit P s'il contient entre 60 et 70 % de A, entre 5 et 10 % de B, le reste en C. Toutes les conditions de ce style se traduisent par des égalités ou des inégalités de premier degré quand on prend pour inconnues les quantités de chacun des constituants. Par exemple, dans le dernier cas :

$$\begin{aligned} 0,60 \leq A : (A + B + C) \leq 0,70 \text{ et} \\ 0,05 \leq B : (A + B + C) \leq 0,10 \end{aligned}$$

qu'on peut encore écrire :

$$\begin{aligned} 40 A - 60 B - 60 C &\geq 0 \\ 30 A - 70 B - 70 C &\leq 0 \\ 5 A - 95 B + 5 C &\leq 0 \\ 10 A - 90 B + 10 C &\geq 0 \end{aligned}$$

Il en est encore de même pour des spécifications apparemment plus compliquées, qu'on rencontre fréquemment dans les productions industrielles chimiques. Il s'agit de propriétés physiques ou chimiques (inflammabilité, densité, volatilité, etc.), caractérisées par certains indices numériques et telles que l'indice d'un mélange soit une moyenne pondérée des indices des constituants. Si, par exemple, les indices respectifs de A, B,

C, D sont 94, 83, 74 et 95 et qu'on exige que l'indice de M soit au moins égal à 80, on écrira :

$$\frac{94 A + 83 B + 74 C + 95 D}{A + B + C + D} \geq 80$$

c'est-à-dire :

$$14 A + 3 B - 6 C + 15 D \geq 0.$$

Supposons que toutes les spécifications imposées aux mélanges M, N, P, Q..., soient de l'un ou l'autre des types présentés ci-dessus, c'est-à-dire que toutes les conditions imposées aux quantités à mélanger soient représentables par des équations ou bien des inéquations du premier degré. Un programme de fabrication pourra être présenté sous forme d'un tableau à double entrée :

	A	B	C	D
M				
N				
P				
Q				

Dans la première case de ce tableau, on fera figurer la quantité de A qui doit entrer dans la composition de M, et ainsi de suite pour chacune des cases.

Les conditions imposées à un pareil tableau de chiffres pour qu'il représente un programme techniquement *acceptable* sont :

1° Que les nombres écrits dans chaque case ne soient pas négatifs.

2° Que la quantité totale employée, pour chacune des matières premières A, B, etc., ne soit pas supérieure à la quantité disponible. C'est-à-dire que la somme des nombres d'une même colonne soit bornée supérieurement :

$$x(AM) + x(AN) + \text{etc.} \leq a.$$

3° Chaque ligne du tableau indiquant la composition d'un mélange doit être soumise à un certain nombre de spécifications telles que celles énumérées plus haut.

Enfin, il faut, pour orienter le choix, indiquer comment on peut décider si tel programme est ou non meilleur que tel autre. Prenons un critère simple, par exemple le prix de vente total de toute la production (les prix de vente unitaires pour chaque mélange étant connus).

Dans le cadre ainsi tracé, on peut bâtir une théorie générale très semblable à celle des transports. On montre d'abord qu'il n'est pas nécessaire de considérer tous les programmes acceptables, mais seulement les programmes admissibles. Et l'on caractérisera les programmes admissibles comme étant des programmes *extrêmes*, c'est-à-dire qui ne permettent aucune variation bilatérale. Ce ne sont plus seulement les zéros du programme qui, comme dans le cas du transport, permettent de reconnaître le caractère extrême. Il faut considérer aussi celle des conditions qui,

exprimées par des inégalités, se trouvent bloquer le programme d'un côté ou de l'autre. Ainsi, pour prendre un exemple très simple, si l'on a écrit parmi les conditions :

$$x + y + z \leq 100$$

et si l'on a :

$$x = 40 \quad y = 60 \quad z = 0,$$

on comprend bien que le programme se trouve bloqué en ce sens que la somme  $x + y + z$  ne peut que diminuer, mais ne peut pas augmenter.

On montre alors qu'un programme est extrême si le nombre de tels blocages est assez grand. On comprend, d'autre part, par un raisonnement en tous points semblable à celui que nous avons fait pour les programmes de transports, qu'un programme qui ne serait pas extrême pourrait certainement être amélioré.

Reste enfin à définir, comme toujours, la notion de programmes voisins (ou « adjacents »). Dans le cas présent, moyennant quelques précisions, on peut dire encore que deux programmes extrêmes sont voisins si les blocages diffèrent aussi peu que possible de l'un à l'autre.

Les opérations de « mélange » auxquelles il vient d'être fait allusion ne sont pas forcément des mélanges physico-chimiques. Le mot de mélange et celui de combinaison peuvent être pris dans un sens métaphorique. Par exemple, une fabrication industrielle peut être considérée comme le résultat de la combinaison, en proportions convenables, d'un certain nombre de facteurs de production (matières premières, quantités de travail d'espèces bien définies, usage d'installations variées, etc.). Dans bien des cas, on peut formuler le problème du programme optimum sous une forme analogue à la précédente : tableau rectangulaire, conditions imposées en lignes ou en colonnes, critère économique convenable. On pourrait multiplier les exemples. Citons seulement un cas intéressant : celui qui concerne un programme de travail saisonnier. On peut penser à l'agriculture et dresser le tableau : chaque colonne sera une période de l'année, par exemple douze colonnes pour les mois, et chaque ligne, un emploi possible (telle ou telle culture). Le nombre d'heures disponibles en un mois est fixé : la somme des nombres en colonne est bornée supérieurement. En ligne, d'autre part, diverses conditions peuvent s'exprimer de façon suffisamment précise par des égalités ou inégalités de proportions, c'est-à-dire finalement par des équations ou inéquations linéaires : on retombe sur le modèle précédent.

Du travail agricole, on peut passer à la production et à la consommation d'énergie, sans trop changer le cadre formel. Supposons qu'on étudie les programmes d'investissement d'un ensemble hydro-électrique. Chaque usine possible, située à tel endroit et construite de telle façon, est capable de produire des puissances variables au long de l'année. D'autre part, l'ensemble doit fournir à chaque moment une puissance totale suffisante. On dressera le tableau : chaque colonne attribuée à une période (ou bien un ensemble de périodes semblables), chaque ligne à un type d'installation. Les sommes en colonne doivent atteindre au moins l'objectif fixé, et les quantités écrites dans une même ligne sont liées entre elles. Il se peut, ici encore, que ces liaisons soient assez bien représentées par des règles de proportions, c'est-à-dire une algèbre du premier degré. Quant au critère économique, ce pourra être le coût de l'opération, comptabilisé d'une façon convenable.

## CARACTERISTIQUES DES PROGRAMMES LINEAIRES

Nous ne pouvons songer à l'exploration systématique de tous les domaines de l'économie où l'on peut trouver des situations analogues à celles qui viennent d'être évoquées. Ce qui en a été dit suffit à faire comprendre l'intérêt d'une théorie générale. C'est la théorie des programmes *linéaires*. Tous les modèles utilisés entrent dans la classe suivante :

1° Le programme à choisir est défini par un ensemble de paramètres  $x, y, z$ , etc.

2° Pour qu'un système de valeurs numériques pour  $(x, y, z, \dots)$  représente un programme *acceptable* (on dit plus souvent : *réalisable*), il faut et suffit que ces nombres satisfassent à un système de conditions linéaires telles que :

$$ax + by + cz + \dots \geq h \quad (\text{inéquation})$$

ou bien :

$$Ax + By + Cz + \dots = H \quad (\text{équation})$$

en nombre d'ailleurs quelconque.

3° Le choix entre deux ou plusieurs programmes se fait au moyen d'un critère économique ou « valeur » susceptible d'une expression linéaire :

$$V = px + qy + rz + \dots$$

Les bases mathématiques de la théorie générale sont alors constituées par une analyse dont on a déjà signalé les éléments principaux.

D'abord, la définition des programmes *extrêmes*, au moyen des blocages (on dit aussi : goulots d'étranglement), c'est-à-dire des cas où une condition telle que  $ax + by + \dots \leq h$  se trouve satisfaite, mais à la limite (le signe  $\leq$  devenant  $=$ ).

On est alors ramené à un schéma combinatoire, analogue à celui que nous avons cité en commençant : un choix à faire parmi un nombre fini, mais en général colossal, de combinaisons admissibles. Et on est tiré d'affaire dès qu'on a aperçu la *connexité* qui règne entre ces programmes admissibles : à tout programme extrême, on associe ses adjacents. Il ne reste plus qu'à organiser un cheminement méthodique.

Bien entendu, tout n'est pas dit, et tout n'est pas si simple. Nous ne nous attarderons pas, cependant, aux détails des proportions mathématiques et des techniques de calcul : je renverrai le lecteur à l'adaptation française que vient d'écrire M. Bouzitat du petit livre de Vajda (*Programmes linéaires et théorie des jeux*, chez Dunod, Paris, 1958).

Mais, sans entrer dans les détails, il nous sera permis de souligner quelques idées importantes dans le développement de la théorie et de ses applications.

### LA THEORIE GENERALE N'ABOLIT PAS LES TECHNIQUES PARTICULIERES

En premier lieu, il ne faut pas oublier que la possibilité d'une théorie générale, englobant tous les cas particuliers, n'abolit pas les techniques particulières et la nécessité d'une classification des problèmes. C'est ainsi que le problème des transports (le problème de Monge) peut bien être considéré comme un cas particulier du problème général : chercher le minimum d'une fonction linéaire de plusieurs variables liées par des équations et des inéquations linéaires. Mais, dans la pratique des calculs, il serait maladroit de fermer les yeux sur les particularités de ces équations.

lions et inéquations. Un juste dosage des artifices particuliers et des propositions générales doit être recherché. Dans l'état actuel de nos connaissances, c'est sur ce point précis que doivent porter les efforts.

Il se peut que la méthode générale soit impraticable, pour un problème déterminé, parce qu'elle entraîne de trop longs calculs, et qu'un examen attentif de la structure propre du problème conduise à des simplifications considérables.

Une classification des problèmes de programmation linéaire est indispensable. Mais il convient de réfléchir un instant aux principes de cette classification. Dans les exemples que nous avons examinés précédemment, nous avons bien aperçu une certaine diversité (affectation, transport, mélanges), mais les nécessités d'une exposition aussi peu savante que possible pourrait donner le change. La *nature* d'un problème de programmation économique ne doit pas être définie par son domaine d'application, par le genre de phénomènes industriels qui le fait naître, mais bien par ses *propriétés mathématiques fondamentales*. Or, il arrive souvent que deux questions apparemment très différentes sont, en fait, et profondément justiciables de la même formulation mathématique et des mêmes procédures de résolution.

### EXPLOITATION DES ANALOGIES

Ce sont des analogies, ou mieux des isomorphies, qu'il convient d'exploiter au maximum. C'est probablement un des traits importants de la Recherche Opérationnelle que la découverte et l'exploitation systématique de ces isomorphies, lesquelles permettent de transférer, à l'un des domaines, une bonne part des connaissances acquises par les praticiens de l'autre.

Donnons un exemple simple de deux problèmes de programmation dont on n'aperçoit pas au premier abord l'isomorphie.

*Le premier problème* concerne l'établissement d'un flux maximum dans un réseau. On peut imaginer un réseau routier ou bien le réseau des rues dans une ville. Chaque rue, chaque carrefour, peut être caractérisé par le débit (nombre de véhicules à l'heure) maximum qui peut y passer (ce qu'on peut appeler la capacité de cette rue ou de ce carrefour).

Supposons alors que l'on nous demande d'organiser la traversée de cette ville par un convoi important : le point d'entrée et le point de sortie étant fixés, nous sommes libres de partager le flux total de toutes les manières que nous voulons, pourvu qu'on respecte les capacités. On comprend aisément que certains programmes doivent être meilleurs que d'autres en ce qu'ils permettent un débit global supérieur. Comment trouver l'optimum ? Sans tenter ici l'analyse détaillée, on peut signaler que les principes généraux sont valables : programmes extrêmes, programmes adjacents, etc. On peut d'ailleurs montrer qu'un choix convenable des paramètres inconnus permet d'écrire toutes les conditions sous forme d'inégalités linéaires. Il s'agit d'un programme linéaire, mais de structure très spéciale (et tout à fait différente du problème de Monge).

*Le second problème* : sur une voie unique (routière ou ferrée), circulent des convois dont les horaires sont bien connus. Le doublement est interdit, sauf en un certain nombre de points où des garages ont été aménagés. On veut utiliser cette voie pour y faire passer des convois supplémentaires de vitesse donnée — plus lente, par exemple, que les convois normaux déjà fixés. Et on demande de construire le programme de façon à utiliser au maximum les possibilités offertes. L'optimum sera

tout naturellement défini par le nombre maximum de convois supplémentaires.

Or, on peut établir que ce second problème est, malgré les apparences, du même type que le précédent. Qui sait résoudre l'un saura résoudre l'autre. Il suffit, pour le voir, de dessiner un diagramme pour le second, analogue aux graphiques en usage dans les chemins de fer, et d'y apercevoir un problème de flux maximum dans un réseau.

### LA DUALITE

On comprend sans peine que l'exploitation systématique des analogies entre problèmes apparemment différents puisse être d'un grand profit en théorie et en pratique. Il faut signaler un cas d'analogie qui joue un rôle très particulier : c'est la correspondance entre deux problèmes que l'on nomme la *dualité*. Ici encore, on voudra bien m'excuser de rester très superficiel : il ne suffirait d'ailleurs pas d'énoncer avec soin les théorèmes, ni même d'en donner les démonstrations rigoureuses — car les phénomènes de dualité présentent de nombreux aspects, fort variés —, et, seul, un véritable « tour de la question » permettrait des conclusions solides. Mais ce serait fort long. Contentons-nous de quelques brèves indications.

On peut, tout d'abord, présenter la dualité la plus traditionnelle, celle des espaces vectoriels et des systèmes d'équations du premier degré.

L'un des modes classiques de résolution des équations du premier degré consiste à éliminer une à une les inconnues. Ce qui se fait commodément en combinant les équations entre elles : on multiplie les deux membres d'une équation par une constante bien choisie et l'on ajoute le résultat aux deux membres d'une autre équation. En poursuivant le processus, on peut tenir registre des multiplicateurs utilisés, et, lorsque l'on a atteint la solution (l'élimination est achevée), on sait alors quels multiplicateurs auraient pu être utilisés pour obtenir, d'un seul coup, le résultat final. Dans des cas simples, cette élimination en un seul coup est facile à imaginer :

$$\begin{array}{rcl} x + y + z & = & 14 \\ x - y & = & 2 \\ y - z & = & 3 \end{array}$$

la combinaison :  $14 - 2 + 3$  donne immédiatement  $y$ , tandis que les combinaisons :

$$\begin{array}{r} 14 + 2(2) + 3 \\ 14 - 2(2) - 3 \end{array}$$

donnent respectivement  $x$  et  $z$ .

Mais un instant de réflexion montre que le problème de choix des multiplicateurs qui procurent l'élimination n'est pas beaucoup moins simple que celui de la résolution — et même, qu'à dire vrai, ce sont là deux faces d'un seul et même problème. Toutefois, comme le problème des multiplicateurs peut être formellement écrit comme système d'équations, il apparaît ainsi deux problèmes de même espèce qui sont intrinsèquement liés.

De pareilles réflexions peuvent être étendues aux systèmes d'inégalités du premier degré. C'est un peu plus délicat ; entre autres raisons, on voit tout de suite que les multiplicateurs d'une inégalité doivent être positifs sous peine de changer le sens de l'inégalité. Quoi qu'il en soit, on peut faire apparaître une théorie analogue qui associe à toute résolution d'un certain problème la résolution d'un autre problème de forme analogue.

Une autre façon d'introduire à la dualité, c'est d'approfondir l'aspect économique des problèmes de programme. Et ce n'est pas un des moindres avantages de la théorie mathématique que cette interpénétration des points de vue.

Nous avons déjà vu que les programmes de « mélanges » peuvent être généralisés de façon à représenter un choix complexe d'activités qu'on veut rendre optimum en distribuant au mieux des ressources limitées. Lorsqu'on a affaire à un problème complètement linéaire, nous savons que le programme optimum sera un programme extrême, c'est-à-dire finalement un programme bien défini par ses goulots d'étranglement. Il est alors tout à fait naturel de « discuter » la solution obtenue, c'est-à-dire de se demander quelle modification du programme et quel avantage pourraient résulter d'un léger élargissement de ces blocages. On est amené à définir ainsi une série d'avantages *marginiaux* : si telle ressource se trouvait augmentée d'une unité, il en résulterait tel bénéfice, etc...

Or, on constate que, pour calculer ces coefficients marginaux, il n'est pas nécessaire d'instaurer une procédure spéciale. Pour peu qu'on y prenne garde, il y a des liaisons simples entre les coefficients multiplicateurs nécessairement introduits par la recherche systématique de l'optimum et les avantages marginaux qu'on désire connaître. En d'autres termes, le tableau des calculs nécessaires fournit, outre la solution optimale demandée, un certain nombre de résultats numériques auxiliaires, et certains de ces « sous-produits » sont susceptibles d'une interprétation économique très utile.

Si l'on veut pousser plus loin, on arrive à construire deux problèmes de décision économique associés de telle sorte que leurs résolutions sont indissolublement liées. La théorie des jeux fournit alors un langage commode pour exprimer cette dualité, ce qui n'étonne guère puisque, en « duel », il y a deux points de vue, ceux de chacun des deux joueurs, qui sont indissolublement associés.

Je sais fort bien que ces quelques généralités disent trop ou trop peu, et qu'il y a loin de ces quelques indications vagues à une intelligence correcte des phénomènes de dualité. Mais si l'on doit souligner l'importance des méthodes « duales » en théorie de la programmation, il ne faut pas dissimuler que tout n'y est pas immédiatement clair. Certains aspects, mathématiquement bien assurés, n'en sont pas moins assez curieux. Voici un exemple intéressant, où la dualité prend une forme essentiellement combinatoire.

Considérons un problème d'affectation, mais un peu plus simple que ceux que nous avons examinés en commençant. Deux ensembles sont donnés : des objets A, B, C, D..., et des emplois possibles : P, Q, R, S... Certaines affectations sont possibles, on en donne la liste :

A peut être en P ou en Q ou en S,  
B peut être en R ou en S,  
C peut être en P seulement.

Toutes les autres sont réputées impossibles. Mais il faut ajouter, bien entendu, les incompatibilités :

- 1° Chaque objet ne peut occuper qu'une seule place.
- 2° Il ne peut y avoir deux objets à la même place.

Ceci dit, on demande de choisir le plus grand nombre possible d'affectations possibles et compatibles. Le langage utilisé fait jouer des rôles assez différents aux « objets » A, B, C..., et aux « emplois » P, Q, R... Mais on peut évidemment échanger les rôles.

Pour souligner cette symétrie, transformons l'énoncé. Considérons une réunion comportant un assez grand nombre de personnes. Chacune a un nom et un prénom (usuel). Il n'est pas impossible que plusieurs aient le même nom de famille, ni que plusieurs aient le même prénom. Représentons les prénoms par des lettres A, B, C... et les noms par P, Q, R... Puis établissons les listes des personnes présentes :

AP, AQ, AS, BR, BS, CP...

Le problème posé devient : choisir dans l'assemblée un groupe comprenant le plus grand nombre possible de personnes et ne contenant pas deux personnes qui aient le même prénom, ni deux qui aient le même nom.

Considérons maintenant avec les mêmes données un problème apparemment tout autre. On peut établir une liste d'appel, formée d'une suite de noms et d'une suite de prénoms, telle qu'à sa lecture chacune des personnes présentes s'entende appeler soit par son nom, soit par son prénom, soit, éventuellement, par les deux. On demande d'établir la liste la plus courte possible.

Or, on peut associer ces deux problèmes de manière fort étroite, les résoudre simultanément, et montrer enfin l'égalité des deux nombres suivant le nombre de personnes du plus grand groupe sans homonymes, nombre de mots dans la plus courte liste d'appel. C'est un phénomène de dualité.

#### LIMITES DES METHODES DE PROGRAMMATION LINEAIRE

Pour avoir quelques idées suffisamment claires sur les méthodes mathématiques de programmation économique, il ne suffirait pas de développer les considérations précédentes et d'exposer le détail des théorèmes et des calculs, appliqués aux modèles que l'on sait traiter complètement et sûrement. Il conviendrait aussi de parcourir les frontières : c'est-à-dire non seulement approfondir notre connaissance des méthodes de résolution, mais prendre contact avec les difficultés non encore résolues. Il n'est pas très facile de définir correctement les deux expressions, apparemment banales, de « problème résolu », « problème non résolu ». Il suffit de se reporter à ce qui a été pris en considération au début de cet exposé même, pour apercevoir à quel point la coupure est conventionnelle.

Le problème des affectations, par exemple, peut, en un certain sens, être considéré comme « résolu » dès qu'il a été énoncé. On peut dire : énumérer toutes les possibilités, calculer la valeur de chacune d'elles, et choisir la meilleure. Tout cela n'exigera qu'une manipulation finie, un « calcul », en droit, toujours possible. Mais, comme nous l'avons vu, ce calcul peut être, dans l'état actuel de nos moyens, impraticable parce qu'il exigerait trop de temps. Pour dire que le problème est résolu, on exige donc davantage : des règles de simplifications qui ramènent le travail à des dimensions humaines. Et on comprend alors aisément quel sens donner à l'expression célèbre : « Il n'y a pas de problèmes résolus, il n'y a que des problèmes plus ou moins résolus. »

Il est possible, sans trop déborder notre cadre actuel, de fournir un exemple d'un problème qui, actuellement, est beaucoup « moins résolu » qu'un autre.

Nous avons déjà examiné l'établissement d'un programme d'affectation simple portant sur  $N$  véhicules à placer en  $N$  points. Finalement, il s'agit de joindre les points donnés deux à deux, de façon que la somme

des distances ainsi distinguées soit la plus petite possible. Or, si l'on a l'habitude de considérer que l'on possède une « solution » pour de pareils problèmes, on a coutume aussi bien d'affirmer le contraire pour un problème assez peu différent. C'est celui de l'organisation d'une « tournée » : on a  $N$  points sur le terrain, et il s'agit de passer par chacun d'eux en revenant au point de départ, et de choisir, parmi toutes les tournées possibles, celle qui est la plus courte.

Contrairement à l'impression première, en effet, les deux problèmes ne sont pas de même nature. On pourrait peut-être éclairer cette irréductibilité par la remarque suivante : pour les programmes d'affectation, on peut passer d'un programme à un autre (un programme « voisin ») en opérant un simple échange : affecter à  $P$  le véhicule qui était destiné à  $Q$  et inversement. Pour les programmes de tournée, si l'on opère des échanges de ce genre, on risque d'obtenir des programmes qui ne remplissent plus les conditions exigées, mais se décomposent en deux tournées sans point commun. En d'autres termes, une notion efficace de programmes « voisins » est plus difficile à définir pour les tournées que pour les affectations.

Cela ne veut pas dire qu'on soit actuellement incapable d'établir un programme de tournée économique. Cela veut dire seulement que ce sera beaucoup plus difficile. Et qu'en particulier, pour vérifier qu'une solution proposée est effectivement la meilleure, il faudra beaucoup plus de précautions et vérifications que dans le cas des affectations ordinaires pour lequel il suffit d'examiner les « voisins ».

Une autre façon de faire saisir la différence consiste à donner une expression algébrique du premier problème (affectation) en montrant qu'on peut le réduire à la résolution d'un système d'inégalités linéaires ; puis de montrer que cette forme algébrique n'est pas suffisante pour exprimer toutes les exigences du problème des tournées.

Il en résulte qu'on sait, à l'heure actuelle, organiser proprement les approximations successives dans le premier cas, alors qu'on reste, dans le second cas, au niveau des tâtonnements très mal ordonnés. C'est, bien sûr, une question de degré : « approximations » signifie seulement « tâtonnements » rationnellement organisés. Mais cette différence a des conséquences pratiques fort importantes. En particulier, on comprend sans peine que l'emploi des machines à calculer exige justement une organisation soignée des tâtonnements ; ce que nous nommons « intuition », qui correspond à une vue d'ensemble, au choix d'un raccourci, à l'abréviation d'une étape, peut difficilement être transféré à la machine.

Dans l'état actuel de la théorie, ainsi qu'on a déjà pu l'apercevoir par les exemples présentés ci-dessus, ce sont les problèmes *linéaires* qui ont été particulièrement travaillés. Mais il est cependant utile de noter que la méthode la plus féconde demeure une méthode *combinatoire*. Toute la logique de la méthode, en effet, peut être résumée au moyen des principes suivants :

1° Ayant à choisir parmi divers plans d'action ou programmes, il faut d'abord standardiser la description de tous les programmes possibles (choix des paramètres) et ensuite indiquer le critère qui permet de comparer deux programmes et de dire quel est le plus avantageux.

2° On procède ensuite à une première élimination qui ne conserve que les programmes « admissibles ».

3° On organise une exploration systématique au sein de l'ensemble des programmes admissibles au moyen d'un cheminement ou amélioration progressive d'un programme admissible initial quelconque.

4° On a établi, au préalable, c'est la justification de la méthode de

cheminement, qu'un programme est optimal si aucune amélioration n'est possible par les voies du cheminement.

Dans le cas des programmes soumis à des conditions linéaires, la sélection des admissibles se ramène à l'étude des polyèdres convexes dans un espace vectoriel (les admissibles ou extrêmes sont les sommets du polyèdre) et le cheminement utilise les connexions qui existent entre ces sommets (arêtes, facettes, etc.). Les propositions mathématiques utilisées sont donc susceptibles d'une interprétation géométrique commode. Le mot de « simplexe » rappelle cette illustration : le simplexe étant l'élément le plus simple de la description des structures polyédriques (segment, triangle, tétraèdre, etc.).

### RECURRENCE, ACTUALISATION, ESPERANCE

Ainsi, la théorie générale ne suffit pas à guider économiquement toutes les applications. Une classification des particularités est indispensable.

Parmi les études de structures particulières, il faut faire une place de choix aux phénomènes de décomposition. On comprendra aisément l'idée directrice sur l'exemple des problèmes d'affectation. Ayant à affecter un ensemble ABCDEF à un ensemble PQRSTU, il peut se faire qu'on aperçoive d'emblée — avant d'avoir la solution complète — la possibilité d'exclure certaines attributions et de dire, par exemple, qu'il faut attribuer ABC à QST et DEF à PRU. On est alors ramené à plusieurs problèmes analogues au problème original, mais plus petits. Ce sont de telles décompositions (les lettres AB..., PQ... désignant des points dans un plan) que cherchait Monge pour résoudre méthodiquement le problème des déblais.

Mais le terrain de choix pour des analyses de ce genre est certainement celui des structures temporelles. Il arrive, en effet, assez souvent qu'un programme économique (fabrication, emploi des ressources, investissements, etc.) se décompose rationnellement en une séquence de programmes partiels, mais enchaînés. Le traitement de ces chaînes est à lui seul un sujet au moins aussi étendu (et probablement plus important) que celui des programmes linéaires.

Nous ne pouvons songer à commencer ici un exposé qui exigerait — même à vol d'oiseau — autant de temps et de mots que tout ce qui précède. Bornons-nous à introduire la notion de *récurrence* qui domine toute la théorie des programmes séquentiels.

Supposons qu'on dispose d'un certain capital et qu'on puisse l'utiliser à différents emplois. Plaçons-nous dans le cas où le partage est possible : la décision élémentaire consiste à faire dans le total donné un certain nombre de parts qui seront affectées à ces divers emplois. Pour choisir entre les partages possibles, on introduira un critère économique. Notre capital (peu importe ici sa nature, pourvu qu'on sache mesurer, partager, etc.) n'a pas le même rendement dans ses divers usages. On pourra étudier les variations du rendement global quand varient les proportions du partage et chercher le maximum. Voilà le mécanisme élémentaire.

Introduisons maintenant un enchaînement temporel. C'est-à-dire supposons que les emplois dont il a été question soient décidés pour un temps limité, et qu'à la fin de la période envisagée un nouveau plan de partage puisse être mis en œuvre. Ceci suppose évidemment que le capital soit libéré à la fin de la période, et il est assez naturel de penser qu'il ne sera pas intact. Il est même vraisemblable que l'usure du capital est d'au-

tant plus forte que le rendement est élevé. Un arbitrage est donc nécessaire.

Si l'on avait quelque moyen (les prix du marché, par exemple) d'estimer la valeur du capital en fin de période, on pourrait traiter séparément chaque période, sous la forme suivante : trouver le mode de partage et d'emploi qui assure le maximum du bilan final, constitué par la somme du profit retiré et de la valeur du capital final. Mais, dans bien des cas, on ne sait pas calculer cette valeur.

C'est alors qu'on peut essayer un calcul récurrent (qui remonte le cours du temps) en raisonnant d'abord sur la dernière d'une suite de  $N$  périodes. On se demande ce qu'il faudrait faire à la fin de la  $(N - 1)^{\text{ème}}$  période pour tirer le meilleur parti possible du dernier exercice. On pourra — c'est le problème élémentaire — établir la « stratégie optimale », c'est-à-dire une règle d'action indiquant ce qu'il convient de faire *en fonction* du capital laissé par les  $(N - 1)$  exercices précédents. Cette règle ayant été déterminée, on pourra de même se demander pour la  $(N - 1)^{\text{ème}}$  période ce qui est le plus avantageux, compte tenu de l'hypothèse que la dernière période sera gérée à l'optimum, selon le calcul qui vient d'être fait. Et ainsi de suite. Pour définir un *optimum* de gestion concernant une période quelconque, il suffit d'avoir, au préalable, calculé un optimum pour chacune des périodes ultérieures.

Bien entendu, pour mettre sur pied de pareils calculs, il faudra, puisqu'il s'agit de l'évolution dans le temps des valeurs économiques, introduire, à bon escient, une technique d'actualisation (taux d'intérêt) et une représentation comptable des incertitudes de l'avenir (probabilisation).

Ces trois principes : récurrence, actualisation, probabilité, permettent la construction d'une « Espérance économique », c'est-à-dire d'une expression correcte, apte à guider la décision, des valeurs économiques soumises à la triple qualification : décision optimale, évolution temporelle, aléas de l'avenir.

Le développement de ces modèles de décisions séquentielles ou en chaîne est, à l'heure présente, l'un des chapitres les plus attachants et les plus prometteurs de la théorie de programmes économiques. Il fallait le signaler : je regrette d'avoir été contraint de me limiter à un *addendum* final beaucoup trop sec. Il faudra y revenir et examiner les méthodes et leurs limites.

\*  
\*\*

Puisse cet aperçu donner goût et curiosité à ceux qu'intéresse la promotion rapide des méthodes scientifiques du calcul économique.

# QUELQUES ASPECTS THEORIQUES ET PRATIQUES DES JEUX DE STRATEGIE

par J. BOUZITAT, *Ingénieur en Chef à la Société Alsacienne  
de Constructions Mécaniques, Ancien Elève de l'É.N.S.*

## 1. — Objet de la théorie des jeux.

Le terme de « *théorie des jeux* » désigne un vaste domaine de réflexion. A certains égards, ce nom peut paraître mal choisi, car il tend à accréditer l'idée d'une sorte d'amusement, ou, en tout cas, l'idée d'une réflexion qui porterait seulement sur « les jeux », c'est-à-dire, selon l'acception commune du mot, les jeux pratiqués dans les casinos, ou les jeux de société.

Mais il faut comprendre la « *théorie des jeux* » dans son contexte historique. C'est autour des jeux de hasard, effectivement, que Pascal en a, le premier, développé les rudiments dans le *Traité du triangle arithmétique* (vers 1654), tout en mettant bien en lumière, dès l'origine, le sérieux d'une telle réflexion. Et, dès 1657, *Huygens* pouvait écrire, à propos des jeux, cette phrase que reprend, en exergue, un livre récent : « ...Mais si l'on se met à examiner plus attentivement nos leçons, on ne tardera pas à découvrir, j'en suis sûr, qu'elles ne traitent point, comme on pourrait croire, du jeu, mais qu'elles établissent les bases d'une réflexion intéressante et très profonde. » (1).

Quel est donc, au juste, le champ d'application de la théorie des jeux ? Pour ne pas le limiter indûment, il faut dire que c'est la *décision et l'action humaines, dans les situations de conflit*. C'est seulement parce que les jeux de société fournissent des modèles accessibles et dépouillés de ces situations qu'ils ont, à l'origine, servi de base à la réflexion théorique. Mais cela ne doit pas masquer la diversité des applications possibles : domaine économique, décisions politiques, art militaire..., sans oublier la ligne de conduite plus humble que nous avons à adopter devant les situations où nous place la vie de chaque jour. L'essentiel est qu'il s'agit ici, non seulement de connaître, mais surtout de décider et d'agir.

---

(1) Christian HUYGENS, *De ratiociniis in ludo aleæ*, 1657, cité par J. D. WILLIAMS, dans son ouvrage *La stratégie dans les actions humaines, les affaires, la guerre, les jeux* (Dunod, 1956, traduction de *The Compleat Strategyst*, New-York, Mc Graw Hill, 1954) Cet ouvrage est à recommander à tous ceux qui veulent s'initier à la théorie des jeux. Les nombreux exemples, judicieusement choisis, dont il est illustré, en rendent la lecture attrayante et instructive. Et l'humour de l'auteur dissimulé à peine le sérieux de ses préoccupations.

La notion de probabilité joue un rôle important en théorie des jeux, et l'on pense tout de suite aux hasards, aux « aléas » du jeu. Mais il faut ici éviter une confusion : la théorie des jeux n'est pas une application du calcul des probabilités ; on peut dire au contraire que *le calcul des probabilités est un instrument au service de la théorie des jeux*, entendu au sens large que nous avons dit. Au second chapitre de son « *Ars conjectandi* », publié en 1713, *Jacques Bernoulli* définit parfaitement le rôle de cet instrument dans la décision et l'action humaines : « Pour ce qui est sûr et hors de doute, nous parlons de connaissance et de compréhension ; pour tout le reste, nous disons seulement *conjecture* ou *opinion*. *Conjecturer* quelque chose, c'est mesurer son degré de probabilité : ainsi le Savoir-Conjecturer ou Stochastique se définit pour nous comme savoir mesurer, le plus exactement possible, les degrés de probabilité, *afin que*, dans nos décisions et nos actions, nous puissions toujours choisir ou accepter ce qui nous aura paru meilleur, plus satisfaisant, plus sûr, plus prudent : seul objet à quoi s'applique toute la sagesse du philosophe, toute la prévoyance du Politique. » (2).

Tel est bien l'objet de la théorie des jeux.

## 2. — Jeux de hasard pur. Règle des partis ; espérance mathématique.

Après cette introduction, reprenons en partie la démarche historique par laquelle s'est constituée la théorie des jeux. Sans chercher à en détailler les circonstances, nous nous attacherons seulement à ce qui peut éclairer le contenu de la théorie.

Les règles de décision, qui sont l'objet propre de la théorie des jeux, ont d'abord été recherchées pour *régler la conduite des joueurs dans les jeux de hasard*. C'est le but que se propose *Pascal* en donnant sa « *règle des partis* », applicable au partage des enjeux.

D'une façon schématique, voici ce dont il s'agit. Deux joueurs s'affrontent dans un jeu de hasard pur, après avoir mis des enjeux sur la table. La règle du jeu fixe les conditions dans lesquelles toute « *partie* » doit se dérouler pour être « régulière », c'est-à-dire, non seulement la succession des coups et la manière dont ils doivent être joués, mais encore les conditions dans lesquelles une partie — jouée selon la règle — peut se terminer et la manière dont les enjeux doivent alors être partagés. (Il est facile d'imaginer par exemple qu'il s'agisse du jeu de pile ou face, ou encore d'un jeu de dés, mais il importe de fixer de façon précise les conditions finales dont il vient d'être question).

La règle du jeu fixe le règlement qui doit intervenir entre les joueurs pour les seules parties conduites à leur terme. *Pascal* s'est demandé quel devrait être ce règlement si la partie se trouvait interrompue pour une raison de force majeure, avant le terme prévu par la règle. On pourrait convenir qu'alors chaque joueur reprendra sa mise,

---

(2) *Jacques Bernoulli* (1654-1705) fut l'aîné d'une dynastie de mathématiciens. Son *Ars Coniectandi* fut publié en 1713 par les soins de son neveu *Nicolas Bernoulli* (1687-1759). Nous le citons d'après G.-Th. *Guilbaud*, « Leçons sur les éléments principaux de la théorie mathématique des jeux », *Stratégies et Décisions économiques*, Paris, C.N.R.S. 1954, à qui nous empruntons beaucoup de la substance de cet exposé.

mais ce serait faire table rase du jeu qui, sans avoir été mené à son terme, a cependant eu lieu, favorisant plus ou moins chacun des deux joueurs. Il faut donc chercher une solution plus « juste ».

Pascal observe alors *l'enchaînement des coups successifs*. Au cours d'une partie, une situation donnée peut être immédiatement suivie d'un certain nombre d'autres dont les apparitions respectives ont certaines probabilités, connues d'après la règle. Si toutes ces situations étaient « finales », c'est-à-dire marquaient la fin régulière de la partie, la règle fixerait les règlements correspondants, mais n'attacherait aucun règlement à la situation antérieure. S'il fallait cependant y arrêter la partie, il serait juste de tenir compte des règlements qui interviendraient après le coup suivant, si celui-ci était joué, et d'en tenir compte en fonction de leurs probabilités respectives. Ainsi Pascal est-il amené à attacher à chaque situation un règlement égal à une moyenne pondérée des règlements attachés aux situations immédiatement ultérieures, en associant à chacun de ces règlements un coefficient de pondération égal à la probabilité qu'a la situation correspondante de succéder immédiatement à la situation considérée. Il peut ainsi remonter de proche en proche, depuis les situations finales, prévues par la règle, jusqu'à une situation quelconque, supposée réalisée au cours d'une partie jouée selon la règle. Et cela lui permet d'attacher un règlement, un partage des enjeux, un « parti », à chacune de ces situations intermédiaires.

Le règlement ainsi défini par la « règle des partis » de Pascal est exactement « *l'espérance mathématique* », au sens actuel de ce mot, que possèdent les joueurs lorsque la partie se trouve dans la situation considérée, c'est-à-dire une moyenne en probabilité des résultats attachés par la règle à chacune des situations finales, chacun de ces résultats étant affecté de la probabilité de réalisation de la situation finale correspondante, à partir de la situation considérée.

Il est intéressant d'appliquer la méthode de Pascal pour remonter jusqu'au début de la partie, et attacher ainsi un règlement à la situation initiale elle-même. Ce règlement est celui qui devrait intervenir si les joueurs, s'étant mis au jeu, ayant posé les enjeux sur la table, étaient amenés à se séparer avant même d'avoir joué le premier coup ; ou, en d'autres termes, c'est l'espérance mathématique que possèdent les joueurs au début de la partie, par le seul fait qu'ils se sont mis au jeu. À ce moment-là, naturellement, les joueurs trouveraient juste de récupérer exactement leurs enjeux respectifs, puisqu'ils se séparent en fait sans avoir joué. De façon précise, on dira qu'un jeu est *équitable* si l'espérance mathématique de chaque joueur au début de la partie est égale à l'enjeu qu'il a mis sur la table pour prendre part au jeu. Dans le cas général, chaque joueur pourra mesurer l'écart qui existe entre l'enjeu qui lui est demandé et l'espérance mathématique qui lui est offerte en échange. Ainsi pourra-t-il décider en connaissance de cause, d'après la règle des partis, s'il doit accepter ou refuser de jouer. On voit qu'il s'agit bien ici de régler la conduite des joueurs.

À vrai dire, la décision à prendre (jouer ou ne pas jouer) ne dépend pas de la seule information donnée par la règle des partis. Sans insister sur le fait que l'espérance mathématique d'une variable aléatoire ne la

défini pas entièrement, il faut souligner la difficulté que présente, en pratique, l'arbitrage entre un versement certain et l'espérance mathématique offerte en contrepartie. Il est permis, et même souvent recommandé, de refuser un jeu équitable, ou même un jeu mathématiquement avantageux, s'il présente de trop grands *risques*, par exemple si l'une des situations finales, peu probable mais possible, entraîne une perte énorme, équivalente à la ruine. Inversement, il est permis, bien que peu recommandé, d'accepter un jeu mathématiquement désavantageux, où l'on risque tout au plus de perdre un enjeu relativement minime, mais où l'on peut espérer, avec une très faible probabilité, gagner une somme considérable. Le rôle des facteurs psychologiques est primordial dans de tels cas : goût du risque, ou prudence, attire plus ou moins fort du gain éventuel, sensibilité plus ou moins grande à telle ou telle publicité, autant d'éléments qui interviennent dans la décision du joueur, et qui lui font même souvent négliger tout calcul d'espérance mathématique.

Cependant, la règle des partis garde toujours sa valeur. Et elle ouvre la voie à toute une série de réflexions sur la conduite humaine.

### 3. — Jeux de réflexion pure. Existence d'un équilibre ; espérance.

Après avoir considéré les jeux de hasard pur, où le seul choix laissé au joueur est d'accepter ou de refuser le jeu, portons maintenant notre attention sur des jeux tout différents, où le hasard ne joue plus aucun rôle. Ce sont les *jeux de réflexion pure*, dont les exemples classiques sont le jeu d'échecs et le jeu de dames.

Ici encore, schématisons. Supposons que deux joueurs A et B s'affrontent dans un jeu de réflexion pure, où chacun d'eux est parfaitement informé, à chaque instant, de tout ce qui s'est passé antérieurement. La règle du jeu fixe encore les conditions dans lesquelles une « *partie* » doit se dérouler, la situation prise comme point de départ, la manière dont l'initiative — le « *trait* » — passera de l'un à l'autre joueur, au cours de la partie, l'éventail des choix dont dispose le joueur qui a le trait pour passer d'une situation à une autre, enfin les conditions dans lesquelles une partie (jouée selon la règle) peut se terminer et quel en est alors le résultat. Nous supposons ici, pour simplifier, qu'il y a seulement deux résultats possibles : le joueur A gagne, ou le joueur B gagne. Une plus grande variété des résultats possibles ne changerait rien d'essentiel à l'analyse qui va suivre, mais en compliquerait l'exposé. Nous laissons donc au lecteur le soin de généraliser, en considérant d'abord par exemple le cas où le jeu comporte trois résultats possibles : A gagne, partie nulle, B gagne (tel est le cas du jeu d'échecs).

Les situations  *finales*  sont ainsi réparties en deux classes : les situations fortes pour A (ou faibles pour B), les situations fortes pour B (ou faibles pour A). Mais que faut-il penser des situations  *intermédiaires* , ou même de la situation  *initiale*  ? Cette question ressemble beaucoup à celle que s'est posée Pascal à propos des jeux de hasard. Mais il a fallu attendre 1896 pour que la réponse fût aperçue par Lucas (3), et quelques

---

(3) E. LUCAS, *Récréations mathématiques*, Paris, 1896, tome II, chap. 1.

années de plus pour qu'elle fût donnée sous forme rigoureuse par Zermelo et Kalmar (4). C'est que cette réponse paraît, au premier abord, assez paradoxale. Pourtant, il nous suffira de reprendre l'analyse de Pascal, en l'adaptant au jeu actuel, pour remonter de proche en proche, depuis les situations finales, prévues par la règle, jusqu'à une situation quelconque, intermédiaire ou même initiale.

Observons en effet *l'enchaînement des coups successifs*. Au cours d'une partie, une situation donnée peut être immédiatement suivie d'un certain nombre d'autres, dont la réalisation dépend seulement du choix fait par le joueur qui, d'après la règle, a le trait. Si toutes ces situations étaient finales, la règle les répartirait entre les deux classes indiquées plus haut ; dans le cas le plus général, certaines seraient fortes pour A, et les autres seraient fortes pour B ; mais il peut arriver aussi qu'elles appartiennent toutes à la même classe. Le joueur qui a le trait (et qui, par conséquent, a toute liberté pour passer de la situation donnée à l'une quelconque des situations immédiatement ultérieures) a, bien entendu, intérêt à porter son choix sur l'une des situations qui sont fortes pour lui, s'il en existe au moins une parmi celles entre lesquelles la règle lui permet de choisir : auquel cas il est juste de considérer que la victoire est à sa portée, et ce joueur doit être réputé vainqueur avant même qu'il ait joué ce dernier coup dont il est seul maître. Si, au contraire, toutes les situations entre lesquelles peut choisir le joueur qui a le trait sont fortes pour son adversaire, aucun choix ne peut le sauver, et il est juste de considérer que la victoire est acquise à son adversaire avant même que le dernier coup ait été joué.

En d'autres termes, une situation est forte pour le joueur qui a le trait si *l'une au moins* des situations immédiatement ultérieures (entre lesquelles la règle lui permet de choisir) est forte pour lui ; et elle est forte pour son adversaire si *toutes* les situations immédiatement ultérieures sont fortes pour cet adversaire.

Ainsi peut-on remonter de proche en proche et classer une situation quelconque, intermédiaire ou initiale, soit comme forte pour A, soit comme forte pour B. Cela revient à dire qu'à chaque instant de la partie, et en particulier dès que les joueurs A et B se sont mis au jeu, l'un d'eux gagnerait à coup sûr, contre toute défense de son adversaire, s'il avait du jeu une connaissance assez complète pour discerner, chaque fois qu'il a le trait, le choix le plus avantageux pour lui. Autrement dit, *il serait inutile de jouer puisque le résultat de la partie serait connu d'avance*, la victoire ne pouvant être enlevée au joueur pour lequel la situation initiale est forte.

Le jeu en question perdrait donc tout intérêt, et c'est ce qui arrive effectivement pour certains jeux de réflexion pure, joués avec des cailloux ou des allumettes : ils amusent ou étonnent, pendant quelques parties, les joueurs non prévenus auxquels un joueur expert les propose ;

---

(4) E. ZERMELO, « Ueber eine Anwendung der Mengenlehre auf die Theorie des Schachspiels », *Proceedings of the Fifth International Mathematical Congress*, Cambridge, 1912, volume 2, p. 501.

L. KALMAR, « Zur Theorie der abstrakten Spiele », *Acta Szeged*, tome IV, p. 65-85, 1928.

mais ce ne sont pas de vrais « jeux », puisque leur résultat devient certain pour ceux qui les pratiquent un peu. Aussi passent-ils au rang de « récréations mathématiques », mais il en est de fort ingénieuses.

*Qu'en est-il donc pour le jeu d'échecs ?* Une analyse tout à fait semblable à la précédente montre ici qu'une situation quelconque, intermédiaire ou initiale, peut être classée, à partir des situations finales, soit comme forte pour les Blancs, soit comme forte pour les Noirs, soit enfin comme neutre (si l'une au moins des situations immédiatement ultérieures est neutre, sans qu'aucune soit forte pour le joueur qui a le trait). Dans ce dernier cas, si chacun des deux joueurs agissait au mieux de ses intérêts chaque fois qu'il a le trait, la situation finale serait encore neutre, c'est-à-dire qu'elle correspondrait à la nullité de la partie. Il est donc certain que deux joueurs ayant une connaissance *complète* du jeu d'échecs, c'est-à-dire capables de classer ainsi une situation quelconque (définie par la position des pièces sur l'échiquier et l'indication du joueur qui a le trait), connaîtraient d'avance le résultat de la partie qu'ils pourraient jouer l'un contre l'autre. Ce résultat (« les Blancs gagnent », ou « les Noirs gagnent », ou « la partie est nulle ») constitue un « *équilibre* » tel qu'aucun des deux joueurs ne pourrait obtenir un résultat plus favorable contre son adversaire parfaitement clairvoyant. Aussi ces deux joueurs renonceraient-ils certainement à jouer et considéreraient-ils le résultat comme acquis.

*L'existence d'un tel équilibre dans tous les jeux de réflexion pure (jeu d'échecs, jeu de dames...) est précisément l'objet du théorème de Lucas, auquel nous faisons allusion.*

On peut dire que cet équilibre constitue *l'espérance* des joueurs dans les jeux de réflexion pure, de même que le règlement défini par la règle des partis constituait l'espérance des joueurs dans les jeux de hasard pur. Mais ici l'espérance n'est plus seulement « mathématique », en ce sens qu'elle ne s'appuie plus sur le calcul des probabilités : chaque joueur est assuré de réaliser au moins son espérance s'il est assez habile pour jouer au mieux de ses intérêts ; le résultat de la partie ne pourra d'ailleurs dépasser l'espérance d'un joueur que si son adversaire commet quelque maladresse.

Cette conclusion est, à première vue, assez paradoxale, et cela explique sans doute qu'elle soit restée cachée jusqu'aux dernières années du XIX<sup>e</sup> siècle. Le paradoxe vient précisément de l'intérêt constant suscité par le jeu d'échecs ou le jeu de dames, aucun de ces jeux ne paraissant menacé de devenir une simple récréation mathématique. C'est que *la complexité d'un tel jeu le met à l'abri d'une analyse complète* : aucun joueur ne peut concevoir, dans leur ensemble, les chaînes de coups par lesquelles la situation initiale peut être reliée aux situations finales (cela n'est possible que pour certaines « fins de partie », objets des problèmes d'échecs ou de dames). De plus, la structure de ces chaînes est trop compliquée pour se prêter à l'analyse mathématique, et il ne peut être question de les passer en revue, même à l'aide d'une machine électronique puissante. Aussi l'homme doit-il sans doute renoncer à savoir quelle est l'espérance des joueurs au début d'une partie d'échecs, bien que cette espérance existe certainement et soit l'une des trois suivantes :

les Blancs gagnent, la partie est nulle, les Noirs gagnent. C'est pourquoi le jeu d'échecs continuera à être pratiqué. On peut d'ailleurs se demander dans quelle mesure les méthodes préconisées par les champions s'approchent de celles qui permettraient à des joueurs parfaitement clairvoyants d'atteindre à coup sûr l'équilibre... et, par là-même, de tuer le jeu d'échecs (5).

Nous voyons ainsi que l'analyse précédente n'est pas d'un grand secours, en pratique, pour régler la conduite des joueurs d'échecs ou de dames ; et, s'il en était autrement, le jeu disparaîtrait. Mais les conditions sont bien différentes s'il s'agit d'un jeu de réflexion pure, à la fois moins compliqué et moins souvent joué, relatif à une situation réelle de conflit. *Le joueur qui parvient à analyser le jeu dispose alors d'une règle de conduite dont la valeur ne saurait être contestée.* Ce joueur, engagé dans une partie, connaît à tout instant son espérance, ce qui lui permet d'agir de manière à ne jamais en compromettre la réalisation, et, s'il arrive qu'une maladresse de son adversaire accroisse cette espérance, de conserver ensuite cet avantage « inespéré » (au sens propre) jusqu'à la fin de la partie.

Contrairement à ce qui se passait dans les jeux de hasard, l'espérance ne s'accompagne ici d'aucun risque pour le joueur capable d'appliquer une telle règle de conduite. Mais il n'en est naturellement pas de même pour le joueur qui, faute d'une parfaite clairvoyance, s'en remet, si peu que ce soit, à son « inspiration ». Ce défaut de clairvoyance est d'ailleurs le seul élément qui puisse ici introduire un risque, rendre le résultat de la partie incertain, redonner au jeu ce caractère de lutte sans lequel il n'est pas de vrai jeu.

L'étude des jeux de réflexion pure, après celle des jeux de hasard pur, ouvre de nouvelles voies à l'analyse de la conduite humaine. Il y aurait lieu d'approfondir leur étude comparée, de regarder de plus près la transformation subie par le jeu quand la *réflexion* y remplace le *hasard*, d'insister en particulier sur le passage de *l'espérance mathématique* à *l'espérance* (tout court) et sur l'existence même d'une espérance unique, qui suppose un *équilibre* entre les volontés antagonistes des deux joueurs. Nous aurons du moins l'occasion de revenir sur cette importante notion d'équilibre.

#### 4. — Jeux de stratégie et duels.

##### 4. 1. — Définition. Distribution de l'information. Intervention du hasard.

Les jeux de réflexion pure, dont nous venons de parler, appartiennent à une classe plus générale de jeux, celle des « *jeux de stratégie* ». Ce sont les jeux dont le résultat dépend de certains choix, en nombre

---

(5) Il est remarquable qu'Edgar Poe parle, dès 1831, de « la laborieuse futilité des échecs » où, dit-il, « la complexité est prise (erreur fort commune) pour de la profondeur ». Tout le passage est à lire (au début du *Double assassinat dans la rue Morgue*). E. Poe y compare les échecs, les dames et le whist, pour la manière dont ces jeux exploitent « la haute puissance de la réflexion » et font travailler « la faculté de l'analyse ». Certains détails laissent à penser qu'il entrevoyait déjà, confusément, le théorème de Lucas.

fini, effectués par les joueurs, dans des conditions fixées par la règle ; il n'est d'ailleurs pas exclu que leur résultat dépende aussi du hasard, dans une certaine mesure.

Nous nous bornerons souvent, dans la suite, à considérer sous le nom de jeux les « *duels* », c'est-à-dire les jeux de stratégie dans lesquels s'affrontent deux joueurs dont les intérêts sont exactement opposés : l'un des joueurs gagne ce que perd l'autre et, plus généralement, leurs ordres de préférence sont inverses.

Les duels que nous avons rencontrés, dans les jeux de réflexion pure, possédaient les caractères suivants : *toute intervention du hasard en était exclue et l'information y était parfaite*, en ce sens que les choix faits par les deux joueurs étaient successifs (le trait passant d'un joueur à l'autre) et que chacun des joueurs était, à chaque instant, parfaitement informé de tous les choix faits antérieurement (ou, ce qui revient au même, de leurs répercussions sur la situation). C'est en tenant compte de ces caractères que nous avons pu montrer l'existence d'un équilibre, au sens indiqué plus haut.

Il est facile d'imaginer des jeux qui ne possèdent pas ces caractères et, en particulier, des *jeux à information imparfaite*, où chacun des joueurs peut faire certains choix sans les révéler à son adversaire, ou en les lui révélant avec un retard plus ou moins long, c'est-à-dire après lui avoir laissé faire, à son tour, un certain nombre de choix. Nous verrons plus loin quelques exemples simples de tels jeux (entre autres, le jeu de « Pair ou Impair »), mais il est clair que ces jeux peuvent prendre des structures variées et complexes, ce qui les rend aptes à représenter de façon assez correcte de nombreuses situations réelles de conflit (par exemple, les situations économiques ou militaires dans lesquelles deux adversaires, disposant de certaines forces globales, peuvent les répartir en secret entre les différents terrains où ils s'affrontent). Il est d'ailleurs possible d'y faire intervenir le *hasard* sous la forme de certains choix aléatoires, ajoutés aux choix volontaires des joueurs (c'est ce qui arrive dans beaucoup de jeux de cartes).

L'étude générale des jeux de stratégie est beaucoup plus délicate que celle des jeux de réflexion pure. La méthode d'analyse pascalienne appliquée à ces derniers, comme aux jeux de hasard pur, ne s'étend facilement qu'aux *jeux à information parfaite*, dans lesquels interviendraient successivement le hasard et la réflexion, mais où chacun des joueurs resterait, à chaque instant, parfaitement informé de tous les choix, aléatoires ou volontaires, antérieurement effectués. L'étude de certains jeux particuliers, à information imparfaite, peut encore suivre la manière dont l'information circule entre les deux adversaires (c'est-à-dire considérer ces jeux sous leur « *forme développée* ») — une telle analyse de la succession et de l'enchaînement des coups présente un grand intérêt — ; mais cette méthode devient rapidement impraticable dès que le schéma de distribution de l'information se complique, et elle ne se prête pas bien à la recherche de résultats généraux sur les jeux de stratégie (6).

(6) On pourra consulter à ce sujet J. C. C. Mc KINSEY, *Introduction to the theory of games*, New-York, Mc Graw Hill, 1952, chap. 5 et 6.

#### 4. 2. — Forme normale d'un jeu de stratégie. Tactiques. Résultats. Ordres de préférence des joueurs.

Pour obtenir un schéma théorique à la fois simple et général, nous réduirons les jeux de stratégie à une forme appelée « *forme normale* », où l'on suppose que chaque joueur profite en une seule fois de toutes les initiatives que lui laisse la règle du jeu, et fixe à l'avance, par un choix global, la manière dont il jouerait dans toutes les circonstances possibles.

Il importe d'observer que ce schéma n'exclut nullement les jeux où les joueurs ont à faire, au cours d'une partie réelle, des choix successifs enchaînés de façon plus ou moins complexe par l'information sur laquelle ils s'appuient (il permet, en particulier, de représenter les jeux à information parfaite et, par exemple, le jeu d'échecs). Il faut seulement supposer que chacun des joueurs dresse à l'avance la liste de tous les choix élémentaires, en nombre fini, devant lesquels il pourrait se trouver placé au cours de la partie, et choisit, avant le commencement de cette partie, la décision qu'il prendrait à propos de chacun d'eux. On peut se représenter la liste ainsi dressée, avec l'indication de tous les choix élémentaires effectués à l'avance, comme exprimant les consignes données par le joueur à un mandataire chargé de le remplacer durant la partie. Tous les cas étant prévus, le mandataire n'aura aucune initiative à prendre, et tout se passera comme si le joueur avait été présent pour prendre personnellement, librement, au cours de la partie, toutes les décisions nécessaires. Car la liberté du joueur n'a été en rien diminuée par le fait que ses choix aient été décalés dans le temps.

Bien entendu, une telle manière de procéder exigerait en général des choix élémentaires préalables beaucoup plus nombreux que les choix à faire au cours d'une partie effectivement jouée, où tous les cas prévus ne se présenteraient pas. Dans les jeux où chaque partie impose déjà une grande variété de choix, le joueur devrait envisager, pour prévoir tous les cas, un nombre de choix élémentaires qui dépasse l'imagination. Il faut noter que chacun de ces choix élémentaires correspond, non pas à une *situation objective* (une situation des forces en présence, par exemple), mais à un ensemble d'informations possédées sur une telle situation par le joueur considéré, ou, en d'autres termes, à une *situation subjective*. Cette distinction ne joue que dans les jeux à information imparfaite, mais elle y est fondamentale, car deux situations objectivement identiques peuvent être subjectivement différentes, et inversement.

Mais, quelle que soit la complexité du choix global que chacun des joueurs devrait faire pour fixer à l'avance, et de façon complète, la manière dont il entend jouer, rien ne nous empêche de considérer le schéma théorique dans lequel un tel choix global représente l'ensemble des choix élémentaires par lesquels un joueur agit sur le résultat de la partie. Nous donnerons à tout choix global de ce type le nom de « *tactique* » ou « *stratégie pure* ».

La règle du jeu permet en principe de définir et d'énumérer toutes les tactiques dont dispose chacun des joueurs (le nombre de ces tactiques est par hypothèse fini, bien qu'il puisse être très grand). *L'initiative d'un joueur dans une partie se réduit alors au choix de l'une des tactiques dont il dispose.* Et il faut considérer que *les différents joueurs effectuent*

*leurs choix respectifs indépendamment les uns des autres, car chaque tactique prévoit toutes les situations subjectives auxquelles conduiraient toutes les tactiques possibles des autres joueurs, et fixe les choix élémentaires correspondants.*

Dès que chacun des joueurs a ainsi choisi l'une des tactiques dont il dispose, la règle du jeu précise *le résultat de la partie*, soit en le définissant de façon certaine si le hasard n'intervient pas dans le jeu, soit en indiquant, avec leurs probabilités respectives, les divers résultats qui restent possibles, si le hasard intervient dans le jeu. Dans ce dernier cas, il reste à jouer la partie pour en déterminer le résultat exact, mais il s'agit alors d'un jeu de hasard *pur*, puisque les joueurs ont déjà choisi leurs tactiques respectives. Il faut en effet retenir que ces tactiques prévoient, en particulier, toutes les circonstances aléatoires possibles, et fixent la ligne de conduite à tenir dans chaque éventualité ; les joueurs n'ont donc, strictement, plus aucune décision à prendre au cours de la partie. Et il reviendrait au même d'effectuer un tirage au sort convenable, pour en déterminer le résultat, parmi tous ceux qui restent possibles, après que les joueurs ont choisi leurs tactiques respectives. Ainsi, *le choix d'une tactique par chacun des joueurs définit complètement le résultat de la partie dans un jeu non aléatoire, et définit ce résultat à un tirage au sort près dans un jeu aléatoire.*

Cette dernière circonstance soulève une double difficulté, qu'il serait d'ailleurs impossible d'éluider, même si l'on écartait les jeux aléatoires. Il est bien clair que *toute réflexion visant à régler la conduite des joueurs doit s'appuyer non seulement sur le schéma de causalité précédent, mais encore sur un schéma de finalité : un ordre de préférence établi par chacun des joueurs* sur les résultats possibles du jeu, et même sur les situations qui résultent des diverses combinaisons de tactiques possibles, avant tout tirage au sort éventuel. Ainsi se trouve-t-on amené à comparer les résultats aléatoires dont il vient d'être question, c'est-à-dire certaines situations caractérisées par la possibilité de voir se réaliser divers résultats avec des probabilités données. Et la difficulté provient d'une part de ce que les résultats réalisables eux-mêmes peuvent être de natures différentes et sans commune mesure, d'autre part de ce qu'il ne suffit pas de définir un ordre de préférence relatif aux résultats réalisables pour qu'en découle, sans autre hypothèse, un ordre de préférence relatif aux distributions de probabilités sur ces résultats.

Si l'on suppose d'abord, schématiquement, que les divers résultats réalisables correspondent seulement à certains règlements monétaires entre les joueurs, chacun des joueurs établira sans peine un ordre de préférence sur ces résultats, d'après le « gain » (positif ou négatif) qui en résulte pour lui. Mais comment classera-t-il les situations aléatoires qui lui laissent l'espérance de réaliser certains gains avec des probabilités données ? Nous retrouvons ici la difficulté signalée à propos des jeux de hasard. La méthode la plus simple, et même la seule méthode objective, est de fonder le classement sur *l'espérance mathématique*, conformément à la règle des partis ; mais il n'est pas sûr, en fait, que le classement ainsi établi corresponde bien aux préférences personnelles du joueur.

C'est, suivant une expression classique, « l'espérance morale » qu'il faudrait atteindre ; il est théoriquement possible d'y parvenir à partir de quelques « équivalences morales » particulières, compte tenu des règles logiques auxquelles on peut soumettre l'ordre de préférence adopté sur les divers résultats aléatoires (7). Mais il y a peu de chances que se rencontre, au plan subjectif de l'espérance morale, l'opposition complète d'intérêts qui est à la base de la théorie du duel.

Plus généralement, ce n'est pas seulement de règlements monétaires qu'il s'agit dans les jeux auxquels nous nous intéressons surtout, ceux qui ont pour objet la décision et l'action humaines. Les intérêts en jeu sont complexes, sans commune mesure entre eux ; les divers résultats réalisables sont des situations concrètes où peuvent intervenir à la fois des valeurs matérielles, immédiates ou potentielles, certaines valeurs psychologiques ou morales (il ne serait pas sage de les négliger), des vies humaines. Un classement préférentiel peut ici devenir difficile, d'autant plus que certains éléments d'appréciation risquent de manquer ou d'être mal connus. Pourtant, *un tel classement est indispensable à qui veut ou doit jouer*. Et, s'il faut classer certaines distributions de probabilités sur les situations concrètes en question, on échappera difficilement à la nécessité, non seulement de *classer* les situations concrètes, mais de *repérer* (nous ne disons pas mesurer) la valeur qu'elles possèdent aux yeux du joueur, par un indicateur numérique qui puisse servir de base à la détermination de l'ordre de préférence adopté sur l'ensemble des résultats, certains ou aléatoires. C'est cet indicateur numérique (ou « utilité ») qui caractérise l'espérance morale ; il est logiquement permis de le calculer comme une espérance mathématique à partir de quelques valeurs de base qui expriment l'attitude psychologique personnelle du joueur par certaines équivalences morales. (Nous en verrons un exemple à la fin de cette étude).

Toutefois, une telle schématisation des préférences des joueurs, pour logique qu'elle paraisse, ne donne pas nécessairement une image exacte de la réalité psychologique, qui risque, en bien des cas, d'être plus souple (voir à ce sujet : R.-D. LUCE et H. RAIFFA, *op. cit.*, appendice I). Aussi la détermination d'indicateurs de préférence pour les différents joueurs peut-elle parfois présenter de sérieuses difficultés qu'il ne faut pas dissimuler. Mais il ne faut pas non plus en exagérer sans raison la portée, pour contester toute valeur pratique à la théorie des jeux.

On voit, en résumé, la nature des questions que pose la représentation d'une situation réelle de conflit par un jeu de stratégie mis sous forme normale :

- définition des *tactiques* dont disposent respectivement les différents joueurs ;
- définition des *résultats*, certains ou aléatoires, qui correspondent

---

(7) Voir au sujet de l'espérance morale :

J. VON NEUMANN et O. MORGENTERN, *Theory of games and economic behaviour*, Princeton, New-Jersey, 1944, 2<sup>e</sup> édition 1947 : « The notion of utility » (p. 15-30).

G. Th. GUILBAUD, *op. cit.*, chap. IV, par. 18.

R. D. LUCE et H. RAIFFA, *Games and Decisions*, New-York, Wiley 1957, chap. 2, « Utility theory », p. 12-38.

aux diverses combinaisons de tactiques possibles (chacun des joueurs étant libre de choisir l'une quelconque de ses tactiques, indépendamment des autres joueurs) ;

— détermination d'un *ordre de préférence* sur ces résultats, pour chacun des joueurs.

Il faut insister sur le fait qu'il est souvent délicat, en pratique, d'établir un tel modèle, et de savoir s'il est bien adapté à l'étude de la situation réelle qu'il doit représenter, compte tenu du contexte où elle s'insère par ses « tenants » et ses « aboutissants » ; car le jeu s'inscrit, par nature, dans un cadre limité (« horizon » temporel, spatial, ou opérationnel, bornant le « champ » d'action des joueurs), et la définition d'un tel cadre comporte toujours une part d'arbitraire. Un bon modèle, pour être à la fois représentatif et maniable, ne doit, ni trahir la réalité, ni en général la copier de trop près. La construction d'un modèle qui possède ces qualités demande de l'expérience et du discernement ; mais l'effort dépensé pour l'établir a déjà prouvé sa fécondité dans de nombreuses situations concrètes. Il serait, au contraire, vain de prétendre fonder une décision ou une action sur une réflexion dont les bases seraient mal assurées. C'est un point sur lequel nous voulions attirer l'attention avant de laisser, apparemment, les situations concrètes à l'arrière-plan de nos préoccupations.

#### 4. 3. — Etude tactique du duel. Existence ou absence d'un équilibre.

Ce que nous avons dit jusqu'ici s'applique à tous les jeux de stratégie, quels que soient le nombre des joueurs et la manière dont peuvent s'opposer ou s'allier leurs intérêts respectifs. Mais nous devons maintenant particulariser l'objet de notre étude et considérer le cas du *duel* (en nous réservant de revenir, pour terminer, sur certains cas plus généraux).

Un duel est, nous l'avons dit, un jeu de stratégie dans lequel s'affrontent deux joueurs dont les intérêts sont exactement opposés. Pour mettre un duel sous la forme normale, nous supposerons énumérées les *tactiques* respectives des deux joueurs A et B, définis les *résultats* correspondants aux divers choix possibles d'une tactique de A et d'une tactique de B, adopté un *indicateur numérique de préférence*, qui peut s'appliquer aux deux joueurs, compte tenu de l'opposition de leurs intérêts. Nous conviendrons que l'ordre des valeurs croissantes de cet indicateur marque l'ordre des préférences croissantes du joueur A et des préférences décroissantes du joueur B.

Afin de simplifier le langage, nous supposerons *schématiquement* que la valeur de l'indicateur de préférence représente une somme d'argent versée par le joueur B au joueur A, tous les résultats possibles, certains ou aléatoires, étant ainsi assimilés à des règlements monétaires certains entre les deux joueurs. Dans cette perspective, le résultat de chaque partie du jeu est un « règlement » qui représente le « gain » du joueur A et la « perte » du joueur B ; mais les gains négatifs (équivalents à des pertes) et les pertes négatives (équivalentes à des gains) ne sont pas exclus.

Le joueur A, ayant intérêt à rendre le règlement aussi grand que possible, est appelé le « *joueur du maximum* », tandis que le joueur B, ayant intérêt à rendre le règlement aussi petit que possible, est appelé le « *joueur du minimum* ».

La manière la plus commode de représenter dans son ensemble le schéma du jeu est de dresser ce que l'on appelle la « *matrice du jeu* ». C'est un tableau rectangulaire dont chaque ligne correspond à l'une des tactiques du joueur A (joueur du maximum) et dont chaque colonne correspond à l'une des tactiques du joueur B (joueur du minimum). Nous repérerons chaque tactique de A, et la ligne correspondante, par un numéro d'ordre  $i = 1, 2, 3, \dots, n$ , chaque tactique de B, et la colonne correspondante, par un numéro d'ordre  $j = 1, 2, 3, \dots, m$ .

A l'intersection de la  $i^{\text{e}}$  ligne et de la  $j^{\text{e}}$  colonne de la matrice du jeu, nous inscrivons le règlement  $r_{ij}$  qui interviendrait, d'après la règle du jeu, si le joueur A choisissait sa  $i^{\text{e}}$  tactique et le joueur B sa  $j^{\text{e}}$  tactique (cela veut dire, dans notre schéma, que B verserait  $r_{ij}$  à A).

*La matrice du jeu ainsi dressée définit entièrement la forme normale du duel considéré.*

Donnons, à titre d'exemple, la matrice d'un jeu où les joueurs A et B disposent respectivement de 5 et 4 tactiques (c'est ce que l'on appelle un jeu  $5 \times 4$ ) :

		B			
		1	2	3	4
A	1	12	3	-1	2
	2	4	-2	15	-20
	3	-3	1	11	-7
	4	6	5	7	8
	5	8	2	9	10

*Comment les joueurs A et B pourront-ils raisonner pour choisir leurs tactiques respectives au mieux de leurs intérêts ?*

La théorie des jeux, au stade où nous l'envisageons, exclut d'abord toutes les considérations psychologiques, d'après lesquelles l'un des joueurs pourrait chercher à déterminer l'attitude de son adversaire, à deviner la tactique qu'il va choisir. Nous ne prétendons pas, bien entendu, que ces considérations soient inutiles : si le joueur A avait de bonnes raisons de penser que le joueur B va se laisser tenter par la possibilité de gagner 20 et qu'il choisira en conséquence sa 4<sup>e</sup> tactique, A aurait un intérêt évident à choisir lui-même sa 5<sup>e</sup> tactique qui lui ferait alors gagner 10 (nous y reviendrons).

Mais la théorie du duel se place à un point de vue différent, plus prudent, et suppose que chacun des joueurs raisonne comme si son adversaire, intentionnellement ou non, opposait à chacune de ses tacti-

ques la riposte qui rend le règlement correspondant aussi peu avantageux que possible pour lui (chacun des joueurs joue, comme on dit, « contre le diable »). Sans discuter pour l'instant ce point de vue, voyons ce qu'il donne dans le cas de notre exemple.

Le joueur A (joueur du maximum) est ainsi amené à considérer, pour chacune de ses 5 tactiques, le règlement minimum inscrit dans la ligne correspondante de la matrice du jeu, soit respectivement — 1, — 20, — 7, 5, 2. Il préfère le plus grand de ces règlements, soit 5, ce qui le conduit à choisir sa 4<sup>e</sup> tactique, grâce à laquelle il est sûr de gagner au moins 5, qui est le « maximin », c'est-à-dire le maximum des minima de lignes.

Le joueur B (joueur du minimum) est de même amené à considérer, pour chacune de ses 4 tactiques, le règlement maximum inscrit dans la colonne correspondante de la matrice du jeu, soit respectivement 12, 5, 15, 10. Il préfère le plus petit de ces règlements, soit 5, ce qui le conduit à choisir sa 2<sup>e</sup> tactique, grâce à laquelle il est sûr de perdre au plus 5, qui est le « minimax », c'est-à-dire le minimum des maxima de colonnes.

Nous constatons immédiatement — mais n'en tirons pas de conclusion générale — que *le maximin est ici égal au minimax* (l'étymologie et le sens de ces deux termes ont été mis en évidence par ce qui précède). Il faut et il suffit, pour qu'il en soit ainsi, que l'un des éléments de la matrice du jeu soit à la fois minimum dans sa ligne et maximum dans sa colonne (le lecteur n'aura aucune peine à l'établir), ce qui est le cas ici de l'élément 5, à l'intersection de la 4<sup>e</sup> ligne et de la 2<sup>e</sup> colonne. On dit qu'un tel élément constitue un « col » (en anglais : *saddle point*) de la matrice du jeu.

Il est très important d'observer que *l'égalité du maximin et du minimax équivaut à l'existence d'un équilibre*, de même nature que celui dont nous avons établi l'existence dans les jeux de réflexion pure ; chacun des joueurs dispose d'une tactique prudente qui lui permet de s'assurer le règlement correspondant, et il ne peut obtenir un résultat plus favorable si son adversaire adopte aussi la tactique prudente qui vient d'être définie. Même s'il savait ou supposait que son adversaire ait adopté cette tactique, il n'aurait aucun intérêt à en adopter lui-même une autre ; il pourrait au contraire y perdre.

La valeur commune du maximin et du minimax constitue ici, comme dans les jeux de réflexion pure, *l'espérance* des deux joueurs. Mais il n'est pas exclu que cette espérance corresponde à un résultat aléatoire, si le hasard intervient dans le jeu.

Le point de vue auquel se place la théorie du duel avait pu nous paraître trop prudent, voire pessimiste. L'analyse précédente montre cependant que, *si le duel considéré admet un équilibre*, ce point de vue prudent permet à chacun des joueurs d'obtenir, contre toute défense de son adversaire, un résultat au moins aussi favorable que celui correspondant à l'équilibre. *Le jeu prudent*, le jeu « contre le diable », est donc aussi, dans ce cas, *le jeu habile contre un adversaire habile*. Mais ce n'est pas nécessairement le jeu le plus avantageux contre un adversaire maladroit.

Lorsqu'un duel admet un équilibre, la valeur commune du maximin et du minimax est appelée la « valeur du jeu ». Toute tactique qui assure au joueur du maximum un gain au moins égal à la valeur du jeu est une « tactique optimale » ; il en est de même pour toute tactique qui assure au joueur du minimum une perte au plus égale à la valeur du jeu. Chaque joueur peut d'ailleurs disposer de plusieurs tactiques optimales. Tout couple formé par deux tactiques optimales adverses constitue une « solution du jeu » et conduit à un règlement ayant pour valeur la valeur du jeu.

Le cas où il existe un équilibre est celui de tous les jeux de réflexion pure, et même de tous les jeux à information parfaite. *Il ne faudrait cependant pas croire que tous les duels admettent un équilibre.*

Considérons un second exemple, apparemment plus simple que le premier, celui d'un jeu  $2 \times 2$ , représenté, sous sa forme normale, par la matrice suivante, dont la signification est maintenant assez claire :

$$\begin{array}{c}
 \begin{array}{c}
 \text{B} \\
 \hline
 \begin{array}{cc}
 1 & 2
 \end{array}
 \end{array} \\
 \begin{array}{c}
 \left. \begin{array}{l}
 \text{A} \\
 \left. \begin{array}{l}
 1 \\
 2
 \end{array}
 \right\}
 \end{array}
 \right\}
 \begin{array}{|cc|}
 \hline
 3 & 6 \\
 \hline
 5 & 4 \\
 \hline
 \end{array}
 \end{array}
 \end{array}$$

Si le joueur A (joueur du maximum) fait ici le même raisonnement que tout à l'heure, il est conduit à choisir sa seconde tactique, grâce à laquelle il est sûr de gagner au moins 4, qui est le *maximin*.

Si le joueur B fait aussi le même raisonnement, il est conduit à choisir sa 1<sup>re</sup> tactique, grâce à laquelle il est sûr de perdre au plus 5, qui est le *minimax*.

Mais, ici, *le maximin n'est plus égal au minimax* ; aucun élément de la matrice n'est à la fois minimum dans sa ligne et maximum dans sa colonne, ou, autrement dit, cette matrice ne présente pas de col.

Dans ces conditions, *le jeu n'admet plus d'équilibre* : il reste une marge entre le gain de 4, le plus grand que le joueur A puisse être sûr de réaliser (autrement dit, *l'espérance de A*), et la perte de 5, la plus petite que le joueur B puisse être sûr de ne pas dépasser (autrement dit, *l'espérance de B*). Il est alors naturel de se demander comment cette marge va se trouver attribuée ou partagée. Si les deux joueurs s'en tiennent à leur ligne de conduite prudente, A choisit sa seconde tactique et B choisit sa première tactique, ce qui a pour résultat de faire gagner 5 à A ; c'est donc A qui bénéficie de la marge en gagnant plus qu'il ne l'avait espéré. Mais si le joueur B savait ou supposait que A ait choisi sa seconde tactique, il aurait intérêt à modifier sa ligne de conduite et à choisir sa seconde tactique au lieu de la première, car il ne perdrait plus alors que 4. Si à son tour A connaissait ce changement de tactique, il aurait intérêt à modifier, lui aussi, sa ligne de conduite et à choisir sa première tactique au lieu de la seconde, car il gagnerait alors 6. Mais c'est maintenant B qui aurait intérêt à revenir à sa première tactique pour ne plus perdre que 3. Après quoi, A aurait intérêt à revenir à sa seconde tactique pour porter son gain à 5. Et ainsi de suite...

On voit donc bien ce que signifie l'absence d'équilibre. Mais on voit moins bien ce qui va se passer, et sur quoi les joueurs vont pouvoir régler leur conduite : car il n'est pas question que l'un d'eux modifie son choix en fonction de celui de son adversaire. Le joueur qui serait autorisé à choisir après avoir été informé du choix de son adversaire serait en fait placé dans les conditions hypothétiquement accordées au « diable » par le raisonnement prudent de cet adversaire. L'équilibre se trouverait évidemment rétabli, mais il ne s'agirait plus du même jeu. Le lecteur pourra dresser les matrices des deux jeux  $2 \times 4$  et  $4 \times 2$  à information parfaite que l'on peut ainsi former à partir du jeu considéré, qui est à information imparfaite.

L'existence d'un équilibre apportait à la fois au joueur une satisfaction et une déception. Les tactiques optimales lui fixaient une règle de conduite, en un sens, « irréprochable » ; il ne pouvait mieux faire que de la suivre, si du moins il se mesurait avec un adversaire à sa taille. Mais alors, le résultat de la partie était connu d'avance ; aucune habileté ne permettait de l'améliorer.

En l'absence d'équilibre, la situation est bien différente. Il n'existe plus de tactique optimale, pas de meilleure manière de jouer. Quelles que soient les tactiques choisies par les deux joueurs, l'un d'eux, au moins, regrettera son choix ; que doit faire chacun pour essayer cependant de jouer au mieux de ses intérêts ? Cette fois, la partie n'est pas jouée d'avance, le risque reparait, la lutte reprend ses droits ; et l'on sent bien que l'habileté des joueurs doit pouvoir s'y exercer.

Si l'étude des jeux possédant un équilibre s'est ouverte dès 1838, par les importants travaux de *Cournot* (8), il a fallu attendre *E. Borel* et *J. von Neumann* pour voir se constituer une véritable théorie des jeux sans équilibre. Pourtant, le problème avait été posé de façon claire et précise par *Montmort* dès 1708 (9), à propos d'une situation de conflit assez proche de celle qui se présente dans le jeu de « pair ou impair » (*Montmort* lui avait seulement donné l'apparence du réalisme). Mais les duels dépourvus d'équilibre ont défié l'analyse pendant plus de deux siècles.

Nous allons maintenant reprendre à leur sujet la démarche dont *E. Borel* a été l'initiateur.

#### 4. 4. — Etude stratégique du duel. Jeu stratégique associé à un jeu tactique. Existence d'un équilibre.

Ceux qui, à la suite de *Montmort*, se sont intéressés aux jeux de stratégie, ont longtemps pensé que, dans une situation analogue à celle présentée par notre second exemple, il était impossible de régler efficacement la conduite des joueurs d'après la seule analyse objective du jeu.

---

(8) A. COURNOT (1801-1877), *Recherches sur les principes mathématiques de la théorie des richesses*, 1838 (réimprimées en 1938, avec introduction et notes par G. LUTFALLA, Paris, Rivière). Le point de vue y est d'ailleurs assez différent de celui que nous venons d'adopter, car la réflexion de *Cournot* porte sur la concurrence plutôt que sur la lutte à l'état pur (voir G. TH. GUILBAUD, *op. cit.*, chap. III).

(9) Pierre-Raymond DE MONTMORT (1678-1719), *Essai d'analyse sur les jeux de hasard*, Paris, 1708, 2<sup>e</sup> édition 1713.

C'est la psychologie qui aurait dû compléter cette analyse et apporter aux joueurs les éléments de décision nécessaires. Dans l'affrontement des deux volontés adverses, rien n'aurait pu, en définitive, remplacer l'engagement personnel des joueurs, la mesure qu'ils prennent l'un de l'autre, non seulement quant à leurs forces et à leurs objectifs, mais aussi quant à leur caractère, dans sa mystérieuse et mouvante complexité. Aussi, les consignes données à quelque mandataire, c'est-à-dire les tactiques prévues par la forme normale du jeu, n'auraient-elles pu suffire en fait à rendre un compte exact de la confrontation humaine constituée par la lutte des deux joueurs.

*Il faut, certes, se garder de méconnaître l'importance des facteurs psychologiques dans la conduite des joueurs, quels que soient les jeux où ils sont engagés, qu'il s'agisse de problèmes économiques, politiques ou militaires. Ceux qui ont à prendre des décisions importantes agiraient légèrement en ne tenant pas le plus grand compte de la psychologie de leurs adversaires. Et, bien qu'il soit possible de faire entrer certaines données psychologiques dans la formulation d'un problème de jeu (dans la construction même de la matrice du jeu), il reste que la théorie des jeux, comme toutes les techniques de Recherche Opérationnelle, ne prétendent nullement imposer une décision à celui qui en a la responsabilité. La théorie des jeux cherche seulement, comme la « Stochastique » au sens de Bernoulli, à réunir des éléments de décision, fondés sur une analyse aussi poussée que possible de la situation. Après quoi, le responsable, informé par la théorie des jeux, mais peut-être aussi par son expérience et son discernement psychologique, choisit ce qui lui paraît, en définitive, « meilleur, plus satisfaisant, plus sûr, plus prudent ».*

On voit donc ce qui peut fonder le point de vue de Montmort et de ceux qui l'ont suivi, jusqu'à E. Borel. Mais, s'il est manifestement avantageux d'exploiter toute information sur la psychologie de l'adversaire, il n'est pas moins avantageux, par conséquent, de ne donner à cet adversaire aucune information psychologique dont il pourrait à son tour tirer parti. Il est donc raisonnable et utile de considérer, à la limite, le cas où l'un des aspects de l'habileté des joueurs est de priver l'adversaire de toute information psychologique, grâce à un comportement sur lequel on pourra aussi s'interroger. La véritable question est de savoir si, *dans ces conditions*, il est possible d'aller au-delà de l'analyse qui nous a conduits à constater, dans certains cas, qu'il n'existe pas d'équilibre, pas de meilleure manière de jouer. *C'est la réponse négative donnée à cette question (jusqu'à E. Borel) qu'il s'agit de remettre en cause.*

Pour introduire cette réflexion, relisons le passage de *La lettre volée* (publiée en 1841) où *Edgar Poe* analyse, de façon simple et profonde, le jeu de Pair ou Impair, en insistant sur les considérations psychologiques :

« J'ai connu un enfant de huit ans, dont l'infailibilité au jeu de pair ou impair faisait l'admiration universelle. Ce jeu est simple, on y joue avec des billes. L'un des joueurs tient dans sa main un certain nombre de ses billes, et demande à l'autre : " Pair ou non ? ". Si celui-ci devine juste, il gagne une bille ; s'il se trompe, il en perd une. L'enfant dont je parle gagnait toutes les billes de l'école. Naturellement, il avait un mode de divination, lequel consistait

dans la simple observation et dans l'appréciation de la finesse de ses adversaires. Supposons que son adversaire soit un parfait nigaud et, levant sa main fermée, lui demande : " Pair ou impair ? ". Notre écolier répond : " Impair ! " et il a perdu. Mais, à la seconde épreuve, il gagne, car il se dit en lui-même : " Le niais avait mis pair la première fois, et toute sa ruse ne va qu'à lui faire mettre impair à la seconde ; je dirai donc : Impair ! " Il dit : " Impair ! ", et il gagne.

Maintenant, avec un adversaire un peu moins simple, il aurait raisonné ainsi : « Ce garçon voit que, dans le premier cas, j'ai dit " Impair ", et, dans le second, il se proposera — c'est la première idée qui se présentera à lui — une simple variation de pair à impair comme a fait le premier bête ; mais une seconde réflexion lui dira que c'est là un changement trop simple, et finalement il décidera à mettre pair comme la première fois. Je dirai donc " Pair ! ". Il dit : " Pair ", et gagne. Maintenant, ce mode de raisonnement de notre écolier, que ses camarades appellent la chance, — en dernière analyse, qu'est-ce que c'est ?

— C'est simplement, dis-je, une identification de l'intellect de notre raisonneur avec celui de son adversaire. »

(Traduction de Ch. BAUDELAIRE).

Cette analyse est intéressante à plusieurs titres, et particulièrement par sa conclusion, sur laquelle Edgar Poe insiste d'ailleurs dans la suite du texte. L'habileté de l'enfant est d'identifier son « intellect » avec celui de son adversaire. Cette identification repose ici sur les observations psychologiques que l'adversaire lui laisse faire. Et si ces observations se révèlent presque toujours justes, c'est que les adversaires considérés ne sont pas assez habiles pour varier leur jeu d'une manière imprévisible.

Dans un duel, en effet, *c'est une règle d'or de chercher à raisonner en s'identifiant à l'adversaire*, afin d'agir au mieux, compte tenu de ce que fera sans doute l'adversaire. Mais il faut, bien entendu, prévoir que l'adversaire raisonnera de la même manière, si du moins il est assez habile. Il s'agit ici, non pas d'exercer sa perspicacité sur l'appréciation d'un *obstacle matériel*, mais d'analyser la démarche d'une intelligence qui sert une *volonté adverse*, de se représenter en un mot l'adversaire comme un « *autre soi-même* », dans le jeu duquel on cherche à entrer.

Lorsqu'un joueur raisonne comme s'il jouait contre le « diable », il amorce une telle analyse, mais avec un certain pessimisme, comme nous l'avons déjà souligné. Pourtant, si le jeu présente un équilibre, il n'est pas nécessaire de pousser plus loin l'analyse ; cela n'a d'ailleurs rien d'étonnant s'il s'agit d'un jeu à information parfaite, puisque l'adversaire perspicace est alors aussi bien informé que pourrait l'être le « diable ».

Mais, dans le cas général où il n'existe pas d'équilibre, faut-il s'en tenir à ce point de vue pessimiste et se contenter en principe du résultat que l'on est sûr de pouvoir atteindre ? Ou bien ne faut-il pas plutôt chercher à profiter de ce que l'adversaire n'est pas le « diable » ? L'exemple donné par E. Poe nous donne à penser qu'il serait sage de *varier son jeu* de telle manière que l'adversaire, cherchant à deviner ce jeu, n'y puisse parvenir. Et cela est naturellement conforme à la conduite d'un bon stratège, qui ruse avec l'ennemi en lui cachant l'emploi qu'il va faire de ses forces, afin de lui ôter toute sécurité.

Ces indications de bon sens, connues depuis fort longtemps, demandent cependant à être précisées, d'autant plus que la « *variation du jeu* »

n'est pas une notion claire si le jeu comporte une seule partie ; et, s'il en comporte plusieurs, l'exemple même du jeu de Pair ou Impair montre qu'il n'est pas si facile de varier judicieusement son jeu.

C'est *Emile Borel* qui, en termes précis et mesurés, ouvre une nouvelle voie, en 1921, par la première note qu'il consacre à ce sujet (10) :

« Dans le cas où cette meilleure manière (de jouer) n'existe pas, on peut se demander s'il n'est pas possible, à défaut d'un code choisi une fois pour toutes, de jouer d'une manière avantageuse en variant son jeu. Si l'on veut formuler une règle précise pour varier le jeu, cette règle ne faisant intervenir que les faits observés dans le jeu, et non pas des remarques psychologiques sur le joueur auquel on est opposé, cette règle équivaut forcément à un énoncé tel que le suivant : *la probabilité pour que, en un moment donné du jeu, (le joueur) A adopte, pour fixer sa conduite à ce moment, le code  $C_k$ , est  $p_k$ .* »

Le duel prend ainsi un nouvel aspect et, à vrai dire, *il s'agit bien d'un nouveau jeu*, dans lequel E. Borel propose de plonger le jeu donné. Au lieu de choisir *une tactique* parmi celles dont il dispose, comme la règle du jeu donné l'y invite, chaque joueur va choisir *une distribution de probabilités* à répartir sur l'ensemble de ses tactiques ; après quoi, rien n'empêche de supposer qu'il effectuera un tirage au sort convenable, conforme à la distribution de probabilités choisie, pour déterminer effectivement la tactique qu'il adoptera.

Les distributions de probabilités ainsi choisies par les joueurs sont appelées des « *stratégies* », ou « *stratégies mixtes* », par opposition aux « *tactiques* », ou « *stratégies pures* », dont ils disposent d'après la règle du jeu primitif. Le nouveau jeu ainsi défini est le « *jeu stratégique* » associé au « *jeu tactique* » donné. Il importe d'observer que les stratégies, contrairement aux tactiques, sont en nombre infini. Les tactiques sont des stratégies particulières.

Le jeu stratégique est toujours un jeu aléatoire, même si le jeu tactique ne l'est pas. Les différents résultats du jeu stratégique sont des distributions de probabilités sur les résultats du jeu tactique. Et, bien entendu, cela pose une question essentielle : celle de la *détermination d'un ordre de préférence sur ces résultats aléatoires*, pour chacun des deux joueurs.

C'est ici que se présente à nouveau la difficulté déjà rencontrée dans la détermination de la forme normale d'un jeu aléatoire, et il n'est pas nécessaire d'y revenir. Mais c'est essentiellement parce que l'ordre de préférence relatif aux résultats du jeu tactique n'impose pas un ordre de préférence relatif aux résultats du jeu stratégique associé, qu'il y a lieu de considérer ce dernier comme un nouveau jeu.

Limitons-nous, pour l'instant, au cas dont la théorie est la plus ache-

(10) E. BOREL, « La théorie du jeu et les équations intégrales à noyau symétrique gauche », *C. R. Ac. Sc.*, t. 173, p. 1304-1308, Séance du 19-12-1921.

Cette note a été reproduite, avec un mémoire qui la développe (« Sur les jeux où interviennent le hasard et l'habileté des joueurs ») dans la 3<sup>e</sup> édition des *Eléments de la théorie des probabilités*, d'E. BOREL, Paris, Hermann, 1924, notes III et IV. p. 199 sqq.

vée, celui où le jeu stratégique associé au duel tactique reste un duel ; autrement dit, les intérêts des deux joueurs y restent exactement opposés. Nous supposons alors que l'indicateur de préférence du jeu tactique peut être choisi de telle manière que celui du jeu stratégique puisse s'en déduire, pour les deux joueurs, par un calcul d'espérance mathématique. (En fait, c'est la complète opposition des intérêts des deux joueurs dans le jeu stratégique qui est contestable pour certains jeux concrets. Nous reviendrons sur le cas de ces jeux à propos d'un exemple typique.)

Ces conventions sont parfaitement claires si l'on conserve ici le schéma dans lequel la valeur de l'indicateur de préférence du jeu tactique représente un règlement monétaire certain entre les deux joueurs : le gain du joueur A (joueur du maximum) et la perte du joueur B (joueur du minimum). L'indicateur de préférence du jeu stratégique est alors l'espérance mathématique du règlement aléatoire qui constitue le résultat d'une partie de ce jeu, compte tenu des distributions de probabilités choisies par chacun des deux joueurs sur l'ensemble de ses tactiques. (S'il s'agissait effectivement de règlements monétaires entre les deux joueurs, il ne s'imposerait nullement que l'indicateur numérique de préférence, calculé comme une espérance mathématique à partir de certaines valeurs de base, coïncide avec l'espérance mathématique du règlement monétaire. Mais il ne s'agit ici que de règlements figuratifs).

Si donc le joueur A choisit la stratégie  $(x_i)$  définie par l'attribution des probabilités respectives  $x_i$  ( $i = 1, 2, \dots, n$ ) aux  $n$  tactiques dont il dispose, avec

$$\sum_{i=1}^n x_i = 1, \quad \text{et } x_i \geq 0,$$

et si le joueur B choisit la stratégie  $(y_j)$  définie par l'attribution des probabilités respectives  $y_j$  ( $j = 1, 2, \dots, m$ ) aux  $m$  tactiques dont il dispose, avec

$$\sum_{j=1}^m y_j = 1, \quad \text{et } y_j \geq 0,$$

le résultat aléatoire correspondant a pour « valeur » l'espérance mathématique du « gain » de A et de la « perte » de B, soit :

$$R(x_i, y_j) = \sum_{i=1}^n \sum_{j=1}^m r_{ij} x_i y_j,$$

$r_{ij}$  étant le règlement certain qui interviendrait si le joueur A choisissait sa  $i^{\circ}$  tactique et le joueur B sa  $j^{\circ}$  tactique.

Dans le second exemple que nous avons considéré, celui du jeu  $2 \times 2$  dépourvu d'équilibre, dont la matrice est  $\begin{pmatrix} 3 & 6 \\ 5 & 4 \end{pmatrix}$ , ce résultat aléatoire a pour valeur

$R(x_1, y_1) = 3x_1y_1 + 6x_1y_2 + 5x_2y_1 + 4x_2y_2 = 4 + 2x_1 + y_1 - 4x_1y_1$ ,  
puisque'ici

$$x_2 = 1 - x_1, \quad y_2 = 1 - y_1.$$

Reprenons alors, pour le jeu stratégique, le point de vue prudent

déjà adopté pour le jeu tactique, et supposons que chacun des joueurs raisonne comme s'il jouait « contre le diable ».

Le joueur A (joueur du maximum) est ainsi amené à considérer, pour toute stratégie  $(x_i)$  qu'il peut choisir, la valeur minimum que peut prendre  $R(x_i, y_j)$  quand les probabilités  $y_j$  varient de toutes les manières possibles, soit ici :

$$\min_{(y_j)} R(x_i, y_j) = \min_{(y_1)} [4 + 2x_1 + y_1(1 - 4x_1)].$$

Et il choisit sa stratégie  $(x_i)$  de manière à rendre cette valeur minimum aussi grande que possible, ce qui lui permet de s'assurer une espérance mathématique de gain au moins égale à

$$\max_{(x_i)} \min_{(y_j)} R(x_i, y_j),$$

soit ici :  $4 + 2x_1 = 5 - 2x_1 = 9/2,$

avec la stratégie :  $(x_1 = \frac{1}{4}, x_2 = \frac{3}{4}).$

Le joueur B (joueur du minimum) est de même amené à considérer, pour toute stratégie  $(y_j)$  qu'il peut choisir, la valeur maximum que peut prendre  $R(x_i, y_j)$  quand les probabilités  $x_i$  varient de toutes les manières possibles, soit ici :

$$\max_{(x_i)} R(x_i, y_j) = \max_{(x_1)} [(4 + y_1) + x_1(2 - 4y_1)].$$

Et il choisit sa stratégie  $(y_j)$  de manière à rendre cette valeur maximum aussi petite que possible, ce qui lui permet de s'assurer une espérance mathématique de perte au plus égale à :

$$\min_{(y_j)} \max_{(x_i)} R(x_i, y_j),$$

soit ici :  $4 + y_1 = 6 - 3y_1 = 9/2,$

avec la stratégie :  $(y_1 = \frac{1}{2}, y_2 = \frac{1}{2}).$

Nous constatons que le remplacement du jeu tactique par le jeu stratégique associé a permis, dans l'exemple considéré, d'obtenir un *équilibre*. Dans le jeu stratégique, le maximin et le minimax ont une valeur commune, soit 9/2, comprise entre le maximin 4 et le minimax 5 du jeu tactique.

La marge qui subsistait, dans le jeu tactique, entre les espérances des deux joueurs, se trouve ainsi, dans le jeu stratégique, partagée entre leurs espérances mathématiques (ce partage se fait ici à parts égales, mais il n'y a pas lieu, bien entendu, d'y voir un résultat général). Chacun des deux joueurs dispose alors d'une stratégie prudente qui lui permet de s'assurer cette espérance mathématique, et il ne peut en obtenir une meilleure si son adversaire adopte aussi la stratégie prudente à laquelle conduit l'analyse qui précède. Même s'il savait ou supposait que son adversaire ait adopté cette stratégie, il n'aurait aucun intérêt à en adopter lui-même une autre ; il pourrait au contraire y perdre.

Ici encore, la *stratégie prudente est aussi la stratégie habile contre un adversaire habile*. Mais ce n'est pas nécessairement la stratégie la plus avantageuse contre un adversaire maladroit.

Ces conclusions, dont nous venons de vérifier l'exactitude sur l'exemple considéré, s'appliquent au cas général. Cela résulte du théorème fondamental d'après lequel *le maximin est égal au minimax dans tous les duels stratégiques*. Autrement dit, avec les notations précédentes :

$$\max_{(x_i)} \min_{(y_j)} \left[ \sum_{i=1}^n \sum_{j=1}^m r_{ij} x_i y_j \right] = \min_{(y_j)} \max_{(x_i)} \left[ \sum_{i=1}^n \sum_{j=1}^m r_{ij} x_i y_j \right],$$

quels que soient les  $mn$  nombres donnés  $r_{ij}$ , les opérations  $\max_{(x_i)}$  et  $\min_{(y_j)}$  étant respectivement effectuées par rapport à toutes les distributions de probabilités :

$$x_1, x_2, \dots, x_n \text{ avec } \sum_{i=1}^n x_i = 1, x_i \geq 0, \text{ et } y_1, y_2, \dots, y_m \text{ avec } \sum_{j=1}^m y_j = 1, y_j \geq 0.$$

Emile Borel a établi ce théorème pour certaines classes de jeux, mais n'a pas réussi à en obtenir la démonstration générale (la genèse de cet échec est d'ailleurs fort instructive, mais son étude nous entraînerait trop loin). C'est *J. von Neumann* (11) qui a donné le premier cette démonstration en 1928, par des méthodes qui sont loin d'être élémentaires. La première démonstration élémentaire, due à *J. Ville* (12), date de 1938. D'autres démonstrations ont été données depuis, en particulier par *J. von Neumann* et *O. Morgenstern* dans leur monumental traité, déjà cité, sur la théorie des jeux. Mais aucune de ces démonstrations n'est simple et cela tient, semble-t-il, à la profondeur même du théorème fondamental. Disons seulement ici que ce théorème est un aspect particulier du théorème fondamental de dualité, dont l'importance est primordiale dans l'étude des programmes linéaires (la théorie du duel stratégique peut d'ailleurs être considérée comme un chapitre de la programmation linéaire).

Nous considérons donc comme établi qu'*un duel stratégique admet toujours un équilibre*. Ici encore, la valeur commune du maximin et du minimax est appelée la « *valeur du jeu* ». Toute stratégie qui assure au joueur du maximum une espérance mathématique de gain au moins égale à la valeur du jeu est une « *stratégie optimale* » ; il en est de même pour toute stratégie qui assure au joueur du minimum une espérance mathématique de perte au plus égale à la valeur du jeu. Chaque joueur peut d'ailleurs disposer de plusieurs stratégies optimales, dont certaines peuvent se réduire à des tactiques. Tout couple formé par deux stratégies optimales adverses constitue une « *solution du jeu* » et conduit à un résultat ayant pour valeur la valeur du jeu.

On montre sans difficulté que toute tactique intervenant, avec une probabilité non nulle, dans une stratégie optimale d'un joueur, conduit aussi à un résultat ayant pour valeur la valeur du jeu, si l'adversaire

(11) J. VON NEUMANN, « Zur Theorie der Gesellschaftsspiele », *Math. Ann.*, 100, p. 295-320, 1928.

(12) J. VILLE, « Note sur la théorie générale des jeux où intervient l'habileté des joueurs », *Traité du calcul des probabilités et de ses applications*, publié par E. Borel, tome IV, fascicule II, p. 105-113, Paris, Gauthier-Villars, 1938.

adopte l'une quelconque de ses stratégies optimales. Mais il ne faut pas en conclure qu'un joueur puisse impunément choisir, à la place d'une stratégie optimale, l'une des tactiques qui y interviennent ; car son adversaire pourrait alors en profiter pour le pénaliser en abandonnant à son tour ses stratégies optimales (ces remarques se vérifient immédiatement sur le jeu  $2 \times 2$  considéré à titre d'exemple).

La détermination des stratégies optimales et de la valeur d'un jeu est facile, comme nous l'avons vu, pour les jeux  $2 \times 2$ . Les remarques qui précèdent (et d'autres qui dépassent le cadre de cet exposé) aident, dans tous les cas, à résoudre les jeux stratégiques. Mais cette résolution peut demander de longs calculs pour les jeux de grandes dimensions (il s'agit en fait de résoudre deux problèmes de programmation linéaire se correspondant par dualité) (13).

Au terme de cette analyse, nous pouvons nous interroger sur la portée des résultats obtenus. L'intuition de Borel s'est révélée féconde en ce qu'elle a permis de dépasser le point de vue suivant lequel, en l'absence d'équilibre, l'analyse objective du jeu devrait s'arrêter à la constatation qu'il n'existe pas de manière optimum de jouer. S'il est vrai que le joueur doit alors, dans une certaine mesure, s'en remettre au hasard pour régler sa conduite, du moins pouvons-nous le guider dans le choix du tirage au sort à effectuer parmi les tactiques dont il dispose, afin que ce tirage au sort soit, non un aveu d'impuissance, mais le résultat d'une analyse objective conduite à son terme. Il ne s'agit pas ici de jouer au hasard, mais d'introduire délibérément et judicieusement le hasard dans sa manière de jouer.

Cette introduction du hasard, par une distribution de probabilités sur l'ensemble des tactiques, a quelque chose de paradoxal au premier abord ; mais comme le souligne Borel, c'est la seule voie qui reste ouverte, hors celle de l'analyse psychologique ; et, à la réflexion, c'est pour le joueur la seule manière vraiment efficace de « varier son jeu ».

S'il s'agit d'un jeu comportant un grand nombre de parties, il est naturel de penser à jouer les diverses tactiques possibles avec certaines fréquences judicieusement choisies ; et l'analyse du jeu global se présentera comme celle du jeu stratégique correspondant, mais à la condition essentielle que l'adversaire ne puisse deviner la manière dont les diverses tactiques vont être successivement employées (faute de quoi on tomberait dans des pièges analogues à ceux que signale Edgar Poe à propos du jeu de Pair ou Impair). Or, il est bien difficile de simuler le hasard, de ne pas introduire de régularité décelable dans la succession des choix ; mieux vaut donc tirer au sort.

Il y a cependant des jeux, des situations de conflit, où la partie à jouer est unique, impossible à rejouer : il n'est plus question d'établir

---

(13) On pourra consulter l'ouvrage synthétique de S. VAJDA, *The theory of games and linear programming*, Londres, Methuen, 1956. Une adaptation française de ce livre, par l'auteur du présent article, a paru chez Dunod en 1959.

On trouvera d'autres détails sur la théorie et la résolution des duels dans :

J. C. C. Mc KINSEY, *op. cit.*, chap. 1, 2, 3 et 14.

R. D. LUCE et H. RAIFFA, *op. cit.*, chap. 3 et 4, et appendices 2 à 6.

C. W. CHURCHMAN, R. L. ACKOFF, E. L. ARNOFF, *Introduction to operations research*, New-York, John Wiley, 1957, 8<sup>e</sup> partie : « Competitive models ».

des moyennes entre des résultats partiels. Tel est souvent le cas des jeux économiques ou militaires. On peut alors penser que, dans ces conditions, chacun des joueurs doit, en définitive, adopter une *tactique*, quelle que soit la manière dont il la détermine, et que l'intervention des *stratégies* devient assez artificielle. Cependant, cette critique n'est pas fondée, car l'habileté du joueur consiste ici à varier son jeu d'une manière non plus effective mais potentielle, de telle façon que son adversaire ne puisse tirer avantage du fait qu'il cherche à calquer son raisonnement sur le sien. Le meilleur moyen d'échapper à l'analyse de l'adversaire est encore de tirer au sort, comme le simple examen du jeu de pair ou impair peut nous en convaincre.

L'introduction contrôlée du hasard dans les jeux stratégiques ne doit donc pas nous surprendre : elle est conforme à la nature même de ces jeux. C'est l'un des moyens les plus efficaces de ruser avec l'adversaire, de le priver, comme nous le souhaitons, de toute information psychologique exploitable sur les variations, effectives ou potentielles, de notre jeu. La théorie des jeux nous permet de nous servir à bon escient de ce moyen, de manière à obtenir, contre toute stratégie de notre adversaire, une espérance mathématique dont la valeur la moins favorable soit aussi bonne que possible.

Une difficulté subsiste cependant, et nous l'avons déjà rencontrée. L'introduction du hasard dans les stratégies oblige à considérer les résultats du jeu stratégique comme aléatoires et à déterminer *un ordre de préférence* sur ces résultats aléatoires pour chacun des deux joueurs. Si les résultats du jeu tactique sont déjà aléatoires, la difficulté se présente, comme nous l'avons vu, dans le jeu tactique lui-même, et n'est pas accrue par le passage au jeu stratégique. Mais si les résultats du jeu tactique sont certains, le passage au jeu stratégique change la nature du jeu à un point tel que certains joueurs pourront s'en tenir au jeu tactique, même s'il ne présente pas d'équilibre.

C'est qu'il s'agit à nouveau ici d'établir un arbitrage entre des résultats certains et des résultats aléatoires, dans lesquels interviennent, avec des probabilités déterminées, différents résultats du jeu tactique : le maximin, le minimax et quelques autres. Si le joueur estime que certains de ces résultats lui font courir un trop grand risque, il peut préférer l'espérance dont il est sûr de voir la réalisation dans le jeu tactique (le maximin pour le joueur du maximum, le minimax pour le joueur du minimum), à l'espérance « mathématique » que pourrait lui donner le jeu stratégique. Car cette espérance mathématique ne caractérise pas la situation concrète qui se réalisera effectivement à la fin d'une partie du jeu, mais seulement une situation fictive correspondant à une distribution de probabilités sur plusieurs situations possibles, dont certaines peuvent rester bien en-deçà de l'espérance promise au joueur dans le jeu tactique.

[Dans le jeu  $2 \times 2$  qui nous a servi d'exemple, la valeur  $9/2$  du jeu stratégique ne se réalisera en aucun cas, mais si les joueurs choisissent leurs stratégies optimales :  $(1/4, 3/4)$  pour A, et  $(1/2, 1/2)$  pour B, la partie pourra donner comme résultat 3, 4, 5 ou 6 avec les probabilités respectives :  $1/8, 3/8, 3/8, 1/8$ .

Cependant, le joueur A peut s'assurer un gain au moins égal à 4 en s'en tenant à sa seconde tactique, et le joueur B peut s'assurer une perte au plus égale à 5 en s'en tenant à sa première tactique. En contrepartie, le joueur le plus prudent devra alors craindre que son adversaire devine sa prudence et en profite pour s'adjuger la marge qui existe entre leurs deux espérances].

Il ne faut pas non plus exagérer la portée de cette difficulté. De nombreuses situations concrètes sont représentables, de façon satisfaisante, par un modèle de jeu conforme à notre hypothèse ; autrement dit, l'indicateur de préférence du jeu tactique peut y être choisi de telle manière qu'il soit raisonnable d'en déduire celui du jeu stratégique par un calcul d'espérance mathématique. L'arbitrage cherché entre les différents résultats, certains et aléatoires, se trouve ainsi établi, et *les stratégies optimales fixent alors au joueur une règle de conduite, en un sens, « irréprochable »*. Que le jeu se joue en une ou en plusieurs parties, il ne peut mieux faire que de la suivre, si du moins il se mesure avec un adversaire à sa taille, auquel cas *le jeu se transforme en fait en un jeu de hasard pur*.

Nous voulons enfin attirer à nouveau l'attention sur le fait que les tactiques ou stratégies optimales auxquelles conduit la théorie du duel définissent une ligne de conduite *prudente* qui n'est vraiment *optimale* que si l'on s'oppose à un adversaire *habile*. Mais nous avons déjà montré que certaines observations psychologiques sur un adversaire maladroit peuvent permettre d'employer avec profit une tactique ou une stratégie non optimale. C'est qu'alors il est possible d'acquérir *a priori* certaines informations sur la ligne de conduite de l'adversaire, d'écartier par exemple certaines des tactiques dont il dispose, ou de limiter l'ensemble de ses stratégies.

Dans le même ordre d'idées, la règle de conduite donnée par la théorie du duel ne se justifie pleinement que si l'on est vraiment engagé dans un duel, c'est-à-dire s'il s'agit bien de l'affrontement de deux volontés adverses. Tel n'est pas le cas si les intérêts en jeu ne sont pas vraiment opposés (nous allons y revenir), ni si l'on se heurte à des « obstacles » plutôt qu'à un véritable « adversaire ».

Aussi les jeux « contre la Nature », où seule notre ignorance peut nous conduire à prévoir le pire (à jouer « contre le diable »), ne devraient-ils pas, en principe, être traités comme des duels.

Il s'agit en fait, dans ces jeux, de prendre une décision ou une suite de décisions, devant une situation sur laquelle on dispose tout au plus d'une information incomplète, mais sans qu'une autre volonté, capable d'intentions et de choix, agisse, à notre connaissance, sur l'événement qui résulte de nos décisions. Peut-être pourrions-nous améliorer notre information au prix de certaines expériences plus ou moins coûteuses, plus ou moins risquées, conduites suivant un plan dont nous devrions aussi décider. En tout cas, il faut décider et agir, non pas face à un adversaire, dont les réactions pourraient être l'objet d'une « analyse », mais devant une nature inconnue, ou mal connue, qui peut seulement se laisser « espionner ».

Ainsi se présente la difficile question de la conduite à tenir devant

une situation incertaine, où il peut être arbitraire, et même dénué de sens, d'introduire des probabilités *a priori* (voir à ce sujet R. D. LUCE et H. RAIFFA, *op. cit.*, chap. 13 : « Individual decision making under uncertainty »).

La partie la plus importante de cette théorie concerne les « décisions statistiques », à l'étude desquelles Abraham Wald a consacré des travaux fondamentaux (14). Disons seulement ici que la théorie des décisions statistiques et la théorie du duel s'éclairent l'une l'autre et se prêtent un mutuel appui pour faire face aux situations concrètes, leur accord profond reposant sur le fait essentiel qu'elles utilisent les probabilités comme *instrument de décision et d'action* beaucoup plus que pour décrire objectivement l'univers matériel.

N. B. — *Quelques exemples* éclaireraient utilement les différents aspects que nous avons essayé de présenter dans le duel tactique et stratégique. Faute de place, et parce que nous ne saurions mieux faire, nous renvoyons le lecteur au livre de J.-D. WILLIAMS, *La stratégie dans les actions humaines*, et particulièrement aux exemples 2, 4, 5, 10, 13, 16, 19, 20, 22, 27, 29 et 34. Si le lecteur ouvre le livre, il est d'ailleurs à présumer qu'il aura du mal à laisser de côté les vingt-trois autres exemples.

#### 4. 5. — Jeux où interviennent à la fois la coopération et la lutte.

Nous venons de parler du duel, où les intérêts des deux joueurs sont, par définition, exactement opposés en toutes circonstances (on dit parfois brièvement que le duel est un jeu « à somme nulle »). Mais le cas du duel n'est qu'un cas extrême, de sorte qu'il importe de considérer aussi les jeux où apparaissent des intérêts dont les relations mutuelles sont plus complexes. Il peut arriver, par exemple, que deux joueurs aient des intérêts opposés sur certains terrains, mais concordants sur d'autres. Si d'ailleurs le nombre des joueurs dépasse deux, des situations de ce genre se présentent nécessairement. Et l'on est ainsi conduit à étudier, à côté de la *lutte*, la *coopération* et les *alliances*.

L'étude des jeux où interviennent à la fois la coopération et la lutte est beaucoup plus complexe que celle du duel. Elle a été particulièrement approfondie par J. von Neumann et O. Morgenstern dans le cas des *problèmes de partage*, où il s'agit, schématiquement, de partager une somme d'argent donnée entre plusieurs participants. Indiquons seulement que la discussion d'un problème de partage repose sur l'analyse des « *préférences efficaces* » dont disposent respectivement les participants et leurs diverses alliances (pour s'opposer efficacement à toute diminution de leurs parts respectives, à partir de certains partages supposés réalisés), et sur l'étude des « *relations d'exclusion* » qui en résultent dans l'ensemble des différents partages *a priori* possibles. On est alors conduit à définir, comme étant les « *solutions* » du problème, des ensembles d'éventualités tels que chacun d'eux ne contient plus aucune

---

(14) A. WALD, *Sequential Analysis* (1947), *Statistical decision functions* (1950), John Wiley and Sons, New-York, Chapman and Hall, London.

On trouvera un exposé complet, mais assez difficile, des travaux de Wald dans D. BLACKWELL et M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, 1954, mêmes éditeurs.

exclusion (stabilité interne), mais ne laisse en dehors de lui aucune éventualité qu'il n'exclue (stabilité externe). Il peut d'ailleurs exister plusieurs types de solutions, et il est intéressant de les rapprocher de certains types bien définis de situations économiques ou sociales.

(Cette théorie généralise celle de *Pareto* (15) qui se contente de définir un ensemble de situations « extrêmes » par le jeu des « préférences unanimes », sans chercher à introduire l'efficacité des diverses préférences individuelles).

L'analyse détaillée des solutions n'a été achevée que dans le cas de trois participants, car elle présente déjà, dans le cas de quatre participants, des difficultés pratiques dues à la grande diversité des cas possibles. Nous devons nous borner ici à ces indications sommaires et renvoyer le lecteur à l'ouvrage fondamental de von Neumann et Morgenstern, à certains traités sur la théorie des jeux, ainsi qu'aux leçons et à un article de G.-Th. Guilbaud (16).

Nous voulons seulement présenter ici quelques remarques générales à propos des *jeux à deux joueurs et à « somme non nulle »*, que l'on rencontre d'ailleurs en étudiant les « solutions dégénérées » des problèmes de partage à trois participants.

Supposons, pour simplifier, que chacun des résultats possibles du jeu tactique corresponde à un règlement monétaire certain, dans lequel chaque joueur donne ou reçoit une somme d'argent, mais sans que le gain réalisé par l'un des joueurs soit nécessairement compensé par une perte équivalente de l'autre joueur (c'est en cela que les intérêts des deux joueurs ne sont plus exactement opposés). Cela suppose implicitement l'existence d'un troisième joueur tout à fait passif, la « Nature » ou la « Banque », chargé de payer ou d'encaisser les différences. On peut imaginer des situations économiques de cette espèce ; et, dans le conflit de deux pays en guerre, les prélèvements opérés par la « Nature » ne sont que trop évidents.

Si les joueurs ignorent la structure du jeu dans lequel ils sont engagés, s'ils ignorent en particulier les intérêts de l'adversaire, chacun d'eux peut estimer prudent de jouer comme s'il était engagé dans un duel, ce qui lui garantira un certain gain, ou une certaine espérance mathématique de gain. Pourtant, il agit ainsi comme si son adversaire défendait habilement des intérêts exactement opposés aux siens ; et l'on conçoit déjà qu'il puisse avoir avantage à modifier sa ligne de conduite en tenant compte des préférences effectives de son adversaire, même s'il n'a pas la possibilité de se concerter avec lui. Cependant, il est souvent délicat de s'engager dans cette voie (nous le vérifierons plus loin à propos d'un exemple un peu différent) ; car une telle tentative de coopération tacite peut échouer et rendre le joueur très vulnérable.

---

(15) Voir, par exemple, l'*Encyclopédie des sciences mathématiques*, publiée sous la direction de Jules Molk, tome I, volume 4, fascicule 4, exposé I, 26 : « Economie mathématique », par Vilfredo PARETO (p. 624).

(16) G. Th. GUILBAUD, « Les problèmes de partage », *Economie appliquée*, 1952, n° 1, p. 93-137, et *op. cit.*, chap. V.

J. C. C. Mc KINSEY, *op. cit.*, chap. 15, 16 et 17.

R. D. LUCE et H. RAIFFA, *op. cit.*, chap. 6 à 9.

Mais, si les deux joueurs ont la possibilité de se concerter, ils ont, en principe, intérêt à s'entendre pour obtenir que la Banque leur verse globalement une somme aussi élevée que possible. Il leur est facile de trouver les tactiques qui leur permettent d'y parvenir, puisqu'il leur suffit de considérer la somme algébrique de leurs gains respectifs pour tous les couples de tactiques dont ils disposent. Rien ne peut d'ailleurs les empêcher de maximiser cette somme, s'ils sont d'accord pour le faire, puisque la Banque reste passive.

(Il peut arriver que, dans ces conditions, chacun des deux joueurs gagne plus, d'après la règle du jeu, que s'il n'avait pas accepté l'alliance ; mais il se peut aussi que, d'après cette règle, l'alliance soit favorable à l'un des joueurs et défavorable à l'autre. Dans tous les cas, l'alliance est globalement avantageuse pour le couple des deux joueurs, qui en tire, pour exploiter la Banque, un surcroît d'efficacité).

Cependant, si les deux joueurs actifs ont réussi à s'entendre pour obtenir de la Banque un versement global maximum, *il leur reste à régler leurs comptes entre eux*. Chacun d'eux peut, à bon droit, prétendre à ce qu'il avait le moyen de s'assurer en jouant individuellement (ce qui ne suffit d'ailleurs pas à garantir le respect de ce *bon droit*) ; mais comment répartir le surplus, le « *boni* » obtenu grâce à l'alliance ? C'est alors que se retrouve, à l'état pur, l'opposition des deux joueurs, et l'on sait bien que ce genre de partage (la redistribution interne des gains entre les alliés) peut prêter à de sérieuses contestations.

Aussi est-il prudent de fixer les conditions de ce partage avant de jouer la partie ; mais cela peut demander un *marchandage serré*, d'autant plus délicat que les moyens de persuasion des deux antagonistes sont souvent des *menaces* qu'ils n'ont pas intérêt à exécuter. Les conditions d'un éventuel *compromis* dépendent donc beaucoup de la psychologie des deux joueurs. Encore faudra-t-il s'assurer de la loyauté de l'adversaire par des *garanties* convenables, si celui-ci a un *intérêt personnel* à ne pas respecter les conditions de l'alliance, soit dans sa manière de jouer, soit dans le partage ultérieur. Un joueur loyal qui négligerait ces garanties risquerait de se voir gravement lésé.

L'élaboration d'une telle entente est donc souvent fort délicate. Si les joueurs ont des exigences incompatibles, ou s'ils se défient l'un de l'autre au point de suspecter toutes les garanties possibles, il y a de fortes chances pour qu'ils ne parviennent pas à s'entendre et renoncent à l'exploitation maximum de la Banque. Il pourrait même arriver qu'un joueur, se trouvant lié par les menaces qu'il a faites, soit alors conduit à jouer contre son intérêt personnel, en même temps que contre celui de son adversaire, pour ne pas perdre la face. Pour éviter d'en venir à un tel résultat, qui les désavantage tous les deux, peut-être les joueurs envisageront-ils de choisir un *arbitre* impartial qui puisse leur proposer les termes d'un accord et en contrôler l'exécution loyale. Mais les modalités d'un tel arbitrage sont elles aussi délicates à préciser ; et il n'est pas sûr que les décisions de l'arbitre soient acceptées par les joueurs (17).

---

(17) R. D. LUCE et H. RAIFFA (*op. cit.*) étudient avec grand soin des situations de ce genre.

On entrevoit, d'après cet exemple élémentaire, le rôle essentiel que peuvent jouer la coopération et les alliances, mais aussi la complexité des questions qu'elles soulèvent dans les jeux où elles interviennent. De nombreux exemples historiques confirment la validité des indications qui viennent d'être données à ce sujet.

Signalons enfin qu'une partie importante de la théorie des jeux où intervient la coopération est l'étude des *décisions collectives* (votes et élections par exemple), qui éclaire précisément le mécanisme des alliances. L'essentiel est ici de dégager un schéma de finalité, c'est-à-dire d'analyser la *formation des systèmes de valeurs collectifs* ou, suivant une expression classique, la *naissance de l'intérêt général*. Mais on y rencontre certaines difficultés logiques, déjà signalées par *Condorcet*, *Laplace* et *Lacroix*, entre 1780 et 1820, retrouvées et profondément analysées par *K.-J. Arrow* en 1951. Nous faisons ici allusion à l'« *effet Condorcet* », suivant lequel la cohérence des jugements au sein de chaque opinion individuelle est insuffisante pour assurer la même cohérence aux jugements collectifs déduits des jugements individuels par l'application d'une règle tenant compte effectivement de plusieurs opinions (une règle majoritaire par exemple).

En conséquence, s'il n'existe aucune considération objective, aucune volonté d'accord qui puisse assurer à l'ensemble des diverses opinions subjectives une certaine harmonie interne, l'intérêt général ne peut naître d'une simple combinaison des intérêts particuliers, où l'autonomie de chacun serait respectée. A moins que la définition de l'intérêt général ne soit remise au jugement d'un *arbitre*, c'est une *lutte* qui doit alors décider, avec les discussions, les marchandages, les menaces, les compromis qui, nous l'avons vu, en sont les épisodes (dès qu'il ne s'agit pas d'un duel). Mais ce qui en résultera ne saurait être considéré comme un jugement de valeur collectif.

Cette question délicate demanderait d'importants développements, pour lesquels nous renvoyons le lecteur aux travaux de *K.-J. Arrow*, au traité déjà cité de *R.-D. Luce* et *H. Raiffa*, ainsi qu'à un article de *G.-Th. Guilbaud* (18).

#### 4. 6. — Analyse d'une situation typique : *Sherlock Holmes* et *Moriarty*.

Avant de conclure, nous voudrions présenter un exemple typique, dont l'intérêt a déjà retenu l'attention de *O. Morgenstern* (voir *VON NEUMANN* et *MORGENSTERN*, *op. cit.*) et de *G.-Th. Guilbaud* (voir les leçons déjà citées, fin du chap. IV). A vrai dire, cet exemple n'entre exactement dans aucun des types de jeux que nous avons présentés jusqu'ici, bien qu'il puisse être rattaché à plusieurs d'entre eux. Il s'agit d'un jeu à deux joueurs, dont une seule partie est décisive, où les résultats ne sont pas réductibles à des règlements monétaires, et où il ne peut être question de

(18) *K.-J. ARROW*, *Social choice and individual values*, Cowles Commission Monograph 12, New-York, Wiley, 1951.

*G. Th. GUILBAUD*, « Les théories de l'intérêt général et le problème logique de l'agrégation », *Economie appliquée*, 1952, n° 4, p. 501-584.

*R.-D. LUCE* et *H. RAIFFA*, *op. cit.* chap. 14.

coopération concertée, bien que les intérêts des deux joueurs, exactement opposés dans le jeu tactique, ne le soient plus dans le jeu stratégique. (Nous avons déjà signalé la possibilité de telles différences entre les préférences personnelles des deux joueurs ; l'exemple que nous allons considérer nous éclairera sur leurs répercussions).

La situation est empruntée à *Conan Doyle* ; c'est la *lutte décisive entre Sherlock Holmes et Moriarty* : « the final problem ». Holmes, traqué par Moriarty, décide de se réfugier sur le continent ; il doit pour cela prendre le train qui va de Londres à Douvres, où il s'embarquera... si du moins Moriarty n'a pas réussi à le rejoindre, auquel cas Holmes serait tué. Holmes est dans le train qui, avant Douvres, s'arrêtera seulement à Canterbury ; il sait que Moriarty, averti de sa fuite, a le temps de le rejoindre, qu'il a même fait chauffer un train spécial. Que faire ? Holmes doit-il descendre à Canterbury, en escomptant que son ennemi le croira à Douvres (ce qui, sans le tirer d'affaire, lui donnerait un répit dont il pourrait peut-être profiter pour s'enfuir) ? Mais il se peut que Moriarty s'attende à cette ruse de Holmes et descende lui-même à Canterbury, auquel cas Holmes ferait mieux d'aller jusqu'à Douvres et de s'embarquer.

Etant donnée cette situation, comment Holmes et Moriarty, qui sont l'un et l'autre des joueurs habiles, et qui ne se sous-estiment pas mutuellement, doivent-ils raisonner et jouer ?

Essayons d'abord de mettre le jeu sous la « forme normale ». Il est clair que Holmes et Moriarty disposent chacun de deux *tactiques* : s'arrêter à Canterbury (tactique C), ou aller jusqu'à Douvres (tactique D). Aux quatre couples possibles de tactiques correspondent d'autre part les *résultats* suivants, désignés par les couples de lettres qui caractérisent les tactiques respectives de Holmes et de Moriarty :

- |   |   |
|---|---|
| { | (CC) et (DD) : mort de Holmes, victoire de Moriarty ; |
| { | (CD) : répit pour Holmes, déception pour Moriarty ;   |
| { | (DC) : fuite de Holmes, échec de Moriarty.            |

Les *ordres de préférence* de Holmes et de Moriarty, sur ces trois résultats distincts, sont absolument manifestes et exactement opposés.

Mais on voit aussitôt que *le duel tactique n'admet pas d'équilibre*. Il faut donc, pour passer au jeu stratégique, définir de plus les ordres de préférence respectifs de Holmes et de Moriarty sur les résultats aléatoires de ce jeu. Il est en particulier nécessaire de savoir classer le résultat moyen (CD) par rapport aux combinaisons aléatoires des résultats extrêmes (CC) ou (DD) d'une part, (DC) d'autre part, ce qui oblige à définir des « *équivalences morales* » de la forme suivante :

— *Holmes* considère comme équivalentes la situation (CD) qui lui procure un répit, et une situation aléatoire dans laquelle il aurait la probabilité  $\lambda$  de s'enfuir et la probabilité  $(1 - \lambda)$  d'être tué.

— *Moriarty* considère comme équivalentes la situation (CD) qui le déçoit dans son désir de tuer Holmes, et une situation aléatoire dans laquelle il aurait la probabilité  $\mu$  de laisser s'enfuir Holmes et la probabilité  $(1 - \mu)$  de le tuer.

La logique paraît d'ailleurs imposer que toute combinaison aléatoire

de résultats équivalents leur soit équivalente, et qu'une combinaison aléatoire quelconque soit équivalente à toutes celles que l'on peut obtenir en y remplaçant un résultat par un résultat équivalent. On en déduit que les deux équivalences précédentes suffisent à définir entièrement les ordres de préférence respectifs de Holmes et de Moriarty (voir, à ce sujet, G.-Th. Guilbaud, *op. cit.*, par. 18). (Il n'est cependant pas sûr que la réalité psychologique soit aussi rigide).

On peut alors faire correspondre aux trois résultats (CC) ou (DD), (CD) et (DC) du jeu tactique les indices de préférence :

0,  $\lambda$  et 1 (avec  $0 < \lambda < 1$ ) pour Holmes,

0,  $\mu$  et 1 (avec  $0 < \mu < 1$ ) pour Moriarty,

et faire correspondre à tout résultat aléatoire du jeu stratégique les deux indices de préférence (un pour Holmes et un pour Moriarty) obtenus à partir de ceux-là par un calcul d'espérance mathématique (la signification de ces indices sera précisée dans la suite).

Le jeu stratégique ainsi défini diffère d'un duel si  $\lambda$  diffère de  $\mu$ . Et il faut bien reconnaître que rien ne permet de supposer que  $\lambda$  soit égal à  $\mu$  ; car les valeurs des paramètres  $\lambda$  et  $\mu$  dépendent essentiellement des attitudes psychologiques respectives de Holmes et de Moriarty devant le risque :  $\lambda$  croît avec la prudence de Holmes, tandis que  $\mu$  croît avec la hardiesse de Moriarty. *Le duel tactique entre Holmes et Moriarty donne donc naissance à un jeu stratégique qui, en général, n'est pas un duel.*

En revanche, *ce jeu stratégique constitue un modèle satisfaisant de la situation de conflit réelle*, car il a suffi de définir deux équivalences morales, qui n'ont rien d'arbitraire, qui semblent s'imposer logiquement (sinon psychologiquement), pour que les ordres de préférence respectifs de Holmes et de Moriarty en résultent d'une manière logique. Et ces ordres peuvent se représenter par deux indicateurs de préférence calculés comme des espérances mathématiques.

(Nous vérifions ainsi, sur cet exemple, les indications générales qui ont été données à propos de l'« *espérance morale* ». Si le lecteur conteste l'exactitude psychologique du modèle ici obtenu, nous sommes persuadés qu'il n'a pas tort. Nous croyons cependant que ce modèle permet une analyse utile de la situation, en vue d'une décision dont, naturellement, Holmes et Moriarty restent les seuls maîtres. D'ailleurs, leur psychologie est sans doute assez imprégnée de logique mathématique pour que notre modèle rende un compte presque exact de leur attitude devant le risque !).

Nous représenterons le jeu en question par la matrice suivante :

		Moriarty	
		(1-q)	(q)
Holmes		(p)	C
		0	1
Holmes		(1-p)	D
		1	0

où figurent les indices de préférence correspondant aux résultats du jeu tactique, la case du résultat (CD), divisée en deux parties, contenant à la fois l'indice de préférence  $\lambda$  de Holmes et l'indice de préférence  $\mu$  de Moriarty.

Si Holmes choisit la stratégie ( $p$ ) définie par l'attribution des probabilités respectives  $p$  et  $(1-p)$  aux deux tactiques C et D, et si Moriarty choisit la stratégie ( $q$ ) définie par l'attribution des probabilités respectives  $q$  et  $(1-q)$  aux tactiques D et C, le résultat aléatoire correspondant est classé par Holmes et par Moriarty suivant les valeurs qu'il donne à leurs indicateurs de préférence respectifs H et M (nous les mettons sous deux formes équivalentes qui aideront le lecteur à comprendre la suite) :

$$H = (1-p)(1-q) + \lambda pq = \frac{\lambda}{1+\lambda} + \frac{1}{1+\lambda} [1 - (1+\lambda)p] [1 - (1+\lambda)q];$$

$$M = (1-p)(1-q) + \mu pq = \frac{\mu}{1+\mu} + \frac{1}{1+\mu} [1 - (1+\mu)p] [1 - (1+\mu)q].$$

Rappelons que les indicateurs de préférence H et M *repèrent* (mais ne mesurent pas, ce qui serait dépourvu de sens) les valeurs que possèdent les différents résultats aléatoires du jeu stratégique aux yeux de Holmes et de Moriarty. De façon plus précise, un résultat qui donne aux indicateurs H et M les valeurs respectives  $h$  et  $m$  est considéré *par Holmes* comme équivalent à une situation aléatoire dans laquelle il aurait la probabilité  $h$  de s'enfuir et la probabilité  $(1-h)$  d'être tué, et *par Moriarty* comme équivalent à une situation aléatoire dans laquelle il aurait la probabilité  $m$  de laisser s'enfuir Holmes et la probabilité  $(1-m)$  de le tuer.

En conséquence, Holmes désire que l'indicateur H prenne une valeur aussi grande que possible, mais Moriarty désire que l'indicateur M prenne une valeur aussi petite que possible. On peut donc considérer, avec les notations adoptées, que *Holmes est le joueur du maximum* et que *Moriarty est le joueur du minimum*.

Le jeu dans lequel sont engagés Holmes et Moriarty étant ainsi parfaitement défini, intéressons-nous à leurs *manières de raisonner et de jouer*.

On peut d'abord estimer que, si Holmes, par exemple, ignore jusqu'à quel point Moriarty risque de pousser la prudence ou la hardiesse, c'est-à-dire s'il ignore la valeur de  $\mu$ , il agira sagement en jouant comme s'il était engagé dans un duel, c'est-à-dire de manière à assurer à son indicateur de préférence H une valeur minimum aussi grande que possible. Ce point de vue le conduit à choisir la stratégie ( $p$ ) qui rend H indépendant de  $q$ , soit  $p = \frac{1}{1+\lambda}$ , ce qui lui assure  $H = \frac{\lambda}{1+\lambda}$  quel que soit  $q$ .

*Cette valeur de l'indicateur H constitue l'« espérance » de Holmes.*

De même, si Moriarty ignore la valeur de  $\lambda$ , on peut estimer qu'il agira sagement en jouant comme s'il était engagé dans un duel, c'est-à-dire de manière à assurer à son indicateur de préférence M une valeur maximum aussi petite que possible. Ce point de vue le conduit à choisir

la stratégie ( $q$ ) qui rend  $M$  indépendant de  $p$ , soit  $q = \frac{1}{1 + \mu}$ ,

ce qui lui assure  $M = \frac{\mu}{1 + \mu}$  quel que soit  $p$ .

*Cette valeur de l'indicateur  $M$  constitue l'« espérance » de Moriarty.*

Mais Holmes et Moriarty sont ennemis de longue date et s'apprécient à leur juste valeur. Il est vraisemblable qu'ils connaissent mutuellement leur prudence et leur hardiesse. Supposons donc qu'ils soient l'un et l'autre capables d'estimer les valeurs numériques de  $\lambda$  et de  $\mu$ , et supposons, pour fixer les idées, que  $\lambda$  soit inférieur à  $\mu$  (ce qui est le cas si Holmes et Moriarty acceptent le risque hardiment).

Chacun d'eux peut alors constater que *les stratégies prudentes indiquées par la théorie du duel ne conduisent pas ici à un équilibre*. Si Moriarty choisissait  $q = \frac{1}{1 + \mu}$ , inférieur par hypothèse à  $\frac{1}{1 + \lambda}$ , Holmes aurait intérêt, pour augmenter  $H$ , à choisir  $p = 0$ , c'est-à-dire à prendre la décision non aléatoire d'aller jusqu'à Douvres.

De même, si Holmes choisissait effectivement  $p = \frac{1}{1 + \lambda}$ , supérieur par hypothèse à  $\frac{1}{1 + \mu}$ , Moriarty aurait intérêt, pour diminuer  $M$ , à choisir  $q = 0$ , c'est-à-dire à prendre la décision non aléatoire de descendre à Canterbury.

Mais, bien entendu, aucun d'eux ne peut s'arrêter à une décision de cet ordre, car il craindrait que son adversaire réussisse à reconstituer le raisonnement par lequel il y serait parvenu. Sa hardiesse ne va pas jusqu'à accepter témérairement un tel risque.

Pendant, *il existe un équilibre dans le jeu stratégique*. Il suffit, pour le trouver, de considérer les réponses optima de chacun des joueurs aux diverses stratégies de l'autre, ce qui conduit aux résultats suivants et à la représentation graphique qui les accompagne.

Tant que Moriarty choisit  $q < \frac{1}{1 + \lambda}$ , la réponse optimum de Holmes est  $p = 0$ ; elle devient  $p = 1$  quand Moriarty choisit  $q > \frac{1}{1 + \lambda}$ .

Tant que Holmes choisit  $p < \frac{1}{1 + \mu}$ , la réponse optimum de Moriarty est  $q = 1$ ; elle devient  $q = 0$  quand Holmes choisit  $p > \frac{1}{1 + \mu}$ .

*Il y a donc équilibre si Holmes et Moriarty choisissent respectivement*

$$p = \frac{1}{1 + \mu} \quad \text{et} \quad q = \frac{1}{1 + \lambda}.$$

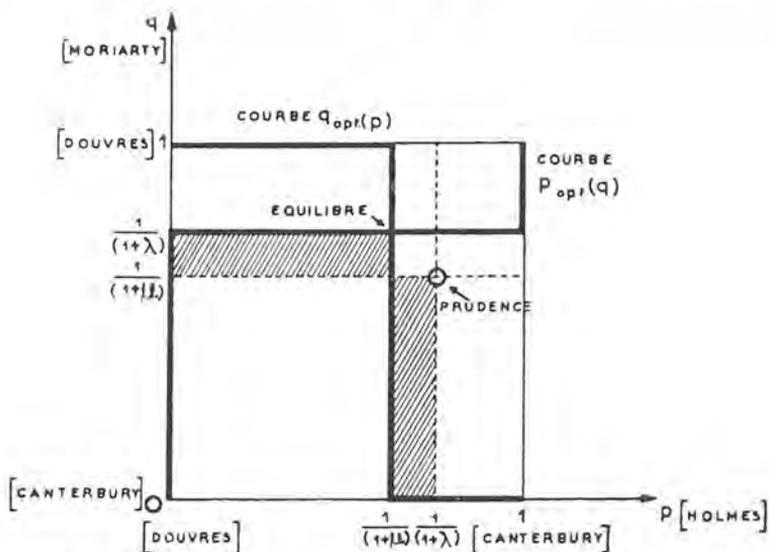
Le résultat aléatoire correspondant donne à leurs indicateurs de préférence respectifs les valeurs

$$H = \frac{\lambda}{1 + \lambda} \quad \text{et} \quad M = \frac{\mu}{1 + \mu} ;$$

ce sont précisément les valeurs qu'assuraient les stratégies prudentes indiquées par la théorie du duel.

Toutefois, le choix  $p = \frac{1}{1 + \mu}$  n'assure pas à Holmes la valeur  $H = \frac{\lambda}{1 + \lambda}$ , car H décroît alors de  $\frac{\mu}{1 + \mu}$  à  $\frac{\lambda}{1 + \mu}$  quand  $q$  varie de 0 à 1.

De même, le choix  $q = \frac{1}{1 + \lambda}$  n'assure pas à Moriarty la valeur  $M = \frac{\mu}{1 + \mu}$ , car M croît alors de  $\frac{\lambda}{1 + \lambda}$  à  $\frac{\mu}{1 + \lambda}$  quand  $p$  varie de 0 à 1.



N.-B. — La figure est relative au cas  $\lambda < \mu$ , avec  $\lambda = \frac{1}{3}$  et  $\mu = \frac{3}{5}$

Ainsi, le fait que le jeu stratégique considéré ne soit pas un duel (la différence entre  $\lambda$  et  $\mu$ ) a ici pour effet de dissocier les stratégies correspondant à l'équilibre de celles qui assurent respectivement à chacun des joueurs un résultat aléatoire au moins équivalent à son espérance, suivant l'ordre de ses préférences. (Nous avons reporté, sur la figure précédente, un point ayant pour coordonnées les valeurs correspondantes de  $p$  et  $q$ . Il y a lieu de remarquer que l'on obtient ces valeurs en échangeant celles qui correspondent à l'équilibre).

Dans ces conditions, *comment Holmes et Moriarty doivent-ils jouer ?* La question est rendue délicate par la dissociation que nous venons de mettre en évidence.

Si chacun des joueurs « joue l'équilibre », aucun d'eux n'aurait lieu de regretter son choix s'il connaissait celui de son adversaire, car il constaterait alors l'impossibilité où il se trouve d'obtenir un résultat préférable par sa seule action (à vrai dire, son propre choix est alors sans effet sur son indicateur de préférence) ; d'ailleurs, l'équilibre est, à son point de vue, équivalent à son espérance. Cependant, un joueur qui joue l'équilibre ne peut être sûr que son adversaire le jouera, de sorte que cette stratégie ne lui assure pas son espérance ; elle lui garantit seulement que son indicateur de préférence restera compris dans un certain intervalle encadrant son espérance, d'autant plus étroit que  $\lambda$  et  $\mu$  sont plus voisins. En revanche, un joueur qui joue l'équilibre assure par là-même à son adversaire un résultat aléatoire strictement équivalent à son espérance (mais cette considération n'agit pas directement sur ses propres préférences).

D'autre part, un joueur qui « joue la prudence », c'est-à-dire un joueur qui choisit la stratégie indiquée par la théorie du duel, s'assure un résultat équivalent à son espérance, contre toute défense de son adversaire. En revanche, il laisse à son adversaire une certaine possibilité de manœuvre (mais cette considération n'agit pas directement sur ses propres préférences).

On peut encore observer que, si Holmes joue l'équilibre, tandis que Moriarty joue la prudence, l'indicateur H prend la valeur  $\frac{\mu^2 + \lambda}{(1 + \mu)^2}$ , supérieure à  $\frac{\lambda}{1 + \lambda}$ , donc préférée par Holmes à son espérance, dès que  $\mu$  diffère de  $\lambda$ .

En revanche, si Moriarty joue l'équilibre tandis que Holmes joue la prudence, l'indicateur M prend la valeur  $\frac{\lambda^2 + \mu}{(1 + \lambda)^2}$ , supérieure à  $\frac{\mu}{1 + \mu}$ , donc surclassée aux yeux de Moriarty par son espérance, dès que  $\lambda$  diffère de  $\mu$ .

*Est-ce à dire que Holmes doit jouer l'équilibre et Moriarty la prudence ?* On peut trouver raisonnable une telle manière de jouer, mais elle ne s'impose pas, et à vrai dire, aucune des manières de jouer que nous avons envisagées ne s'impose. La raison profonde de cet état de choses est qu'il y aurait théoriquement place, dans le jeu stratégique, pour une coopération entre Holmes et Moriarty ; car certains couples de stratégies donnent un résultat aléatoire qui serait préféré par chacun des joueurs à celui qu'il peut espérer.

(Il est facile de vérifier qu'il en est ainsi si l'une des probabilités  $p$  et  $q$  est comprise entre  $\frac{1}{1 + \lambda}$  et  $\frac{1}{1 + \mu}$ , l'autre étant comprise entre 0 et  $\frac{1}{1 + \mu}$  si  $\lambda < \mu$ , entre  $\frac{1}{1 + \mu}$  et 1 si  $\lambda > \mu$ . Nous avons hachuré les régions correspondantes sur la figure, qui est relative au cas  $\lambda < \mu$ . Le lecteur n'aura aucune peine à faire, s'il le désire, la figure relative au cas  $\lambda > \mu$ ).

Il faudrait donc qu'un marchandage puisse s'instaurer entre Holmes et Moriarty pour qu'une véritable solution du jeu puisse se dégager, sur des bases qui dépendraient d'ailleurs de l'habileté avec laquelle chacun d'eux défendrait ses préférences (chacun pourrait éventuellement menacer l'autre de lui imposer son espérance en jouant l'équilibre).

On voit aisément que Holmes et Moriarty auraient tous les deux intérêt, d'après leurs préférences respectives, à s'entendre pour que l'une des probabilités  $p$  et  $q$  soit égale à 0 si  $\lambda < \mu$ , et à 1 si  $\lambda > \mu$ , l'autre étant comprise entre  $\frac{1}{1+\lambda}$  et  $\frac{1}{1+\mu}$ . (Chacun des résultats aléatoires correspondants est « extrême » au sens de Pareto). Mais c'est dans la fixation de cette dernière probabilité qu'un marchandage difficile devrait intervenir ; et l'on imagine le danger que courrait, par exemple, Sherlock Holmes s'il se laissait convaincre de descendre à Douvres ( $\lambda < \mu$ ) ou à Canterbury ( $\lambda > \mu$ ) sans avoir de solides garanties sur la loyauté avec laquelle jouerait alors Moriarty !).

Si, comme on le conçoit, un tel marchandage et un tel compromis sont impossibles, nous aurions tendance, pour notre part, à préférer jouer la prudence, mais c'est une affaire de goût. *Il n'existe, en fait, pas de manière optimum de jouer le jeu si toute possibilité de coopération en est exclue.* Nous ne croyons cependant pas qu'il soit vain, pour Holmes et pour Moriarty, d'analyser la situation comme nous venons de le faire.

Il est intéressant de remarquer qu'au début du jeu, l'espérance de Holmes est  $\frac{\lambda}{1+\lambda} < \frac{1}{2}$  ; cette espérance est d'autant plus grande que Holmes est plus prudent, mais elle est en tout cas moins favorable que s'il avait une chance sur deux de s'enfuir. En contrepartie, l'espérance de Moriarty est  $\frac{\mu}{1+\mu} < \frac{1}{2}$ , ce qui veut dire qu'elle est en tout cas plus favorable que si Moriarty avait une chance sur deux de laisser Holmes s'enfuir.

*Holmes est donc désavantagé par les conditions initiales du jeu.* (Toutefois, il ne lui serait pas interdit, en principe, d'essayer de retourner la situation par un marchandage habile, dans le cas  $\lambda > \mu$ , en obtenant que Moriarty lui laisse une espérance voisine de  $\frac{\lambda}{1+\mu}$ . En revanche, dans le cas  $\lambda < \mu$ , Holmes ne pourrait obtenir mieux que  $\frac{\mu}{1+\mu}$ ).

Cependant, si Holmes et Moriarty jouent l'équilibre, ou la prudence, ou une stratégie intermédiaire (ce qui paraît raisonnable en l'absence de coopération),  $p$  et  $q$  sont certainement supérieurs à  $1/2$ , ce qui veut dire que le résultat le plus probable est (CD), c'est-à-dire le répit pour Holmes et la déception pour Moriarty. Dans le livre de Conan Doyle, Holmes descendait effectivement à Canterbury, tandis que Moriarty filait à toute vapeur jusqu'à Douvres !

## 5. — Conclusion.

Nous avons essayé de présenter la théorie des jeux en reprenant les grandes lignes de la démarche historique par laquelle elle s'est lentement constituée. Nous avons ainsi rencontré différents résultats théoriques, plus ou moins satisfaisants pour l'esprit, plus ou moins directement applicables à l'objet multiforme de notre réflexion — la décision et l'action humaines —, objet que nous nous sommes cependant efforcés de ne pas perdre de vue. Mais nous ne saurions prétendre que cet exposé soit complet, même sur l'essentiel. Tout au plus voulions-nous indiquer les thèmes principaux des recherches entreprises.

Le lecteur aura peut-être été déçu de ne pas trouver dans cet article l'exposé bien ordonné d'une théorie achevée. Le caractère fragmentaire des résultats évoqués, les questions laissées en suspens, les difficultés rencontrées dès que la situation analysée n'entre pas dans le cadre rigide du duel, lui auront peut-être donné l'impression que la théorie des jeux n'est pas à la hauteur de ses ambitions.

Peut-être y aurait-il plutôt lieu d'admirer qu'elle ait réussi à maintenir ces ambitions sans en être écrasée, que l'unité à laquelle elle prétend ne lui soit pas contestée, que ses succès dans l'étude de certaines actions collectives aient suffi pour faire admettre la fécondité d'une collaboration si étendue entre les Mathématiques et les Sciences humaines, que la théorie du duel enfin ait pu analyser de façon si profonde et si complète l'affrontement de deux volontés.

Il n'est pas question, cependant, de dissimuler qu'il reste beaucoup à faire, particulièrement en ce qui concerne la coopération et les alliances, où les résultats obtenus sont déjà très prometteurs. Mais si, dans ce jeu contre la Nature, la partie n'est jamais achevée, il ne paraît pas téméraire de miser sur l'habileté de ceux qui s'y exercent. Peut-être aussi n'est-il pas si naïf de s'intéresser à leur jeu ; peut-être même ne perdrait-on pas son temps en coopérant à cette action collective, si l'on en a le goût et les moyens, car il semble qu'ici les intérêts particuliers puissent s'accorder, et se fondre dans un intérêt général.



# LA STRUCTURE MATHÉMATIQUE SIMPLE DES GRANDES MACHINES À CALCULER

par J. VILLE

*Professeur à la Faculté des Sciences de Paris*

L'apparition, puis le développement des grandes calculatrices électroniques offrent aux Mathématiques des possibilités nouvelles, mais posent également des exigences nouvelles aux mathématiciens.

Les possibilités sont connues ; la plus commune est la rapidité et le bon marché avec lesquels on peut procéder à des opérations classiques, ce qui permet d'entreprendre des calculs auxquels on aurait jadis renoncé faute de temps et de moyens. Plus une machine va vite, plus elle remplace de calculateurs, et par conséquent plus elle est économique. Mais ce n'est pas tout ; à supposer qu'une machine travaille au même prix que des calculateurs manuels, si elle en remplace un grand nombre, elle rend possibles des calculs qui restaient irréalisables parce que le travail simultané d'un grand nombre de calculateurs manuels sur un même problème posait des problèmes d'organisation presque insurmontables. Une autre ressource des machines est apparue lorsque l'on a constaté que l'on pouvait leur faire traiter des problèmes non classiques ; ce fut le cas des problèmes où interviennent de très nombreuses inégalités ; lorsqu'on essaie de résoudre par un calcul « à la main », on est surpris par le développement qu'il prend, tellement il est compliqué.

Parlons maintenant des exigences qui vous intéressent particulièrement, puisque vous aurez sans doute à préparer des élèves qui devront y satisfaire, même si peu d'entre vous sont appelés, en tant que professeurs, soit à utiliser directement les machines, soit à être les bénéficiaires de leurs possibilités. Ces exigences sont dues à la structure des machines, structure que le titre de cet exposé annonce comme mathématiquement simple, mais qui, justement à cause de cette simplicité, requiert d'assez grands efforts pour être mise effectivement en œuvre.

Le premier principe de structure est que les machines ne savent actuellement travailler qu'en système binaire, c'est-à-dire dans le système de numération à base deux où chaque entier  $n$  est mis sous la forme :

$$n = \sum \alpha_k 2^k \quad \text{où } \alpha_k = 0 \text{ ou } 1 \text{ et } k \in \mathbb{N}.$$

Second principe de structure : chaque chiffre binaire est représenté par une impulsion électrique, c'est-à-dire par un courant pendant un temps très court. Si le courant est présent, cela représente le chiffre 1 ; s'il est absent, cela représente le chiffre 0.

Ces deux premiers principes n'ont pas, en eux-mêmes, de grandes conséquences. Mais un troisième principe intervient alors : un chiffre binaire ne peut se trouver qu'en deux états : soit *en circulation*, pour être traité par le calcul, soit *enregistré*, pour attendre d'être traité.

Ce dernier principe, qui paraît une évidence, a des conséquences importantes, parce qu'il fait intervenir une notion nouvelle, et même entièrement nouvelle, celle d'*adresse*. En effet, si nous admettons qu'une opération doit intervenir, telle que :

$$(1) \quad a + b = c,$$

l'essentiel n'est pas de connaître les valeurs numériques de  $a$  et  $b$ , puisque la machine sait faire l'addition, mais de savoir où il faut aller chercher  $a$  et  $b$ .

Pour comprendre ce qui va suivre, il faut insister sur la signification de l'égalité (1), qui paraît tout à fait simple. Elle signifie en réalité : « J'ai appelé  $a$  un certain nombre ; j'ai appelé  $b$  un certain nombre ; je calcule leur somme ; cette somme, je l'appelle  $c$ . »

Or, pour une machine numérique, il n'existe que des nombres. Donc, la façon dont on appelle un nombre ne peut être qu'un nombre. Ceci est d'ailleurs beaucoup plus souple. En effet, s'il était imposé de représenter des nombres par des lettres toujours de la même façon, un traité d'algèbre ne pourrait porter que sur des nombres en nombre égal à celui des lettres de l'alphabet. A partir du moment où l'on fait intervenir des indices, on commence à entrer dans un système de numération. Ces faits passent souvent inaperçus parce que, dans les Mathématiques classiques, l'esprit complète l'exposé. Par exemple, la phrase : « Je considère la suite des  $x_i$ ,  $i$  variant de 1 à  $n$  », est parfaitement claire, mais pour une machine elle est terriblement compliquée. Elle serait également très compliquée pour quelqu'un qui refuserait de faire l'effort d'interprétation qu'on lui demande, et qui exigerait qu'on l'exprime à partir d'axiomes logiques.

Pour en revenir au « nom » des nombres, le plus simple est de nommer un nombre par l'emplacement qu'il a dans la machine. Par exemple, l'égalité :

$$\langle 15 \rangle + \langle 37 \rangle = \langle 203 \rangle$$

signifie : le nombre qui est à l'emplacement numéro 203 doit être la somme du nombre qui est à l'emplacement n° 15 et du nombre qui est à l'emplacement n° 37. Ce n'est qu'une généralisation des procédés employés couramment pour établir les feuilles de calcul ou même les déclarations d'impôt ; on y lit : inscrire dans la colonne 4 la somme de ce qui est inscrit dans la colonne 1 et dans la colonne 2, ce qui peut donc se noter :

$$\langle 1 \rangle + \langle 2 \rangle = \langle 4 \rangle .$$

Une adresse est une notion assez facile à comprendre sur le schéma suivant, où nous voyons apparaître deux éléments :  $a$ ) un registre R ;  $b$ ) un élément arithmétique EA (voir fig. 1).

Sur le registre sont inscrits les éléments (les nombres). Chacun est à un emplacement. L'élément arithmétique sait faire un certain nombre d'opérations. La conduite de la machine consiste à prélever des couples

de nombres dans le registre (choix  $\beta$ ), à choisir l'opération (choix  $\gamma$ ), — après quoi celle-ci est effectuée —, et à choisir l'endroit où l'on place le résultat (choix  $\alpha$ ).

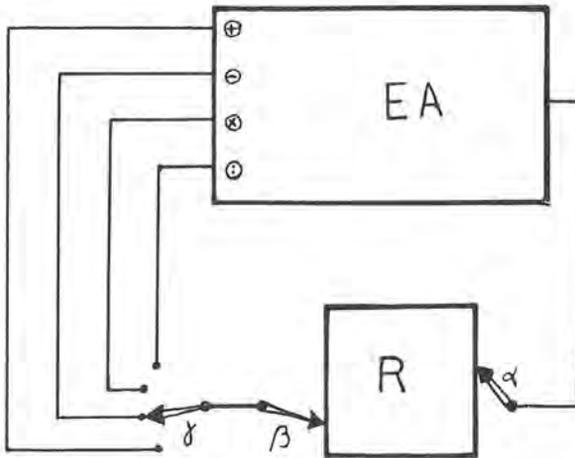


FIG. 1

Un ordre se présenterait ainsi sous la forme  $(\beta, \gamma, \alpha)$  et, pour un choix convenable des adresses des données, on peut toujours trouver une suite d'ordres qui amène le résultat à des adresses données. Si, par exemple, j'écris :

$$\begin{aligned} ax + by &= c \\ a'x + b'y &= c' \end{aligned}$$

sous la forme :

$$\begin{aligned} \langle 1 \rangle \langle 7 \rangle + \langle 2 \rangle \langle 8 \rangle &= \langle 3 \rangle \\ \langle 4 \rangle \langle 7 \rangle + \langle 5 \rangle \langle 8 \rangle &= \langle 6 \rangle . \end{aligned}$$

ce qui revient à mettre les données dans les cases numérotées de 1 à 6 et à porter les inconnues dans les cases de numéros 7 et 8, on peut trouver une chaîne de calculs et d'ordres amenant au résultat.

Ceci suppose que l'on adopte une méthode de calcul (c'est-à-dire un ordre d'opérations) et que l'on choisit à l'avance toutes les adresses. C'est là qu'interviennent les servitudes propres aux machines et les exigences relatives aux calculs. Il est un fait que l'on peut résoudre sur machine des systèmes de cent équations du premier degré à cent inconnues. On sait que le déterminant d'un tel système contient  $100!$  termes. Si rapide que soit une machine et si grande que soit la capacité de ses registres, il est impossible de calculer tous ces termes et de les enregistrer à des adresses différentes.

Il faut donc mener le calcul de la façon la plus économique, tant en nombre d'opérations qu'en nombre de symboles intermédiaires. Ces deux exigences font appel à des notions simples, mais parfois négligées. Par exemple, s'il s'agit de calculer la valeur numérique d'un polynôme :

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n,$$

pour une valeur  $\alpha$  de  $x$ , on constate que l'on a besoin de moins d'opérations en procédant à la division de  $f(x)$  par  $x - \alpha$  qu'en procédant à la substitution, et que l'on a moins de registres à utiliser. De plus l'opération se met sous forme séquentielle simple, en posant :

$$\begin{aligned} f_0 &= a_n \\ f_1 &= f_0\alpha + a_{n-1} \\ f_2 &= f_1\alpha + a_{n-2} \\ &\dots\dots\dots \\ f_n &= f_{n-2}\alpha + a_0 = f(\alpha). \end{aligned}$$

Nous avons procédé à  $n$  multiplications et  $n$  additions et nous n'avons eu besoin que de deux mémoires, car nous n'avons pas besoin de conserver les valeurs  $f_i$  pour  $i < n$  ; dès que  $f_i$  est calculé,  $f_{i-1}$  peut être effacé.

En ce qui concerne la résolution d'un système d'équations linéaires, la méthode que l'on peut utiliser consiste à tirer une inconnue d'une équation et à porter sa valeur dans les autres équations. Par exemple, le système :

$$\left\{ \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots\dots\dots \\ a_{n1}x_1 + \dots\dots\dots &= b_n \end{aligned} \right.$$

est ramené au système :

$$\left\{ \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ \left( a_{22} - \frac{a_{21}a_{12}}{a_{11}} \right) x_2 + \dots &= b_2 - \frac{a_{21}b_1}{a_{11}} \\ \dots\dots\dots \end{aligned} \right.$$

ce qui exige  $(n - 1)^2$  fois le calcul de termes tels que :

$$a_{ij} - \frac{a_{i1}a_{1j}}{a_{11}} ;$$

les quotients  $\frac{a_{ij}}{a_{11}}$  sont au nombre de  $n - 1$ . Nous aurons donc  $(n - 1)$  divisions,  $(n - 1)^2$  multiplications,  $(n - 1)^2$  soustractions, c'est-à-dire  $(n - 1)(2n - 1)$  opérations.

Pour les seconds membres, nous aurons une division  $\left(\frac{b_1}{a_{11}}\right)$ ,  $(n - 1)$  multiplications,  $(n - 1)$  soustractions, soit  $(2n - 1)$  opérations. Cette première étape de la réduction du système aura donc exigé  $n(2n - 1)$  opérations. La réduction totale ne fera donc intervenir que :

$$n(2n - 1) + (n - 1)(2n - 3) + \dots + 6 \sim \frac{n^3}{3} \text{ opérations.}$$

La remontée du procédé en fait intervenir la moitié. En tout, le système est résolu en  $\frac{n^3}{3}$  opérations. On est loin des  $n!$  termes du déterminant. Quant au déterminant lui-même, son calcul peut également se faire par récurrence.

Si l'on veut donner des ordres à la machine pour effectuer le calcul, et que l'on écrive *tous* ces ordres, on s'aperçoit que cela est compliqué et le choix des mémoires est également compliqué. Il y a évidemment beaucoup d'intermédiaires, leur nombre est du même ordre que celui des opérations. Mais si on écrit les relations de récurrence sous la forme :

$$a_{ij}^{(k)} = \frac{a_{ik}^{(k-1)} a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}}$$

$k$  désignant le numéro du pas de réduction, on remarque que les  $a_{ij}^{(k)}$  n'ont pas besoin d'être tous conservés, et que leur adresse naturelle est  $(i, j)$ . Quant aux ordres, ils ne dépendent pas de  $(k)$  en réalité, puisque l'opération est toujours la même ; ce qui dépend de  $k$ , c'est l'adresse des nombres à opérer. Et, d'un pas de réduction à l'autre, ces adresses varient d'une unité. On conçoit donc qu'à côté des opérations purement numériques, on ait à opérer sur des choses qui ne sont pas des nombres, mais des *noms* (écrits sous forme numérique), ces noms pouvant désigner des emplacements, l'emplacement étant occupé par un nombre proprement dit, un ordre, ou même le nom d'un autre emplacement.

Les ordres eux-mêmes peuvent ne pas être des ordres d'opération, mais des ordres à aspect logique. Pour calculer une suite convergente avec une certaine précision, par exemple, on pourra convenir de l'ordre suivant : « Si  $x_n = x_{n-1}$ , s'arrêter. Si  $x_n \neq x_{n-1}$ , passer à l'exécution de l'ordre  $(n + 1)$  ». C'est là qu'intervient une algèbre logique binaire, les alternatives étant simples, en général.

Ces choix sont assez rares dans les calculs classiques, mais ils peuvent intervenir à tout instant dans les calculs non classiques pour lesquels des machines très puissantes deviennent nécessaires. Parmi ces calculs, il faut citer ceux où interviennent des inégalités ; elles demandent des opérations de choix très nombreuses. Les systèmes d'inégalités linéaires de la forme

$$\sum a_{ij} x_j \leq b_i$$

sont encore très peu étudiés. Les méthodes classiques d'élimination conduisent à un nombre énorme d'opérations ; de plus, elles obligent à conserver des éléments trop nombreux.

L'emploi des machines contraint donc à des études spéciales portant principalement sur l'algèbre et où interviennent un très grand nombre de décisions. Ces problèmes d'algèbre sont élémentaires, mais ils n'ont pas encore été beaucoup étudiés.



# LES CALCULATEURS ELECTRONIQUES ET LE DEVELOPPEMENT DES MATHEMATIQUES NUMERIQUES

par François GENUYS

(Compagnie I.B.M.-France, Paris)

Le développement actuel des Mathématiques numériques n'a été possible qu'avec le développement des moyens de calcul. Le calcul à la main, avec ou sans machine de bureau, tend à être remplacé par le calcul automatique. Deux types de machines le réalisent :

a) *Les calculateurs analogiques* : les divers nombres considérés sont représentés par des grandeurs physiques (courant électrique, niveau d'eau...), et l'on utilise des lois physiques pour simuler les calculs à effectuer.

b) *Les calculateurs digitaux* : ceux-ci travaillent sur des nombres de développement limité dans une certaine base de numération.

Je donnerai d'abord quelques indications sur les calculateurs digitaux. Les divers nombres cités correspondent aux performances atteintes par les machines les plus puissantes.

## A. — DESCRIPTION D'UN CALCULATEUR DIGITAL

### 1° Codage de l'information :

Les constructeurs n'utilisent actuellement pour coder l'information que des dispositifs binaires, c'est-à-dire à deux états (courant ou pas courant, sens d'une magnétisation, présence ou absence de trou dans une carte), ceci pour des raisons essentiellement technologiques. Je désignerai par position binaire ou par bit (abréviation de l'anglais binary digit) tout atome d'information à deux états ; les deux états seront notés 0 et 1.

Au moyen de ces organes élémentaires, il faut pouvoir coder les nombres (les facteurs) sur lesquels opère la machine et les opérations (les instructions) que celle-ci devra exécuter. Souvent, ce codage est redondant, de manière à dépister des erreurs éventuelles : par exemple, le nombre de 1 dans le codage est pair ou, encore, un nombre est toujours accompagné de son reste modulo 7.

### 2° Représentation des nombres :

Si le système binaire est utilisé, il est naturel de représenter un chiffre au moyen d'un bit. En revanche, si l'on utilise le système décimal,

il faudra au moins 4 bits pour représenter un chiffre, mais, puisque l'on peut former  $2^4 = 16$  combinaisons avec 4 bits, cette représentation est redondante (certaines combinaisons seront à exclure). On utilise de nombreux codes ; la remarque que  $C_n^2 = 10$  a conduit à utiliser le code 5 dont 2 : code à cinq bits dont deux, et deux seulement, sont égaux à 1 ; par exemple :

Chiffre	Code
0 .....	0 0 0 1 1
1 .....	0 0 1 0 1
2 .....	0 0 1 1 0
3 .....	0 1 0 0 1
4 .....	0 1 0 1 0
5 .....	0 1 1 0 0
6 .....	1 0 0 0 1
7 .....	1 0 0 1 0
8 .....	1 0 1 0 0
9 .....	1 1 0 0 0

Le signe d'un nombre nécessite un bit. Pour coder un caractère alphabétique, il faut au moins 5 bits (32 possibilités).

Le codage d'un nombre sera la réunion des codages des chiffres formant ce nombre. On le fait soit en *virgule fixe*, soit en *virgule flottante*. En virgule fixe, les facteurs sont des nombres entiers ; l'utilisateur peut évidemment imaginer qu'une virgule est placée entre deux des chiffres, mais, celle-ci n'étant pas codée, il devra suivre avec soin la position de la virgule dans les diverses opérations. En virgule flottante, en revanche, les nombres sont mis la sous la forme :

$$B^n f \quad \text{avec } B^{-1} \leq f < 1$$

où B est la base de numération ; n s'appelle l'exposant et f la partie fractionnaire ; le codage d'un nombre comprend alors son exposant et sa partie fractionnaire, sur lesquels le calculateur opère automatiquement et simultanément au moyen d'instructions convenables. Exemple :

$$0,483 \times 10^{-2} - 0,472 \times 10^{-2} = 0,11 \times 10^{-3},$$

6

$$\text{ou } (-2,483) - (-2,472) = (-3,11).$$

La virgule flottante est d'un usage plus commode, mais elle est plus lente, plus onéreuse de réalisation, et nécessite plus de bits pour un nombre donné de chiffres significatifs ; son usage tend cependant à se généraliser.

Les calculs d'erreur sont différents suivant le mode employé ; ce sont souvent eux qui guident le choix.

### 3° Schéma général :

Un calculateur comporte :

a) *Une unité centrale*, qui contrôle le déroulement du « programme » et exécute les instructions ; elle comprend un certain nombre de registres où sont envoyés les résultats des opérations, où sont signalées cer-

taines anomalies telles que dépassement de capacité (trop de chiffres dans un résultat), etc...

b) *Des unités de mémoire*, qui emmagasinent une très grande quantité d'informations ; le calculateur peut « écrire » dans une mémoire et la lire sans intervention d'un opérateur humain.

c) *Des unités d'entrée et de sortie*, qui permettent au calculateur de recevoir de l'information extérieure (programme, données numériques), ou inversement de fournir de l'information à l'extérieur (résultats, renseignements permettant de reprendre un calcul interrompu).

Pour permettre au calculateur l'utilisation de la mémoire, on divise celle-ci en *mots* d'un nombre fixe de bits (par exemple 40) et l'on affecte à chaque mot un nombre que l'on appelle son *adresse*. On fait tenir dans un mot un nombre ou une instruction.

Le schéma de fonctionnement est alors le suivant. Lorsque le calculateur vient d'exécuter l'instruction d'adresse  $\alpha$ , il lui faut connaître l'adresse  $\beta$  de l'instruction suivante. On a adopté deux solutions :

a)  $\beta$  est indiqué dans l'instruction  $\alpha$ .

b)  $\beta = \alpha + 1$  ; on dit alors que la machine est *séquentielle*.

Mais, même dans une machine séquentielle, on doit pouvoir changer l'ordre d'exécution des instructions suivant le résultat de tests ; la solution *a* est employée alors pour certaines instructions (de *saut* ou de *transfert*).

Les premières instructions (à exécuter) d'un programme sont en général toujours les mêmes et sont conçues pour lire le programme (c'est-à-dire le mettre en mémoire) ; elles sont exécutées dès que l'on appuie sur le bouton « Départ ». A partir de ce moment, il n'y a plus besoin d'opérateur.

#### 4° Description des instructions :

L'ensemble des bits codant une instruction est décomposé en deux :

a) une partie sert à désigner la fonction de l'instruction (il peut y avoir plus d'une centaine d'instructions différentes) ;

b) d'autres parties indiquent les adresses des facteurs sur lesquels opèrent l'instruction et (ou) l'instruction à exécuter immédiatement après.

Selon leur nature, les instructions se répartissent en :

a) *Instructions arithmétiques* (binaires ou décimales suivant le calculateur) ; ce sera, par exemple, l'addition en virgule flottante d'un registre et d'une mémoire (résultat dans le registre). Les temps d'exécution peuvent atteindre quelques microsecondes.

b) *Instructions logiques* : exécution d'opérations logiques bit à bit, telles que *et*, *ou*, *non*,  $\Rightarrow$ , etc...

c) *Instructions de test* : leur rôle est de faire prendre des décisions au calculateur ; rechercher l'instruction suivante en  $\alpha$ ,  $\beta$  ou  $\gamma$  selon qu'un registre est négatif, nul, ou positif, etc...

d) *Instructions d'entrée-sortie* : ce sont celles qui relient le calculateur avec le monde extérieur : lecture d'une carte, etc...

### 5° Les mémoires :

D'une mémoire, il faut surtout retenir la *capacité* et le *temps d'accès*. Ce dernier est la durée qui sépare l'instant où un mot est demandé (par le calculateur) et celui où le mot est effectivement disponible pour le calcul.

a) *Mémoires à ferrites* : de petits anneaux de ferrites à deux états d'aimantation représentant un bit ; capacité de l'ordre de  $10^6$  bits ; temps d'accès de quelques microsecondes.

b) *Tambours magnétiques* : ce sont des tambours recouverts d'oxyde magnétique tournant à grande vitesse ; capacité de l'ordre de  $2 \cdot 10^5$  bits ; temps d'accès variable dépendant de la position du tambour quand un mot est demandé ; temps moyen de quelques millisecondes.

c) *Bandes magnétiques* : ce sont des bandes similaires à celles utilisées en enregistrement sonore, capacité de  $10^7$  bits ; le temps d'accès est variable, peut aller jusqu'à 3 minutes. Pratiquement, on classe l'information sur une bande, de manière à éviter les sauts d'un bout à l'autre de celle-ci.

d) *Disques magnétiques* : ce sont des disques empilés recouverts d'oxyde magnétique. La capacité est de  $10^7$  bits et le temps d'accès moyen de 0,5 seconde.

Le mathématicien doit prévoir une organisation judicieuse de son calcul, de manière à éviter les temps d'attente d'information.

### 6° Unités d'entrée-sortie :

a) *Carte perforée* : une carte contient  $12 \times 80 = 960$  positions perforables, donc 960 bits, en général utilisés avec grande redondance (une colonne de 12 bits pour un chiffre ou signe). La vitesse de lecture ou de perforation est de l'ordre de 4 000 bits/seconde.

b) *Ruban perforé* : vitesse de 2 400 bits/seconde.

c) *Bande magnétique* : outre leur fonction de mémoire, les bandes magnétiques peuvent servir d'entrée ou de sortie à une vitesse de  $10^5$  bits/seconde.

d) *Impression* : 500 lignes/minute ; directement consultable par l'homme, mais non exploitable postérieurement par une machine.

e) *Entrée-sortie analogique* : on a enfin imaginé des dispositifs permettant d'utiliser directement les résultats de mesure ou de fournir à l'utilisateur les résultats de calcul sous forme de courbe.

## B. — UTILISATION D'UN CALCULATEUR

Je n'étudierai que l'utilisation en calcul scientifique, laissant de côté les applications considérables dans le domaine de la gestion.

**1° L'organigramme :**

Un calcul à effectuer se compose d'une suite (dans le temps) d'opérations et de tests ; avant de procéder au codage, il faut faire un schéma détaillé de ceux-ci. C'est l'organigramme. Par exemple, si l'on veut résoudre  $ax^2 + bx + c = 0$ ,  $a, b$  et  $c$  étant lus par carte, l'organigramme sera le suivant :

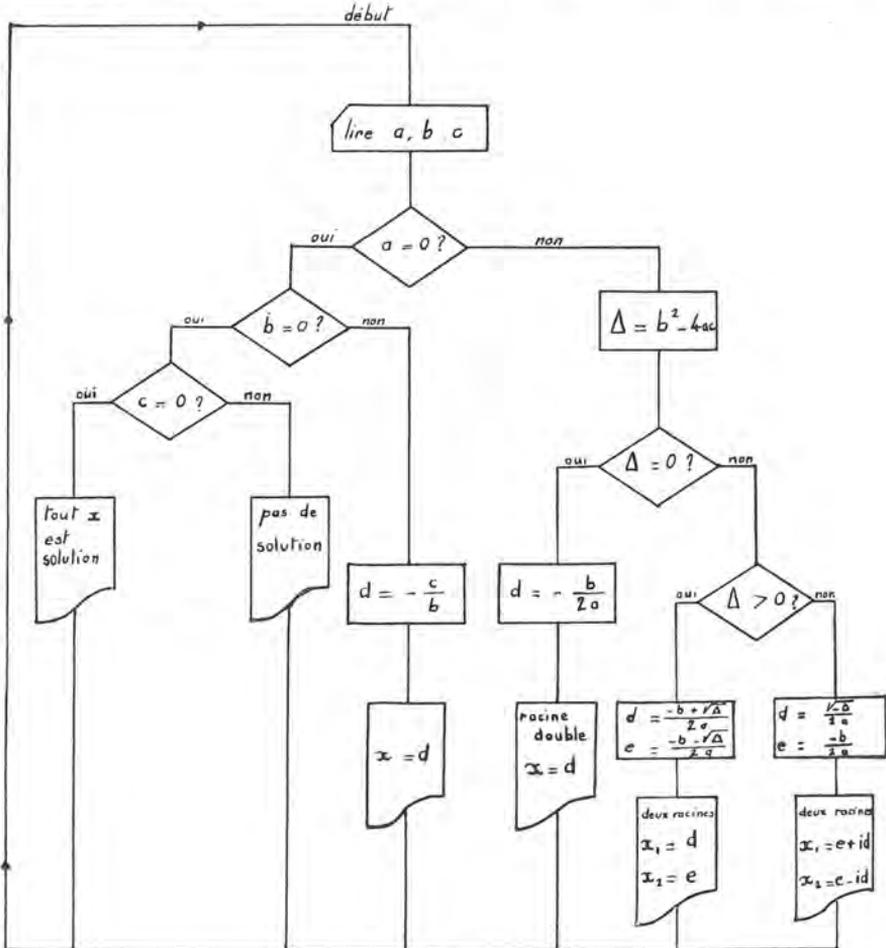


FIG. 2

Ici, la résolution d'une équation est suivie de la lecture de nouvelles données.

Cet organigramme n'est d'ailleurs pas le meilleur ; par exemple, il y aurait intérêt à calculer  $-\frac{b}{2a}$  avant de faire le test  $\Delta > 0$ , ceci afin de diminuer le nombre d'instructions.

Je crois que, même au niveau scolaire, la confection d'organigramme est un exercice excellent dans toutes les discussions de calcul ; il oblige

à étudier tous les cas et ne permet pas de laisser dans l'ombre le point délicat.

Remarquons enfin que l'organigramme doit prévoir également l'utilisation rationnelle des mémoires ; on peut très bien ainsi affecter de la même adresse des quantités différentes, à la condition que l'on n'ait plus besoin de la première quantité au moment où la seconde est introduite. C'est là une différence avec le calcul à main où l'on n'est pas limité par le nombre des symboles.

### 2° Les sous-programmes :

Dans le programme ci-dessus, nous voyons apparaître une racine carrée, qui n'est pas généralement une instruction. Il faut donc écrire une suite d'instructions effectuant la racine carrée ; il faudra d'ailleurs écrire cette suite en tous les points où l'on calcule cette racine carrée. Pour économiser du temps et des mémoires, on réalisera une fois pour toutes un *sous-programme* de racine carrée étudié avec soin et qui ne se trouvera qu'en un seul emplacement du programme principal (ici, résolution de l'équation du second degré). Le programme se branchera sur ce sous-programme toutes les fois qu'une racine devra être calculée ; le retour se fera (par des techniques appropriées) au point où l'on aura quitté le programme principal.

Les divers sous-programmes mis au point sur un calculateur sont réunis en une bibliothèque (sur bande ou sur carte perforée) ; celle-ci contient des sous-programmes mathématiques (calcul de fonction, méthodes d'intégration, etc...), ou non-mathématiques (tri, conversion binaire-décimal, lecture de cartes, etc...).

Enfin, au moyen de sous-programmes appropriés, le programmeur peut réaliser toute opération non prévue par le constructeur (calcul sur le corps des complexes, sur des nombres comportant plus de décimales que le nombre standard, etc...). L'évolution des machines conduit souvent à réaliser comme nouvelles instructions des sous-programmes d'usage courant ; par exemple, la virgule flottante est apparue d'abord sous forme de sous-programme.

### 3° La programmation automatique :

L'organigramme étant écrit, il faut alors coder le problème, c'est-à-dire remplacer chaque case de l'organigramme par la séquence d'instructions convenable, affecter aux instructions et aux variables des adresses et, par exemple, perforer des cartes pour l'introduction dans la machine.

Cette partie du travail est très fastidieuse, d'où l'idée de la faire exécuter par le calculateur. Sous sa forme la plus évoluée, la programmation automatique prend la forme suivante :

a) Le « programmeur » écrit le programme sous forme d'une suite d'instructions écrites suivant un code se rapprochant de l'écriture mathématique (ou usuelle pour les décisions logiques) ; par exemple :

$$X = (AXC1 - AXC) / (AXB1 - AXB)$$

ALLER EN 1 SI  $X > 0$  .

Ces instructions sont représentées par une suite linéaire de symboles (nécessité technique).

b) Les instructions sont transcrites caractère à caractère sur un support lisible par la machine (carte, bande, etc...).

c) La machine, ayant lu le programme ainsi écrit, utilisera le « programme à programme » (programme d'assemblage) et créera un nouveau programme, dans le langage propre de la machine, qu'il suffira alors d'utiliser. Ces programmes d'assemblage sont, on le conçoit, d'une grande complexité.

Je signale enfin que les constructeurs tentent de se mettre d'accord sur une symbolisation unique appelée ALGOL (au niveau *a*). Ainsi, un programme donné serait utilisable par toutes les machines (après assemblage convenable).

### C. — LES MATHÉMATIQUES NUMÉRIQUES

Je vais maintenant sur quelques exemples essayer de faire ressortir les particularités du calcul sur machine.

#### 1° Calcul d'une fonction :

Lorsqu'un programme utilise une fonction  $f$ , le sous-programme de calcul de  $f$  peut utiliser :

a) une table de  $f$  calculée par la machine ou introduite directement (à partir de cartes par exemple), sur laquelle une interpolation est faite si les valeurs ne sont pas assez rapprochées, ou

b) une formule que l'on sait représenter assez bien  $f$  dans l'intervalle considéré.

On envisage souvent des solutions intermédiaires ; dans le cas *b* par exemple, on peut utiliser des formules différentes suivant l'intervalle où se trouve  $x$ .

La solution *a* est plus rapide, mais nécessite une plus grande place en mémoire. Dans un programme très important, on penchera pour la solution *b*.

Soit à calculer par exemple  $e^x$  :

$$e^x = 2^{\rho} (e^z)^x$$

$$z = \frac{\text{Log } 2}{2} \left\langle \frac{x}{\text{Log } 2} \right\rangle; \rho = \left[ \frac{x}{\text{Log } 2} \right], \text{ si } 0 \leq \left\langle \frac{x}{\text{Log } 2} \right\rangle \leq \frac{1}{2}$$

$$z = \frac{\text{Log } 2}{2} \left\langle \frac{x}{\text{Log } 2} - 1 \right\rangle; \rho = \left[ \frac{x}{\text{Log } 2} \right] + 1, \text{ si } \frac{1}{2} \leq \left\langle \frac{x}{\text{Log } 2} \right\rangle < 1.$$

Remarquons que  $|z| \leq \frac{\text{Log } 2}{4}$ . Pour calculer  $e^z$ , on utilise :

$$e^z \sim \sum_0^s \frac{z^k}{k!},$$

le programme déterminant le nombre de termes à utiliser.

Remarquons que, sur une machine ne connaissant que les quatre opérations, on ne peut calculer que des fractions rationnelles. C'est ce qui conduit à rechercher des approximations rationnelles ou polynômiales des fonctions. Mais, plutôt que d'utiliser, ainsi que nous venons de le faire, des formules d'origine théorique, on préfère rechercher numériquement les coefficients inconnus de la fonction approchée.

Le problème s'énonce ainsi dans le cas de l'approximation polynômiale : trouver  $N, a_0, a_1, \dots, a_N$  tels que :

$$|f(x) - \sum_0^N a_k x^k| \leq \varepsilon \quad \text{pour } x \in I$$

et que  $N$  soit le plus petit possible (de manière que le calcul soit le plus rapide possible) ; accessoirement, on peut imposer que, pour le  $N$  ainsi déterminé, le maximum de l'erreur soit le plus petit possible. Ce problème d'optimisation se traite par des méthodes comparables à celles utilisées pour minimiser des dépenses dans les problèmes économiques. On peut utiliser la théorie de la programmation linéaire.

Il existe des programmes dont l'objet est précisément la recherche du polynôme optimum dans un certain intervalle. L'établissement d'un sous-programme donnant  $f$  se fait alors en deux temps :

a) faire un programme précis au besoin assez compliqué, calculant  $f$  (par exemple pour  $e^x$  le programme indiqué ci-dessus) ;

b) utiliser ce programme pour trouver le « meilleur » polynôme (ou fraction rationnelle) donnant  $f$ .

Ceci peut être long, mais est fait une fois pour toutes ; la suppression d'une seule opération dans un sous-programme justifie un calcul préalable, même long, sur machine.

## 2° Méthodes de Monte-Carlo :

Si l'on utilise la méthode des trapèzes avec 10 intervalles pour chaque variable, le calcul d'une intégrale multiple d'ordre  $p$  exige le calcul de  $10^p$  valeurs de la fonction. Même si le calculateur est très rapide, le temps exigé devient facilement supérieur à un siècle. Nombreux sont d'ailleurs les problèmes d'apparence bénigne, surtout si des dénombrements interviennent, qu'il est vain actuellement d'essayer de résoudre numériquement.

Soit par exemple à calculer l'intégrale :  $I = \int \dots \int_V f(P) dv$ .

Les méthodes classiques reviennent à calculer  $\sum_1^N \alpha_i f(P_i)$ , les  $\alpha_i$  et les  $P_i$  étant choisis suivant une loi bien déterminée (dans les trapèzes, les  $P_i$  sont les sommets d'un réseau de points).

Dans une méthode de Monte-Carlo, on considérera la variable aléatoire  $f(P)$ ,  $P$  étant un point aléatoire de distribution uniforme dans  $V$ , son espérance  $E$  est  $\frac{1}{I} \int \dots \int_V f(P) dv$  ; si l'on connaît le volume de  $V$ , on fait

un nombre suffisant de « tirages » de  $P$ , la moyenne des  $f(P)$  obtenus est, avec une certaine probabilité  $p$ , comprise entre  $E + \varepsilon$  et  $E - \varepsilon$  ; si la probabilité est assez grande, on admet que  $E$  est égale à cette moyenne à  $\varepsilon$  près.

D'une manière générale, lorsque l'on désire un nombre  $a$ , l'on recherche une variable aléatoire  $X$  dont l'espérance est égale à  $a$ , et l'on fait un nombre suffisant de tirages pour l'estimation de  $a$ .

Souvent avec dix mille tirages l'on peut espérer trois chiffres significatifs ; on a donc réduit le calcul d'intégrale indiqué au début à une taille raisonnable.

Remarquons que l'on doit disposer en machine d'un programme à tirer des échantillons d'une variable aléatoire. Deux solutions ont été retenues :

a) un phénomène physique, type bruit de fond, fournit des « nombres au hasard » qui sont enregistrés et fournis le cas échéant à la machine ;

b) un sous-programme calcule des pseudo-nombres au hasard, utilisables, comme des séries de tests le montrent, pour ce type de calcul. Une loi fréquemment utilisée est :

$$u_{n+1} = ku_n \pmod{p}.$$

Un des premiers calculs de Monte-Carlo (avant la lettre) a été l'estimation de  $\pi$  par le jet d'une aiguille.

L'idée est toujours de trouver un modèle probabiliste à un problème mathématique et de simuler ledit modèle. Quelquefois même, ce mode de résolution revient à simuler le phénomène physique de nature probabiliste (mouvement de particules, etc...), ou un phénomène simplifié, qui a donné naissance aux équations à résoudre. Ici, la mise en équation du physicien n'aura été d'aucun secours pour la résolution numérique.

Remarquons que la confiance dans le résultat obtenu dépend de la variance de la variable considérée. Faute de mieux, cette variance est souvent elle-même estimée expérimentalement.

### 2° Résolution d'équations linéaires :

Pour résoudre un système d'équations linéaires d'ordre  $n$ , les formules utilisant des déterminants sont à proscrire (un déterminant d'ordre  $n$  est la somme de  $n!$  termes eux-mêmes produits de  $n$  facteurs) ; la méthode la plus rapide en général, pour une précision donnée, est l'élimination dont de nombreuses variantes sont utilisées. On voit que les constantes occupent sensiblement  $n^2$  mémoires, le temps de calcul varie comme  $n^3$ . Des considérations d'encombrement font utiliser d'autres méthodes, qui peuvent faire gagner du temps dans le cas où de nombreux coefficients sont nuls.

a) *Méthode de Monte-Carlo* : problème du parieur.

Soit deux parieurs A et B de richesses  $a$  et  $b$ , faisant une suite de paris de mise  $h$  ; à chaque pari, A a la probabilité  $p$  de gagner, et B la probabilité  $q = 1 - p$ . On recherche la probabilité que A ruine B (événement où le jeu s'arrête, de même que si B ruine A) ; nous supposons  $a$  et  $b$  multiples entiers de  $h$ . Nous représentons l'état du jeu à un certain instant par le schéma :



OE représentant la fortune actuelle de A et EF celle de B,  
(OF =  $a + b = C$ ).

On peut imaginer que E est une particule parcourant  $+h$  avec la probabilité  $p$  et  $-h$  avec la probabilité  $q$  ; on recherche alors la probabilité qu'a la particule d'atteindre F.

Soit  $v_m(x)$  la probabilité que A ruine B en au plus  $m$  coups, la fortune de A étant  $x$  au départ ; ou A gagne le premier pari et ruine ensuite B en au plus  $m-1$  coups, ou A le perd et ruine ensuite B en au plus  $m-1$  coups, ce qui se traduit par :

$$(1) \quad v_m(x) = pv_{m-1}(x+h) + qv_{m-1}(x-h) \quad \begin{matrix} 0 < x < a+b \\ m \geq 1 \end{matrix}$$

d'autre part :  $v_m(0) = 0 \quad v_m(a+b) = 1.$

$$\left\{ \begin{array}{ll} v_0(x) = 0 & \text{si } 0 \leq x < a+b \\ v_0(x) = 1 & \text{si } x = a+b. \end{array} \right.$$

La relation (1) et  $v_1(x) \geq v_0(x)$  entraînent par récurrence que

$$v_m(x) \geq v_{m-1}(x) ;$$

d'autre part  $v_m(x) \leq 1$ , donc  $v_m(x)$  admet une limite  $v(x)$  lorsque  $m \rightarrow \infty$ .

C'est cette limite que nous définirons comme probabilité que A ruine B en un nombre fini de paris ;  $v(x)$  sera solution de :

$$\left\{ \begin{array}{l} v(x) = pv(x+h) + qv(x-h) \\ v(0) = 0 \\ v(1) = 1. \end{array} \right.$$

Si l'on remarque que  $x$  prend un nombre fini de valeurs, on voit que nous avons un système, cramérien d'ailleurs, d'un nombre fini d'équations linéaires pour déterminer  $v(x)$ . On peut trouver la valeur explicite de  $v(x)$  en utilisant les méthodes de résolution d'équations aux différences :

$$v(x) = \frac{1 - \left(\frac{p}{q}\right)^{-\frac{x}{h}}}{1 - \left(\frac{p}{q}\right)^{-\frac{a+b}{h}}}$$

On peut voir également que  $v(x) + u(x) = 1$ ,  $u(x)$  étant la probabilité que B ruine A en un nombre fini de paris ; autrement dit, le jeu se termine presque sûrement en un nombre fini de paris.

On peut donc simuler le jeu sur machine et estimer par la méthode de Monte-Carlo ; on est « presque » assuré que chaque partie demande un nombre fini de coups, donc que le calcul lui-même est réalisable en un temps fini. Dans le cas particulier, le calcul ne présenterait pas d'intérêt pratique. En revanche, en remplaçant certaines équations différentielles (ou aux dérivées partielles) par des équations aux différences, on tombe sur des équations d'un type voisin de (2) en si grand nombre qu'il est raisonnable d'appliquer une méthode de Monte-Carlo.

### b) Méthodes itératives.

Utilisant la notation matricielle, un système d'équations linéaires, que nous supposons cramérien, peut se mettre d'un grand nombre de manières sous la forme :

$$(3) \quad x = Ax + b.$$

Considérons la suite :  $x_0$  quelconque,

$$x_{k+1} = Ax_k + b.$$

Si elle tend vers une limite  $\alpha$ ,  $\alpha$  est la solution du système proposé. On montre que, pour que cette suite ait une limite, il faut et il suffit que la plus grande valeur propre de  $A$  soit en module inférieure à 1 et que la convergence est d'autant plus rapide que le module maximum est plus petit.

Cette méthode de résolution, bien que plus longue que l'élimination si  $A$  est une matrice pleine (pas de coefficient nul), est de programmation très simple et surtout nécessite beaucoup moins de mémoires si la matrice  $A$  est très creuse. Quand on l'utilise, on cherche à écrire l'équation (3) de manière que la convergence soit la plus rapide possible.

#### 4° Conclusion :

Je n'ai pas voulu développer ici tous les domaines d'application du calcul numérique ; pratiquement, dès qu'un problème mathématique a pour solution un nombre fini de nombres, il n'est pas vain d'essayer d'en rechercher une solution approchée numérique. Remarquons que l'on se contente, dans un problème où la solution est un ensemble infini de nombres, de ne donner qu'une partie de la solution (par exemple, une fonction de deux variables sera calculée aux sommets d'un réseau).

En revanche, je voudrais, en conclusion de ces quelques exemples, dégager ces deux points fondamentaux :

1° Une solution explicite d'un problème est rarement à utiliser ; elle ramène le problème à d'autres (calcul d'intégrale, de fonctions spéciales, sommation d'une série), qui sont souvent, numériquement, d'un traitement plus difficile que le problème initial, ou beaucoup plus long. Ainsi, pour résoudre une équation du second degré, une méthode itérative peut souvent être préférée aux formules classiques. L'étude théorique d'un problème ne sert souvent qu'à interpréter certains accidents de calcul ; par exemple, si l'équation du second degré n'a pas de racines, on conçoit qu'une méthode itérative ne converge pas.

2° Les méthodes employées sur calculateurs rapides ne sont pas généralement une adaptation de ce qui est fait à la main. Lorsqu'un problème déjà résolu à la main est à résoudre sur machine, mais à une échelle beaucoup plus grande (exemple : système de mille équations linéaires à mille inconnues), le temps de calcul, ou l'encombrement, devient souvent prohibitif. On doit alors se tourner vers des méthodes dont le temps de calcul, ou l'encombrement, croît moins vite avec la taille du problème, et qui seront par conséquent, à partir d'une certaine taille, à préférer à la méthode initiale. Il faut également remarquer qu'avec des mémoires à bandes et, dans une certaine mesure, des mémoires à tambours, il y a intérêt à utiliser l'information en séquence sur la mémoire, ce qui peut donner la préférence à une méthode utilisant des nombres toujours dans le même ordre.

---

CAHORS, IMP. A. COUSSLANT. — 97.714. — Dépôt légal : IV-1961  
*Printed in France*

---

En préparation : pour paraître au début de 1962.

André et Germaine REVUZ

# COURS DE L'A.P.M.

## I. GROUPE, ANNEAUX, CORPS

*Brochure de l'A.P.M., n° 6*

Ce volume, entièrement inédit, de 180 pages, aurait pu s'intituler « Les grands commençants » ou encore « Les professeurs de Mathématiques parlent aux professeurs de Mathématiques ». Ces titres « ironiques » auraient souligné certains aspects originaux de l'ouvrage.

A la demande de nombreux collègues, A. Revuz, président de l'A.P.M., fut sollicité d'organiser un cours suivi pour les professeurs désireux de reprendre, on pourrait dire à la base, leur formation mathématique. Concevant sa tâche de président de façon concrète, Revuz ne fit pas de discours ; un cours *seulement*. Les auditeurs en gardent le vif et savoureux souvenir et ils en redemandent...

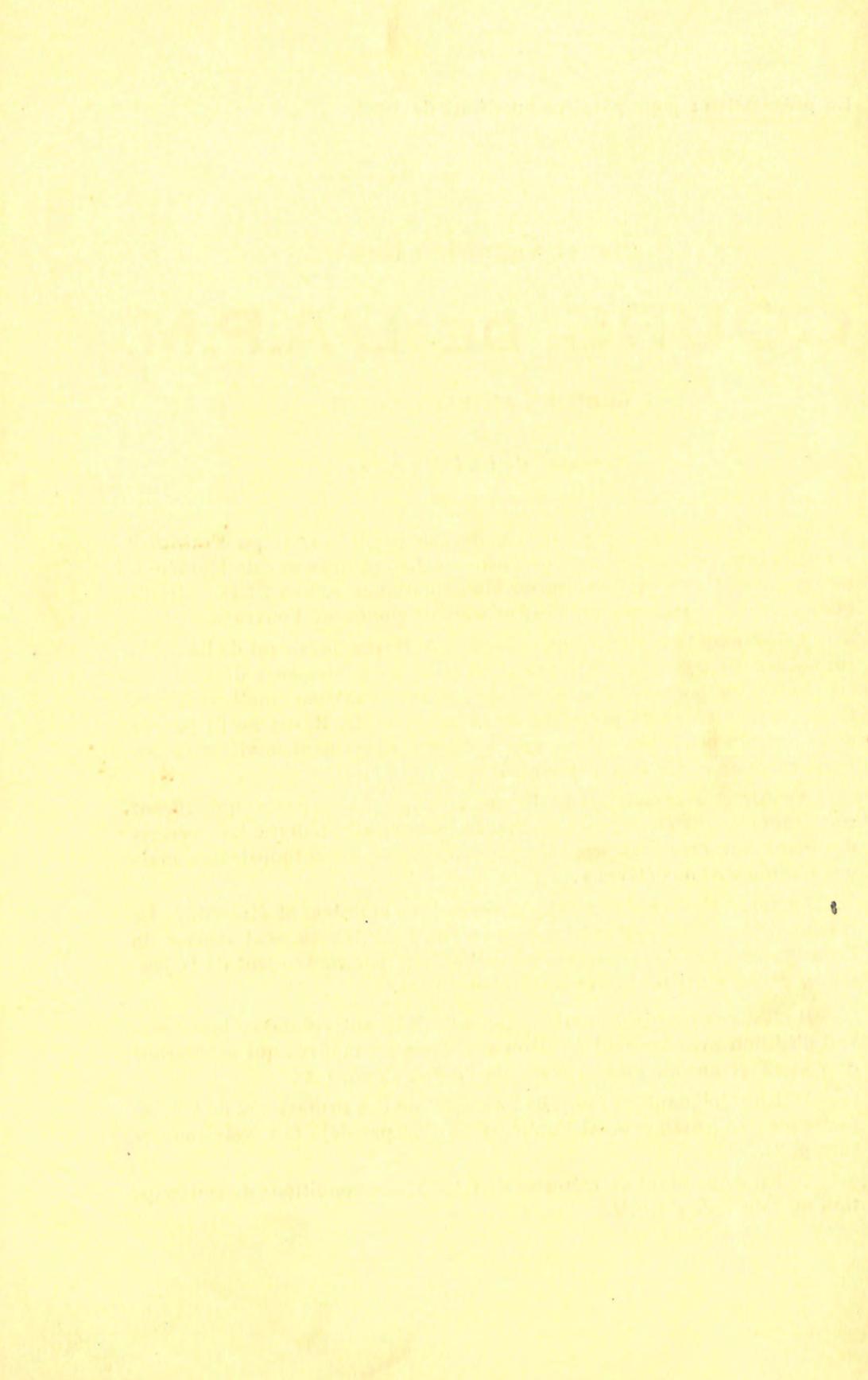
Au fur et à mesure, Mme Revuz a rédigé les exposés qui furent ronéotypés et distribués aux auditeurs. Mieux, elle rédigea les corrigés des nombreux exercices proposés par le maître à ses enthousiastes, mais pas toujours « bons élèves ».

Le texte plusieurs fois relu et corrigé va maintenant être édité. Le travail d'éducation mutuelle entrepris par l'A.P.M., au seul service de l'enseignement et de la science, au seul service par conséquent de la jeunesse, prend ainsi un nouveau développement.

Il n'est pas inutile de souligner que l'A.P.M. entreprend ce lourd travail d'édition avec ses seules ressources. Tous les maîtres qui profiteront de son effort auront donc à cœur de l'aider. Comment ?

1° En rejoignant les rangs de l'Association des Professeurs de Mathématiques de l'Enseignement Public, si ce n'est pas déjà fait (voir couverture p. 2).

2° En demandant au trésorier de l'A.P.M. les conditions de souscription au *Cours de l'A.P.M.*





# L'APMEP en quelques mots...

## **Fondée en 1910, l'APMEP est une association :**

- totalement indépendante, politiquement et syndicalement, et bénévole ;
- qui représente les enseignants de mathématiques de la maternelle à l'université.

## **L'APMEP se préoccupe simultanément :**

- des contenus des programmes ;
- des compétences requises des élèves ;
- des méthodes d'enseignement et de formation ;
- des horaires et effectifs, en particulier des dédoublements de classes ;
- de l'harmonisation entre les cycles ;
- de la valorisation des mathématiques comme instrument de formation et non de sélection.

## **L'APMEP est un lieu de :**

- libre parole et de confrontation d'idées ;
- démarches coopératives d'auto-formation ;
- propositions pour une politique d'enseignement des mathématiques.

## **L'APMEP intervient pour :**

- défendre ses positions ;
- intégrer les nouveaux outils (calculatrices, logiciels de géométrie, de calcul...) ;
- faciliter les évolutions et les démarches d'équipe (formation initiale et permanente, laboratoires de maths...).

## **L'APMEP agit pour préserver, donner ou redonner aux élèves :**

- le goût des mathématiques ;
- le plaisir d'en faire.

## **Pour l'APMEP, faire des mathématiques, c'est :**

- identifier, formuler un problème ;
- expérimenter sur des exemples ;
- conjecturer un résultat ;
- bâtir une démonstration ;
- mettre en œuvre des outils théoriques ;
- contrôler les résultats et leur pertinence ;
- communiquer une recherche, une solution ;
- développer simultanément :
  - o le travail individuel et le travail collectif des élèves ;
  - o le sens de l'écoute et du débat ;
  - o la persévérance ;
  - o les capacités d'imagination, d'esprit critique, de cohérence et de rigueur.

## **Faire des mathématiques, c'est œuvrer pour :**

- la formation de l'esprit ;
- l'intégration dans la vie sociale, culturelle et professionnelle.

***Plus d'informations sur : [www.apmep.fr](http://www.apmep.fr)***