Perspectives de l'enseignement de la statistique et des probabilité

Michel HENRY (*Université de Franche-Comté, IREM*)

1. Les objets d'un débat abusivement présenté comme une querelle entre anciens et modernes

La publication au B.O. du nouveau programme de seconde, applicable à la rentrée 2000, dont la partie statistique a été adoptée « sans réserve » par le conseil national des programmes, a soulevé une émotion visible (pétitions...) dans la communauté des enseignants de mathématiques. Un débat a posteriori s'est développé sur la faisabilité de l'enseignement proposé par ce programme et des comptes rendus de nombreuses expérimentations commencent à être diffusés. Leur analyse pourrait apporter un peu de sérénité, tout en ne réglant ni le problème de fond sur la place relative de l'initiation à l'observation statistique et de l'enseignement plus théorique des probabilités, ni celui de la formation massivement nécessaire des enseignants confrontés à des démarches assez nouvelles pour eux, dans un contexte de réductions horaires et d'alourdissement des contraintes de l'enseignement des mathématiques dans des classes de seconde très hétérogènes.

Le caractère très limité de cet article ne permet pas de développer comme il serait souhaitable cette problématique qui oppose une approche purement expérimentale de cet enseignement à une conception plus théorisante de l'initiation aux probabilités. Il ne semble pas y avoir de divergence profonde quant à l'appréhension du statut théorique de la probabilité, en tant qu'objet mathématique défini dans le cadre de la modélisation de phénomènes aléatoires. La discussion porte surtout sur les objectifs qu'il convient d'assigner à cet enseignement au lycée.

L'option retenue par le programme de seconde est de se limiter à l'observation de l'aléatoire, sous la forme des « fluctuations d'échantillonnage » et à l'étude des « distributions des fréquences » des issues d'une expérience répétée, variables suivant les échantillons obtenus. Seules les situations d'équiprobabilité, ramenées à des jeux de hasard simples (pièces, dés...) sont envisagées comme exemples de manipulations, simulées ensuite grâce aux performances des outils informatiques.

J'ai peur d'avoir par trop caricaturé l'esprit de cette introduction à la statistique, une lecture du programme lui-même me semble tout à fait nécessaire. Un des éléments les plus importants de la controverse, me semble-t-il, est l'évitement délibéré et systématique du concept de probabilité, remplacé par la notion floue de « chances », interdisant la possibilité d'une prise de recul un peu théorique dans l'explicitation de ces faits d'observation. Il semble que ce point de vue s'impose également dans les projets actuellement en débat des programmes de première et terminale S, où la notion de loi de probabilité est dégagée de l'étude empirique des distributions de fréquences. Nous divergeons donc sur la place et la nature de l'enseignement des probabilités au lycée.

La dialectique statistique-probabilité me semble devoir faire l'objet d'un enseignement coordonné, prenant ses racines dès le collège. C'est pourquoi, je ne suis pas vraiment satisfait du programme actuel de probabilités, introduites tardivement en première, sans lien explicite avec les démarches de la décision statistique. Les travaux contemporains sur cet enseignement[1] mettent en avant l'importance de l'initiation aux processus de modélisation à partir de situations aléatoires simples, soit d'équiprobabilité légitimement postulée, soit sous des hypothèses issues de l'observation des fréquences.

Mais le nouveau programme de seconde me semble prendre un contre-pied trop systématique à l'enseignement actuel, fruit d'une évolution progressive positive et non achevée. Pour replacer ces changements dans leur perspective, il n'est pas inutile de revenir sur trente cinq années d'enseignement des probabilités au lycée.

2. Évolution de l'enseignement des probabilités dans le secondaire français de 1965 à 1991

L'enseignement des probabilités dans le secondaire est relativement récent[2]. À l'exception de certaines séries techniques économiques de la fin des années 50, il remonte au milieu des années 60. Pour se faire une idée de son évolution au cours de ces 30 dernières années, on peut reprendre quelques extraits des programmes des séries scientifiques classiques (terminales C et D), qui ont précédé le changement de point de vue des programmes de 1991, adapté ensuite aux filières actuelles L, ES et S en 93 :

> *— 1965 (TD) : « 1° Préliminaires d'analyse combinatoire... 2° Principe du calcul des probabilités. Variable aléatoire... 3° Statistique appliquée ».*
> *— 1970 (TC) : « 1° Espaces probabilisés finis* $(\Omega, \mathrm{B}(\Omega), p)$*. Applications mesurables... 2° Espérance mathématique,... 3° loi faible des grands nombres ».*
> *— 1982 (TD) : « a) Combinatoire. b) Exemples de situations où le hasard intervient. Ensemble fini d'épreuves* Ω*... Probabilité uniforme sur* Ω*, calcul des probabilités par dénombrement... c) Aléa numérique... d) statistiques ».*
> *— 1986 (TC) : « 1. Combinatoire ; probabilités... En probabilités, l'objectif est d'entraîner les élèves à décrire grâce au langage élémentaire des événements, quelques expériences aléatoires simples, et à employer les techniques de dénombrement pour calculer des probabilités ».*

On peut y remarquer un dénominateur commun : mis à part les années 70 des « maths modernes », où la structure d'espace probabilisé servait de toile de fond à l'introduction des probabilités comme objets abstraits définis dans un modèle mathématique, le calcul des probabilités est a priori présenté comme application de la combinatoire. La notion de probabilité est en effet introduite

---

[1] Notamment publiés dans le livre coordonné par R. Kapadia et M. Borovcnik *Chance encounters : Probability in Education*, Kluwer Academic Publishers, 1991.

[2] Pour une étude plus complète sur l'évolution des programmes de l'enseignement secondaire français, on pourra se reporter à l'article de Bernard Parzysz *Les probabilités et la statistique dans le secondaire d'hier à aujourd'hui*, publié dans le livre de la commission inter-IREM *« Statistique et probabilités » : enseigner les probabilités au lycée*, ed. IREM de Reims, 1997.

par le premier principe de Laplace qui donne comme définition de la probabilité d'un événement décrit par un système de cas supposés équiprobables, le rapport $\frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$.

Cette définition est restrictive, adaptée aux situations d'équiprobabilité idéales des jeux de hasard. En posant que toute situation aléatoire peut se ramener à un système de cas équiprobables, elle réalise difficilement le lien avec l'observation de la réalité : phénomènes d'attente, fiabilité en technologie, modèles économiques... Sur le plan de l'enseignement, elle a eu surtout pour effet d'enfermer les élèves dans des exercices délicats de combinatoire, ce qui fut la source de l'échec de cet enseignement pour bon nombre d'entre eux.

Citons Jacques Bernoulli[3] :

> *« On en est ainsi venu à ce point que pour former selon les règles des conjectures sur n'importe quelle chose, il est seulement requis d'une part que les nombres de cas soient soigneusement déterminés, et d'autre part que soit défini combien les uns peuvent arriver plus facilement que les autres. Mais c'est ici enfin que surgit une difficulté, nous semble-t-il : cela peut se voir à peine dans quelques très rares cas et ne se produit presque pas en dehors des jeux de hasard... En effet lorsqu'il s'agit de tous les autres résultats, dépendant pour la plupart soit de l'œuvre de nature soit de l'arbitre des hommes, cela n'a pas du tout lieu ».*

Et Bernoulli cite les phénomènes climatiques, la propagation des épidémies, les choix stratégiques entre adversaires.

3. L'enseignement des probabilités dans les années 90 au lycée

La réforme du programme de première en 1991 tente un rapprochement avec l'observation de la réalité aléatoire pour introduire la notion de probabilité à partir de l'observation des fréquences :

> *« L'objectif est d'entraîner les élèves à décrire quelques expériences aléatoires simples et à calculer des probabilités. On évitera tout développement théorique. Pour introduire la notion de probabilité, on s'appuiera sur l'étude de séries statistiques obtenues par répétition d'une expérience aléatoire, en soulignant les propriétés des fréquences et la relative stabilité de la fréquence d'un événement donné lorsque cette expérience est répétée un grand nombre de fois ».*

Puis le concept mathématique de probabilité est introduit par la définition empruntée au deuxième principe de Laplace :

> *« La probabilité d'un événement est définie par addition de probabilités d'événements élémentaires ».*

Cette définition, très générale dans le cas discret, laisse la porte ouverte sur l'introduction des probabilités élémentaires : hypothèses de modèle (équiprobabilité ou répartition théorique donnée), estimation fréquentiste, ajustements de lois, appréciations subjectives...

---

[3] *Ars Conjectandi*, chapitre IV de la 4e partie : La double manière de rechercher les nombres de cas...

Quelques remarques sur ce programme de 1991[4] :

— Le choix d'introduire la notion de probabilité par l'observation de la « relative » stabilisation des fréquences lors de la répétition d'une même expérience aléatoire, induit un regard expérimental sur cette notion. Cela suppose la mise en œuvre dans la classe d'expériences concrètes (répétées par l'accumulation des observations de chacun des élèves) et de simulations sur ordinateur. Il faut reconnaître que cette partie de la découverte de cette nouvelle notion n'est pas suffisamment prise en compte par les enseignants qui sont souvent en manque de temps (mais ce n'est pas la seule raison : des arguments épistémologiques et didactiques plus profonds seraient éclairants).

— La définition proposée est la bonne, car elle correspond à l'idée plus générale de mesure : sous-additivité de la probabilité. Le cas particulier de l'équiprobabilité est introduit seulement ensuite dans le programme. Cela n'empêche pas les manuels de le mettre massivement en avant dans les « activités introductives » et les exercices proposés, et les enseignants de se rabattre très vite sur des exercices de combinatoire simple.

— La rédaction de ce programme comporte certaines faiblesses, voire ambiguïtés. En faisant l'impasse sur tout processus de modélisation (on s'en tient à la « description d'expériences aléatoires simples » sans proposer de l'interpréter en un modèle théorique), le programme autorise la confusion entre fréquence expérimentale et probabilité conçue alors à tort comme une fréquence limite (quelle sorte de limite entre des relevés d'observations et un nombre théorique ?).

— L'existence même de la probabilité (son unicité objective) est source de questionnement épistémologique. Si elle n'est généralement pas niée dans le cadre de situations relevant de la « géométrie du hasard », elle suppose une inférence dans les cas où l'équiprobabilité n'a pas de sens[5].

En fait la confusion entre modèle et réalité est omniprésente dans l'enseignement des mathématiques et à l'origine de difficultés didactiques essentielles. Le projet de programme de première L de 1993 avait tenté de pallier à cet obstacle, mais il n'a pas été retenu :

> *« Il s'agit, comme dans les autres programmes, d'aborder la notion de probabilité à partir de la fréquence, mais on a choisi dans cette série d'affiner l'explicitation du processus de modélisation. L'objet de cette partie de la formation est donc de faire découvrir, en s'appuyant sur l'expérimentation numérique, quelques notions qualitatives et quantitatives liées à la modélisation mathématique des phénomènes aléatoires ».*

— Le programme de 1991 présente une réelle disjonction entre l'introduction expérimentale de la notion et l'usage ultérieur théorique qui en est fait. La simulation n'est pas évoquée, le vocabulaire ensembliste, délaissé jusque là, est appelé pour combiner formellement les événements sans références à leurs significations logiques, ni retour au contrôle expérimental. La notion de loi est

[4] On trouvera une critique plus étoffée sur les plans épistémologiques et didactiques dans *L'enseignement des probabilités*, M. Henry, ed. IREM de Besançon, ainsi que dans *Enseigner les probabilités au lycée*, Commission Inter-IREM « Statistique et probabilités », ed. IREM de Reims.

[5] Cf. le raisonnement d'Hélène Ventsel cité dans l'article : *L'enseignement du calcul des probabilités dans le second degré*, M. Henry, Repères-IREM n° 14, janvier 1994.

introduite formellement en terminale, sans la dégager de situations concrètes lui donnant sens et induisant une certaine cohérence dans la répartition des probabilités élémentaires. La mise hors programme de toute loi standard enlève à cette notion toute utilité à ce niveau.

Ainsi, de nombreuses critiques, notamment dans les publications des IREM, émergeaient après sept ans d'enseignement de ces programmes de probabilités. En 1991 ils avait représenté une réelle avancée, mais ils demandent aujourd'hui à être améliorés pour mieux atteindre les objectifs de formation scientifique dans laquelle l'enseignement de la statistique et des probabilités prend toute sa place.

4. L'enseignement des outils de la statistique descriptive

Les programmes des collèges de 1985 introduisent très progressivement les outils de la description statistique sous le titre « organisation et gestion des données » : populations statistiques et caractères, séries statistiques, séries classées, effectifs et fréquences, représentations graphiques, diagrammes, paramètres de position (médiane, moyennes), notion de dispersion. L'ancien programme de seconde proposait une synthèse des notions et du vocabulaire de base de cette description statistique, prolongée dans certaines séries de premières et terminales par les outils d'étude de statistiques à deux caractères conjoints.

L'acquisition de ces notions et leur utilisation critique est essentielle pour la formation des futurs citoyens : lecture de tableaux, de graphiques, d'analyses économiques, sociologiques... Or il apparaît que cet apprentissage n'est pas spontané, il s'appuie sur une bonne compréhension des pourcentages et de la proportionnalité, ainsi que sur une certaine pratique des articulations logiques (disjonction, conjonction, négation) mettant en jeu un peu du vocabulaire des ensembles. Il me semblerait désastreux de faire l'impasse sur cette partie des connaissances de base, qui sont loin d'être maîtrisées à l'entrée en seconde.

Cependant, les programmes des collèges et de seconde des années 85–99 évitaient soigneusement de présenter les données statistiques comme provenant d'un échantillonnage (aléatoire ou systématique) ou plus généralement d'un ensemble de résultats aléatoires. Elles étaient présentées comme exhaustives, ce qui ne permet pas dans ces conditions de soulever des questions d'inférence, de jugement statistique. Ainsi les élèves ne pouvaient réellement comprendre la fonction d'aide à la décision de la statistique moderne. On peut donc penser que limiter cet enseignement aux seuls outils de la description statistique, occultait le sens profond des démarches en statistique. Dans le domaine de l'enseignement de la statistique, une mise à jour paraissait également s'imposer.

5. Un regard critique sur le nouveau programme de seconde

Les intentions affichées dans le projet de document d'accompagnement me semblent excellentes :

— acquérir une expérience de l'aléatoire ;

— comprendre ce qu'est une question statistique et le type de réponse que l'on peut proposer ;

— voir dans un cas simple ce qu'est un modèle probabiliste.

Concernant l'expérience de l'aléatoire, on peut remarquer qu'il est grand temps en classe de seconde de travailler sur un sujet qui semble tabou au collège : l'aléatoire dont les élèves sont familiers depuis longtemps dans leur vie quotidienne. Des expérimentations (doctorat en cours) montrent que la variabilité des issues d'une expérience aléatoire est facilement acceptée, et dans les situations simples de succès-échec, peut être modélisée par des tirages dans une urne de Bernoulli dont le statut théorique ne semble pas poser de problème.

Mais aborder l'aléatoire par l'observation des fluctuations d'échantillonnage me semble ambitieux pour un démarrage. À l'étape actuelle, en attendant que les futurs programmes de collèges intègrent cette familiarisation avec l'aléatoire, ne vaudrait-il pas mieux se limiter en seconde à l'observation des fréquences de réalisation d'un événement dans plusieurs séries d'épreuves, comme les programmes des classes de première le proposaient jusqu'à présent ?

Cela amène à s'interroger sur les conditions didactiques permettant de travailler sur la simulation d'expériences aléatoires, sans introduire la notion de probabilité en tant que concept théorique. Quel lien en effet établir entre une simulation et l'expérience réelle sans passer par la référence à un modèle théorique commun ?

Je partage entièrement la critique du risque des programmes de première actuels de faire passer la probabilité pour une fréquence limite, ce qui effectivement n'a pas beaucoup de sens. Or n'est-ce pas multiplier ce risque que de proposer l'étude et même la comparaison des distributions de fréquences d'événements associés à une épreuve répétée, comme préfigurant la « distribution théorique » de la probabilité ?

Je pense qu'il conviendrait de limiter la complexité de ces premiers pas : les travaux en didactique des probabilités montrent que la notion même d'expérience aléatoire nécessite un certain travail, faute duquel des confusions rédhibitoires sont monnaie courante. De ce point de vue, il me semble essentiel de bien séparer dans le vocabulaire ce qui relève de l'observation statistique de ce qui permet la description d'un modèle probabiliste interprétatif. Quel sens donner par exemple à la locution « événement associé à une série statistique » ? De manière générale, il ne me semble pas judicieux de réduire l'outil statistique à la seule description des issues d'une épreuve aléatoire répétée.

Pour revenir au nouveau programme de statistique en seconde, si la variabilité des résultats obtenus dans un échantillonnage en est un des objectifs, les moyens d'exprimer cette variabilité doivent en être les outils. L'argumentation présentée pour l'étendue ne me paraît pas convaincante, ne traduisant pas suffisamment l'idée de dispersion. L'intervalle interquartile, très facilement accessible aussi bien pratiquement que conceptuellement (sans interpolations !), en donne une meilleure idée, tout en relativisant le poids des valeurs extrêmes. Sa représentation simple en boîtes à pattes (ou à moustaches) permet des comparaisons visuelles. Le concept étant ainsi introduit, la critique du caractère grossier de l'intervalle interquartile étant faite, son raffinement peut aboutir à l'écart type.

Mais le lien de ce paramètre avec les répartitions gaussiennes me semble impossible à présenter en seconde. Qu'est-ce qu'un « phénomène gaussien » ? Qu'en est-il pour un élève, mais aussi pour un enseignant qui n'a reçu qu'une formation élémentaire en probabilités ? Remarquons, pour répondre à certains arguments, que l'écart type intervient aussi comme outil théorique en dehors de situations gaussiennes (par exemple dans l'inégalité de Tchebychev) : si Bernoulli en avait disposé, sa démonstration de son « théorème d'or » en aurait été grandement raccourcie.

J'en viens à ce qui a fait le plus réagir les collègues qui enseignent en classe de seconde : l'introduction en thèmes d'étude aux situations de sondages (répartition d'un fort pourcentage des fréquences observées sur des échantillons de taille suffisante dans un intervalle calculable a priori). Ce thème d'étude introduit la notion de « fourchette de sondage », toujours en l'absence de la notion de probabilité. L'intention me semble tout à fait intéressante, ne serait-ce que pour parler sérieusement des résultats d'un sondage par exemple. Mais elle suppose beaucoup de travaux didactiques pour aboutir à l'installation d'un savoir à ce niveau. Je ne sais pas si cela est accessible. Sans une claire compréhension du caractère aléatoire des échantillons observés et du caractère probabiliste des indications que l'on peut en tirer, on ne peut s'en tenir qu'à un aperçu empirique. Il faudra beaucoup d'exemples pour dégager une idée transférable et en faire un outil de décision dans le cadre de l'observation statistique. Le programme va jusqu'à faire remarquer que l'écartement de cette « fourchette » est en $\frac{1}{\sqrt{n}}$. Je ne vois pas ce qui dans ce contexte empirique peut le valider, pourquoi pas en $\frac{1}{\ln(n)}$? Comment, dans ces conditions, éviter un enseignement dogmatique auquel les élèves ne peuvent donner du sens ?

D'autres thèmes d'étude proposés (promenades aléatoires) me semblent maladroits par l'effet d'incompréhension qu'ils provoquent chez les lecteurs du programme que sont les enseignants. Intéressants en soi, ils me semblent plutôt convenir pour l'illustration de modèles probabilistes à un niveau post-bac que pour un enseignement de la statistique.

Je conclurai ces quelques remarques introductives à un débat d'ordre épistémologique et didactique, par des préoccupations de nature à la fois psychologiques et institutionnelles.

Nous avons déjà connu (période des maths modernes) cette sorte de volontarisme prétendant pousser les enseignants à compléter leur formation par l'obligation de s'adapter à une réforme des programmes. Pour louable qu'il soit, il se heurte aux réalités : on n'enseigne vraiment que ce qui est profondément compris. Il faut que l'objet d'enseignement puisse vivre dans une relation didactique contrainte par une culture scolaire dont l'évolution est lente et complexe. Combien de professeurs de seconde auront compris les objectifs de ce programme au point d'en faire leur affaire personnelle ? Combien sauront aménager ces objectifs avec les réductions d'horaires dont ils disent être victimes en fonction de leur propre conception de l'enseignement des mathématiques ? Il y aura à vaincre des résistances sur le plan même du rapport au savoir.

L'enjeu profond de ce programme, qu'au fond je partage, est de prendre en charge la dimension expérimentale des outils mathématiques et d'en tirer une didactique. La meilleure méthode n'est pas forcément de bousculer les pratiques enseignantes au point de déstabiliser un système de transposition, même sous la promesse d'une formation continue dont la mise en œuvre laisse aujourd'hui sceptiques la plupart des collègues, qui par ailleurs vivent leur métier avec beaucoup d'investissement. Le projet de programme proposé dans sa formulation a provoqué d'emblée une levée de boucliers, ce qui n'induit pas un contexte très favorable pour engager un progrès souhaitable de l'enseignement de la statistique et des probabilités.

Pour tout ce qui touche à l'enseignement, on ne peut ignorer les contraintes didactiques et la prise en compte des travaux de recherche sur le sujet ne peut être négligée. C'est un des messages que j'ai souhaité faire passer à nos collègues chargés de la réforme des programmes et au-delà à tous ceux qui se sentent concernés par l'avenir de l'enseignement des mathématiques dans notre pays.