LA LIAISON STATISTIQUES-PROBABILITES DANS L'ENSEIGNEMENT

Jean-Claude GIRARD
IUFM et Irem de Lyon

1. Introduction

Si la théorie des probabilités a toujours été considérée comme une partie des mathématiques, depuis ses origines (Pascal lui donnait déjà le nom de *Géométrie du Hasard* en 1654) jusqu'à son axiomatisation par Kolmogorov en 1933, il n'en a pas toujours été ainsi en ce qui concerne la Statistique. Ceci est particulièrement vrai au niveau de l'enseignement puisque beaucoup de professeurs ont refusé pendant longtemps à la Statistique le statut de mathématiques. Une illustration peut en être donnée par ce livre de lycée [2] de 1966 intitulé « *Mathématiques **et** statistique* ». Cette conception n'a pas complètement disparu [3] de nos jours et ceci n'est pas sans conséquence sur certains choix qui sont faits, à un moment ou à un autre, dans l'enseignement.

Pour cette raison, et pour d'autres en rapport avec le mode de présentation préconisé par les programmes de l'enseignement secondaire, la liaison Statistique-probabilités a présenté différents visages, de réforme en réforme, jusqu'au dernier changement de programme des lycées (2000-2002).

1 Cet article reprend le texte d'une intervention au 2ème séminaire international sur l'enseignement des mathématiques, Santos (SP), Brésil, 29 octobre-1er novembre 2003.
2 Cluzel R., Vissio P. et Chartier F., Mathématiques et Statistique, 1ère D, Delagrave, 1966.
3 Certains pensent encore que l'enseignement de la Statistique devrait être assuré par ses utilisateurs : économistes, biologistes, etc… Il n'est même pas sûr qu'il y ait même un consensus sur le sujet chez les professeurs de mathématiques. Si on en croit le projet de programme de mathématiques pour la Terminale S (Groupe d'Experts pour les Programmes du Secondaire, version du 08/01/01) : Une partie (et une partie seulement) de l'enseignement de la Statistique trouve naturellement sa place dans le cadre d'un enseignement de mathématiques. Reste à définir quelle partie ?

2. Heurs et malheurs de la liaison Statistique-probabilités

L'enseignement des probabilités a été introduit dans les années 60, c'est-à-dire de façon relativement récente dans le cursus secondaire (Parzysz, 1997 et 2003), mais il a déjà connu bien des changements de cap.

Jusqu'en 1970, l'enseignement des probabilités et de la Statistique est réservé aux élèves se destinant aux études de biologie ou d'économie. La démarche est celle que l'on retrouve encore actuellement dans l'enseignement supérieur : statistique descriptive (en première), probabilités (en terminale). La liaison entre probabilités et Statistique est établie (toujours en terminale) par la statistique inférentielle (estimation, intervalle de confiance)[4].

La généralisation de l'enseignement des probabilités à (presque) tous les élèves accompagne la réforme dite « des maths modernes » (1970). C'est sans surprise une forme axiomatique qui est proposée en première et en terminale. On introduit la notion d'espace probabilisé[5]. La Statistique peut, elle-aussi, « bénéficier » d'une présentation « moderne ». Par exemple, un caractère statistique est défini comme une relation d'équivalence définie sur la population[6]. C'est *la* mathématique moderne qui assure la liaison, ou au moins l'uniformisation, entre probabilités et Statistique.

A cette époque, et même après la réforme de 1981 qui tente de redresser les excès de vocabulaire et de formalisation des maths modernes, la définition pratique de la probabilité est celle qui fut érigée en premier principe par Laplace[7] (mais utilisée bien avant, par Pascal par exemple) c'est-à-dire :

$$P(A) = \frac{\text{nombre de cas favorables à l'événement A}}{\text{nombre de cas possibles}}$$

Cela suppose de réduire les événements considérés à un système de cas équiprobables[8] et le calcul des probabilités se ramène à cette

La théorie des hasards consiste à réduire tous les événements du même genre à un certains nombre de cas également possibles, c'est-à-dire que nous soyons également indécis sur leur existence et à déterminer le nombre de cas favorables à l'événement dont on cherche la probabilité. Le rapport de ce nombre à celui de tous les cas possibles est la mesure de cette probabilité qui n'est autre qu'une fraction dont le numérateur est le nombre de cas favorables et dont le dénominateur est le nombre de cas possibles.
Pierre Simon de Laplace
Théorie Analytique des Probabilités, 1814.

4 Voir par exemple : Lespinard V., Pernet R., Mathématiques – Terminale B, Desvigne 1966.
5 Voir par exemple : Statistique et Probabilités, 1ères ABCDE et terminale B, R. Cluzel et P. Vissio, Delagrave, 1970.
6 Voir par exemple : Mathématique Terminale B, Collection Queyzanne Revuz, Nathan, 1971.
7 C'est ce qui permet à certains ouvrages (comme Déclic, Maths, Première ES, Hachette, 2001, page 190) de parler de « Formule de Laplace » pour introduire la probabilité d'un événement dans le cas d'une loi équirépartie.
8 mais sur quel ensemble ?

Ce qu'il n'est pas donné de savoir a priori l'est du moins a posteriori.

Jacob Bernoulli

Ars Conjectandi, 1713

époque à un jeu de logique ensembliste et à des dénombrements. Certaines séries (la série C, scientifique, par exemple) ne font d'ailleurs que cela. Le danger pour les élèves des autres séries est d'assimiler probabilités et dénombrements. Cette confusion se retrouve également chez les enseignants puisque l'abandon des formules de combinaisons, arrangements, permutations dans les réformes ultérieures a laissé de nombreux professeurs très perplexes sur ce qu'on pouvait faire en probabilités sans ces formules. De plus, de nombreuses études ont montré que ce type de présentation renforce le « biais d'équiprobabilité » c'est-à-dire l'idée, chez les élèves, que l'on peut toujours se ramener à une situation d'équiprobabilité.

A partir de 1986, la statistique (descriptive) fait son entrée, progressivement, dans toutes les classes du collège sous la dénomination « Organisation et Gestion des Données ». Cette réforme est prolongée au lycée et ceci aboutit, en 1991, à une nouvelle approche, dite « fréquentiste », des probabilités. La justification épistémologique est alors la référence à Bernoulli (cf. ancadré cidessus). Cette approche se veut expérimentale et tous les lycéens se mettent à lancer des punaises [9] (au moins virtuellement !). Ceci permet de parler de loi non équirépartie. La combinatoire est renvoyée en terminale puis disparaît peu à peu. Dans cette nouvelle optique, la probabilité d'un événement est appréhendée par l'observation de la stabilisation de la fréquence de réalisation de cet événement dans la répétition d'une même expérience aléatoire c'est-à-dire qu'elle est complètement liée à l'observation statistique. Le risque didactique est alors que les élèves assimilent fréquence observée et probabilité (théorique).

La dernière réforme (pour le moment) est celle de la « révolution statistique » (2000-2002). D'abord en terme de volume horaire (1/8 de l'année, en seconde, dit le programme officiel) puis en raison des nouveautés présentées. Le programme de seconde introduit ainsi l'observation des fluctuations d'échantillonnage et la simulation [10] . On n'y parle pas encore de

9 Pour estimer la probabilité qu'elles ont de tomber « pointe en haut », voir par exemple : Math Analyse 1ère S et E, Collection Terracher, Hachette, 1991.

10 Le programme de seconde propose une démarche expérimentale utilisant les TICE (technologies d'information et de communication pour l'enseignement) dans chacun des trois chapitres de cette classe (statistique, calcul et fonction, géométrie). L'utilisation de l'outil informatique (tableur, logiciel de géométrie dynamique) ou des calculatrices graphiques « multiplie …les possibilités d'expérimentation…Cet outil élargit les possibilités d'observation et de manipulation….Il donne la possibilité d'étudier une même notion sous uns plus grande diversité d'aspect ; cela contribue à la démarche d'abstraction propre au mathématiques et conduit à une meilleure compréhension ».

probabilité mais de « chances ». En première, une loi de probabilité est définie comme une distribution de fréquences théorique intervenant dans la modélisation d'une situation aléatoire. En terminale, les lois de probabilités continues font leur apparition de même qu'un test d'adéquation d'une distribution observée à une loi équirépartie s'inspirant du test du χ^2. Il y a ainsi une volonté de lier Statistique et probabilités par la loi des grands nombres (énoncée sous une forme « vulgarisée » ou « naïve » dit le programme). Mais, deux difficultés didactiques apparaissent cette fois, dont l'origine se trouve dans les activités de modélisation et de simulation.

3. Les difficultés liées au processus de modélisation

L'hypothèse des concepteurs du nouveau programme de lycée est que les élèves, confrontés en seconde aux fluctuations d'échantillonnage (réelles ou simulées) seront *« aussi familiers avec les objets « distributions de fréquences » qu'ils le sont par exemple en sixième avec les objets « cubes »* [11]

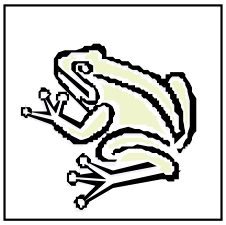
On voit donc dans cette démarche une analogie avec l'enseignement de la géométrie. Plus précisément, l'idée est que : *« après avoir effectivement manipulé des cubes en classes primaires, on peut leur parler de l'objet cube sans qu'ils en aient un devant les yeux : le cube est (devenu) un objet mental familier* [12] *».*

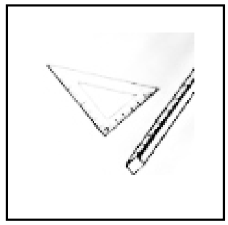
Mais suffit-il d'être familier avec les distributions de fréquences pour comprendre le concept de probabilité ? Si l'enseignement des probabilités présente une ressemblance avec celui de la géométrie (Henry,1999) (Girard, 1999), on peut toutefois pointer une différence essentielle.

En effet, la construction du modèle euclidien de géométrie est un long processus (qui s'étend sur une dizaine d'années) et dont le point de départ est l'observation et la manipulation d'objets concrets puis leur représentation par des dessins à l'aide de matériel. Ces objets sont reconnus globalement (à l'école maternelle et au début de l'école primaire), puis progressivement leurs propriétés sont repérées concrètement à l'aide d'instruments (à la fin de l'école primaire). Les objets mathématiques correspondants sont ensuite définis à partir de l'interprétation abstraite ces propriétés (au collège). On passe ainsi du cube au carré reconnu de façon perceptive puis au concept de carré défini uniquement pas ses propriétés théoriques, même si on en fait encore un dessin. On connaît la difficulté que ce saut conceptuel pose encore en quatrième malgré le nombre d'années sur lequel il s'étend, le passage de la réalité au modèle n'étant pas toujours explicite pour les élèves. On peut illustrer la progression de la façon schématisée en haut de la page suivante. Le programme prévoit pour les probabilités une démarche de modélisation du même type mais celle-ci s'étend au mieux sur deux ans (puisqu'on ne parle pas d'expérience aléatoire, ni même de hasard au collège). Le point de départ est l'observation de fréquences dans la répétition d'une expérience réelle puis simulée. Les probabilités sont alors définies comme fréquences théoriques liées à l'expérience (schéma du bas de page [13]).

11 Annexe 2 A propos des probas-stat de première et terminale S, projet de document d'accompagnement des nouveaux programmes de lycée, Ministère de l'Education nationale 2001.
12 ibid.

13 Voir par exemple : Déclic, Maths, Première ES, Hachette, 2001, page 188.





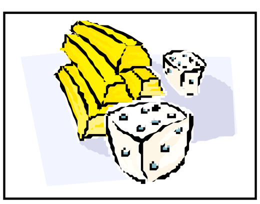
Soit ABCD un quadrilatère avec 4 côtés de même longueur et 4 angles droits.

Ecole maternelle (2-6 ans) ⟶ Ecole primaire (6-11 ans) ⟶ Collège (11-16 ans)

De plus, cette présentation a lieu de façon assez tardive, c'est-à-dire quand de nombreuses conceptions plus ou moins adéquates, voire erronées, se sont installées chez les élèves. Cet enseignement tardif n'arrange pas toujours les choses surtout quand il est de type dogmatique.

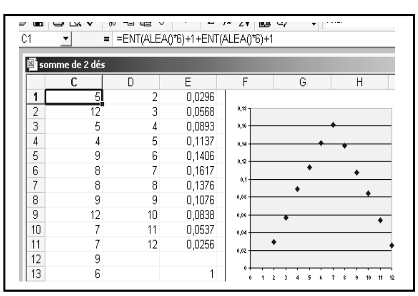
En effet, la perception du hasard n'est pas univoque et met quelquefois en jeu des croyances. Elle n'est pas l'objet d'un consensus comme cela peut être le cas dans l'observation des configurations géométriques (carré ou cube, par exemple). La modélisation dans ce cas risque donc d'être fragile et peu étayée sauf à commencer plus tôt et s'étaler sur un temps plus long.

Il existe un autre différence entre les deux situations. Le modèle euclidien de la géométrie étant construit, on peut se poser des problèmes dans le modèle (et on ne s'en prive pas) ou utiliser ce modèle pour traiter des problèmes concrets (ou faussement concrets) comme ceux bien classiques de la hauteur de la pyramide de Chéops ou du rayon de la



somme	2	3	4	5	6	7	8	9	10	11	12
fréquence	0,03	0,06	0,09	0,11	0,14	0,17	0,14	0,11	0,08	0,05	0,02

Seconde (15-16 ans) ⟶



s	2	3	4	5	6	7	8	9	10	11	12
p	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

⟶ Première (16-17 ans)

terre. Ce type de problème est toutefois rare, en particulier en situation d'examen, sauf à expliquer en détail la modélisation qu'il faut en faire. On observe ainsi deux types de modélisation en géométrie. Le premier (construction du modèle euclidien) qui s'étend sur 10 ans et qui permet de passer de l'espace sensible à l'espace mathématique et un autre qui permet (ou permettrait) de résoudre des problèmes concrets, le modèle euclidien étant maîtrisé (en partie).

La modélisation du deuxième type (utilisation d'un modèle), rare en géométrie, est par contre présente dans *tous* les exercices de probabilités. Ceux-ci se présentent en effet toujours avec un habillage concret. On fait l'hypothèse ici que les situations proposées seront assez proches des situations d'apprentissage (c'est-à-dire de référence) pour que les élèves puissent, sans aide, effectuer le transfert (d'abord du modèle équiréparti, plus tard de la loi binomiale ou de certaines lois continues). Le danger est alors que tout ceci n'ait pas beaucoup de sens pour beaucoup compte tenu du court temps d'assimilation.

4. L'ambiguïté de la simulation

Le programme de seconde utilise la simulation pour initier les élèves à l'aléatoire et les amener (en première) à la notion de probabilité.

Mais, d'après le programme de lycée [14]

« *Simuler une expérience, c'est choisir un modèle de cette expérience, puis simuler ce modèle* » et « *Modéliser une expérience à valeurs dans un espace E, c'est choisir une loi de probabilité P sur E* ».

En conséquence, pour simuler une expérience en seconde, il faut avoir un modèle, c'est-à-dire une loi de probabilité que l'on introduira en première seulement. Dans cette classe, les lois de probabilité seront alors définies par analogie avec les distributions de fréquences obtenues dans la simulation. Il y a là comme un cercle (didactique) vicieux.

D'autres problèmes se posent :

— Comment garantir l'équivalence d'une expérience aléatoire réelle avec sa simulation qui met en œuvre une autre expérience concrète ou pseudo-concrète (c'est-à-dire déjà idéalisée) ou informatisée (programmée à partir d'un modèle) ? Ce qui le permet, c'est que les deux expériences relèvent du même modèle probabiliste, mais encore une fois on ne peut faire aucune référence à un quelconque modèle en seconde. Comment interpréter alors les fluctuations d'échantillonnage observées dans la répétition de l'expérience simulée et non dans la répétition de l'expérience réelle ?

— Le programme [15] ajoute encore

« *Pour déterminer et/ou valider un modèle probabiliste, le premier outil dont on dispose est un théorème de mathématiques que l'on appelle loi des grands nombres, dont un énoncé intuitif est : dans le monde théorique défini par une loi de probabilité P sur un espace E, les fréquences des éléments de E dans une suite de n expériences identiques et indépendantes tendent vers*

14 Accompagnement des programmes de lycée, Mathématiques, rentrée 2002, Ministère de la jeunesse, de l'éducation et de la recherche, CD-ROM.

15 ibid.

leurs probabilités quand n augmente indéfiniment ».

La loi des grands nombres est effectivement un théorème interne à la théorie des probabilités, elle découle directement de son axiomatique. Les fréquences des éléments de E obtenues dans la simulation sont-elles des objets du modèle ou dans la réalité ? Dans le premier cas, on ne valide rien du tout puisqu'on ne sort pas du modèle, dans le deuxième cas comment les fréquences, objets du monde réel peuvent-elles « tendre » vers les probabilités mathématiques ? Il y a donc là une ambiguïté épistémologique qui fait que l'on ne sait plus si on est dans le modèle ou dans la réalité. Il est malencontreux que cette difficulté fréquente (Girard, 2001 et 2003) soit présente au cœur de l'apprentissage.

— Pour finir, s'il est vrai que

« *l'esprit statistique naît lorsque l'on prend conscience des fluctuations d'échantillonnage* [16] »,

l'observation des fluctuations de la distribution des fréquences lors de plusieurs simulations d'une même expérience aléatoire (simulée ou non) conduit-elle naturellement à accepter l'idée d'une loi théorique fixe liée à cette expérience ?

Enseigner les probabilités par la loi des grands nombres et la simulation n'est donc pas un long fleuve tranquille. Cela nécessite de toute façon un temps d'apprentissage beaucoup plus long que le temps d'enseignement prévu pour le moment.

16 ibid.

5. Conclusion

La formation d'images mentales à propos de l'aléatoire est plus délicate et demande encore plus de temps que dans le cas de la géométrie. Il faut donc proposer aux élèves des activités propres à créer ces images mentales bien avant le lycée. Des recherches récentes ont montré que des élèves de collège pouvaient utiliser la simulation pour construire des expériences équivalentes (Bordier, 1991), pour peu qu'ils soient confrontés à des expériences aléatoires concrètes pendant un certain temps, et qu'ils pouvaient modéliser des situations par analogie avec des urnes de Bernoulli (Coutinho, 2001). Mais pour l'instant, rien n'est prévu, en France, concernant l'aléatoire au collège (Girard et al., 2001). On peut bien sûr le regretter.

Le nouveau programme de l'école primaire [17] envisage par contre cette possibilité :

« *Quelques exemples de phénomènes aléatoires peuvent être proposés dans la perspective de faire apparaître des régularités (par exemple, lancers d'une pièce ou d'un dé, lancers de deux dés dont on fait la somme)* ».

Les prochains programmes de collège combleront-ils le trou entre l'école primaire et le lycée au sujet de l'aléatoire ? Se rapprocheront-ils des enseignements délivrés à ce niveau dans de nombreux pays étrangers ? On peut l'espérer.

Si c'était le cas, la liaison statistique-probabilité serait présente tout au long du pro-

17 Documents d'accompagnement des programmes, Mathématiques, cycle 3. Ministère de la Jeunesse, de l'Education et de la Recherche, 2002.

cessus d'apprentissage de la modélisation, de l'école primaire à la seconde, dans un apprentissage allant du perceptif au théorique. Par analogie avec ce qui est proposé pour la géométrie, le point de départ pourrait être alors l'observation, la construction, la reproduction, la description et la représentation d'expériences aléatoires.

L'enseignement de la géométrie établit à partir d'objets concrets une construction intellectuelle qui n'utilise que les propriétés d'objets définis mathématiquement c'est-à-dire de concepts. Le passage se fait sur une très longue période d'abord par une reconnaissance globale et perceptive des objets concrets puis une reconnaissance instrumentée des propriétés sur les objets ou leur représentation dessinée avant de parvenir au raisonnement hypothético-déductif sur les propriétés elles-mêmes c'est-à-dire sur les objets mathématiques. Comme le mot « carré » fait partie du langage courant avant de faire partie du langage mathématique et d'être un concept, il pourrait en être de même pour le mot « probabilité ». Qu'on le veuille ou non, ce mot est utilisé dans le langage courant, autant faire en sorte que les élèves lui donnent un sens le moins éloigné possible de la définition mathématique.

Les techniques de statistique descriptive du collège (moyennes, pourcentages, graphiques en barres, circulaires, en boîtes, en tiges et feuilles…) trouveraient naturellement leur place pour communiquer et résumer les résultats de ces expérimentations. Les exercices de statistique au collège ne porteraient plus uniquement sur des populations mais sur des échantillons tirés de ces populations. La liaison entre le point de vue statistique et le point de vue probabiliste pourrait être l'étude d'un caractère (qualitatif ou quantitatif) sur les individus d'un échantillon « tiré au hasard » dans une population. Ceci pourrait se faire en tirant des étiquettes correspondant aux individus dans un chapeau, ensuite avec une table de nombre au hasard et enfin avec la touche Random de la calculatrice.

Ces expériences illustreraient d'une façon perceptible la variabilité des résultats et les fluctuations d'échantillonnage. La notion d'expériences équivalentes se construirait alors en acte et le passage au modèle sous-jacent aurait plus de chance de prendre sens plus tard si toutefois ce passage est clairement identifié comme c'est le cas (ou comme cela devrait être le cas) pour la géométrie. Pour le moment, toutes ces étapes sont concentrées en seconde (souvent en fin d'année, quelquefois pas du tout) et en première (éventuellement avant la définition formelle). Il y a peu d'études sur les difficultés des élèves par rapport à cette nouvelle présentation mais on en a un aperçu dans les difficultés des professeurs qui n'ont pas reçu d'enseignement sur le sujet pendant leurs études au cours des formations qui ont été mises en place (parcimonieusement) à la suite du changement de programme.

La conclusion évidente est qu'il y a matière à développer des recherches sur la liaison statistique-probabilités aboutissant à des propositions curriculaires sur l'ensemble de la scolarité depuis le primaire jusqu'au secondaire car c'est sûrement là un point crucial de l'enseignement. En effet : *La Statistique sans la théorie des Probabilités est aveugle et la théorie des Probabilités sans la Statistique est vide.* [18]

18 Hanss Schupp, « Appropriate teaching and learning of stochastics in the middle grades (5-10) », in Studies in mathematics education. Vol 7. The teaching of statistics. Editions UNESCO, Paris, 1994.

Bibliographie

Bordier J., Un modèle didactique utilisant la simulation sur ordinateur, pour l'enseignement de la probabilité. Thèse de doctorat, Université Paris-7, 1991.

Coutinho C., Introduction aux situations aléatoires dès le collège : de la modélisation à la simulation d'expériences de Bernoulli dans l'environnement informatique cabri-Géomètre 2. Thèse de doctorat, Université Joseph Fourier, Grenoble, 2001.

Girard J. C., « Le professeur de mathématiques doit-il enseigner la modélisation ? », in *Repères-IREM n° 36*, p. 7-14, Topiques Editions, 1999.

Girard J. C. « Un exemple de confusion modèle-réalité » in *Autour de la modélisation en probabilités,* p. 145-148, M. Henry (Coord.), Commission inter-IREM Enseignement de la statistique et des probabilités, Presses Universitaires Franc-Comtoises, 2001.

Girard J. C., « Modélisation et simulation » in *Probabilités au lycée*, Commission inter-IREM Enseignement de la statistique et des probabilités, brochure APMEP n° 143, p. 143-151, B. Chaput (Coord.), 2003.

Girard J. C., Henry M., Parsysz B., Pichard J. F., « Quelle place pour l'aléatoire au collège ? », in *Repères-IREM n° 42,* p. 27-43, Topiques Editions, 2001.

Henry M., « L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie », in *Repères-IREM n° 36*, p. 15-34, Topiques Editions, 1999.

Parsysz B., « L'enseignement de la statistique et des probabilités dans l'enseignement secondaire, d'hier à aujourd'hui », in *Enseigner les probabilités au lycée*, p. 17-38. Commission inter-IREM Enseignement de la statistique et des probabilités, IREM de Reims, 1997. Article remanié et actualisé sous le titre *L'enseignement de la Statistique et des probabilités en France : évolution au cours d'une carrière d'enseignant (période 1965-2002)* in *Probabilités au lycée*, Commission inter-IREM Enseignement de la statistique et des probabilités, brochure APMEP n° 143, p. 9-34, 2003.