« 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

(*) Avaient 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 conjectandi » (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 = (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.

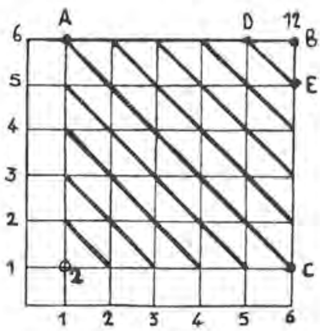


Fig. 1

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 :

$$5 \leqslant S \leqslant 8.$$

La réalisation de E_1 a une probabilité $\Pr(E_1)$ 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 :

$$6 \leqslant S \leqslant 8,$$

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

$$\Pr(E_2) < \Pr(E_1)$$

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

$$E_2 \subset E_1 \Longrightarrow \Pr(E_2) < \Pr(E_1).$$

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 :

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

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 :

$$\Pr(A) = \sum_1^p \Pr(A_i).$$

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 $\overline{E}_1$. Ces deux événements constituent un système complet. Si l'on note $E_1 \cup \overline{E}_1$ l'événement certain « E_1 ou $\overline{E}_1$ » et $E_1 \cap \overline{E}_1$ l'événement impossible « E_1 et $\overline{E}_1$ », on peut écrire :

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

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

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

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

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 ;

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

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

III. — Evénement $E_1 \cap E_2$; 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 :

$$E_1 \cap E_2, \; E_1 \cap \overline{E}_2, \; \overline{E}_1 \cap E_2, \; \overline{E}_1 \cap \overline{E}_2.$$

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 :

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

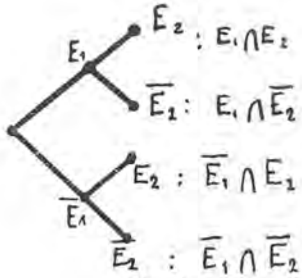
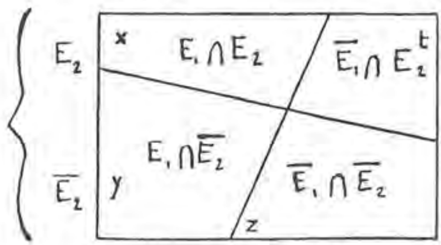


Fig. 2

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 :

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



E_1 Fig. 3 $\overline{E}_1$

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 : $\Pr(E_1/E_2) = \dfrac{x}{x+t}$

d'où résultent les formules :

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

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 agite avant de faire le second tirage !).

Exercice. — Soient n événements indépendants $E_1, E_2, ..., E_n$, 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 :

$$\Pr(E_1 \cap E_2 \cap E_3 \ldots \cap E_n) = \Pr(E_1) . \Pr(E_2) \ldots \Pr(E_n) \qquad (2) \qquad (*).$$

(*) 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 jarretières 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'invraisemblable, 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. — Evénement $E_1 \cup E_2$.

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 $E_1 \cap \overline{E}_2$; ce qui permet d'écrire :

$$\Pr(E_1) = \Pr(E_1 \cap E_2) + \Pr(E_1 \cap \overline{E}_2) \qquad (3)$$
et de même :
$$\Pr(E_2) = \Pr(E_1 \cap E_2) + \Pr(\overline{E}_1 \cap E_2).$$

Un dénombrement direct facile conduit à la formule :
$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2) \qquad (4).$$

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

Appliquant (3), on peut écrire :
$$\Pr(\overline{E}_1 \cap \overline{E}_2) = \Pr(\overline{E}_1) - \Pr(\overline{E}_1 \cap E_2).$$

Mais les événements $\overline{E}_1 \cap \overline{E}_2$ et $E_1 \cup E_2$, E_1 et $\overline{E}_1$ sont contraires ; l'égalité précédente peut donc s'écrire :
$$1 - \Pr(E_1 \cup E_2) = 1 - \Pr(E_1) - \Pr(\overline{E}_1 \cap E_2),$$
ou :
$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(\overline{E}_1 \cup E_2),$$
et comme, en utilisant de nouveau (3) :
$$\Pr(\overline{E}_1 \cap E_2) = \Pr(E_2) - \Pr(E_1 \cap E_2),$$
on obtient finalement la formule (4).

En général, on a donc :
$$\Pr(E_1 \cup E_2) \leqslant \Pr(E_1) + \Pr(E_2),$$
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 $E_1, E_2, ..., E_n$. 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 :

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

d'où :

$$\Pr(\mathrm{E}_i) \, . \Pr(\mathrm{A}/\mathrm{E}_i) = \frac{x_i}{\mathrm{N}}$$

d'où enfin :

$$\Pr(\mathrm{A}) = \sum \frac{x_i}{\mathrm{N}} = \sum \Pr(\mathrm{E}_i) \, . \Pr(\mathrm{A}/\mathrm{E}_i) \qquad (5).$$

C'est la formule de la probabilité complète.

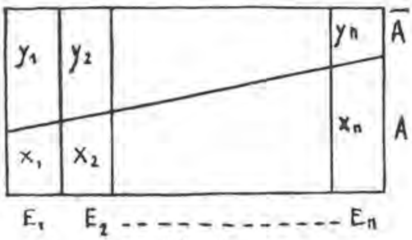


Fig. 4

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

$$\Pr(\mathrm{E}_i \cap \mathrm{A}) = \Pr(\mathrm{E}_i) \, . \Pr(\mathrm{A}/\mathrm{E}_i) = \Pr(\mathrm{A}) \, . \Pr(\mathrm{E}_i/\mathrm{A}),$$

d'où la formule de Bayes :

$$\Pr(\mathrm{E}_i/\mathrm{A}) = \frac{\Pr(\mathrm{E}_i) \, . \Pr(\mathrm{A}/\mathrm{E}_i)}{\Pr(\mathrm{A})}$$

c'est-à-dire, en utilisant (5) :

$$\Pr(\mathrm{E}_i/\mathrm{A}) = \frac{\Pr(\mathrm{E}_i) \, . \Pr(\mathrm{A}/\mathrm{E}_i)}{\Sigma \Pr(\mathrm{E}_k) \, . \Pr(\mathrm{A}/\mathrm{E}_k)} \qquad (6).$$

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

$$\Pr(\mathrm{E}_1) = 0{,}96, \qquad \Pr(\overline{\mathrm{E}}_1) = 0{,}04.$$

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

$$\Pr(\mathrm{E}_1/\mathrm{A}) = \frac{0{,}96 \times 0{,}97}{0{,}96 \times 0{,}97 + 0{,}04 \times 0{,}04} \backsimeq 0{,}998.$$

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 $\Pr(E_i)$ qui vérifie les propriétés :

$$\begin{cases} 0 \leqslant \Pr(E_i) \leqslant 1 \quad \text{pour tout } E_i \subset E, \\ \Pr(E_i) = 0 \text{ si et seulement si } E_i = \emptyset, \\ \Pr(E) = 1, \\ E_1 \subset E_2 \Longrightarrow \Pr(E_1) < \Pr(E_2), \\ E_1 \cap E_2 = \emptyset \Longrightarrow \Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2). \end{cases}$$

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 $x_1, x_2, ..., x_n$, avec les probabilités $p_1, p_2, ..., p_n$. On appelle espérance mathématique l'expression :

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

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

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

On jette deux dés ; les résultats connus nous montrent que si X est la somme des points, nous avons : $\mathrm{E(X)} = 7$. 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 :

$$\mathrm{E(X + Y) = E(X) + E(Y)}.$$

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 :

$$\mathrm{X} \begin{cases} x_1 x_2 \ldots, x_n, \\ p_1 p_2 \ldots, p_n \end{cases}$$

$$\mathrm{Y} \begin{cases} y_1 y_2 \ldots, y_m, \\ q_1 q_2 \ldots, q_m. \end{cases}$$

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 :

$$p_{ik} = \mathrm{Pr}\,(x_i) \,.\, \mathrm{Pr}\,(y_k/x_i) = \mathrm{Pr}\,(y_k) \,.\, \mathrm{Pr}\,(x_i/y_k),$$

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

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

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 :

2.

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

$$Y = y_1, Y = y_2, ..., Y = y_m$$

forment un système complet, et nous avons :

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

d'où :

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

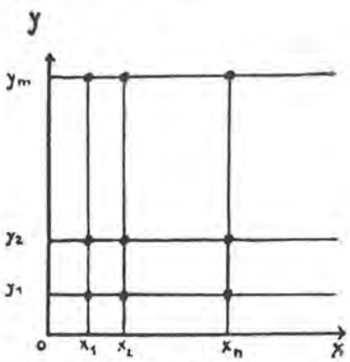


Fig. 5

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

$$\sum x_i . \operatorname{Pr}(x_i) = \mathrm{E(X)}.$$

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 :

$$\mathrm{E}(\sum \mathrm{X}_i) = \sum \mathrm{E}(\mathrm{X}_i).$$

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

$$\mathrm{E}(\mathrm{X} + a) = \mathrm{E}(\mathrm{X}) + a,$$

et, comme il est clair que :

$$\mathrm{E}(a\mathrm{X}) = a\mathrm{E}(\mathrm{X}),$$

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

$$\mathrm{E}(\sum a_i \mathrm{X}_i) = \sum a_i \mathrm{E}(\mathrm{X}_i).$$

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 :

$$\mathrm{E(X)} = 0 . 5/6 + 1 . 1/6 = 1/6.$$

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_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 :

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

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 :

$$E(XY) = E(X) . E(Y) ;$$

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

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

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 :

$$E(XY) = E(X) . E(Y),$$

comme on le voit facilement, en remarquant que la probabilité de l'événement : « $X = x_i$ et $Y = y_k$ » est : $\Pr(x_i) . \Pr(y_k)$, 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 ».

La solution classique apportée par Pascal au « problème des parties » que lui avait proposé de Méré revient à utiliser l'espérance mathématique. Mais citons d'abord la lettre de Pascal à Fermat :

« Voici à peu près comme je fais pour savoir la valeur de chacune des parties, quand deux joueurs jouent, par exemple, en trois parties, et chacun a mis 32 pistoles au jeu :

Posons que le premier en ait deux et l'autre une ; ils jouent maintenant une partie, dont le sort est tel que, si le premier la gagne, il gagne tout l'argent qui est au jeu, savoir 64 pistoles ; si l'autre la gagne, ils sont deux parties à deux parties et, par conséquent, s'ils veulent se séparer, il faut qu'ils retirent chacun leur mise, savoir chacun 32 pistoles.

Considérez donc, Monsieur, que si le premier gagne, il lui appartient 64 ; s'il perd, il lui appartient 32. Donc, s'ils veulent ne point hasarder cette partie et se séparer sans la jouer, le premier doit dire : " Je suis sûr d'avoir 32 pistoles, car la perte même me les donne ; mais, pour les 32 autres, peut-être je les aurai, peut-être vous les aurez ; le hasard est égal ; partageons donc ces 32 pistoles par la moitié et me donnez, outre cela, mes 32 pistoles qui me sont sûres. "

Il aura donc 48 pistoles et l'autre 16. »

Pascal envisage encore d'autres cas : celui où l'un des joueurs a gagné deux parties et l'autre aucune, pour lequel il propose le partage des 64 pistoles en 56 et 8 ; celui où l'un des joueurs a gagné une partie et l'autre aucune, pour lequel il propose le partage 44 et 20.

Il s'agit donc d'un jeu dans lequel, à chaque coup, chacun des deux joueurs a une probabilité 1/2 de gagner. Le jeu est gagné par le joueur qui, le premier, a gagné trois coups. Chaque joueur risque 32 pistoles, et les 64 pistoles sont prises par le gagnant.

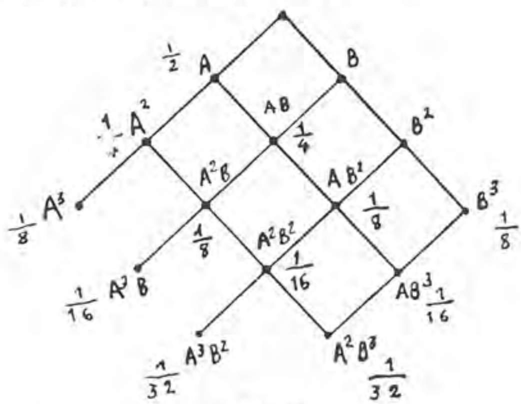


Fig. 6

Les probabilités sont $\frac{1}{2^n}$, n étant la somme des exposants dans chaque situation A^iB^j.

La figure 6, dans laquelle la notation A^2B signifie que le joueur A a gagné deux coups et le joueur B un coup, donne le déroulement de la partie. Le problème est de déterminer le partage des enjeux, lorsque la partie est interrompue en A^2B par exemple.

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 :

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

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 :

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

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 $E(Z^2)$:

$$V = E(Z^2) = E\left[(X-m)^2\right] = \sum p_i(x_i - m)^2 = \sum p_i x_i^2 + \sum p_i m^2 - 2\sum p_i x_i m$$
$$= \sum p_i x_i^2 + m^2 \sum p_i - 2m \sum p_i x_i = \sum p_i x_i^2 + m^2 - 2m^2 = \sum p_i x_i^2 - m^2$$

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 :

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

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 :

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

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

$$\sigma_z^2 = E\left[\left(\sum a_k(X_k - m_k)\right)^2\right] = E\left[\sum a_k^2 (X_k - m_k)^2 + 2\sum a_i a_k (X_i - m_i)(X_k - m_k)\right]$$
$$= \sum a_k^2 E\left[(X_k - m_k)^2\right] + 2\sum a_i a_k E\left[(X_i - m_i)(X_k - m_k)\right]$$

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

$$E[(X_i - m_i)(X_k - m_k)] = E(X_i - m_i) . E(X_k - m_k) = 0,$$

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 :

$$\begin{matrix} 1 - p & -p \\ p & 1 - p \end{matrix}$$

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, σ_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. σ 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 σ. Etant donné un nombre positif t, cherchons :

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

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

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

or, Σ ne s'étend qu'à une partie des termes de σ^2 ; donc :

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

En résumé :

$$\Pr(|X - m| \geqslant t\sigma) \leqslant \frac{1}{t^2}, \qquad \Pr(|X - m| \leqslant t\sigma) \geqslant 1 - \frac{1}{t^2}.$$

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 :

$$\Pr(|X - m| \leqslant 5\sigma) \geqslant 0{,}96.$$

C'est la probabilité pour que toutes les valeurs de X soient contenues dans un intervalle de longueur 10σ 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 :

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

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 σ_F tend vers zéro quand n croît).

Exemple. — Soit une suite de coups « pile ou face » ; $p = 1/2$, $\sigma_F = \dfrac{1}{2\sqrt{n}}$. 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 ε 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 ε, 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 » ; $\overline{E}$ est l'événement, « on n'obtient pas de 6 ». On a évidemment : $\Pr(\overline{E}) = 5/6$. 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 « quasicertitude » ? 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^{er} lancé, X_2 : point donné par le 2^e 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 :

$$\Pr(E_1 \cap E_2) \leqslant \Pr(E_1) \; ; \; \Pr(E_1 \cap E_2) \leqslant \Pr(E_2) \; ;$$
$$\Pr(E_1 \cup E_2) \geqslant \Pr(E_1) \; ; \; \Pr(E_1 \cup E_2) \geqslant \Pr(E_2).$$

6. — Démontrer :

$$E_1 \subset E_2 \Longrightarrow \Pr(E_1/A) < \Pr(E_2/A),$$
$$\Pr(E_1/A) + \Pr(\overline{E}_1/A) = 1.$$

7. — Quand a-t-on :

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

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

9. — Démontrer :

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

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[er] coup donne pile ; E_2 : le 2[e] coup donne pile ; E_3 : les deux coups donnent le même résultat. Calculer :

$$\Pr(E_1) \; ; \; \Pr(E_3/E_1) \; ; \; \Pr(E_3/E_2) \; ; \; \Pr(E_3/E_1 \cap E_2) \; ; \; \Pr(E_3/E_1 \cup E_2).$$

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 :

$$\Pr(\overline{E}_1 \cap \overline{E}_2 \cap \ldots \cap \overline{E}_n) = \Pr(\overline{E}_1).\ \Pr(\overline{E}_2) \ldots \Pr(\overline{E}_n) \; ;$$
$$\Pr(E_1 \cup E_2 \ldots \cup E_n) = 1 - [1 - \Pr(E_1)] \ldots [1 - \Pr(E_n)].$$

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 Σx_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 $X \leqslant 1/4$. 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 :

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

étant un nombre positif donné et σ 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 σ_i, *les variables étant de plus indépendantes deux à deux.* Soit X la variable aléatoire Σx_i. En supposant que tous les écarts-type σ_i sont majorés par un nombre positif donné b, montrer que l'on a :

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

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.