La liaison statistique-probabilités dans les nouveaux programmes français de lycée

Le statut de la modélisation et de la simulation

Jean Claude Girard*

** IUFM de Lyon*
Centre local de St-Etienne
90, rue de la Richelandière
F-42023 St-Etienne Cedex
jean-claude.girard @lyon.iufm.fr

Thème : Enseignement des probabilités et de la statistique

RÉSUMÉ : Les derniers changements dans les programmes de lycée en France (2000-2002) ont donné une importance beaucoup plus grande à l'enseignement de la statistique et ont introduit une présentation de la théorie des probabilités en tant que modélisation des situations aléatoires. La méthode utilisée fait une large place à la simulation.
Cette nouvelle approche pose de nouveaux problèmes didactiques liés à la durée nécessaire à l'activité de modélisation et à l'ambiguïté de la simulation. Elle ne résout pas pour autant les difficultés bien connues comme l'assimilation fréquence-probabilité.
La liaison entre l'enseignement de la statistique et celui des probabilités reste donc à construire. Elle nécessite sans doute de s'étendre sur de nombreuses années et pourrait s'inspirer de l'enseignement de la géométrie.

1. Introduction

Si les probabilités ont toujours été considérées, dans l'enseignement comme ailleurs, comme une partie des mathématiques, depuis Pascal jusqu'à Kolmogorov, il n'en a pas toujours été ainsi pour ce qui concerne la statistique, témoin ce livre de lycée[1] de 1966 intitulé " *Mathématiques **et** statistique* ". Pour cette raison, et pour d'autres en rapport avec le mode de présentation préconisé par les programmes officiels, la liaison statistique-probabilité a présenté différents visages, de réforme en réforme, jusqu'au dernier changement de programme des lycées (2000-2002).

[1] Cluzel R., Vissio P. et Chartier F., *Mathématiques et Statistique,* 1ère D, Delagrave, 1966.

2. Heurs et malheurs de la liaison statistique-probabilité

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), mais il a déjà connu bien des changements de cap. Les probabilités pour tous ont accompagné la réforme dite " des maths modernes " (1970). C'est sans surprise une forme axiomatique (espace probabilisé) qui est proposée en première et en terminale. A cette époque, et même après la réforme de 1981 qui tente de redresser les excès des maths modernes, la définition pratique de la probabilité est celle qui fut érigée en premier principe par Laplace. Cela suppose l'équiprobabilité des événements élémentaires et le calcul des probabilités se ramène alors à des dénombrements. Certaines séries (C, par exemple) ne font d'ailleurs que cela. Pour les autres, cela conduit souvent à assimiler les deux concepts.

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 ". Ceci aboutit, en 1991, à une approche fréquentiste de la probabilité. On passe de Laplace à Bernoulli. Cette approche est expérimentale et la France lycéenne lance des punaises. La probabilité est appréhendée par l'observation de la stabilisation de la fréquence dans la répétition d'une expérience aléatoire. On peut alors parler de loi non équirépartie. La combinatoire est renvoyée en terminale et disparaît peu à peu. 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[2] . On n'y parle pas encore de probabilité mais de

2. 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*

“ chances ”. Les élèves étudient les quartiles et les boîtes à moustaches en première et on y définit une loi de probabilité 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 probabilité. Mais, deux difficultés didactiques apparaissent alors, dont l’origine se trouve dans les activités de modélisation et de simulation.

3. Le paramètre temps dans la 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 ” : 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 un objet mental familier*[3] *”.*

L’analogie est ainsi suggérée avec la géométrie. 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.

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 d’objets concrets et leur représentation par des dessins à l’aide d’instruments. Ces objets sont reconnus globalement (à l’école maternelle et au début de l’école primaire), puis progressivement leurs propriétés sont repérées concrètement à

---

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. ”

3. Annexe 2 *A propos des probas-stat de première et terminale S*, projet de document d’accompagnement des nouveaux programmes de lycée, MEN 2001.

l'aide d'instruments (à la fin de l'école primaire). Les objets mathématiques correspondant sont ensuite définis à partir de l'interprétation abstraite ces propriétés (au collège). On passe ainsi du carré reconnu de façon perceptive 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'année sur lequel il s'étend, le passage de la réalité au modèle n'étant pas toujours explicite pour les élèves.

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 de hasard ni d'expérience aléatoire au collège) et de façon assez tardive (15-16 ans) 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[4].

4. Il existe un autre différence entre les deux situations. L e 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 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

D'après le programme de lycée[5] " *Simuler une expérience, c'est choisir un modèle de cette expérience, puis simuler ce modèle* " mais " *Modéliser une expérience à valeur 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.

Comment garantir l'équivalence d'une expérience aléatoire réelle avec sa simulation par une autre expérience concrète ou pseudo-concrète (c'est-à-dire déjà idéalisée) ou informatisée (programmée à partir d'un modèle) ? Ce qui garantit l'équivalence, c'est que les deux expériences relèvent du même modèle, mais encore une fois on ne fait 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 de l'expérience étudiée ?

Le même programme ajoute " *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 leurs probabilités quand n augmente indéfiniment* ".

La loi des grands nombres est donc un théorème interne à la théorie des probabilités. Les fréquences des éléments de E obtenues dans la simulation sont-elles dans le 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 premières peuvent-elles " tendre " vers les secondes ? Il a donc là une

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

ambiguïté épistémologique qui fait que l'on ne sait plus si on est dans le modèle ou dans la réalité, et s'il est vrai que " *l'esprit statistique naît lorsque l'on prend conscience de fluctuation d'échantillonnage* ", l'observation des fluctuations de la distribution de fréquences lors de plusieurs simulations d'une même expérience aléatoire conduit-elle naturellement à accepter l'idée d'une loi théorique fixe liée à cette expérience ? L'apprentissage des probabilités par la loi des grands nombres et la simulation n'est donc pas un long fleuve tranquille. Il nécessite de toute façon un temps d'apprentissage beaucoup plus long que le temps d'enseignement prévu pour le moment.

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) et qu'ils pouvaient modéliser des situations par analogies avec des urnes de Bernoulli (Coutinho, 2001). Mais pour l'instant, rien n'est prévu concernant l'aléatoire au collège (Girard et al., 2001).

Le nouveau programme de l'école primaire 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 ?

Si c'était le cas, la liaison statistique-probabilité serait alors présente tout au long du processus d'apprentissage de la modélisation, de l'école primaire à la seconde. Par analogie avec ce qui est proposé pour la géométrie, le point de départ pourrait être 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 ”. De toute façon, ce mot est utilisé dans le langage courant, alors autant lui donner un sens le moins faux possible.

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 tirées de ces populations. La liaison entre les deux points de vue 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 perceptive 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 cela devrait l'être 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.

6. 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 ? ”, *Repères-IREM 36*, p. 7-14, Topiques Editions, 1999.

Girard J. C., Henry M., Parsysz B., Pichard J. F., “ Quelle place pour l'aléatoire au collège ”, *Repères-IREM 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 ”, *Repères-IREM 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.