

6 Sur l'enseignement des probabilités au lycée

par Colette BLOCH
Université de Poitiers

PREMIERE PARTIE

Eléments de réponse à une question concernant les programmes
(Première et Terminale)

Du 5 au 12 août 1974 s'est tenue à Talence la XXVIème réunion annuelle de la Commission Internationale pour l'Amélioration de l'Enseignement des Mathématiques (CIAEM), consacrée cette année aux probabilités.

150 participants représentaient de nombreux pays d'Europe, d'Afrique et des deux Amériques et l'accueil de l'IREM de Bordeaux a été remarquable à tous égards. Mon propos n'est pas de rendre compte de ce colloque ; d'ailleurs la chose n'est pas de mise si j'ai bien compris, ces rencontres étant traditionnellement informelles, comme on dit.

Je voudrais seulement relever une question qui a été posée par un collègue, au cours d'une des premières séances, question qui m'avait déjà été posée à peu près dans les mêmes termes, et que se posent tous les professeurs français enseignant en second cycle du second degré.

Voici la question que m'avaient posée des collègues enseignant en Terminale D :

- (1) $\left\{ \begin{array}{l} \text{"Le programme définit un espace probabilisé fini} \\ (E, \mathcal{B}(E), P), \text{ où } \mathcal{B}(E) \text{ est une sous-algèbre de Boole de } \mathcal{T}(E). \\ \text{"Les situations concrètes où le modèle est } (E, \mathcal{T}(E), P) \\ \text{sont classiques (problèmes de jeux, ...), mais les situations} \\ \text{concrètes où } \mathcal{B}(E) \neq \mathcal{T}(E) \text{ sont plus difficiles à trouver.} \\ \text{"Nous désirons quelques exercices ou exemples illus-} \\ \text{trant un espace probabilisé non discret, où } \mathcal{B}(E) \neq \mathcal{T}(E), \\ \text{correspondant à une situation concrète"} \end{array} \right.$

Et voici la question posée à Talence :

- (2) “Pourquoi le programme de Probabilités en classe de Première définit-il la probabilité sur $\mathcal{T}(E)$, et celui de Terminale sur une sous-tribu $\mathcal{B}(E)$ incluse dans $\mathcal{T}(E)$? Justification de l'utilisation des tribus *dans le cas fini* ? ”

Naturellement, je n'ai pas qualité pour répondre à la première partie de cette question (2), mais, puisque tribu il y a, je crois qu'on peut très simplement “justifier l'utilisation des tribus” — sans qu'il soit nécessaire de prononcer le mot — par le fait qu'une algèbre de Boole (*) existe, et s'introduit logiquement d'elle-même, dès qu'on a fait la liste des questions auxquelles on s'intéresse, pour une catégorie d'épreuves donnée.

Rappelons brièvement de quoi il s'agit.

Soit E l'ensemble des résultats possibles d'une expérience aléatoire ; on l'appelle quelquefois ensemble des éventualités ; en Probabilités, une telle expérience s'appelle généralement *épreuve* ; E est l'univers correspondant à une catégorie d'épreuves donnée (on appelle aussi quelquefois E : ensemble des épreuves, pour : ensemble des résultats d'épreuves). Exemples :

— l'épreuve consiste à lancer un dé à jouer et on regarde la face supérieure du dé à l'arrêt ; $E = \{1, 2, 3, 4, 5, 6\}$;

— l'épreuve consiste à tirer sur une cible circulaire ; E peut être l'ensemble (non dénombrable) des points du disque, ou un ensemble fini de couronnes ou de secteurs du disque, ou de carrés d'un quadrillage-repère.

Nous supposons dorénavant que E est un ensemble fini.

Les conditions contrôlables de l'expérience étant fixées, E étant bien défini, on peut se demander si le résultat d'une prochaine épreuve aura telle ou telle propriété (point du dé plus grand que 3, ... ; impact dans le quart nord-est de la cible, ...). Appelons événement aléatoire toute question pertinente : une question est pertinente si le résultat d'une épreuve (un élément de E) lui apporte une réponse par OUI ou NON ; OUI : en ce point l'événement en question est réalisé ; NON : il n'est pas réalisé.

Deux questions sont équivalentes si en tout point précis de E on répond de la même façon aux deux questions. Toute classe de

(*) qui n'a pas besoin d'être appelée σ -algèbre (ou tribu) quand l'univers est fini.

questions équivalentes ($A?$) définit donc un sous-ensemble A' de E — celui des points où l'on répond OUI — analogue à l'ensemble solution d'une fonction propositionnelle sur un référentiel, en Logique.

D'où immédiatement les notions : d'événements identiques (ou : égaux) ; d'événement contraire de A , $\bar{A}$, représenté par $\bar{E}A'$; d'événement certain, représenté par E ; d'événement impossible, représenté par ϕ ; d'implication : A implique B correspond à $A' \subset B'$; d'incompatibilité : $A' \cap B' = \phi$.

Si A et B sont des événements, représentés respectivement par A' et B' , alors $A' \cap B'$, $A' \cup B'$, $A' \setminus B'$, $B' \setminus A'$, $A' \Delta B'$ représentent aussi des événements dont la définition est évidente.

En général on assimile l'événement A à l'ensemble A' des points de E où il est réalisé : on ne distingue plus A et A' (*).

Inversement on peut être tenté de dire que tout sous-ensemble de E correspond à un événement (plus ou moins facile à définir) et prendre $\mathcal{T}(E)$ pour l'ensemble des événements liés à la catégorie d'épreuves correspondante. E étant fini, il n'y aura aucune difficulté à définir sur $\mathcal{T}(E)$ la mesure-probabilité : en fait, dans les cas classiques (jeux, ...), elle est définie comme uniforme sur les singletons, d'où la mesure de tout sous-ensemble par simple dénombrement.

Mais il y a, à mon sens, deux inconvénients à prendre $\mathcal{T}(E)$ tout entier. Le premier c'est que ce n'est pas réaliste, le second c'est que cela masque la structure logique de l'ensemble des événements.

- Ce n'est pas réaliste : si on joue avec deux dés indiscernables, il y a 21 résultats possibles, donc $2^{21} = 2\,097\,152$ événements. Si les deux dés sont discernables, il y en a près de 69 milliards.
- Cela masque la structure : à preuve la question posée par les collègues.

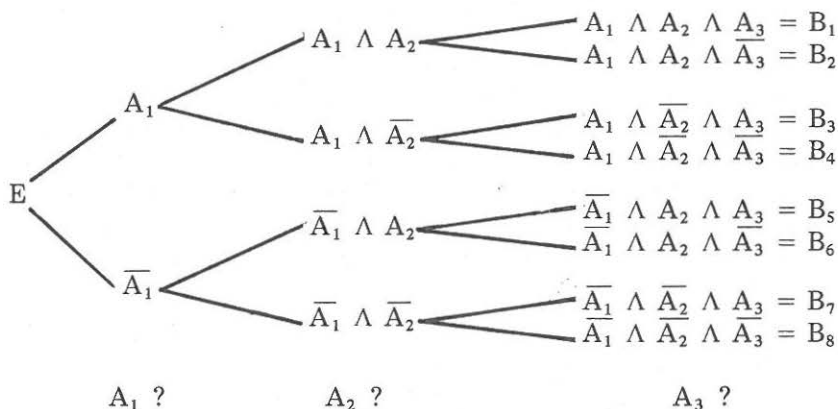
(*) Cette confusion peut s'avérer gênante, J'y reviendrai. Comme, bien souvent, on ne fait pas non plus expressément la distinction entre l'événement aléatoire en tant que question ($A?$) et l'événement réalisé A , les élèves sont excusables quand, montrant l'ensemble A' écrit ou figuré par un dessin, ils disent — puisqu'ils l'ont sous les yeux — que A est toujours réalisé, voulant dire, sans doute, réalisable.
Si on veut introduire un début de théorie mathématique des probabilités (ce qui n'est peut-être pas nécessaire pour ce qu'on en fait au Lycée), il faut procéder correctement. Cela peut se faire en liaison avec la Logique ; les Probabilités sont même une occasion privilégiée de se servir des connecteurs logiques et des opérations ensemblistes en leur donnant un contenu concret.

Or pour rencontrer cette structure il n'y a pas besoin d'aller bien loin.

Soit E une population finie dont tout individu possède — ou ne possède pas — les caractères A_1, A_2, A_3 , supposés distincts en ce sens que deux quelconques des questions $A_1 ?$, $A_2 ?$, $A_3 ?$, ne sont a priori ni équivalentes ni "opposées".

Publier la statistique de la population, relativement à ces caractères, c'est donner les effectifs de toutes les sous-populations d'individus vérifiant soit A_1 et A_2 et A_3 ; soit A_1 et A_2 et non- A_3 ; soit ... etc ... A partir de ces trois questions on a réalisé une partition de E (en supposant qu'aucun effectif n'est nul), la plus fine possible étant donné l'information dont on dispose.

Voici l'arbre correspondant :



Ici $\wedge$ est le signe de la conjonction logique, à laquelle correspond l'intersection d'ensembles.

$B_1, B_2, \dots, B_8$ sont les *atomes* de l'algèbre de Boole engendrée par les questions $(A_1 ?, A_2 ?, A_3 ?)$.

En langage d'ensembles, une algèbre de Boole $\mathcal{B}$ incluse dans $\mathcal{T}(E)$ est une famille, non vide, de parties de E , stable par intersection et complémentation (d'où : réunion, différence, différence symétrique), ce qui est logiquement nécessaire, comme on l'a vu dans le cas où les parties considérées représentent des événements. L'algèbre engendrée par une famille quelconque de parties est par définition la plus petite algèbre contenant ces parties : nous avons trouvé les atomes de l'algèbre engendrée par A'_1, A'_2, A'_3 ; tout élément de cette algèbre est réunion d'atomes ou ϕ .

Le lecteur vérifiera facilement l'existence et l'unicité de l'algèbre engendrée, et le fait qu'on l'obtient par la construction indiquée.

Publier une statistique, c'est donc donner une mesure (effectif ou pourcentage) sur les atomes d'une algèbre de Boole, d'où l'on déduit la mesure de tout sous-ensemble *qui s'exprime en fonction des caractères considérés*. On peut aussi ne pas donner les effectifs des atomes, mais le moyen de les calculer de proche en proche ; si on connaît l'effectif total et si n caractères sont en jeu, il faut $2^n - 1$ données compatibles puisqu'il y a 2^n atomes. (Remarquons qu'il y a donc $2^{(2^n)}$ éléments dans l'algèbre).

Exemple.

Sur 1 000 étudiants, 300 se sont déclarés capables de lire l'anglais, 100 l'allemand, 200 l'espagnol, 50 l'italien, 40 l'anglais et l'allemand, 150 l'anglais et l'espagnol, 40 l'anglais et l'italien, 10 l'allemand et l'italien, 5 l'espagnol et l'italien, 10 l'allemand et l'espagnol, 2 l'anglais, l'espagnol et l'italien, 3 l'anglais, l'allemand et l'espagnol, 5 l'anglais, l'allemand et l'italien, 2 l'allemand, l'espagnol et l'italien ; aucun ne lit les quatre langues.

Combien y en a-t-il qui n'en lisent aucune ?

Réponse : 593. Dessiner l'arbre en posant 4 questions, A ? , D ? , E ? , I ? , une pour chaque langue, dans n'importe quel ordre ; on obtient tous les atomes avec leurs effectifs, dont l'un se trouve être nul : $A \wedge D \wedge E \wedge I$ est vide mais non par nécessité logique, en général c'est bien un atome (*).

Si, concernant la même population, on pose des questions d'un autre type, relatives par exemple à l'origine sociale et géographique, on construira une algèbre non comparable en général à la précédente, en ce sens qu'aucune d'elles ne sera incluse dans l'autre.

Si par contre, dans l'exemple donné, on connaissait de surcroît la répartition garçons-filles dans chaque sous-ensemble, on obtiendrait une algèbre plus fine que la précédente (la contenant) mais on serait encore loin de l'ensemble de toutes les parties, qui apparaît bien comme une absurdité en Statistique.

(*) Dans un cas comme celui-ci, on trouve plus rapidement la réponse sur un diagramme de Venn ou en appliquant la formule d'inclusion et d'exclusion qui donne le cardinal d'une réunion quand on connaît celui de chaque intersection.

La situation est pratiquement la même, dans le cas discret, en Probabilités. Pour un problème donné on n'a pas à envisager tous les événements possibles mais seulement ceux qui sont fonction des questions posées. Cependant, pour les ensembles E finis, dès qu'on parle de probabilité (mesure), $\mathcal{I}(E)$ réapparaît sournoisement pour la raison indiquée plus haut : on connaît la mesure d'un atome (théoriquement insécable) par le nombre d'éléments de E qu'il contient.

En fait c'est très artificiel car, dans le genre d'exercices qu'on traite au niveau du baccalauréat, on *raisonne* en cas favorables et cas possibles équiprobables, sans toujours parler explicitement de l'ensemble des résultats *constatables* de l'épreuve, ensemble qui d'ailleurs n'est pas forcément le bon : par exemple, si on lance ensemble deux pièces de monnaie indiscernables, on a, à l'arrivée, 3 configurations possibles : $\{ \text{pile, pile} \}$, $\{ \text{pile, face} \}$, $\{ \text{face, face} \}$, alors qu'en fait il faudrait considérer l'ensemble-produit $E^2 = \{ \text{pile, face} \} \times \{ \text{pile, face} \}$ — pour cela, il suffit de dessiner un arbre — mais aussi, si l'on veut être conséquent, le produit des algèbres (qui n'est pas un produit cartésien) : $\mathcal{I}(E) \otimes \mathcal{I}(E)$, et la probabilité-produit $P \otimes P$, ce qu'évidemment on ne fera pas explicitement à ce niveau !

Toutefois, puisque les élèves de Terminale sont jugés dignes d'accéder à la tribu, que ce ne soit pas une initiation magique. La notion d'algèbre de Boole est simple, se justifie fort bien en langage d'événements, c'est une structure qu'on retrouve dans des situations variées (une mine d'exercices !), et enfin je crois qu'elle suggère facilement, à partir d'exemples statistiques, les *axiomes de la probabilité sur une algèbre finie* :

E et $\mathcal{B}$ étant définis respectivement : E comme ensemble des résultats possibles de l'épreuve, $\mathcal{B}$ comme ensemble des événements, muni de sa structure d'algèbre de Boole, une probabilité P est une application de $\mathcal{B}$ dans $\mathbb{R}^+$ telle que

$$(i) \quad P(E) = 1$$

$$(ii) \quad \text{si } A \in \mathcal{B}, B \in \mathcal{B} \text{ et si } A \text{ et } B \text{ sont incompatibles, alors } P(A \vee B) = P(A) + P(B).$$

($\vee$ est le "ou non-exclusif" de la disjonction logique, à laquelle correspond la réunion d'ensembles).

De plus, la définition d'une probabilité conditionnelle, et par suite celle de l'indépendance en probabilité, sont tout à fait intuitives sur l'arbre des atomes (et d'abord sur le diagramme de

Carroll, quand deux événements seulement sont en jeu). On voit très bien qu'on ne sort pas de l'algèbre engendrée par les seules questions considérées, et les définitions prennent concrètement leur sens. On voit aussi que deux événements peuvent être indépendants pour une probabilité P et ne pas l'être pour une autre probabilité Q définie sur la même algèbre (deux échantillons statistiques différents).

Je crois que, par l'intermédiaire de $\mathcal{B}$ incluse dans $\mathcal{T}(E)$, on distingue mieux l'ensemble E , sans structure, liste des résultats possibles d'une épreuve, de l'ensemble structuré des événements auxquels on s'intéresse.

J'ai dit qu'il pouvait être gênant de confondre l'événement aléatoire ($A?$) avec l'ensemble A' des points de E qui le réalisent. Les difficultés apparaissent dès qu'on a affaire à un espace probabilisé produit. Si par exemple on couple deux épreuves, soit E_1 l'ensemble des résultats possibles de la première, E_2 celui de la seconde. Soit A l'événement $A_1 \wedge A_2$, réalisé si et seulement si pour la première épreuve $A_1 \in \mathcal{B}_1(E_1)$ est réalisé, et pour la seconde $A_2 \in \mathcal{B}_2(E_2)$ est réalisé; cela n'a aucun sens d'écrire $A' = A'_1 \cap A'_2$, puisque A'_1 et A'_2 sont respectivement parties de deux univers distincts E_1 et E_2 ; il faudrait écrire :

$$A' \subset E_1 \times E_2 \quad A' = (A'_1 \times E_2) \cap (E_1 \times A'_2),$$

ce qui est facile à comprendre mais pénible à écrire. En langage d'événements on écrira

$$A = A_1 \wedge A_2 \quad \text{ou} \quad A = A_1 \cdot A_2$$

$A \in \mathcal{B}_1(E_1) \otimes \mathcal{B}_2(E_2) = \mathcal{B}$, le signe $\otimes$ indiquant que l'ensemble $\mathcal{B}$ des événements liés à $E_1 \times E_2$ est obtenu non pas comme ensemble des couples (A_1, A_2) , $A_1 \in \mathcal{B}_1$, $A_2 \in \mathcal{B}_2$, mais comme ensemble des résultats d'opérations d'algèbre de Boole entre éléments de $\mathcal{B}_1$ et de $\mathcal{B}_2$: c'est dire que les éléments de $\mathcal{B}$ ne sont pas tous représentés par des ensembles de la forme $A'_1 \times A'_2$. $\mathcal{B}$ est l'algèbre engendrée par ces ensembles.

Pour la probabilité, si P_1 est définie sur $\mathcal{B}_1$ et P_2 sur $\mathcal{B}_2$, $P = P_1 \otimes P_2$ est définie sur $\mathcal{B}$, et (avec les mêmes notations que ci-dessus), si A' n'est pas un rectangle $A'_1 \times A'_2$, on n'a pas en général $P(A) = P_1(A_1) \cdot P_2(A_2)$.

En résumé, si l'on n'a à faire que des dénombrements, il est plus simple (et sans doute plus sûr) de se servir de diagrammes (arbres, Venn, Carroll) sans introduire de théorie autre. Les

propriétés de la probabilité découlent tout naturellement des règles observées pour les fréquences relatives ; de même la fréquence conditionnelle se met très simplement en évidence sur les diagrammes de situations statistiques concrètes, et donne sans mystère la notion de probabilité conditionnelle.

Jusque là, et pour traiter les cas “classiques” (équiprobabilité), on peut opérer directement sur les ensembles. Maintenant, si l’on veut aller plus loin et parler en langage d’événements, alors il faut expliquer pourquoi l’ensemble des événements dont on s’occupe constitue une algèbre de Boole — en général strictement incluse dans $\mathcal{I}(E)$ — sur laquelle on sait définir une probabilité ; les événements apparaissent alors comme les ensembles mesurables pour cette mesure. Peut-être faut-il en effet le faire en Terminale, mais il peut sembler inutilement pédant de formaliser, et il est apparemment troublant de le faire en deux étapes. On risque soit de compliquer des situations simples, soit d’introduire des notions dont on ne se servira pas.

DEUXIEME PARTIE

“Faut-il enseigner une axiomatique des Probabilités
et laquelle ? ”

Le titre ci-dessus est celui qui avait été attribué au groupe de travail constitué, à Talence, à partir de la question (2) (cf. page 96). Effectivement le problème de l'axiomatique est immédiatement sous-jacent à cette question.

Dans la première partie, nous avons présenté les axiomes de la probabilité comme suggérés par l'expérience commune — et souvent implicitement utilisés avant d'être posés comme axiomes. La probabilité classique apparaît ainsi comme une théorie physique, de même que, par exemple, la géométrie classique. Disons tout de suite que cette conception ne fait pas l'unanimité (nous reviendrons sur ce point), mais admettons-la pour l'instant, au moins comme un point de départ historique.

Les axiomes de la probabilité (mesure positive finie) étant particulièrement simples, il n'y a aucune difficulté à les introduire. Nous avons vu que l'algèbre de Boole, support de la probabilité dans le cas fini, est également à portée de la main. Maintenant, faut-il présenter, dans toute sa généralité, aux élèves de l'enseignement secondaire, l'espace noté traditionnellement par les probabilistes $(\Omega, \mathcal{A}, P)$? A mon avis ce n'est pour eux qu'une notation de plus, qui alourdit sans rien apporter. Mais voici des opinions plus autorisées :

W. Feller termine le chapitre introductif de son livre célèbre par cette phrase : “Nous suivrons [l'axiomatique de Kolmogorov], mais le terme d'axiome paraît ici bien solennel, dans la mesure où le présent volume (*) ne s'occupe que du cas simple des probabilités discrètes.” Pourtant ce volume est bien loin de ne traiter que des problèmes élémentaires et des applications triviales ! Il est donc possible d'aller assez loin en Probabilités, par des méthodes analytiques, peut-être artisanales mais fort instructives, sans grosse machinerie théorique.

A Talence, le professeur Freudenthal s'est exprimé nettement contre la présentation aux débutants de “l'axiomatique des $(\Omega, \mathcal{A}, P)$ ”. Il a fait remarquer que les produits de tels espaces

(*) Il s'agit du Tome I de “An Introduction to Probability Theory and its Applications”, Wiley & sons, 2nd ed. 1962.

sont impossibles à manier et que d'ailleurs, après avoir introduit ce fameux triplet, "on passe aux choses intéressantes, c'est-à-dire aux variables aléatoires, et on ne parle plus de $(\Omega, \mathcal{A}, P)$."

Pourquoi cela ? *Parce qu'en sciences expérimentales (astronomie, physique, chimie, biologie) on peut, sur un grand nombre d'observations, soupçonner l'existence d'un phénomène aléatoire, c'est-à-dire où le hasard semble obéir à certaines règles. L'expérimentateur connaît ce phénomène aléatoire par des mesures, des résultats numériques, autrement dit les valeurs d'une grandeur aléatoire avec les fréquences qui leur sont associées, ponctuellement ou par intervalles. C'est la répartition empirique de probabilité, d'où l'on induit une répartition théorique. Bien entendu le spécialiste peut avoir a priori quelque idée sur les raisons, ou au moins le mécanisme, de cette régularité statistique, peut-être même sur le type de loi qui la régit.*

En retour la loi théorique, une fois vérifiée, peut aider à analyser le phénomène et permettre de mieux connaître l'ensemble des éventualités. Elle assigne une probabilité aux événements comme images inverses des points ou des intervalles. Le triplet $(\Omega, \mathcal{A}, P)$ est donc en général, en admettant qu'on en ait besoin, construit a posteriori.

Exemples

- (1) A partir des observations de Mendel, on conclut à l'équiprobabilité de l'hérédité des gènes dominants ou récessifs, et à l'indépendance réciproque des caractères (forme et couleur, par exemple).
- (2) Ayant observé que la durée de vie des corps radioactifs suit une loi exponentielle, on peut admettre que le nombre de désintégrations d'atomes au cours du temps suit une loi de Poisson (cf. Rényi, Calcul des Probabilités, Dunod, 1966, ch. 3, § 13).
- (3) Laplace avait d'abord supposé que la répartition des erreurs de mesure était exponentielle en $-|x|$, avant de reconnaître la loi en $-x^2$ dite de Laplace-Gauss (cf. Borel, Deltheil, Huron, Probabilités. Erreurs, ch. V, collection Armand Colin).

Dans tous ces cas, on adopte une loi théorique dont il faut vérifier la validité sur les expériences ultérieures. *Les méthodes de la Statistique mathématique permettent alors de tester l'adéqua-*

tion de la loi, ou, la loi étant admise, de déceler une circonstance parasite.

Certains (Fisher et d'autres) soupçonnaient Mendel d'avoir truqué ses résultats, qui leur paraissaient s'écarter étonnamment peu de la loi théorique. Van der Waerden a eu la curiosité de calculer la probabilité d'une amplitude des écarts inférieure ou égale et a trouvé 97 % : autrement dit, si les lois de Mendel sont vraies, des écarts supérieurs à ceux observés n'ont qu'une probabilité de 3 % de se produire. Un travail analogue a été fait par Bessel sur 470 observations astronomiques de Bradley, vieilles de plus d'un siècle, pour vérifier la loi de répartition des erreurs de mesure. Le résultat a entièrement confirmé la loi de Laplace-Gauss (cf. le livre cité en (3) ci-dessus).

Il semble que nous nous soyions beaucoup éloignés de la théorie des Probabilités en direction des applications du Calcul des Probabilités et de la Statistique mathématique, contournant ainsi le problème épistémologique. Mais c'est aussi un moyen de l'aborder qui me semble propre à développer l'intuition probabiliste des élèves et à susciter leur intérêt, bien plus que le parachutage d'une construction axiomatique. Cette approche aurait aussi l'avantage de justifier, en l'utilisant et en le prolongeant, le programme de Statistique qui, pris isolément et réduit à la Statistique descriptive, est bien dépourvu d'intérêt.

*

* *

Dans l'optique précédente, on se borne à l'aspect de la probabilité qu'on appelle "statistique", ou fréquentiel, ou empirique. C'est le seul qu'envisagent, par exemple, Feller et Rényi, après Fisher et von Mises qui avaient contribué, avec beaucoup d'autres auteurs, à préciser ce concept, à partir des années 20, avant l'édification complète de la théorie axiomatique par Kolmogorov (1933). Cependant ce point de vue a de tous temps été critiqué comme restrictif, et de surcroît non fondé, par les tenants de la probabilité subjective, qui nient la réalité d'une probabilité physique, ou qui estiment n'avoir pas besoin de cette hypothèse, au sens où Laplace n'avait pas eu besoin de Dieu pour sa cosmogonie.

Quant à Feller, il pense que les mathématiciens, physiciens et statisticiens n'ont pas plus à se préoccuper de la vraie signification du mot *probabilité*, que les physiciens du sens réel des mots *masse*

et *énergie*, ou que le géomètre n'a à discuter de la nature du point. "Au lieu de cela, nous allons établir des théorèmes et montrer comment ils s'appliquent." ... "La philosophie des fondements de la probabilité doit être séparée de la mathématique et de la statistique, exactement comme l'analyse de notre conception intuitive de l'espace est maintenant séparée de la géométrie." (*)

Cependant cette attitude pragmatique est aussi une prise de position philosophique. Il est indéniable que le point de vue fréquentiel sous-entend l'existence, indépendamment de l'observateur, d'une probabilité physique, existence suggérée par la stabilité des fréquences. Il est tout aussi indéniable que la vraie valeur de cette probabilité est inaccessible, mais ni plus ni moins que celle de n'importe quelle grandeur physique. De même on admet que les erreurs de mesure sont des valeurs prises par une hypothétique variable aléatoire gaussienne centrée, résultante d'un grand nombre de causes infimes, ce qui se justifie par le théorème de la limite centrale.

On démontre que, si une pièce de monnaie a la probabilité p de montrer pile, alors la fréquence relative du nombre de "pile" en n lancers s'approchera aussi près qu'on veut de p (pour n assez grand), avec une probabilité aussi voisine de 1 qu'on veut (théorème de Bernoulli, loi faible des grands nombres). On démontre même que cette fréquence tend presque sûrement vers p (loi forte des grands nombres). *On démontre* que, si des valeurs observées peuvent être considérées comme prises par des *variables aléatoires* indépendantes de même loi, alors la répartition empirique converge uniformément presque partout vers la répartition théorique (théorème de Glivenko-Cantelli). Mais il subsiste les "si" ...

"Entre 1925 et 1950 (**), presque tous les statisticiens professionnels n'ont fait officiellement usage que de probabilités "statistiques". Mais, durant à peu près la même période, d'autres, tels que Keynes, Ramsey, de Finetti et Jeffreys, ont affirmé que toute description de l'usage de la probabilité statistique est nécessairement donnée en termes de probabilité intuitive, et qu'il est impossible d'utiliser directement une définition partant de la limite d'une fréquence de succès. Car, comme disait Keynes, à la

(*) Feller, loc. cit.

(**) I.J. Good, compte rendu de la 3e édition du livre de H. Jeffreys : *Theory of Probability*, in *JRSS, A*, 125 (1962).

limite nous serons tous morts. Ramsey et de Finetti sont des subjectivistes, tandis que Keynes et Jeffreys penchent en faveur des crédibilités.”

Il est hors de question de décrire ici toutes les acceptions du concept de probabilité : chaque auteur a son propre vocabulaire, traduisant sa propre philosophie, elle-même sans aucun doute marquée par le rôle que jouent les probabilités dans sa spécialité. Pour éclairer ce qui précède, je vais cependant énumérer quelques conceptions citées par I.J. Good, subjectiviste néo-bayésien, dans son petit livre sur l'estimation des probabilités (chap. 2) (*). Good distingue cinq espèces de probabilité qu'il présente sous une forme volontairement “neutre”, en tant que notions opératoires, aspect qui intéresse le statisticien, en dehors de toute appréciation philosophique “afin de réduire autant que possible les sujets de controverse”.

Citons d'abord la probabilité *physique* (= statistique), ou matérielle, ou intrinsèque, celle que Jeffreys appelle “chance” en anglais ; la plupart des subjectivistes refusent de la prendre en considération.

Il y a une probabilité *tautologique*, c'est-à-dire déduite mathématiquement d'une hypothèse statistique simple. (En statistique, on appelle “hypothèse simple” celle qui consiste, par exemple, à supposer que la probabilité de pile est 0,5 — une hypothèse composite serait : supposer que cette probabilité est comprise entre 0,48 et 0,52).

Entre les trois autres espèces citées : *psychologique*, *logique* et *intuitive*, les distinctions sont très subtiles. Elles sont toutes plus ou moins *subjectives*.

Ramsey le premier a donné les axiomes de la probabilité subjective et de l'utilité : il s'agit d'un ensemble partiellement ordonné de probabilités *cohérentes*, en ce sens qu'on ne peut trouver une stratégie contre cet ensemble qui soit gagnante dans tous les cas.

Les tenants de la probabilité psychologique se fondent sur un “degré de croyance” pour parier ou décider, du point de vue de l'utilité. Ceux de la probabilité logique, ou “crédibilité”, parlent

(*) I.J. Good, *The Estimation of Probabilities : An Essay on Modern Bayesian Methods*, The M.I.T. Press (Research Monograph n° 30). Dans l'introduction, parmi les raisons de s'intéresser à l'estimation des probabilités, l'auteur cite celle “de faire mieux comprendre la philosophie probabiliste et statistique, et la nature des jugements de probabilité”.

de l'“intensité de conviction *rationnelle*” basée sur l'information qu'on possède.

D'après Good, la probabilité intuitive est soit logique, soit subjective ou “personnelle” (cas particulier psychologique).

Pour Savage, comme pour de Finetti, les probabilités subjectives sont seules à considérer. Ils refusent les crédibilités, qui me paraissent être — sous toute réserve — des probabilités subjectives auxquelles on a tenté de donner une justification objective. Good remarque qu'on utilise sans doute, le plus souvent, un compromis entre probabilités logiques et subjectives. “Il y a une hiérarchie : psychologique → subjective → logique → physique, mais les frontières entre ces quatre domaines ne sont pas nettement tracées.”

Savage et de Finetti affirment que la probabilité subjective permet de définir la probabilité physique et l'indépendance statistique en tant que concepts mathématiques. Mais on peut se demander si les subjectivistes n'utilisent pas quelquefois implicitement les principes classiques sous une forme déguisée : Good pense, à juste titre selon moi, “qu'on n'adopterait pas le postulat [subjectif] de permutabilité [d'une séquence bernoullienne quelconque de r succès et s échecs] si l'on n'avait déjà derrière la tête la notion de probabilité physique et d'indépendance statistique approchée.” Il remarque également qu'il y a quelque incohérence de la part de Jeffreys “à nier la validité d'une définition des probabilités physiques comme limites de fréquences et à employer ensuite un argument de limite pour justifier l'induction” (*).

Good met à plusieurs reprises les crédibilistes en difficulté, quand ils essayent de *légaliser* une notion intuitive ; on peut tout au plus la *légitimer*. Dans son compte rendu du livre de Jeffreys, il écrit à peu près : le fait qu'une opinion subjective soit émise par une personnalité célèbre ne rend pas cette opinion objective.

Ce trop bref aperçu est évidemment mutilant et sans doute tendancieux. Qu'on ne voie aucune prétention de ma part à m'exprimer sur des sujets si controversés, après tant d'illustres savants et philosophes : la moindre bibliographie sur ces questions comporte quelque 150 titres (livres et articles). Je pense qu'on ne peut enseigner les Probabilités et la Statistique sans au moins

(*) Harold Jeffreys, *Theory of Probability*, Clarendon Press, Oxford, 3rd edition, p. 43.

prendre conscience des problèmes qu'elles n'ont cessé de poser, depuis trois siècles, à leurs créateurs et utilisateurs. Elles ont du même coup fait naître beaucoup d'espoirs — souvent déçus — dans le domaine de la Logique et de l'épistémologie (*), et dans les sciences humaines : l'économie, la politique, la psychologie économique et la psychosociologie, la morale et l'art de la guerre (**).

Ce qui va suivre n'est que mon opinion personnelle et ne veut qu'inciter à la réflexion.

*

* *

Bien entendu, les résultats mathématiques obtenus par les différentes écoles ne sont pas contestables. En particulier les méthodes des subjectivistes sont souvent très ingénieuses ; mais dans bien des cas, en sciences expérimentales, après avoir, par de savants détours, évité toute allusion à une probabilité physique, elles aboutissent aux mêmes recettes que les méthodes classiques ; car en fin de compte les fréquences sont les seules données dont on dispose pour évaluer une probabilité (jointes à des considérations de symétrie physique s'il y a lieu). Dans d'autres cas, quand on ne dispose pas d'informations antérieures, quand une expérience n'est pas répétable à volonté, on est bien obligé de faire des évaluations subjectives : en sciences humaines (linguistique, économie), même en sciences naturelles (espèces rares). Mais pourquoi vouloir faire de nécessité vertu et imposer une théorie de la probabilité psychologique, ou une théorie des utilités, ou une théorie des degrés de croyance, là où on n'en a pas systématiquement besoin ? Les subjectivistes justifient leur attitude, indépendamment des raisons épistémologiques, par le fait que dans la vie courante on est pratiquement toujours en situation de décision dans l'incertitude, de "jeu" contre "la Nature" ou contre un adversaire. Mais *pourquoi ne pas reconnaître qu'il y a en réalité des problèmes très différents, dans lesquels ce qu'on appelle le hasard intervient pour des raisons diverses, et que ces problèmes requièrent évidemment des traitements différents ? Si le hasard*

(*) Sur les tentatives pour fonder une logique inductive probabilitaire, voir le livre très complet de M. Boudot : *Logique inductive et probabilité*, Armand Colin, 1972.

(**) Cf. *Psychologie sociale et théorie des jeux*, article de Michel Plon, *La Pensée*, n° 161, février 1972.
Dans ces deux documents, nombreux renseignements historiques et bibliographiques.

est le nom que nous donnons à notre ignorance[⊕], il est bien certain qu'il y a des ignorances de nature variée et des degrés dans l'ignorance. Il y a des degrés aussi dans ce que nous avons besoin de connaître et de prévoir.

*

* *

On lira en annexe la lettre adressée au colloque de Talence par le Professeur de Finetti. Je suis tout à fait d'accord pour que l'enseignement des Probabilités aux enfants ne se réduise pas à des problèmes artificiels de dés, de cartes ou d'urnes, car ce ne sont formellement que des dénombrements, et si les élèves n'expérimentent pas eux-mêmes et ne travaillent pas sur des documents réels, ils ne peuvent comprendre la nature des phénomènes aléatoires (Les machines permettent aujourd'hui de traiter un grand nombre de données, comme on l'a vu à l'école de Talence). Glaymann et Varga dans leur petit livre (*) ont ouvert des horizons intéressants pour l'enseignement, en particulier sur la simulation.

Il est bon aussi de faire comprendre de bonne heure aux enfants que tout modèle simplifie la réalité, qui n'est pas en blanc et noir, mais je pense par contre qu'il serait mauvais de leur laisser croire qu'il n'y a *jamais* de raison objective de se prononcer pour le blanc plutôt que pour le noir : il peut être dangereux, à mon sens, de leur présenter la probabilité comme un jugement personnel (... et que le plus chanceux gagne !) car ils risquent de n'en retenir que l'arbitraire apparent, *en oubliant les exigences scientifiques de cohérence qui caractérisent les "décisions admissibles"*.

Je pense donc qu'il faut d'abord familiariser les enfants avec les lois *observables* du hasard, et attendre qu'ils soient assez mûrs pour utiliser à bon escient des techniques déjà éprouvées, à propos de problèmes ouverts. Rencontrer de telles situations dans la vie courante ne prouve pas qu'il soit facile de les aborder rationnellement ; au contraire, elles sont d'autant plus grevées d'idées préconçues.

Par exemple, les étudiants ont parfois du mal à admettre que le statisticien auquel on soumet des résultats *scientifiques* peut, *soit infirmer une hypothèse, soit la confirmer* (ou plus exactement

[⊕] Ce n'est pas *ma* définition.

(*) M. Glaymann et T. Varga, *Les Probabilités à l'école*, CEDIC, 1974.

la conforter, comme on dit maintenant), soit réclamer plus d'information pour en décider, mais ne peut en aucun cas assurer que l'hypothèse est vraie, malgré tout le désir qu'on en aurait. De toutes façons il incombe toujours au spécialiste d'interpréter ses résultats.

Qu'on soit subjectiviste ou non, on peut admettre, comme M. Fréchet aimait à le souligner, qu'une probabilité est, en fait, toujours conditionnelle. Mais tout dépend du niveau où on se place.

Le mathématicien pose des axiomes et construit la théorie par déduction — le praticien procède par induction. Mais, d'une part les axiomes, dans ce cas précis, sont en partie suggérés par l'expérience commune, d'autre part le praticien se sert de la théorie mathématique ; il n'y a pas opposition entre les deux démarches et en général on souhaite qu'elles s'accordent, mais leurs domaines et leurs buts ne sont pas identiques.

“Le problème de l'adéquation d'un modèle de probabilité à l'univers des phénomènes n'est pas différent par nature de celui de l'adéquation de n'importe quel modèle conceptuel en sciences. C'est le problème de la validité de l'induction”... “La croyance (consciente ou inconsciente) en la validité de l'induction et de la déduction est un préalable nécessaire à tout travail scientifique” (*).

* * *

En ce qui concerne l'enseignement, ma conclusion (on peut d'ailleurs l'appliquer à d'autres domaines que les Probabilités) sera que la théorie axiomatique doit venir couronner et ordonner des connaissances acquises, pour permettre d'aller plus loin et en mathématiques et dans la pratique (comme cela se produit toujours historiquement), mais qu'elle est *tout à fait inutile aux débutants sous son aspect formel*. Et cela quelle que soit la conception qu'on ait de la probabilité — psychologique ou statistique — ; pour moi il s'agit plutôt de deux usages différents d'un même concept.

* * *

Ci-après, en annexe, la lettre adressée en français au Colloque de Talence par M. le Professeur de Finetti et transmise aux congressistes par la délégation italienne.

(*) J.O. Irwin, Compte rendu de “Essays on Probability and Statistics” par M.S. Bartlett, JRSS, A, 125 (1962).

ANNEXE

Bruno de Finetti

Juillet 1974

L'introduction au raisonnement probable dans l'école.

Je regrette beaucoup de ne pouvoir participer à votre Colloque, car je suis doublement intéressé au sujet choisi cette année. En effet, la probabilité est la notion dont je m'occupe spécifiquement, et les questions didactiques ont pour moi une importance extrême (car l'existence de "vérités" serait une vide abstraction s'il n'y avait pas des personnes qui les comprennent). Ces deux aspects se mêlent de façon tout à fait particulière dans le cas de l'enseignement concernant les probabilités, étant donné que les mêmes questions psychologiques embrassent soit la notion dont il s'agit, soit la façon d'en apprendre la valeur et les propriétés.

C'est pourquoi je suis en même temps favorable et douteux lorsque on essaie d'introduire les probabilités dans l'enseignement, avec le danger d'en donner une idée trop étroite et stéréotypée.

C'est ce qui se passe, à mon avis, lorsqu'on présente la théorie des probabilités en se bornant à des cas trop particuliers où des circonstances secondaires conduisent à identifier simplement la probabilité avec des grandeurs objectives, comme une fraction de "cas également probables" ou une fréquence sur des "épreuves" assujetties à des conditions verbales semblables.

Ce serait, à mon avis, non seulement nécessaire au point de vue conceptuel, mais aussi plus simple et plus clair même pour des enfants, de définir la probabilité d'un événement donné E (bien déterminé) pour un individu donné comme le degré de confiance qu'il a dans son arrivée. Comme *définition opérationnelle*, la probabilité est le prix p (pour cet individu) d'un Franc conditionné au fait que E soit vrai. C'est, avec les précisions que l'on tire aisément des conditions de "cohérence", la notion sur laquelle s'appuient, d'une part, les paris (sur des faits de sport, d'élections politiques, etc.), et — moins futillement — les assurances (vie, accidents, incendie, etc.) et en général toutes les décisions que nous prenons sans cesse, plus ou moins consciemment, au cours de la vie.

Le calcul des probabilités est la théorie mathématique qui nous apprend à nous comporter de façon “cohérente”, c’est-à-dire telle qu’il n’en existe pas d’autre qui donne des résultats meilleurs “en tout cas possible”. (Avec la terminologie de Abraham Wald, il s’agit de caractériser les “décisions *admissibles*”).

Les propriétés des probabilités (additivité, multiplicité, ...) en découlent très simplement. Des *schèmes* de “cas également probables”, dés, extractions, etc... peuvent être utilisés correctement comme *méthodes pour aider* (par comparaison) l’évaluation d’une probabilité (par exemple : jugez si cela vous semble plus ou moins probable que de tirer un billet avec numéro ≤ 283 parmi 1000 billets numérotés de 1 à 1000).

Ce même procédé peut s’appliquer avec référence aux fréquences (par exemple : jugez si cela vous semble plus ou moins probable que si des événements “semblables” se produisent environ dans 283 cas sur 1000). Comme on voit, il se s’agit pas de ne pas utiliser ces schémas, mais de les utiliser raisonnablement, comme aide pour l’évaluation d’un prix, sans attribuer à la notion de “cas également probables” ou de “fréquence (théorique ?)” une valeur apodictique qui serait déraisonnable.

Après ça, il n’y a pas de difficulté à exemplifier et illustrer tout genre de problèmes et théorèmes indépendamment de toute circonstance spéciale : il suffit de donner des exemples où les mêmes formules s’appliquent à des probabilités qui, dans une première interprétation, peuvent être acceptées comme rapport de cas favorables et possibles, dans une autre comme “fréquences théoriques” (notion très insatisfaisante !), et enfin dans une troisième comme probabilités “indubitablement subjectives” (c’est-à-dire : même de l’avis de ceux qui soutiennent qu’il y a des probabilités “objectives”).

La notion de probabilité, les possibilités de l’appliquer, sont ainsi élargies et perfectionnées sans limite, et en évitant tout doute métaphysique. On n’est plus obligé de se borner à des “schémas” simplifiés, mais on peut exprimer sur chaque problème l’opinion propre de chaque individu. On ne peut pas reprocher d’accepter une évaluation de probabilités “fausse”, mais on peut indiquer à chaque individu quelles sont les conséquences de ses évaluations (c’est-à-dire : quelles autres évaluations en découlent par “cohérence”), de façon à faire juger à lui-même si réellement son opinion est celle-là avec tout ce qui est implicitement impliqué.

Le même procédé de *modification d'opinion* se produit au fur et à mesure que des nouveaux faits ont lieu (ou, mieux — car c'est cela qui compte — que ces faits sont *connus* par cet individu) : il ne s'agit pas d'une "correction" d'opinions "contredites par l'expérience", mais de la présentation des mêmes opinions apparemment modifiées du fait que les éventualités que dans le nouvel état d'information on connaît comme non-réalisées ont été effacées et les probabilités des autres (tout à l'heure possibles) ont été proportionnellement augmentées. Ceci n'est pas autre chose que *le raisonnement bayésien*.

De cette façon tous les *vrais* problèmes du raisonnement probable, du raisonnement inductif, du comportement rationnel dans des situations d'incertitude (théorie des décisions, etc.) sont considérés dans leur vraie signification, comme il semble bien nécessaire pour tout individu.

Les exemples schématisés (dés, extractions, etc.) ne cessent pas d'être utiles, comme ils le sont en tant qu'exemples simplifiés ; au contraire ils sont dangereux s'ils suggèrent que la théorie des probabilités est essentiellement bornée à ces cas. Ce serait comme suggérer que la mesure d'une longueur n'a de sens que lorsqu'on considère des segments formés par une suite de dés parfaitement égaux et parfaitement alignés et serrés.

Peut-on espérer qu'il y ait une évolution dans cette direction dans la façon d'introduire les probabilités dans l'école ?

Roma, 18.7.74