Zeitschrift: L'Enseignement Mathématique

Herausgeber: Commission Internationale de l'Enseignement Mathématique

Band: 22 (1976)

Heft: 1-2: L'ENSEIGNEMENT MATHÉMATIQUE

Artikel: RÔLE, PLACE, ET CONTENU D'UN PREMIER ENSEIGNEMENT DÉDUCTIF DES PROBABILITÉS

Autor: Breny, H.

DOI: https://doi.org/10.5169/seals-48177

RÔLE, PLACE, ET CONTENU D'UN

PREMIER ENSEIGNEMENT DÉDUCTIF DES PROBABILITÉS

par H. BRENY

1. QUESTION D'EXISTENCE

a. Depuis son introduction dans le *Rapport sur l'enseignement de la géométrie* de la Mathematical Association en 1923 [9], la distinction des stades A (expérimental et intuitif), B (déductif), et C (systématique) a été étendue à l'enseignement mathématique en général, encore que l'on tende souvent à confondre partiellement les stades A et B (par l'introduction, au stade A, de « noyaux déductifs » où les élèves « sont amenés à constater que, de certaines propositions considérées — au moins provisoirement — comme acquises on peut déduire d'autres propositions par le raisonnement » [5]).

Assez curieusement, il semble que certains didacticiens de la mathématique — s'occupant surtout, il est vrai, de l'enseignement primaire — estiment qu'un enseignement déductif de la théorie des probabilités n'est pas de mise au niveau secondaire, ce qui revient à dire qu'il doit être réservé aux seuls spécialistes de la mathématique (et peut-être, de la physique) [1]; leur argument essentiel est qu'en cette matière ce qui compte c'est la perception d'un certain type de situations non mathématiques (les situations aléatoires) et de la façon de les mathématiser. Or, c'est précisément parce que la théorie des probabilités n'est pas un chapitre de mathématique mais, au sens plénier du mot, une *théorie physique* [2]) qu'un enseignement déductif est indispensable à une formation équilibrée de tous les futurs utilisateurs de la mathématique.

---

1) En dehors de ces spécialités il n'y a virtuellement plus d'enseignement mathématique *déductif* au niveau supérieur: la masse d'applications à étudier s'y oppose efficacement, et y fait de la mathématique un pur outil.

2) Voir toutefois le paragraphe 3 b. ci-dessous.

b. Les « théories physiques » existent depuis longtemps (dès avant les « Éléments » d'Euclide, Eudoxe avait construit une « théorie physique » des mouvements des astres). P. Duhem (lui-même excellent théoricien de la physique et remarquable historien des sciences) en a analysé la structure, mettant clairement en évidence leur aspect construit (ce sont des *modèles théoriques*) et leur forme hypothético-déductive (voir [2] et [8]): la réflexion sur les faits observables disponibles conduit à l'invention d'hypothèses, celles-ci servent d'axiomes à un édifice logique, dont les conséquences sont comparées aux observations; si cette comparaison fait apparaître un désaccord, la théorie est rejetée. K. Popper (notamment dans [11]) a beaucoup insisté sur la « falsifiabilité » des théories physiques, car il s'opposait aux logiciens comme Carnap pour qui des considérations de type probabiliste [1]) pouvaient fournir un « degré de confirmation » d'une théorie par l'expérience. La manière de comparer une théorie à l'expérience a toujours été un point faible de l'analyse épistémologique; des progrès sérieux ont été faits dans ce domaine depuis les travaux de Kuhn sur la « science normale » et les « révolutions scientifiques » ([6], [7], [13]; un bon état de la question se trouve dans [14], chap. 8 et 9). Ces progrès n'ont nullement altéré la thèse de Duhem, mais l'ont complétée: une théorie physique est à la fois une construction conceptuelle et un outil; elle comprend deux composantes: une structure logico-mathématique et un ensemble d'applications envisagées (défini, ou plutôt décrit [«en intention»], par son paradigme, ensemble d'applications particulièrement réussies et (ou?) d'importance historique particulièrement grande) [2]. Ces deux composantes sont essentielles, on ne peut négliger ni l'une ni l'autre.

c. C'est exactement cela qu'est la théorie des probabilités: *la théorie physique des phénomènes fortuits*. Et c'est précisément cela que l'enseignement pré-déductif de la théorie n'est pas en mesure de montrer. Certes, cet enseignement est indispensable: avant tout essai de théorisation il faut familiariser les élèves avec les situations aléatoires et les premiers moyens théoriques de leur analyse; si, comme beaucoup de didacticiens le pensent, un tel enseignement est possible au niveau primaire, c'est à ce niveau qu'il faut l'entreprendre; il faut alors le poursuivre — et, sinon, l'entreprendre — durant le premier cycle du niveau secondaire. Mais, si

[1]) En un autre sens du mot! voir paragraphe 3 b. ci-dessous.

[2]) Accessoirement, ils ont mis en évidence la thèse de la non-falsifiabilité (« immunité ») des théories physiques, dont l'observation dans l'histoire est à mettre à l'actif de Kuhn.

utile qu'il soit, un tel enseignement laisse échapper presque entièrement la structure logico-mathématique de la théorie des probabilités: celle-ci ne peut être saisie que dans un enseignement *déductif*. La situation, dans ce cas, est un peu différente de celle de la géométrie. Certes, la géométrie est aussi, pour une part, une théorie physique (la « théorie physique des relations spatiales »); mais cet aspect est relativement peu important, vu le caractère extrêmement élémentaire des notions « physiques » mises en jeu; les « structures » géométriques sont d'ailleurs perçues (et édifiées) par le jeune enfant bien avant qu'il perçoive l'aléatoire [10]. Un enseignement de la géométrie qui en resterait au stade *A* renforcé par quelques noyaux déductifs ne permettrait pas l'usage plénier de l'outil géométrique et ne donnerait de la géométrie qu'une idée insuffisante; mais pour la théorie des probabilités, un tel enseignement, laissé seul, en donnerait une idée véritablement fausse, parce qu'il n'en ferait pas ressortir l'aspect hypothético-déductif, qui est pourtant essentiel et fondamental (d'ailleurs, en géométrie, il est rarissime que l'on recoure à l'observation pour confirmer ou infirmer a posteriori un modèle théorique).

d. Une autre raison d'introduire un enseignement déductif de la théorie des probabilités est que c'est sans doute l'unique occasion que l'on ait, dans le secondaire, de faire connaître aux élèves la structure de théorie physique, qui est l'une des pierres d'angle — peut-être la plus importante — de la « culture scientifique » d'aujourd'hui. Cette structure est le moyen obligé de toute mathématisation d'un domaine réel quelconque. Mais quand nos élèves ont-ils l'occasion de la voir en action? Pas au cours de physique, assurément, sauf peut-être pour les éléments de la mécanique. Au niveau secondaire, la physique se veut — à juste titre — avant tout empirique, voire expérimentale [1]): la construction théorique abstraite d'un chapitre de physique dépasse les possibilités du niveau secondaire, à cause de la complexité des phénomènes mis en jeu et des outils mathématiques de leur analyse (exemple: seuls quelques spécialistes sont réellement à même d'aborder — par exemple au niveau d'une « sixth form » anglaise — l'édifice newtonien conduisant déductivement aux lois de Kepler, qui est pourtant l'exemple le plus élémentaire possible). L'enseignement déductif de la théorie des probabilités reste en fait le seul où les élèves puissent vraiment rencontrer la structure de théorie physique.

---

[1]) Là où elle reste livresque, elle prend un faux air de discipline déductive, sans base expérimentale réelle: c'est la pire des solutions; mais ce point n'est pas de la compétence des didacticiens de la mathématique.

2. L'époque convenable

Les considérations du paragraphe précédent répondent aussi à la question « quand se place le premier enseignement déductif de la théorie des probabilités? » Il est clair en effet qu'il doit se situer au niveau secondaire. Il est non moins clair qu'il est d'un degré d'abstraction tel que seuls des élèves déjà formés peuvent l'aborder. Il faut donc le placer aux environs des classes 11 et 12 de l'enseignement général (élèves de 16 à 18 ans). Il y aurait pourtant intérêt à le placer plus tôt: en effet, si ses premières notions font l'objet d'un enseignement au stade A dès l'école primaire (ou les toutes premières années du secondaire), il faut alors s'attendre que les enseignants des branches utilisatrices (notamment la biologie) s'en serviront librement avec les élèves de 13-14 ans et plus; il va donc se former, de ce fait, un « tas » passablement informe de notions et d'applications, tas qu'il importe d'organiser et de rationaliser sans trop attendre.

Une solution possible est de commencer, dès la classe de 15-16 ans, une systématisation de la statistique descriptive en termes tels que son transfert à la théorie des probabilités soit immédiat et cohérent: ceci nécessite l'introduction — qui, à ce niveau, ne peut être que bénéfique — de définitions et de méthodes « intrinsèques »; le seul problème que cela pose est d'ordre didactique: un tel enseignement a pour but de préparer l'avenir, et on peut donc craindre qu'il ne soit que faiblement motivé, puisque les seules applications possibles *à ce moment-là* ont déjà été rencontrées sous une forme plus élémentaire.

3. Esquisse d'un contenu

a. Le contenu du premier enseignement déductif des probabilités est déterminé par son but: faire connaître aux élèves la structure de *théorie physique* de la théorie des probabilités, en mettant l'accent *à la fois* sur les situations envisagées (situations aléatoires) et sur les moyens formels (hypothético-déductifs) mis en œuvre pour les analyser. Il faut donc remarquer que si cet enseignement était uniquement formel, s'il sacrifiait « les situations » à leur analyse mathématique, il manquerait totalement son but. C'est là, peut-être, la raison de l'opposition que certains marquent à son égard: s'ils confondent « déductif » avec « purement formel », cette opposition se comprend; et peut-être y a-t-il en effet des exemples de cette confusion. Il convient donc de tout faire pour l'éviter; mais cela n'entraîne pas que l'on renonce à tout enseignement déductif — bien au contraire!

b. Le *cadre formel* de la théorie physique des phénomènes fortuits est, pour virtuellement tous les auteurs, un cadre ensembliste: une situation aléatoire est décrite par l'*ensemble* des résultats possibles; les parties de cet ensemble sont les *événements*; la probabilité est une fonction à valeurs réelles, définie sur un certain ensemble d'événements.

Cette façon de faire a pourtant été critiquée (voir p. ex. [4]): l'appareil ensembliste ne serait rien de plus qu'une feuille de vigne, mise en avant au premier chapitre pour se conformer à une mode, mais oubliée ensuite, dès que l'on aborde sérieusement des problèmes réels, au profit d'une vue « classique » reposant sur les notions primitives de « probabilité d'un énoncé » et de « variable aléatoire ». Cette critique est bien difficile à comprendre. Il faut, avons-nous dit, partir de la notion même de *situation aléatoire*; or, qu'est-ce qui caractérise une telle situation? C'est qu'une constellation bien déterminée, fixe, de conditions de réalisation a, non pas une, mais plusieurs conséquences possibles; il est donc clair que l'ensemble des conséquences possibles, Ω, est un élément tout indiqué de la mathématisation de la situation. (Bien entendu, si on part brutalement d'un ensemble Ω sans référence à une situation aléatoire, on retombe dans le défaut, signalé ci-dessus, d'un enseignement purement formel.) D'autre part, on peut, à propos d'une situation aléatoire donnée, envisager toute une série de circonstances dont on peut dire, pour toute réalisation de cette situation, si elles ont lieu ou non: ce sont les *événements* liés à cette situation. Il est contre-indiqué d'appuyer fortement sur le fait qu'un événement se décrit par une proposition; en effet, la théorie des probabilités comme théorie physique des phénomènes fortuits n'est qu'une des interprétations possibles de la structure formelle générale de « calcul des probabilités » (voir [11], chap. 8); d'autres interprétations possibles ont comme base un ensemble de propositions [interprétation personnaliste, où la probabilité est un degré de croyance (de Finetti, Savage); interprétation logiciste, où la probabilité (conditionnelle) est une inférence partielle (Ramsey, Keynes)]. Ces diverses interprétations sont nettement distinctes, la distinction entre elles va beaucoup plus loin qu'une « réinterprétation du langage des événements en un langage de propositions (à savoir, quant à l'occurrence d'événements) »; en fait, leur confusion a été la plaie de la théorie pendant de nombreuses décennies, elle n'est même pas entièrement éliminée à l'heure actuelle et est, par exemple, à la base de la controverse très aiguë qui oppose, en matière de physique quantique, les tenants et les adversaires de l'interprétation « de Copenhague » (voir p. ex. [11] chap. 9 et [1] parag. 2.4 et chap. 5). Il est donc essentiel d'éviter toute confusion entre ces diverses interpétations possibles; mais si, dans l'exposé de la théorie physique des phénomènes fortuits, on définit un événement par une proposition, ou si — pire encore — on prend comme domaine de la fonction « probabilité » un ensemble de propositions, on rend cette confusion presque inévitable. Tout au contraire, la définition d'un événement comme partie de l'ensemble des possibles ne prête à aucune confusion; bien entendu, il importe, ici aussi, de bien montrer (par l'analyse d'exemples déjà étudiés d'un point de vue intuitif) le lien entre la situation et son modèle mathématique; on peut être raisonnablement assuré que les élèves ont perçu ce lien s'ils comprennent l'énoncé, véritablement fondamental, que voici:

> dans toute réalisation d'un phénomène fortuit $\mathfrak{F}$ [dont l'ensemble des possibles est Ω] un événement (partie de Ω) *se produit*
> si et seulement si
> le résultat réalisé de $\mathfrak{F}$ est un élément de cette partie de Ω.

[Si, dans ces conditions, certains élèves s'imaginent que, p. ex., l'événement impossible (ϕ) est l'absence de toute réalisation, la faute n'en incombe pas à la théorie mais à l'enseignant; « c'est un bien mauvais artisan, qui rejette la faute sur son outil ».]

Les termes *réalisation, résultat réalisé* sont tout entiers du côté de la situation; un enseignement formaliste, qui n'envisagerait pas de façon franche et explicite la situation aléatoire sous-jacente à tout problème probabiliste, serait dans l'impossibilité d'utiliser ces termes, et donc dans l'impossibilité de faire clairement comprendre la notion d'événement. Notons que la notion intuitive d'événement est passablement vague, et n'acquiert un sens précis que par sa définition mathématique (état de chose qui n'est pas sans rappeler les rapports entre « vitesse » et « dérivée »: la notion intuitive de vitesse ne prend un sens précis que par la définition de la dérivée; inversement, l'interprétation d'une dérivée comme une vitesse n'est possible que par référence à une situation concrète d'un type bien déterminé; seulement, l'analyse mathématique n'est pas la « théorie physique » du mouvement local!).

Toutefois, la définition d'un événement comme « partie de Ω » peut dépasser son but: il peut arriver qu'un événement (défini comme partie de Ω) soit d'une complication telle que son occurrence (ou sa non-occurrence) ne soit pas observable, dans la situation envisagée. C'est pourquoi il faut bien mettre en évidence que

> la description mathématique d'une situation aléatoire comprend non seulement l'ensemble Ω des résultats possibles mais encore l'ensemble $\mathscr{T}$ des événements dont l'occurrence est observable (du moins en principe) dans *cette* situation.

Il n'est pas nécessaire d'insister fortement sur ce dernier point dans un premier enseignement déductif (d'autant moins que des exemples élémentaires ne sont pas aisés à trouver); il faut pourtant faire voir que l'ensemble des événements d'occurrence observable est stable pour les opérations ensemblistes $\cap$, $\cup$, et $\complement$. Faut-il aller au delà et introduire la notion de tribu? Ce serait là sans doute un excès de formalisme. Pourtant, certains exemples élémentaires que l'on traite volontiers au stade *A* introduisent tout naturellement des événements infinis; supposons par exemple que l'on étudie le jeu suivant:

> pile ou face jusqu'à la première apparition soit de *PPF* soit de *PFP*;

ce jeu serait représenté (voir par exemple [12]) par le diagramme suivant:

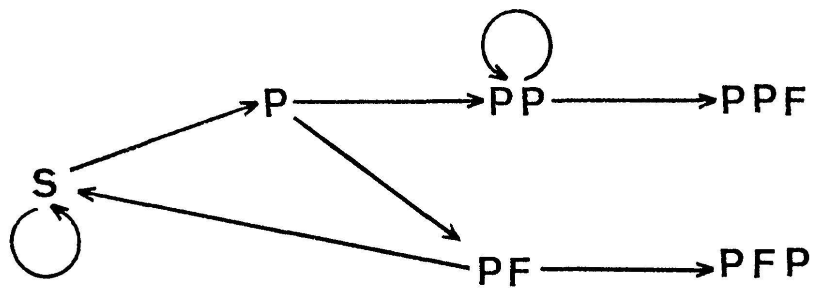


Fig. 1

L'événement (manifestement observable)

> *X*: « entre la *première* occurrence de l'état *PP* et la fin du jeu il y a exactement trois parties »

est un événement infini, car les boucles (S) et (S, P, PF, S) peuvent être parcourues un nombre quelconque de fois:

$$X = \{(S, P, PP, PP, PP, PPF),\ (S, P, PF, S, P, PP, PP, PP, PPF),\\ (S, P, PF, S, P, PF, S, P, PP, PP, PP, PPF), \dots\}$$

Il ne faut donc pas éviter systématiquement ce type d'événements. Mais, dans un premier enseignement déductif, il suffit très largement de considérer de temps à autre une suite décroissante d'événements (et de faire appel au troisième axiome de Kolmogorov; voir parag. d. ci-dessous).

c. Bien entendu, c'est la notion même de probabilité qui est le nœud essentiel d'un premier enseignement déductif. A ce propos plus encore que pour les points précédents, il faut redire qu'un enseignement déductif de la théorie physique des phénomènes fortuits ne peut négliger ni la situation, ni la formalisation mathématique; présenter la probabilité simplement comme une mesure sur un ensemble de parties, c'est négliger la situation; faire de la probabilité un élément de description d'une « variable aléatoire » considérée comme notion première, c'est faire bon marché de la construction hypothético-déductive; plaquer une fonction additive sur un ensemble d'événements n'est pas analyser une situation.

Sans doute n'y a-t-il plus, à l'heure actuelle, aucun partisan de la « définition » par le rapport des nombres de cas favorables et de cas possibles. Par contre, la plupart des exposés déductifs à l'usage des débutants font abondamment usage, lorsqu'ils abordent la « définition » de la probabilité, de la notion de fréquence. C'est là, à n'en pas douter, une conséquence d'un désir louable de ne pas faire fi de la situation sous-jacente; mais c'est une façon déplorable de présenter les choses. Il semble bien, que pour la plupart des auteurs, le message — pourtant si clair — de K. Popper a été perdu: la fréquence n'a aucun rôle à jouer dans la *définition* de la probabilité; un phénomène fortuit pourrait être totalement dépourvu de structure répétitive, la notion de probabilité s'y appliquerait quand même, sans la moindre restriction, alors qu'aucune fréquence ne serait en jeu. Si l'enseignement intuitif (le stade *A*) des probabilités a été bien fait, il a mis les élèves en présence de ce fait fondamental: dans toute situation aléatoire, chaque événement est doué d'une *tendance* plus ou moins forte à se produire.

[La langue commune dispose, pour signifier cette tendance, des mots « chance » et « risque », mais ceux-ci ne sont pas neutres: « chance » est presque toujours utilisé lorsqu'il s'agit de conséquences favorables, « risque » toujours pour des conséquences défavorables; la construction de la théorie nécessite l'emploi d'un mot neutre, et c'est pour satisfaire à cette nécessité que Popper a emprunté au vocabulaire des sciences sociales le terme (excellent) de « propension »; mais « tendance à se produire » fait parfaitement l'affaire, et est moins ésotérique.]

Une fraction notable du temps consacré à un premier enseignement déductif des probabilités doit être employé à faire ressortir cette notion de « tendance à se produire » de l'analyse des exemples déjà vus par les élèves et des mots de la langue commune qui y sont utilisés.

Ce point est sans doute celui où un premier enseignement déductif (stade *B*) bien fait aurait le plus d'influence sur l'enseignement intuitif qui le prépare (stade *A*); les maîtres chargés de l'enseignement au stade *A* ont fort souvent en vue un enseignement *ultérieur* essentiellement fréquentiste, et se laissent ainsi entraîner dans cette direction; alors que l'on attend d'eux, tout au contraire, que leur introduction aux situations aléatoires fasse clairement percevoir la variabilité (d'une réalisation à l'autre) des résultats réalisés, et les diverses tendances à l'occurrence des événements considérés. Cette variabilité et cette diversité des tendances ne peuvent se constater *empiriquement* que si on dispose de plusieurs réalisations d'une même situation aléatoire: il est évident qu'une réalisation unique ne fournit qu'un seul résultat ! C'est là le vrai rôle de la répétition. D'autre part, il existe de nombreuses situations élémentaires où des considérations de symétrie suggèrent l'hypothèse [1]) que certains événements (formant une partition de Ω) ont chacun la même propension, et donc la même probabilité; il est évidemment tout indiqué de déduire les conséquences de cette hypothèse et de se demander ensuite si on a des raisons *empiriques* de la croire justifiée; la répétition intervient ici aussi: car si l'hypothèse de symétrie est réalisée, on constate empiriquement que les événements équivalents ont des fréquences à peu près égales; c'est par ce biais que s'installe peu à peu, *a posteriori*, l'idée que, dans des conditions convenables, la fréquence réalisée d'un événement est une estimation empirique de sa probabilité. Cette idée reste d'ailleurs purement intuitive, aussi longtemps que l'on n'a pas déduit des axiomes de la théorie une « loi des grands nombres » (par exemple celle de Bernoulli). Quand une telle loi a été déduite, elle permet d'avancer d'un pas de plus, et de constater que, pour *une seule* réalisation d'une situation aléatoire suffisamment répétitive, la fréquence moyenne d'une suite d'événements *de même probabilité* est aussi une estimation empirique de cette probabilité. En résumé: les notions fondamentales sont celles de variabilité anarchique des résultats réalisés, de tendance plus ou moins forte à l'occurrence, et de *mesure* (additive) de cette tendance (au sens où on dit «mesure des grandeurs»; voir d. ci-dessous); l'observation des fréquences n'a qu'un rôle (important certes mais) subordonné.

d. Une fois acquise la notion intuitive de propension (tendance à l'occurrence) et le fait qu'elle est susceptible de plus ou de moins, il s'indique d'essayer de la *mesurer*. Un premier enseignement déductif doit rappeler que la mesure quantitative précise (de grandeurs *d'abord* intuitivement perçues comme susceptibles de plus ou de moins) repose toujours sur *deux* définitions, celle de l'additivité et celle de l'unité de mesure (un excel-

---

[1]) Voir paragraphe 6 ci-dessous.

lent exemple est la mesure des débits de courants électriques, qui s'ajoutent en parallèle, ou des tensions électriques, qui s'ajoutent en série). Cela conduit tout droit aux deux premiers axiomes de Kolmogorov, sous une forme équivalente (en moins formel, peut-être) à

$$\Pr : \mathscr{T} \to \mathbf{R}^+ : A \mapsto \Pr A$$
$$K1) \ \forall \{A, B\} \subset \mathscr{T} ; \ [A \cap B = \phi] \Rightarrow [\Pr(A \cup B) = \Pr A + \Pr B]$$
$$K2) \ \Pr \Omega = A$$

et aux conséquences qui en découlent immédiatement (telles que $\Pr \phi = 0$, monotonie, additivité finie, relation et inégalités de Boole, etc). Le troisième axiome:

$$K3) \ \forall \{A_1, \ldots, A_n, \ldots\} \subset \mathscr{T} ; \ (A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots \& \bigcap_1^\infty A_i = \phi) \Rightarrow \lim \Pr A_n = 0$$

paraît tellement évident que la plus grosse difficulté à son sujet est de persuader les élèves qu'il doit être énoncé; il est d'ailleurs impossible de trouver un exemple contraire qui soit à la fois élémentaire et intéressant; mais il faut le mentionner car, dans la suite, on est amené à s'en servir [1]. [Si le maître, pour son usage personnel, se contente d'un exemple élémentaire mais non intéressant, qu'il prenne pour Ω l'ensemble $\mathbf{N}$, avec $\mathscr{T}$ et P définis par

$$\forall X \subset \mathbf{N} ; \ [X \in \mathscr{T} \Leftrightarrow (X \text{ est fini}) \vee (\complement X \text{ est fini})]$$
$$P(X) = \sum \left(\frac{1}{(i+2)(i+3)} \middle| i \in X \right) \text{ si } X \text{ est fini}$$
$$= 1 - P(\complement X) \qquad \text{si } \complement X \text{ est fini.}$$

Si on prend alors $A_n = \mathbf{N} \backslash \{0, 1, \ldots, n\}$, la suite $(n \mapsto A_n)$ ne satisfait pas au troisième axiome de Kolmogorov.] Par contre, on peut trouver des exemples, non dépourvus d'intérêt, de l'usage d'une suite infinie d'événements. Ainsi, dans le jeu « pile ou face jusqu'à la première occurrence de *PPF* ou de *PFP* » (parag. 3b. ci-dessus), on constate que l'événement

$$X = \text{« le jeu se prolonge indéfiniment »}$$

n'est pas vide [il contient notamment l'élément

$$S, P, PF, S, P, PF, S, P, PF, S, P, PF, \ldots \text{ (ad inf)}] ;$$

bien entendu, les élèves sont absolument persuadés que cet événement ne peut pas se produire; on renforce donc leur confiance dans la mathéma-

[1]) Voir paragraphe 5 e. ci-dessous.

tisation adoptée en démontrant que Pr $X = 0$ (on suppose que Pr a été défini comme dans les exemples du stade A: on adjoint le nombre ½ à chaque flèche, et on multiplie les nombres relatifs à toutes les flèches parcourues). [Cette démonstration n'est possible qu'à l'aide des probabilités conditionnelles, mais rien n'interdit de les utiliser avant d'en avoir fait un exposé déductif.] Si A_n est l'événement « le jeu dure au moins n parties », il est clair (puisque chaque état peut conduire à un état final en trois coups au plus) que

$$\Pr(\complement A_{n+3} \mid A_n) \geqslant \frac{1}{8}, \quad \Pr(A_{n+3} \mid A_n) \leqslant \frac{7}{8}; \tag{1}$$

d'autre part, et bien évidemment,

$$\Pr(A_{n+3} \mid \complement A_n) = 0 \quad [\text{car } A_{n+3} \subset A_n] \tag{2}$$

donc

$$\Pr(A_{n+3}) \leqslant \left(\frac{7}{8}\right) . \Pr(A_n) .$$

Or

$$\Pr(A_1) = 1$$

et donc

$$\Pr(A_4) \leqslant 7/8 , \quad \Pr(A_7) \leqslant (7/8)^2 , \ldots, \quad \Pr(A_{3k+1}) \leqslant (7/8)^k ; \tag{3}$$

alors

$$\begin{aligned} &\forall k \in \mathbf{N}_0 ; \quad X \subset A_{3k+1} \Rightarrow \\ &\quad \Rightarrow \forall k \in \mathbf{N}_0 ; \quad \Pr X \leqslant (7/8)^k \Rightarrow \Pr X = 0 , \qquad \text{c.q.f.d.} \end{aligned} \tag{4}$$

e. A côté de la notion (fondamentale) de probabilité, celle de conditionnement (et, en conséquence, celle de probabilité conditionnelle) est d'une importance extrême. Mais on a ici un exemple particulièrement frappant de la nécessité de prendre en considération à la fois la situation et le formalisme mathématique. La définition purement formelle

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr B}$$

ne sert *strictement* à rien; ce qu'il faut faire, c'est reprendre dans les exemples vus par les élèves toutes les situations où on a été amené (le plus souvent de façon tout implicite) à *contracter* le phénomène fortuit $\mathfrak{F}$ à l'un de ses événements, disons A, obtenant ainsi le phénomène contracté $\mathfrak{A}$; à rappeler ensuite qu'une probabilité n'est autre que la mesure d'une propension, avec la propension de l'événement certain comme unité de mesure; distinguant alors (pour un temps seulement) les fonctions-propabilité pour $\mathfrak{F}$, $\Pr_{\mathfrak{F}}$, et pour $\mathfrak{A}$, $\Pr_{\mathfrak{A}}$, rappelant en outre la notion de rapport de deux grandeurs et ce qui se passe quand on change d'unité de mesure, on a

$$\forall\, X \subset A;\ \Pr_{\mathfrak{A}} X = \frac{\text{propension de X}}{\text{propension de A}} \tag{6a}$$

$$= \frac{\text{propension de X}}{\text{propension de } \Omega}\ \frac{1}{\dfrac{\text{propension de A}}{\text{propension de } \Omega}} \tag{6b}$$

$$= \frac{\Pr_{\mathfrak{F}} \mathrm{X}}{\Pr_{\mathfrak{F}} \mathrm{A}}\,. \tag{6c}$$

Cette formulation est la seule véritablement utilisée, mais elle est dissimulée par la notation usuelle,

$$\Pr(Y \mid A) \quad \text{pour} \quad \Pr_{\mathfrak{A}}(Y \cap A)\,,$$

qu'il faut évidemment se résigner à adopter. [$\Pr(. \mid A)$ étant la fonction-probabilité de $\mathfrak{A}$, il est immédiatement évident que c'est une probabilité !]

La *catégorisation* d'un phénomène fortuit doit être bien distinguée de sa contraction: il s'agit, cette fois, de réduire non pas Ω, mais $\mathscr{T}$. Dans un premier enseignement déductif, cette réduction de $\mathscr{T}$ est définie par une partition finie ou dénombrablement infinie: on ne considère plus comme observables les événements de $\mathscr{T}$, mais seulement ceux de

$$\mathscr{A} = \{ A_1, \ldots, A_n \} \quad [\text{ou: } \ldots, A_n, \ldots \}]$$

et leurs unions (il y a donc bien réduction de $\mathscr{T}$). C'est une étape indispensable dans la définition de l'indépendance. Mais, ici encore, la considération *simultanée* de la situation et du formalisme est indispensable. En effet, le point de départ est celui-ci: si les partitions $\mathscr{A}$ et $\mathscr{B}$ décrivent *deux* catégorisations du phénomène fortuit $\mathfrak{F}$, ces phénomènes réduits (notons-les $\mathfrak{A}$ et $\mathfrak{B}$) s'influencent-ils l'un l'autre, ou non? (C'est là une question qui n'aurait même aucun sens si on ne se référait pas explicitement à la situation elle-même.) La réponse s'exprime en termes de probabilités conditionnelles:

A n'influence pas *B* si et seulement si

$$\forall A_i \in \mathscr{A}\,,\ \ \forall B_k \in \mathscr{B}\,;\ \Pr(B_k \mid A_i) = \Pr(B_k)$$

(et ce en vertu de la signification, *pour la situation*, des probabilités conditionnelles en question). On tire alors aisément de là l'exposé élémentaire usuel.

[Notons qu'il n'est pas question, au départ, d'*événements* indépendants, mais de *phénomènes* (réduits) indépendants; les « événements » indépendants correspondent en fait aux partitions du type $\{ A, \complement A \}$; en termes de situation: le phénomène $\mathfrak{A}$ est $\mathfrak{F}$ réduit au point qu'on n'observe plus rien que l'occurrence ou la non-occurrence de A; il en est de même pour $\mathfrak{B}$, et on exprime que $\mathfrak{A}$ et $\mathfrak{B}$ sont sans interaction mutuelle].

Il est clair que, ainsi référées à la situation sous-jacente, les notions de probabilité conditionnelle et d'indépendance sont autrement riches de sens que les simples définitions formelles

$$\operatorname{Pr}(A \mid B) = \frac{\operatorname{Pr}(A \cap B)}{\operatorname{Pr} B}, \qquad \operatorname{Pr}(A \cap B) = \operatorname{Pr} A \,.\, \operatorname{Pr} B \,.$$

C'est cette richesse qui permet leur utilisation; s'en tenir à un exposé formel, et s'attendre néanmoins que les élèves seront à même de l'appliquer, c'est se fier à une pure et simple régression du stade B au stade A: il vaudrait certes mieux, dans ce cas, supprimer le stade B.

4. Étude d'un exemple

a. Revenons à l'exemple du paragraphe 3.b; le phénomène fortuit $\mathfrak{F}$ considéré est donc le suivant:

une suite de parties de « pile ou face » poursuivie jusqu'à la première apparition soit de *PPF* soit de *PFP*.

Voilà une situation aléatoire qu'il s'agit de mathématiser en faisant apparaître, successivement, l'ensemble des possibles, Ω, l'ensemble des événements à considérer, $\mathcal{T}$, et la fonction-probabilité.

b. L'ensemble des possibles peut être « décrit » en extension, sous la forme

$$\begin{aligned}\Omega = \{ & PPF, FPPF, PPPF, FFPPF, FPPPF, PPPPF, PFFPPF, \ldots \\ & PFP, FPFP, FFPFP, FFFPFP, PFFPFP, \ldots \}\end{aligned}$$

mais cette manière est bien peu « déductive »!; il peut aussi être décrit en compréhension:

ensemble des « *mots* » formés des seules lettres P et F et terminés à la première apparition soit de la séquence *PPF* soit de la séquence *PFP*

et des *suites* formées des seules lettres P et F et qui ne contiennent aucune séquence *PPF* ou *PFP*.

A ces descriptions, correctes mais peu intéressantes, on préférera sans doute une représentation géométrique. La plus immédiate est l'arbre qui correspond à la description en extension signalée ci-dessus:

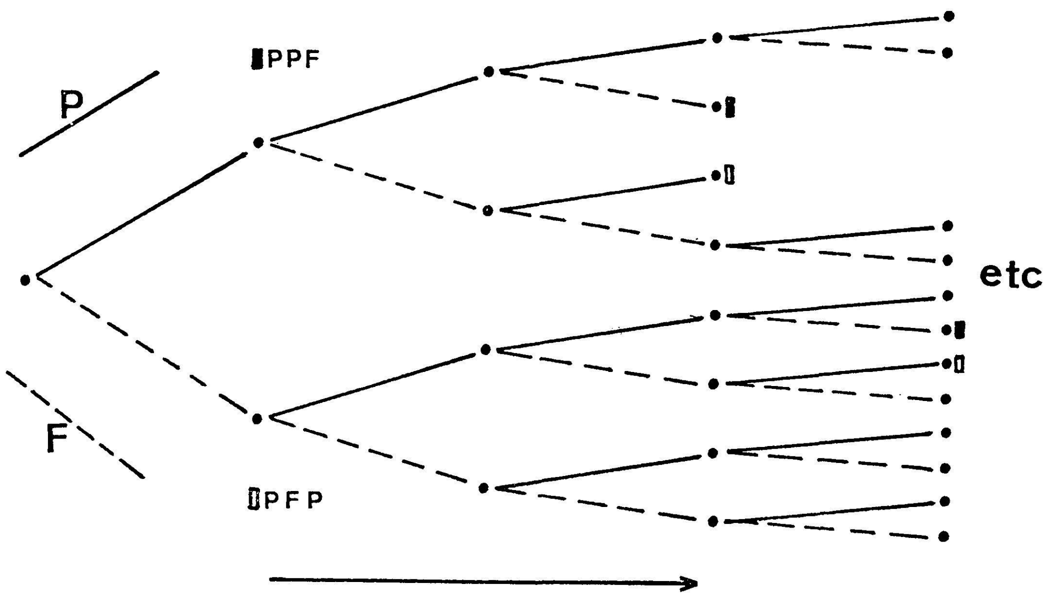


Fig. 2

Cette représentation est, elle aussi, fort compliquée. On trouvera peut-être convenable de la simplifier en tenant compte du degré de réalisation des états finals, de sorte que le système ne peut être que dans les états *S*, *P*, *PP*, *PF*, *PPF*, *PFP*; on a ainsi la représentation géométrique que voici:

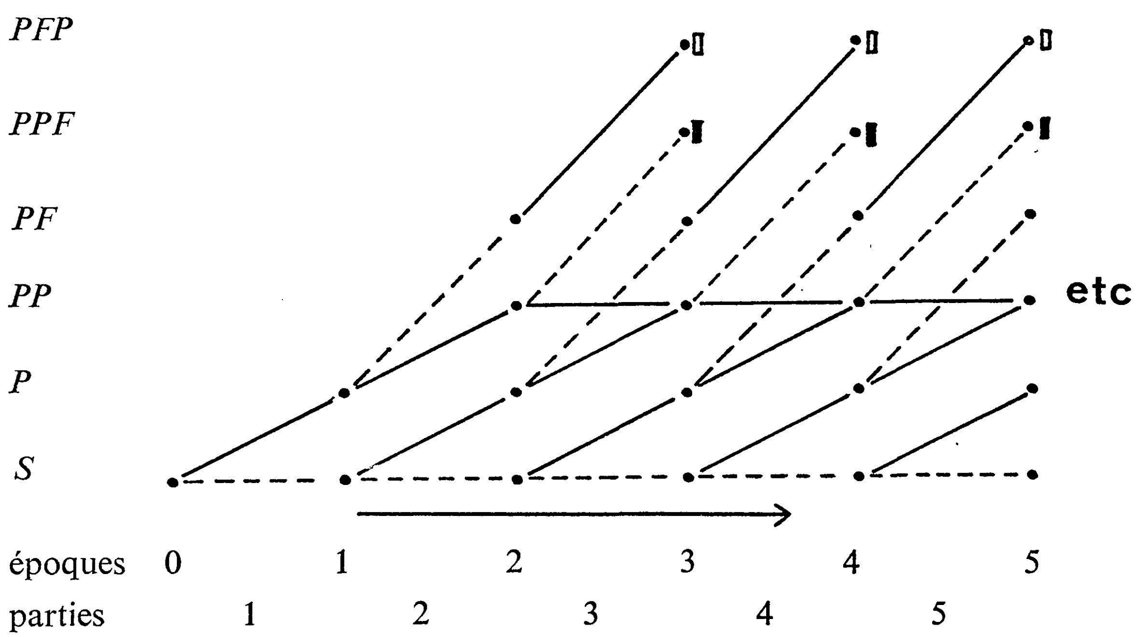


Fig. 3

Au stade A, on habitue volontiers les élèves à simplifier davantage encore, en faisant abstraction des époques auquelles les états sont atteints: on a ainsi la représentation de la figure 1.

Il semble que, comme représentation de Ω, les trois figures (2), (3), et (1) sont également acceptables, et que, par raison de simplicité, (1) est meilleure que (3) et (3) meilleure que (2). C'est là une grave et dangereuse illusion, comme on le verra en d. ci-dessous.

c. L'ensemble $\mathscr{T}$ des événements observables n'a pas besoin d'être décrit en détail: il suffit de remarquer qu'il contient tous les « débuts », c'est-à-dire tous les ensembles d'éléments de Ω qui ont une partie initiale donnée; par exemple,

deb (*FFFPFP*) [qui est un singleton de Ω]

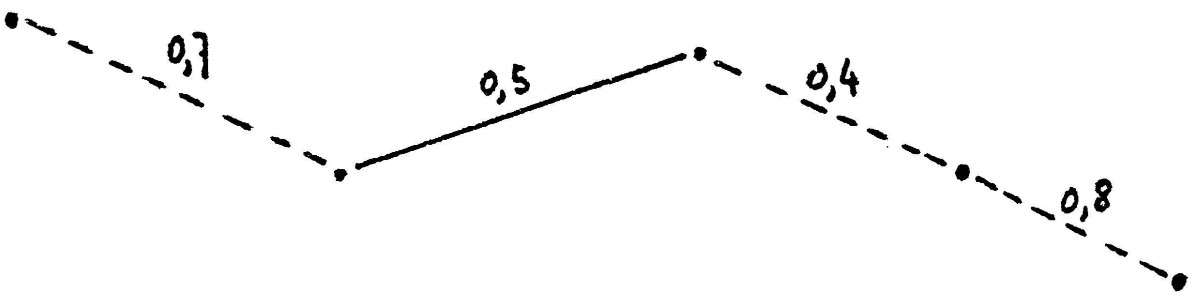
deb (*FFPFFF*) [qui contient une infinité de singletons de Ω, par exemple *FFPFFFPPF* et *FFPFFFPFP*]

deb (*FFPPPP*) [qui contient les singletons *FFPPPPF*, *FFPPPPPF*, *FFPPPPPPF*, *FFPPPP* ... *PPF*, etc.]

ainsi que leurs complémentaires, unions, et intersections.

d. En ce qui concerne la probabilité définie sur un tel ensemble de possibles, il est très important qu'un premier enseignement déductif fasse clairement ressortir les points suivants:

1) il y a beaucoup de définitions de Pr qui sont *mathématiquement* possibles
2) chacune d'elles peut s'obtenir en plaçant sur chaque branche de l'arbre (fig. 2) un nombre compris entre 0 et 1 et définissant la probabilité d'un début (singleton ou non) comme produit des nombres portés par les branches qui le constituent; ainsi on pourrait avoir:



$0{,}7 = p_F$ $\quad$ $0{,}5 = p_{(F)P}$ $\quad$ $0{,}4 = p_{(FP)F}$ $\quad$ $0{,}8 = p_{(FPF)F}$

Pr deb *(FPFF)* $= 0{,}7 \times 0{,}5 \times 0{,}4 \times 0{,}8 = 0{,}112$

Fig. 4

(les deux branches issues d'un même point doivent porter des nombres dont la somme est 1: c'est une conséquence de K1)

3) le choix de l'une de ces probabilités n'est possible que par une analyse de la situation elle-même, et plusieurs choix différents sont manifestement possibles, même en s'en tenant aux situations simples.

Montrer que le procédé multiplicatif mentionné en 2) donne bien une probabilité peut être l'occasion d'un commentaire éclairant sur la portée d'une axiomatique: pour qu'une application de $\mathscr{T}$ dans $\mathbf{R}^+$ soit une probabilité, il suffit qu'elle satisfasse aux axiomes K1, K2, et K3 du paragraphe 3d. ci-dessus; or, on a

$$\operatorname{Pr} \Omega = \operatorname{Pr}[\operatorname{deb}(P) \cup \operatorname{deb}(F)] = \operatorname{Pr} \operatorname{deb}(P) + \operatorname{Pr} \operatorname{deb}(F) = 1$$

et aussi (quels que soient $x, y, ..., u$):

$$[\operatorname{deb}(x, y, \ldots, u, P) \cup \operatorname{deb}(x, y, \ldots, u, F)] = \operatorname{deb}(x, y, \ldots, u)$$
$$[\operatorname{deb}(x, y, \ldots, u, P) \cap \operatorname{deb}(x, y, \ldots, u, F)] = \phi$$

$$\begin{aligned} &\operatorname{Pr} \operatorname{deb}(x, y, \ldots, u, P) + \operatorname{Pr} \operatorname{deb}(x, y, \ldots, u, F) \\ &= p_x \times p_{(x)y} \times \cdots \times p_{(x,y,\ldots)u} \times p_{(x,y,\ldots,u)P} \\ &+ p_x \times p_{(x)y} \times \cdots \times p_{(x,y,\ldots)u} \times p_{(x,y,\ldots,u)F} \\ &= p_x \times p_{(x)y} \times \cdots \times p_{(x,y,\ldots)u} = \operatorname{Pr} \operatorname{deb}(x, y, \ldots, u) . \end{aligned} \tag{7}$$

Quant à l'axiome K3, il n'est pas facile de démontrer qu'il est vérifié; le conflit ainsi mis à jour entre l'intuition d'un procédé (multiplicatif) auquel les élèves sont accoutumés et l'apparente évidence de K3 est, lui aussi, très instructif. D'autre part, montrer que le procédé multiplicatif définit *toutes* les probabilités possibles résulte de la théorie des probabilités composées; par exemple

$$\begin{aligned} \operatorname{Pr} \operatorname{deb}(x, y, z) &= \operatorname{Pr} \operatorname{deb}(x) \,.\, \operatorname{Pr}[\operatorname{deb}(x, y) \mid \operatorname{deb}(x)] \,. \\ &\quad .\ \operatorname{Pr}[\operatorname{deb}(x, y, z) \mid \operatorname{deb}(x) \cap \operatorname{deb}(x, y)] \\ &= \operatorname{Pr} \operatorname{deb}(x) \,.\, \operatorname{Pr}[\operatorname{deb}(x, y) \mid \operatorname{deb}(x)] \,. \\ &\quad .\ \operatorname{Pr}[\operatorname{deb}(x, y, z) \mid \operatorname{deb}(x, y)] \\ &\qquad [\text{car } \operatorname{deb}(x, y, z) \subset \operatorname{deb}(x, y) \subset \operatorname{deb}(x)] . \end{aligned}$$

Des situations concrètes correspondant à ces diverses possibilités devraient être montrées aux élèves dès le stade *A*. Il semble bien qu'à l'heure actuelle on l'oublie fréquemment. Cela est dû à la vogue, parmi les didacticiens, des problèmes où l'ensemble des possibles prend la forme d'un arbre que l'on peut condenser en un graphe fléché et bouclé (comme celui de la fig. 1); or, cette condensation n'est possible que pour *certains* choix de la probabilité (schémas à transitions markoviennes stationnaires); les graphes ainsi condensés sont donc inaptes à représenter l'ensemble des possibles, puisque celui-ci est logiquement antérieur à la définition de la probabilité (preuve: le domaine de Pr est une partie de $\mathscr{P}\Omega$). Il y a là une situation qu'il importe de corriger au stade *B*, avec l'espoir que la correction se propagera au stade *A*.

5. Variables aléatoires

a. Depuis que la réforme de l'enseignement mathématique au niveau secondaire a amené à sa place (l'une des toutes premières) la notion de fonction, il n'est vraiment pas difficile de faire voir aux élèves, dès le stade *A*, que de nombreux éléments intéressants d'une situation aléatoire quelconque sont des applications de l'ensemble des possibles (Ω) dans l'ensemble des réels ($\mathbf{R}$). Au stade *B*, une révision de ces exemples conduit à la définition explicite.

b. Le terme « variable aléatoire » a été critiqué comme impropre, et il est bien vrai qu'une application de Ω dans $\mathbf{R}$ n'a rien de variable ni d'aléatoire. Si on estime que cette critique est justifiée, rien n'empêche de changer de terme, et de dire par exemple « aléa numérique réel ». Mais on peut estimer aussi que cette critique vient de mathématiciens « purs », qui ont privilégié le formalisme et oublié la situation; en effet, d'un point de vue concret, ce qui est en jeu est bel et bien une grandeur qui, *d'une réalisation à l'autre* [1]), peut prendre diverses valeurs — c'est donc bien une *variable* — et ce de façon anarchique — c'est donc bien une variable *aléatoire*. Il y a là une querelle qui n'a ni importance ni intérêt.

c. Il est beaucoup plus important, au stade *B*, de faire ressortir ce qui suit: pour un élément aléatoire $\boldsymbol{x}$ à valeurs réelles, il est de la plus haute importance que, pour tout intervalle I, l'événement « la valeur réalisée de $\boldsymbol{x}$ appartient à I » soit observable; autrement dit (en notant $\mathscr{I}$ l'ensemble des intervalles de $\mathbf{R}$), l'application

$$\boldsymbol{x}: \ \Omega \to \mathbf{R} : \omega \mapsto \boldsymbol{x}(\omega)$$

ne mérite le nom de « variable aléatoire » que si elle satisfait à la condition

$$\forall I \in \mathscr{I}; \qquad \underset{[= \{\omega \setminus \boldsymbol{x}(\omega) \in I\}]}{\overset{-1}{\boldsymbol{x}}\, I} \qquad \in \mathscr{T}.$$

[1]) Nous avons déjà dit, plusieurs fois, que la notion même de réalisation est étrangère au formalisme probabiliste.

Dès lors, la définition de la fonction de répartition de x ne recèle plus aucun mystère:

$$F_x : \mathbf{R} \to \mathbf{R} : a \mapsto \Pr(\overset{-1}{x}\,]-\infty;a]) .$$

d. Les spécialistes contemporains de la théorie mathématique des probabilités manifestent une désaffection très nette vis-à-vis de la fonction de répartition, et lui substituent volontiers la « loi de probabilité ». Celle-ci est, somme toute, aisée à définir; d'une part,

$$\mathscr{X} = \{ A \subset \mathbf{R} \mid \overset{-1}{x} A \in \mathscr{T} \}$$

et d'autre part,

$$L_x : \mathscr{X} \to \mathbf{R} : A \mapsto \Pr(\overset{-1}{x} A) .$$

Mais néanmoins, cette tendance ne doit pas affecter le premier enseignement déductif de la théorie; d'une part, parce que la loi de probabilité ne peut, à ce stade, jouer aucun rôle [1]); d'autre part, parce que la fonction de répartition joue un rôle essentiel et irremplaçable: celui d'une description géométrique des propriétés de la variable aléatoire.

e. Un enseignement déductif des probabilités se doit de justifier son titre en démontrant, en toute rigueur et en toute généralité, non seulement que toute fonction de répartition (F) est croissante (c'est facile) mais encore qu'elle jouit des propriétés

$$\lim_{\to -\infty} F = 0 \qquad \lim_{\to +\infty} F = 1$$

$$\forall a \in \mathbf{R} ;\ \Pr(\overset{-1}{x}\,]-\infty;a[) = \sup\{ F(t) \mid t < a \},$$

ce qui nécessite l'intervention de l'axiome K3.

f. On peut estimer qu'un *premier* enseignement déductif doit s'en tenir aux variables aléatoires à nombre fini de valeurs (encore que cette opinion soit sujette à de très sérieuses objections; par exemple, la situation décrite par la figure 1 met en jeu des variables aléatoires à une infinité de valeurs). Mais il serait déplorable que les définitions et énoncés rencontrés dans cet enseignement soient applicables au seul cas fini. Bien au contraire, pour que ce premier enseignement déductif ne soit pas un

---

[1]) Pour les mathématiciens professionnels, elle permet la considération de l'énoncé, en effet essentiel (où g est une fonction de $\mathbf{R}$ vers $\mathbf{R}$)

$$\int_\Omega (g \circ x)\, dPr = \int_{\mathbf{R}} g\, dL_x .$$

obstacle à un développement ultérieur, il faut que les définitions soient parfaitement générales et que les énoncés soient aussi généraux que possible, quitte à ne les démontrer que pour le seul cas fini. En ce qui concerne les variables aléatoires et leurs valeurs typiques, on y arrive en exploitant les propriétés géométriques de la fonction de répartition; par exemple, on prend comme définition de la moyenne μ de x la relation *géométrique* (égalité d'aires)

$$\int_{-\infty}^{\mu} F(t)\,\mathrm{d}t = \int_{\mu}^{+\infty} [1 - F(t)]\,\mathrm{d}t\,, \tag{8a}$$

tandis que l'écart-moyen V et la variance σ^2 sont définis par les expressions *géométriques* (sommes de deux aires)

$$V = \int_{-\infty}^{\mu} F(t)\,\mathrm{d}t + \int_{\mu}^{+\infty} [1 - F(t)]\,\mathrm{d}t \tag{8b}$$

$$\frac{\sigma^2}{2} = \int_{-\infty}^{\mu} \mathrm{d}t \int_{-\infty}^{t} F(s)\,\mathrm{d}s + \int_{\mu}^{+\infty} \mathrm{d}t \int_{t}^{+\infty} [1 - F(s)]\,\mathrm{d}s\,. \tag{8c}$$

[L'aspect *géométrique* (en termes d'aires) des expressions (8) est particulièrement important: il s'agit de faire « voir » la signification des paramètres μ, V, σ; un traitement analytique basé sur les formules (8) est, à ce niveau, entièrement à rejeter (bien qu'il soit parfaitement correct).]
Il est alors possible de démontrer en toute généralité que V est la moyenne de la variable aléatoire $|\, x - \mu \,|$; par contre, démontrer que σ^2 est la moyenne de la variable aléatoire $(x - \mu)^2$ n'est possible, avec ces moyens, que pour le cas des variables aléatoires dont l'ensemble des valeurs est soit fini soit inclus dans **N**.

De cette façon, un premier enseignement déductif atteint parfaitement son but: il met en place et il organise les notions intuitives acquises au stade antérieur, sans rendre plus difficile, mais au contraire en préparant, l'enseignement plus théorique qui, pour certains élèves, lui fera suite.

6. Introduction a la statistique

a. S'il convient, ou non, d'introduire à la fin de l'enseignement secondaire un premier enseignement systématique de la statistique inférentielle est une question controversée; il y a des arguments en sa faveur (p. ex., que dans l'enseignement supérieur ces éléments sont souvent utilisés avant d'être systématiquement envisagés), il y a aussi de puissants arguments contre (c'est une théorie trop délicate et trop abstraite pour des élèves de cette classe d'âge; on ne peut valablement l'exposer qu'à partir d'exemples réels, et ceux-ci sont trop complexes; on risque de « former » de pseudo-spécialistes). La question restera, ici, non décidée.

b. Mais on ne peut absolument pas échapper à la question, plus générale: «dans quelle mesure tel ou tel modèle mathématique d'une situation donnée est-il une représentation adéquate de cette situation?». Puisque le premier enseignement déductif des probabilités doit tenir compte *à la fois* de la situation et du formalisme, il ne suffit pas que l'examen attentif de la situation ait suggéré des hypothèses, et que l'emploi judicieux du formalisme en ait déduit des conséquences; il *faut* que l'analyse se termine par un retour à la situation, et la comparaison des conséquences déduites aux résultats observables.

c. De ce point de vue, les situations hautement symétriques jouent un rôle d'une grande importance. La tendance actuelle est, au stade A, de ne pas s'en tenir à ces situations-là, mais de familiariser les élèves avec des situations plus compliquées: c'est une tendance fort heureuse; mais néanmoins, c'est à partir des situations hautement symétriques que se fait le premier examen du mode de comparaison de la théorie à l'expérience. Voici par exemple la situation aléatoire composée de quatre tirages successifs d'une même urne $U(3V, 3R)$ [1]). L'ensemble des possibles, Ω, se compose des 16 « mots » formés de 4 lettres V ou R, et on a $\mathcal{T} = \mathcal{P}\Omega$ [on va employer des notations du type

$$(*x*y) = \{(abcd) \in \Omega \,\big\}\, b = x \,\&\, d = y\,\}] .$$

L'examen attentif des 6 billes de l'urne montre qu'elles sont toutes sphériques, qu'elles ont toutes le même diamètre et la même masse, que leurs états de surface sont indiscernables. Ceci suggère avec force que V et R jouent des rôles interchangeables, et donc que l'application Pr doit être invariante vis-à-vis de l'échange de V et R. Cette *hypothèse* de symétrie entraîne donc que

$$H_1 \left\{ \begin{array}{ll} \Pr(V{*}{*}{*}) = \Pr(R{*}{*}{*}), & \left[= \frac{1}{2}, \text{ car } V{*}{*}{*} \cup R{*}{*}{*} = \Omega\right] \\ \Pr({*}V{*}{*}) = Pr({*}R{*}{*}) = 1/2, & \text{et ainsi de suite.} \end{array} \right.$$

---

1) C'est-à-dire contenant 3 billes Vertes et 3 billes Rouges.

D'autre part, on mélange avec soin les billes de l'urne entre les essais successifs; ceci suggère avec force *l'hypothèse* d'absence d'interaction mutuelle entre essais successifs, traduite formellement par

$$H_2 \begin{cases} \text{les partitions } \{R^{***}, V^{***}\}, \{{}^*R^{**}, {}^*V^{**}\}, \{{}^{**}R^*, {}^{**}V^*\}, \\ \{{}^{***}R, {}^{***}V\} \text{ sont indépendantes.} \end{cases}$$

Les hypothèses H_1 et H_2 prises ensemble conduisent à la conclusion

$$\forall (a, b, c, d) \in \Omega\,;\ \Pr\{(a\,b\,c\,d)\} = \frac{1}{16}.$$

C'est là un exemple simple d'analyse probabiliste d'une situation donnée. Mais il est **indispensable** de ne pas en rester là, et de se demander si, empiriquement, les *hypothèses* H_1 et H_2 se justifient. Or, si on procède à un grand nombre de réalisations de la situation, et si on observe les fréquences f_1 de (V^{***}), ..., f_4 de $({}^{***}V)$, on constate empiriquement que

$$f_1 \sim f_2 \sim f_3 \sim f_4 \sim \frac{1}{2}.$$

Cette constatation confirme empiriquement l'hypothèse H_1. De même, on peut observer des fréquences relatives [p. ex. la fréquence relative de $f_{2|1}$ (VV^{**}) par rapport à (V^{***})] et constater qu'elles sont approximativement égales: ceci confirme empiriquement l'hypothèse H_2.

Cela étant, si l'on considère ensuite 4 essais successifs avec une urne $(2V, 4R)$ bien mélangée, il est tout naturel de considérer, par analogie avec le cas précédent, *l'hypothèse*

$$H_1' : \Pr(V^{***}) = \Pr({}^*V^{**}) = \Pr({}^{**}V^*) = \Pr({}^{***}V) = \frac{1}{3}.$$

Si on considère un grand nombre de réalisations de cette situation, l'hypothèse H_1' est empiriquement confirmée par la constatation que

$$f_1 \sim f_2 \sim f_3 \sim f_4 \sim \frac{1}{3}.$$

Il en est de même dans beaucoup de cas. On en vient ainsi à disposer d'une théorie [1]) dont le noyau structurel est constitué par l'axiomatique de

[1]) « Disposer d'une théorie » est, épistémologiquement parlant, un concept relativement compliqué; on trouve des explications raisonnablement complètes, sinon toujours fort claires, dans [14], pp. 189-194.

Kolmogorov et ses conséquences (y compris les lois des grands nombres), qui peut être spécialisée en divers modèles par l'adjonction de contraintes supplémentaires (p. ex. des conditions de symétrie), dont les applications envisagées sont les situations où est en jeu un mécanisme aléatoire formé d'un ensemble fini d'éléments interchangeables, et dont le paradigme comprend les jeux de hasard classiques (dés, cartes), les schémas d'urnes, et la théorie chromosomique de l'hérédité.

d. Toutes les situations étudiées au stade *A* relèvent de cette théorie (voir p. ex. [3]). Il incombe à l'enseignement déductif du stade *B* de mettre très clairement en évidence le *double* rôle des situations concrètes dans ce développement:

au départ, l'analyse de la situation suggère des *hypothèses* relatives à la fonction Pr;

à l'arrivée, l'examen de la situation révèle des éléments empiriques étroitement parallèles aux éléments théoriques qui résultent de ces hypothèses.

En outre, les lois des grands nombres accentuent ce parallélisme, et permettent d'étendre la théorie à des situations aléatoires douées d'une structure répétitive propre, pour lesquelles il n'est plus nécessaire de considérer de multiples réalisations d'une *même* situation.

[Incidemment, on note que, pour beaucoup d'auteurs, des réalisations multiples (en nombre n) d'une situation aléatoire S doivent être considérées comme formant *une* réalisation d'une situation « d'ordre supérieur » $\mathscr{S}$: si la situation S est représentée par $(\Omega, \mathscr{T}, \mathrm{Pr})$ alors $\mathscr{S}$ a comme ensemble de possibles le produit de n exemplaires de Ω, et comme probabilité le produit de n mesures Pr, compte tenu, disent-ils, de l'indépendance des réalisations successives de S. Une telle position est inadmissible, pour plusieurs raisons:

a. s'il en était ainsi, on ne disposerait jamais que d'une seule réalisation (de $\mathscr{S}$), et le caractère aléatoire de la situation ne serait pas empiriquement observable;

b. lorsqu'un physicien, un chimiste, un biologiste répète une expérience, personne ne songe à lui contester le droit de considérer que c'est bien la *même* expérience qui est ainsi faite plusieurs fois; nul ne songe à lui opposer l'aphorisme d'Héraclite, « tout change, on ne se baigne pas deux fois dans le même fleuve »; pourquoi devrait-on contester ce droit aux seuls probabilistes?

c. quiconque prend vraiment au sérieux l'aphorisme d'Héraclite ne peut qu'abandonner tout dessein scientifique; la science n'est possible que dans un univers de « natures » suffisamment stables; mais un mécanisme aléatoire est une « nature stable » exactement au même titre qu'un mécanisme physique, chimique, ou biologique; si je dispose d'une urne U ($2V$, $4R$) et que j'en extraie (avec remise) 20 billes, c'est là un mécanisme stable, avec des propriétés constantes, que l'on peut « réaliser » à diverses reprises sans pour cela devoir craindre que chaque réalisation modifie le mécanisme.]

Pourquoi les études du stade A sont-elles, le plus souvent, si étrangement indifférentes au rôle a posteriori de l'observation empirique, sinon parce que l'enseignement déductif qui les suit oublie lui aussi ce rôle? Il y a là une situation malsaine, à laquelle on croit remédier en imposant à la majorité des élèves un enseignement de statistique. Ce remède — difficile à administrer et peut-être dangereux — n'est nullement nécessaire: il suffit — mais il *faut* — que le premier enseignement déductif de la théorie des probabilités la traite exactement comme il convient à une théorie physique: chacun de ses modèles part d'une situation, et y retourne; ce retour, et lui seul, permet de discriminer entre divers modèles possibles; en ce sens, les débuts de la statistique inférentielle font partie intégrante de cet enseignement, mais c'est pour la seule et unique raison qu'ils font partie intégrante de la théorie elle-même.

BIBLIOGRAPHIE

[1] BUNGE, M. *Philosophie de la physique.* Paris, 1975.
[2] DUHEM, P. *La théorie physique* (2e éd.). Paris, 1914.
[3] ENGEL, A. *L'enseignement des probabilités et de la statistique.* Paris, 1975.
[4] FREUDENTHAL, H. a) Der Warscheinlichkeitsbegriff als angewandte Mathematik, in: *Les applications nouvelles des mathématiques et l'enseignement secondaire* (ouvrage collectif); Echternach, 1973.
b) The crux of course design in probability. *Educ. Studies in Math.* 5 (1974), pp. 261-277.
[5] Groupe de travail ministériel sur les contenus de l'enseignement mathématique. *Bull. Association Prof. Math. Ens. Public*, 53/300. (1975), p. 557.
[6] KUHN, T. S. *La structure des révolutions scientifiques.* Paris, 1972.
[7] LAKATOS, I. and A. MUSGRAVE. *Criticism and the growth of knowledge* (ouvrage collectif). Cambridge, 1970.
[8] MARITAIN, J. *Les degrés du savoir* (3e éd.), Paris, 1932.
[9] Mathematical Association. *The teaching of geometry in schools.* London, 1923 (3d ed., reprinted, 1963).

[10] PIAGET, J. et B. INHELDER. a) *La représentation de l'espace chez l'enfant*, Paris, 1948.
b) *La genèse de l'idée de hasard chez l'enfant*, Paris, 1951.
[11] POPPER, K. *Logik der Forschung.* Tübingen, 4te Aufl., 1971. Traduction française: *Logique de la découverte scientifique.* Paris, 1971.
[12] RÅDE, L. Probability and flow graphs. *Intern. J. Math. Educ. Sc. Technol.* 4/4 (1973), pp. 363-372.
[13] SNEED, J. D. *The logical structure of mathematical physics.* Dordrecht, 1971.
[14] STEGMÜLLER, W. *Probleme und Resultate der Wissenschaftstheorie.* Bd. II: Theorie und Erfahrung. Berlin, 1973.

(Reçu le 20 novembre 1975)

H. Breny

Institut de Mathématique
Avenue des Tilleuls, 15
B-4000 Liège