

Les enjeux de l'enseignement de la Statistique et des Probabilités aujourd'hui

Michel Henry, IREM de Franche-Comté

I - Leur place dans la production scientifique contemporaine

La théorie des probabilités occupe aujourd'hui une place essentielle dans les mathématiques contemporaines où les publications de recherche occupent une place prépondérante. L'outil statistique est de plus en plus performant, même surprenant : prises de décisions, contrôles fins, gains de production importants, sondages, connaissances inaccessibles directement, poids de différents facteurs dans les comportements de systèmes complexes (analyse factorielle). Le développement de la statistique théorique se nourrit des problématiques issues de la réalité et bénéficie des avancées de la recherche en probabilités : utilisation des savoirs probabilistes en vue des applications (analyse, théorie des nombres, codage), production de théorèmes. On a assisté au XX^{ème} siècle, surtout à partir de 1950 à un foisonnement de la théorie statistique. Certains résultats enseignés aujourd'hui étaient secret militaire il y a 30 ans.

Ces connaissances et techniques sont-elles accessibles à l'éducation au-delà des pratiques professionnelles souvent empiriques ? Quelles transpositions didactiques, à quels niveaux ? Quels rapports entre connaissance empirique ou pratiques empiriques en technologie et savoirs théoriques dans l'enseignement ? La mise à jour dans l'enseignement de la statistique devient aussi importante que la maîtrise de l'outil informatique dans les années 90. Ces deux domaines ne sont d'ailleurs pas indépendants et la puissance des NTIC contemporaines permet un enseignement de la statistique complètement renouvelé.

II - Statistique ou Probabilités ?

1 - La Statistique

On a tendance à appeler "statistique" un vaste domaine à la fois de connaissances empiriques et théoriques, de techniques de maîtrise de l'information et d'aides à la prise de décisions relatives à de vastes populations. En globalisant de la sorte on oublie ce qui fait la spécificité de la statistique comme outil de description de la réalité. À confondre l'aléatoire avec le déterminé, même si ce dernier est trop vaste ou complexe pour permettre une

approche directe. La pensée probabiliste reste essentielle pour la compréhension des phénomènes socio-économiques par exemple, pour comprendre aussi les fondements théoriques de la décision statistique. La formation ne peut pas faire l'impasse sur la dualité entre statistique et probabilités.

La statistique, traditionnellement science des populations, est constituée de grands domaines de savoirs et de techniques. Très schématiquement, j'en dégage deux :

1) Des outils de la description des populations, qui donnent lieu à diverses pratiques empiriques, issues d'une pratique sociale et pas nécessairement d'une compréhension scientifique : Les programmes des collèges et en partie de la seconde introduisent un vocabulaire et des modes de représentations élémentaires relatifs à « l'organisation et la gestion des données ». D'autres outils beaucoup plus sophistiqués et performants comme l'analyse des données, l'analyse en composantes principales, utilisent des connaissances mathématiques de haut niveau pour obtenir des informations plus difficiles à atteindre sur des populations connues exhaustivement.

2) L'inférence statistique qui permet de dégager la meilleure information d'une connaissance partielle d'une population statistique, qui donne des outils performants de traitement de cette information et qui peut indiquer la confiance à attribuer à une décision. Toute déclaration au sujet d'une population est entachée d'un risque de se tromper dû aux aléas de l'échantillonnage, c'est à dire d'une probabilité que les informations dégagées de cette connaissance partielle soient trop éloignées de la réalité de la population étudiée.

2 - La probabilité

C'est une notion abstraite au caractère dual : calcul des chances (calcul a priori) et approche fréquentiste (valeur expérimentale a posteriori). La théorie des probabilités ouvre un champ conceptuel spécifique. Elle a en vue la modélisation mathématique des situations de hasard (expériences aléatoires), et donne lieu à des développements théoriques de haut niveau. Le savoir probabiliste est une connaissances de type mathématique, organisée en théorie : concepts de base, définitions, axiomes, propriétés abstraites et théorèmes. Ces connaissances sont conçues pour opérer sur des situations modèles pour déterminer les évolutions de processus où le hasard intervient .

La notion de probabilité est une construction de l'esprit humain, qui a fondamentalement un caractère abstrait. Selon les approches, différentes appréhensions sont possibles :

- D'une part, ce que Pascal appelait la « géométrie du hasard » : la probabilité est conçue comme calculable a priori par dénombrement des issues « favorables », avant la réalisation d'une épreuve, en tenant compte des symétries autorisant des hypothèses d'équiprobabilité posées a priori.

- D'autre part l'appréhension fréquentiste développée par Bernoulli, qui explique qu'on peut avoir une approche de la probabilité (au sens précédent) d'un événement par une mesure de la fréquence stabilisée des réalisations de cet événement au cours de la répétition un grand nombre de fois de l'épreuve.

Ces deux approches vont se nourrir l'une l'autre et vont poser du point de vue didactique des questions de coordination et de relation dialectique. Se pose au niveau de l'enseignement secondaire la question de la relation entre cet outil théorique et l'étude des situations où le hasard intervient dans un contexte concret.

La question de la modélisation est donc fondamentalement posée.

Voici deux extraits pour illustrer ces propos :

a) **Pierre Simon Laplace** dont l'œuvre (*la théorie analytique des probabilités*, 1812) me le fait considérer comme « l'Euclide des probabilités », a fait la synthèse de tous les travaux produits depuis 150 ans et a poussé la théorie aussi loin que les outils mathématiques de l'époque, qu'il a contribué à développer, le lui permettaient. Dans sa préface qui constitue *l'essai philosophique sur les probabilités*, il écrit :

“la probabilité est relative en partie à notre ignorance, en partie à nos connaissances”.

C'est une prise de position épistémologique : la probabilité dans les applications serait ainsi de nature fondamentalement **subjective**. Il ajoute :

“La théorie des hasards consiste à réduire tous les événements du même genre à un certain nombre de cas également possibles, c'est à dire tels que nous soyons également indécis sur leur existence”.

Et il conclut à la définition de la probabilité, son *premier principe*, comme :

“le rapport du nombre des cas favorables à celui de tous les cas possibles.”

b) **Alfred Renyi**, probabiliste hongrois, auteur de manuels universitaires dans les années 60, écrit :

“Nous appellerons probabilité d'un événement le nombre autour duquel oscille la fréquence relative de l'événement considéré”.

Cette définition est totalement différente de celle de Laplace. Comment gérer cela dans l'enseignement secondaire ? Il précise, adoptant ainsi une position **objectiviste** :

“Nous considérons la probabilité comme une valeur indépendante de l'observateur qui indique approximativement [car considérée comme une mesure de grandeur] avec quelle fréquence l'événement considéré se produira au cours d'une longue série d'épreuves”.

Ce point de vue fonde donc sa « définition » sur une longue série d'épreuves réalisées et non sur une seule expérience à venir. Renyi précise :

“Une théorie mathématique des probabilités [entendez celle de Kolmogorov] ne s'occupe pas de jugements subjectifs. Elle concerne les probabilités objectives, qui peuvent être mesurées comme des grandeurs physiques. La définition ci-dessus n'est pas une définition mathématique, c'est une description du substrat concret du concept de probabilité.”

L'enseignement des probabilités doit intégrer la dualité de ce concept et rester ouvert aux deux appréhensions,

- subjectiviste reposant sur l'évaluation a priori « des chances » par le probabiliste (géométrie du hasard), soit qu'il postule l'équiprobabilité quelque part (point de vue laplacien), soit qu'il fonde ses calculs sur le choix d'une loi qu'en tant que spécialiste il pressent et contrôle (démarche bayésienne),

- objectiviste se basant a posteriori sur l'expérience renouvelée et sa description statistique, pour estimer une probabilité par une fréquence stabilisée (théorème de Bernoulli), ou ajuster une répartition par une fonction de fréquences cumulées (théorème de Glivenko-Cantelli), appliquant un principe de « loi des grands nombres » (approche fréquentiste).

Profitons d'avoir invoqué la loi des grands nombres de Bernoulli pour faire une petite remarque historique. Ce théorème dit que :

la probabilité que la fréquence d'une issue observée dans une expérience « Succès-Echec » répétée s'écarte de la probabilité théorique de cette issue de plus d'un ε donné, tend vers 0 quand le nombre d'expériences tend vers l'infini.

Ce résultat crée la dualité fondamentale de la notion de probabilité. Jacques Bernoulli démontre ce théorème à l'entrée du XVIII^{ème} siècle, d'une manière éblouissante, introduisant le langage des limites bien avant Cauchy, après avoir constaté les limites (!) d'un raisonnement à base de calcul infinitésimal.

III - Enjeux et défis pour l'enseignement secondaire :

Le problème didactique dans le secondaire est donc de faire le lien entre l'observation statistique de la réalité et les résultats issus d'une théorie mathématique dans laquelle la probabilité est introduite, selon les choix de transposition :

- par une gestion combinatoire de la définition de Laplace,
- ou bien, ainsi que l'indique Renyi comme fréquence stabilisée,
- ou encore, selon la théorie de Kolmogorov enseignée à l'université, comme mesure positive bornée définie sur un ensemble abstrait.

1 - Questions didactiques posées par l'enseignement des probabilités

A l'université, particulièrement en second cycle, l'enseignement des probabilités ne soulève pas de difficulté, c'est une théorie mathématique comme une autre.

Dans l'enseignement secondaire, cet enseignement est plus délicat. L'objectif est la maîtrise de concepts de base en vue de leur opérationnalité. On est alors amené à distinguer l'observation du réel (fréquence stabilisée) de la notion de probabilité. Celle-ci est introduite depuis 1991 par l'observation de la stabilisation des fréquences mais est définie autrement : c'est la somme des probabilités des événements élémentaires qui constituent l'événement considéré.

Le problème est de gérer la distance entre l'introduction d'une notion pour sa compréhension opératoire et la définition théorique qu'on va en donner.

Autre problème didactique posé par les programmes de 1991 : la probabilité peut-elle être donnée comme valeur limite de la suite des fréquences observées ? Ce serait une erreur profonde du point de vue épistémologique, qui produit comme obstacle majeur la confusion entre réalité et modèle théorique.

Ainsi, dans la tête des élèves, la probabilité revêt plusieurs significations. C'est, suivant les niveaux :

- un nombre relatif de chances (appréciation subjective du réel : quelles chances j'ai de ...),
- un rapport déterminé a priori (nombre de cas favorables sur nombres de cas possibles),
- une valeur limite, estimée à partir de l'observation des fréquences d'un événement dans une épreuve répétée.

- une application (distribution ou loi de probabilités) définie sur les parties d'un espace abstrait (l'univers Ω) qui donne comme une image un nombre compris entre 0 et 1,

- une mesure sur un ensemble numérique, définie par une densité (loi) par rapport à la mesure uniforme.

Comment préparer les élèves à cette diversité de sens donnés au même concept ?

Comment les relier aux problèmes concrets posés dans les applications ?

Comment intégrer les investigations à la fois pratiques (données statistiques) et théoriques (risque d'une décision) dans une compréhension probabiliste dont l'exemple prototypique est la notion d'intervalle de confiance, subrepticement introduite dans les thèmes d'étude de la classe de seconde de l'an 2000 ?

2 - Quels rapports entre Statistique et Probabilités ?

La symbolique mathématique est particulièrement adaptée à la description statistique (variables indicées, représentations graphiques), et à la manipulation des résumés statistiques (moyennes, médianes, dispersion ...).

Les mathématiques contemporaines (combinatoire, analyse, algèbre linéaire...) sont performantes pour, à partir de données complexes et très nombreuses, extraire des informations inaccessibles directement: (corrélations, études de tendances, analyse factorielle ou en composantes principales, analyse des données...). Mais ces techniques sont actuellement hors de portée de l'enseignement secondaire.

Pour étudier une population inaccessible globalement (trop vaste, trop complexe, trop coûteux...), on procède par étude d'échantillons (pas nécessairement représentatifs !), prélevés **au hasard** (avec quel générateur de hasard ?) dans la population. Ce hasard est la condition pour pouvoir appliquer les résultats théoriques du calcul des probabilités et obtenir des théorèmes sur les comportements des échantillons en fonction des caractéristiques de la population mère et dégager la possibilité d'**inférences**.

Car il n'y a pas de probabilités s'il n'y a pas de hasard !

A partir de là, les connaissances probabilistes vont être omniprésentes dans le traitement des échantillons aléatoires : lois théoriques pour ajuster des distributions de caractères quantitatifs, connaissance des propriétés des résumés statistiques, comportements asymptotiques, théorèmes limites pour valider des inférences, **nature probabiliste** de toute affirmation concernant la population.

Sur le plan historique, la (ou les) statistique(s) a précédé toute idée ou concept de probabilité. On faisait des statistiques à Babylone, chez les Égyptiens, chez les Romains... Les outils de la description statistique ont dû être développés au moyen âge. Au XIII^{ème} siècle

avec les transports maritimes, on a dû créer des compagnies d'assurance intéressées par les données statistiques et motivées par la prévision. Mais, on n'avait pas de mathématiques pour y voir clair dans ces recueils de données.

Au collège, en France, ces outils de la description statistique servent aux élèves à disposer d'un vocabulaire et de méthodes de représentation qui leur permettent d'interpréter des observations sur de vastes populations. Mais, comme au XIII^{ème} siècle, il n'y a pas de hasard au collège, c'est une question complètement exclue des programmes actuels, donc pas de probabilité. Cela n'empêche pas les élèves d'avoir une certaine pratique du hasard, car ils vivent continuellement des situations où il intervient.

Si l'on souhaite que l'outil statistique permette de porter des jugements sur des populations inaccessibles globalement, il faut se familiariser avec l'échantillonnage aléatoire consistant en prélèvements au hasard dans cette population, afin de faire fonctionner les connaissances de nature probabiliste dont on dispose. Le lien entre statistique et probabilités va se situer à cet endroit : mieux connaître les populations statistiques inaccessibles globalement ou systématiquement, seulement connues au travers d'échantillons aléatoires. C'est ce caractère aléatoire qui, par applications de résultats probabilistes, permet de prendre les décisions les plus performantes.

Jusqu'à présent, toute étude de l'inférence statistique était exclue de l'enseignement secondaire (avant le bac) où l'on avait d'un côté des outils de description statistique, et d'un autre côté, de façon complètement déconnectée, une définition théorique de la probabilité (somme de probabilités élémentaires), susceptible de fonctionner dans des situations artificielles d'événements qui se combinent logiquement et simplement pour obtenir des résultats issus de calculs combinatoires. Dans ce contexte, il n'y a aucune raison de rencontrer les applications à la réalité et de faire la connexion entre statistique et probabilités.

Mais aujourd'hui on se pose la question de l'introduction de cette pratique sociale dans l'enseignement, de maîtriser des situations d'inférence statistique et de décision à propos de populations. On se pose alors un nouveau problème didactique complètement occulté jusqu'à présent : établir cette liaison entre statistique et probabilités.

C'est la tentative du programme de seconde de l'an 2000 et des programmes de premières et terminales à venir. Vous comprenez que je suis totalement acquis à cet objectif là. Cela ne m'empêche pas d'émettre sur le plan didactique de nombreuses remarques ou critiques.

3 - Un exemple historique de lien entre statistique et probabilité

Mais avant de passer à ces remarques, à titre de TP, il m'a semblé intéressant de vous proposer l'exemple des

Tables de mortalité et espérance de vie de Christiaan Huygens

Ce travail des frères Huygens⁷, notamment Christian bien connu comme physicien, est le premier exemple historique du passage de l'observation statistique à la création d'un concept probabiliste.

Il faut dire ici quand est née la notion de probabilité : c'est en 1654 dans la correspondance entre Pascal et Fermat⁸. Christian Huygens fréquentait Pascal, et en 1657 il publiait un petit traité⁹ d'interprétation des principes que Pascal avait avancés dans cette correspondance. La notion d'espérance mathématique est notamment présente explicitement chez Huygens. En 1669, Huygens, qui a donc un peu maîtrisé la notion d'espérance mathématique, va chercher à conceptualiser et à théoriser l'observation des tables de mortalité de John Graunt qui avaient été dressées à Londres durant le siècle précédent.

Huygens pose un axiome :

"Dans un jeu, la chance qu'on a de gagner quelque chose a une valeur telle que si l'on possède cette valeur on peut se procurer la même chance par un jeu équitable"

Puis il énonce des « théorèmes » :

Théorème 1 : *"Avoir des chances égales d'obtenir a ou b me vaut $(a+b) / 2$ ".*

Théorème 3 : *"Avoir p chances d'obtenir a et q chances d'obtenir b, les chances étant équivalentes, me vaut $(pa + qb) / (p+q)$ ".*

C'est la formule de l'espérance mathématique.

Disposant de ces définitions théoriques, Huygens les applique à l'étude des tables de mortalité :

"À Londres, de 100 personnes conçues, il en meurt :

36 au bout de 6 ans,	9 entre 26 et 36 ans,	3 entre 56 et 66 ans
24 entre 6 et 16 ans,	6 entre 36 et 46 ans,	2 entre 66 et 76 ans
15 entre 16 et 26 ans,	4 entre 46 et 56 ans,	1 entre 76 et 86 ans."

C'est l'observation statistique. Huygens l'interprète en disant :

"qui gagerait qu'un enfant conçu vivrait jusqu'à 6 ans peut mettre 64 contre 36".

Il probabilise la situation :

"donc un enfant conçu a 36 chances pour vivre 3 ans (centre de la classe [0 ; 6]),

⁷ Correspondance des frères Huygens (1669), cf. Bernard Parzysz : atelier APMEP B19 de Marseille (1997), bulletin n° 416.

⁸ C'est très rare qu'on puisse désigner aussi précisément l'acte de naissance d'un concept mathématique.

⁹ *De cunctis in ludo aleae.*

24 chances pour vivre 11 ans, 15 chances pour vivre 21 ans. etc."

Et il conclut :

"Donc par ma règle des jeux de hasard, il faut multiplier chaque nombre des chances par les ans qu'elles donnent et diviser la somme des produits par la somme de toutes les chances."

Ce quotient est l'espérance de vie.

Résumons ces quelques objectifs didactiques qui me semblent se dégager des remarques précédentes, pour un enseignement rénové du collège à l'université :

- maîtriser les outils de la description statistique,
- pouvoir analyser et décrire des situations familières où le hasard intervient,
- introduire assez tôt la notion de probabilité dans ses différentes appréhensions,
- comprendre et mesurer les effets du hasard dans les fluctuations d'échantillonnage,
- traduire en termes probabilistes des situations simples de la réalité (modélisation),
- introduire aux principes et méthodes des investigations statistiques,
- donner des outils théoriques (concepts de base et raisonnements),

en vue de valider les démarches en statistique ou de porter des appréciations critiques sur les interprétations de leurs résultats (sondages par exemple), ces différents objectifs peuvent être résumés en un seul :

Renforcer les liens entre méthodes statistiques et théorie des probabilités

IV - Un état des lieux dans l'enseignement secondaire français

1 - La situation avant la réforme des années 2000

Jusqu'à présent, en France, il existe un grand cloisonnement entre la description statistique et l'initiation aux probabilités.

a) En collège, on aborde successivement :

- l'organisation et gestion des données,
- les notions de populations, caractères, séries statistiques, séries classées,
- les outils de la description : effectifs, fréquences, représentations graphiques et diagrammes,
- les résumés statistiques, médianes, moyennes et une idée de la dispersion.

Pas de hasard au collège

b) Au lycée (1991-1999), on a la progression :

- En Seconde : synthèse des outils de la description statistique.

- En Première : notion de probabilité, appréhension fréquentiste et définition théorique. Probabilité de l'événement contraire et de la réunion de deux événements incompatibles. Cas de l'équiprobabilité.

- En Terminale : probabilités conditionnelles, arbres probabilistes, variables aléatoires et notion de loi.

Pas d'applications concrètes, pas de liens avec l'observation statistique.

2 - Évolution des programmes¹⁰

Le calcul des probabilités a été introduit dans l'enseignement secondaire en 1954 dans la série TE (technique économique), il passe dans la série classique D en 1960

- 1960 : *Calcul des probabilités par dénombrement des cas, combinatoire dans les situations simples d'équiprobabilité (jeux de hasard).*

On remplace un problème de probabilité par un problème de combinatoire (c'est une mine impressionnante d'exercices du fait de la complexité de certaines situations de dénombrements). Pas de notion de loi.

- 1971 : *Espaces probabilisés, Ω fini et répartition uniforme* pour se placer encore dans les cas d'équiprobabilité, *combinatoire*.

On explicite le modèle probabiliste abstrait (Mathématiques modernes), mais en réalité, il n'y a pas de changement de point de vue, seulement l'introduction d'une structure mathématique et du formalisme pour se ramener encore à l'équiprobabilité dans les exercices.

- 1982 : *Expérience aléatoire. Ensemble des issues, calcul des événements, situations d'équiprobabilité.*

Un virage est négocié avec l'introduction du mot « *expérience aléatoire* », à la base de l'enseignement des probabilités. Pour un probabiliste, c'est un terme primitif qu'on ne définit pas. Mais pour parler d'expérience aléatoire aux élèves du Lycée, on est placé devant un problème de présentation et on s'aperçoit que ce n'est pas simple du tout. Malgré cette approche un peu appliquée, on en reste au calcul des probabilités par situations d'équiprobabilité.

- 1991 : *Expérience aléatoire, description, répétition, interprétation probabiliste de la stabilisation des fréquences. Définitions et propriétés de base* (en Première).

Probabilité conditionnelle, notion de loi (en Terminale).

Malgré l'approche fréquentiste préconisée pour introduire la notion, pas d'applications, sinon dans des exercices habillés de combinatoire ou d'application directe interne des formules de définition, pas de lien avec l'observation statistique.

¹⁰ Cf Bernard Parzenz, Les probabilités et la statistique dans le secondaire d'hier à aujourd'hui, *Enseigner les probabilités au lycée*, livre de la CII Statistique et Probabilités, 1997.

Les mots "hasard" et "expérience aléatoire" avaient été introduits en 1986 dans les programmes dans leur sens commun. En 1991, l'expérience aléatoire va intervenir comme outil d'introduction à la notion de probabilité. « *On étudiera la stabilisation des fréquences lors de la répétition d'une expérience aléatoire* ». La définition générale de la probabilité est donnée ensuite, sans limitation aux situations d'équiprobabilité.

Parenthèse historique : Laplace, dans son premier principe, posait :
"la probabilité est le rapport du nombre de cas favorables au nombre de tous les cas possibles."

Mais dans son deuxième principe, il écrit :

"Mais cela suppose les divers cas également possibles. S'ils ne le sont pas, on déterminera d'abord leurs possibilités respectives dont la juste appréciation est un des points les plus délicats de la théorie des hasards. Alors la probabilité sera la somme des possibilités de chaque cas favorable".

Laplace donne donc la définition générale, celle que propose le programme de Première de 1991. Remarquons sa difficulté d'expression, utilisant tour à tour les mots « probabilité » et « possibilité ». Cette définition donne l'impression (pas entièrement fausse) d'une circularité, définissant la probabilité par la possibilité.

Ma thèse est que sur le plan didactique, on ne peut sortir de ce cercle vicieux que par la notion de modèle : les probabilités élémentaires (les possibilités) sont des valeurs issues d'hypothèses de modèle, d'appréciations éventuellement subjectives ou d'estimations fréquentistes, et interviennent comme données dans le modèle probabiliste construit, validé ou non par ses applications.

Mais Laplace pose aussi le problème de l'inférence, il suggère en gros :

"Si vous n'êtes pas en mesure de calculer a priori la probabilité par une réduction à des cas également possibles, vous devez la mesurer concrètement, par l'expérience. En quoi votre mesure va effectivement être proche de la valeur théorique que vous devez utiliser ? *C'est un des problèmes les plus délicats de la théorie des hasards.*"

C'est le problème de l'estimation statistique que Laplace n'ignore pas mais dont il dit que qu'il est très difficile (problème résolu seulement au XX^{ème} siècle).

Fermons cette parenthèse historique pour revenir aux programmes actuels.

V - Le programme de Seconde de l'an 2000

1 - L'option d'une présentation perceptive et statistique des fluctuations d'échantillonnage

Le programme de Seconde de l'an 2000, dans sa deuxième partie, est caractérisé par une approche purement expérimentale des fluctuations d'échantillonnage. Les élèves sont confrontés au fait que quand on observe un certain nombre de fois les résultats d'une expérience aléatoire répétée, on n'observe pas tout à fait la même chose à chaque fois : le hasard intervient. Ce programme propose l'introduction du hasard directement avec l'observation des distributions de fréquences d'un ensemble d'issues dans des expériences répétées par la comparaison d'échantillons. Mais le hasard et la notion d'expérience aléatoire ne font pas l'objet d'un travail explicite particulier.

Cette démarche n'est pas évidente pour tous les élèves. Je cite très souvent cette remarque qu'un élève de Première nous a faite dans une classe, et qui n'en a pas démordu. On disait :

- *"On recommence la même expérience aléatoire"* (jet d'un dé).
- *"Mais non", a-t-il rétorqué, "vous n'avez pas recommencé la même expérience !"*
- *"Mais si, regarde, on jette le même dé, de la même manière..."*

À ce moment là, on pensait que cet élève demandait à préciser les conditions de l'expérience, mais il s'agissait en fait de la confusion entre la réalité et le modèle, ce que l'élève pointait immédiatement :

- *"Mais non, vous n'avez pas recommencé la même expérience, la preuve, c'est que vous n'obtenez pas le même résultat !"*

À partir de cet exemple, vous comprenez que si vous ne parlez pas d'expérience générique (de modèle), vous ne pourrez pas empêcher cet élève de penser qu'on ne peut pas refaire la même expérience aléatoire.

Cette remarque nous renvoie à la proposition faite en 93 pour le programme de la section L qui n'a pas abouti :

"Il s'agit d'aborder la notion de probabilité à partir de la fréquence mais on a choisi 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. L'objectif est de dégager de façon qualitative, éventuellement critique quelques notions probabilistes."

Les notions probabilistes étaient donc appelées comme outils de cette modélisation des phénomènes aléatoires. Ce projet reste un projet sur lequel on peut encore travailler.

2 - Objectifs, intentions du programme de Seconde et commentaires

J'en viens à quelques remarques d'ordre didactique sur ce programme de Seconde.

a) La partie statistique descriptive

L'ancien programme de Seconde proposait une synthèse des notions et du vocabulaire de base de cette statistique descriptive, 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 nous semblerait dommage de faire l'impasse sur cette partie des connaissances de base, qui sont loin d'être maîtrisées à l'entrée en seconde.

Le nouveau programme n'a pas ces ambitions, il se limite à une consolidation insuffisante des outils de la description statistique (moyennes, médianes et classes modales), notamment sur la notion de dispersion, trop abusivement réduite à l'étendue d'une série statistique. L'introduction de l'intervalle interquartile et de sa représentation par boîtes à pattes, pourtant à la portée de la main, est renvoyée en Première.

b) Échantillonnage par simulation informatique et réalité

L'intention affichée dans la deuxième partie du programme est véritablement une **approche de type expérimental** de l'investigation statistique. Pour cela, on donne des outils : définition des **fréquences**, observation d'échantillons et des **fluctuations d'échantillonnage**, introduction de l'informatique pour se donner les moyens d'avoir des échantillons de taille suffisante pour observer quelque chose.

Jusqu'à présent, sans ordinateur, on arrivait à exhiber concrètement le phénomène de stabilisation d'une fréquence, quand tous les élèves de la classe s'y mettaient, on finissait par avoir des données en assez grand nombre. Mais pour étudier plus globalement les comportements des distributions de fréquences lorsqu'on réalise un millier d'expériences pour chaque échantillon, là, l'ordinateur devient important.

On pourrait dire beaucoup de choses sur le terme de « simulation » qui est utilisé dans le programme. Il l'est peut-être improprement si cette « simulation » se limite à générer des données sans rapport avec une expérience concrète dont on aurait implanté le modèle en informatique. Laissons cette question de côté.

Mais cette remarque ne disqualifie pas l'utilisation de l'ordinateur, qui peut être considéré comme un générateur puissant de hasard, même si ce n'en est pas un en réalité. D'ailleurs il est expliqué pour les enseignants que, certes, l'ordinateur ne produit pas de hasard proprement dit, mais qu'en réalité il donne des nombres « pseudo-aléatoires » en quantité faramineuse, imprévisibles et correctement équirépartis, ce qui permet de les considérer comme la manifestation du hasard. L'ordinateur est donc considéré comme une boîte noire génératrice de hasard.

Du point de vue pédagogique, la question est plus précisément : comment va-t-on manipuler l'aléatoire qu'il produit pour singer (plutôt que simuler) des situations de lancers de dés, de tirages d'urnes etc. ?

Se pose alors la question didactique du rapport avec la réalité. Comment interpréter les situations concrètes, y compris au niveau du Collège, quand par exemple les élèves vont au super marché regarder les files d'attente aux caisses, ou qu'ils écoutent un bulletin météo ? Comment prendre en compte ce contact avec l'aléatoire dans la réalité ? Regarder l'aléatoire quand on "simule" le jet d'un dé (par principe non pipé) n'est pas plus porteur d'applications que l'ancien calcul des probabilités basé sur la définition de Laplace.

c) Introduction à l'aléatoire et à l'échantillonnage

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 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 le programme propose d'introduire à l'aléatoire dans des situations complexes de répétitions multipliées d'expériences aléatoires à plusieurs issues, dont l'ensemble des observations constituent un « **échantillonnage** », lui-même répété pour mettre en évidence les « **fluctuations** » que présentent ces divers échantillons recueillis. Cela en excluant toute introduction explicite de l'idée de probabilité, interprétée implicitement sous le vocable flou de « chances ».

La notion de fluctuation ne me pose pas de problème : c'est la mise en évidence de l'intervention du hasard dans la détermination des issues d'une expérience aléatoire.

Par contre, le terme d'usage courant d'« échantillon » (prélèvement, qui peut éventuellement être aléatoire, dans une population statistique) ne correspond pas au sens introduit dans la locution « fluctuation d'échantillonnage » adoptée dans le programme. Celle-ci vise l'ensemble des données recueillies (au moyen éventuel d'une simulation informatique) à l'issue de la répétition d'une même expérience aléatoire. De quelles populations sont alors extraits ces « échantillons » ? Le raisonnement traditionnel des statisticiens considérant une donnée statistique comme la réalisation fictive d'un prélèvement au hasard dans une population (qui ici reste à préciser), n'est pas immédiat, ni pour les élèves, ni pour leurs professeurs en général. Il y a là une difficulté didactique, qui n'est certes pas insurmontable, mais qui montre que ce premier contact avec l'aléatoire en classe n'est pas des plus simples.

Aborder l'aléatoire par l'observation des fluctuations d'échantillonnage me semble donc ambitieux pour un démarrage. A 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 d'ailleurs les programmes des classes de Première de 1991 le proposaient ?

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 au **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 en limite la portée. 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, dans la démarche proposée par le programme de Première ?

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 » ?

Au passage, il ne me semble pas judicieux de réduire l'outil statistique à la seule description des issues d'une épreuve aléatoire répétée.

Le deuxième point marquant de cette partie du programme est l'introduction de la notion de **distribution des fréquences** des issues (puis d'un événement) associées à l'expérience, distributions variables d'une série d'épreuves à l'autre, sans que les élèves puissent disposer des outils de comparaison de ces distributions pour en tirer des conclusions transférables.

Le professeur ne peut s'en tenir alors qu'à une **observation naturaliste** sans pouvoir faire appel à une connaissance de nature théorique (probabiliste) pour dégager une généralisation. Notons que le programme exclut toute étude de stabilisation des fréquences alors que, incontournable dans l'environnement informatique de ce cadre expérimental, elle donnerait du sens à cette observation et préparerait l'introduction de la notion de probabilité, repoussée en Première.

Nous sommes en présence d'un vrai problème didactique, rencontré par ailleurs en physique ou en économie :

- comment introduire un regard de type scientifique sur le réel quand les connaissances déjà installées ne permettent pas de formuler en termes clairs et raisonnés des faits observés afin de les transformer en phénomènes ?

- Quel sens peut prendre une activité de simulation en l'absence de la notion de modèle ? Quelle signification et statut ont alors les nombres pseudo-aléatoires issus de la touche "random" ?

- Comment éviter la « **leçon de choses** » à ce niveau déjà avancé des études secondaires, et comment présenter cette partie du cours de statistique aux élèves en termes de savoirs, condition didactique de sa pérennité ? Bref, peut-on délivrer un enseignement limité à l'expérimental et à sa description ?

Il y a là un vide didactique ingérable pour le professeur s'il ne peut faire appel à des concepts théoriques élémentaires lui permettant le **saut épistémologique** qui seul emmène les élèves vers **l'acquisition d'un savoir** au-delà d'une **révélation perceptive**.

Par exemple, comment éviter le mot « probabilité » quand on jette un dé ? Comment échapper à la présentation de la notion d'« expérience aléatoire » pour théoriser les situations de la réalité explorées ? Les expressions « hypothèses de modèle » (même si l'on s'en tient à l'équiprobabilité), « univers », « événements », « issues », « distribution de probabilité » sont autant d'outils théoriques qui permettent l'approche scientifique et le transfert de cette perception à des classes de situations.

Je n'ai pas de réponse toute faite, les expérimentations en cours pourront donner des éléments d'appréciation pour, je l'espère, nuancer ces réserves.

Pour résumer, je partage pleinement l'appréciation portée par les membres de la commission de réflexion sur l'enseignement des mathématiques dite J.P. Kahane, qui

"un bon enseignement de la statistique ne peut se concevoir sans une bonne maîtrise des probabilités. Mais de manière plus précise, c'est la dialectique "statistique-probabilité" qui permet de donner un sens profond aux belles activités proposées" (communiqué n° 4 du 27/11/99).

d) Dérapages dans les thèmes d'étude

Les objectifs du programme restent donc au niveau du perceptif, voire du rapport naïf à la réalité complexe, marquée par le refus du concept théorique de probabilité. D'où au niveau des thèmes d'étude des impasses didactiques contournées par cette locution dénuée de sens : "les chances de se tromper quand"... , avec cette ambiguïté portée par le terme de chances, signifiant tantôt issues (favorables ou observables), tantôt probabilité. D'où la réduction des exemples proposés à des jeux de dés, pièces, cartes ... et à des situations simples d'équiprobabilité.

Le premier thème propose une introduction élémentaire aux sondages, ce qui vise un fait massif de nos sociétés développées. Il paraît évident que les élèves des lycées doivent pouvoir disposer des notions de base leur permettant d'avoir des éléments d'appréciation, face à un déferlement médiatique. La notion de "**fourchette de sondage**" peut être présentée comme un encadrement d'une mesure entachée d'erreur en physique, mais cela pose le problème du statut de la grandeur à mesurer.

Cet encadrement est associé a priori à une certaine probabilité d'être correct, ou, si l'on veut éviter ce point de vue, à un certain **risque** d'observer une certaine proportion des fréquences recueillies en dehors de cet intervalle. Le parachutage de la formule $[p - 1/\sqrt{n} ; p + 1/\sqrt{n}]$, impossible à justifier au niveau des élèves (ce qui ne serait pas un drame) ainsi qu'au niveau des connaissances de base des enseignants (ce qui pose plus de problèmes), ne permet pas de cerner clairement sa signification et les conditions de son application.

Le problème didactique est : comment introduire une initiation aux sondages (pourtant souhaitable) et à leur interprétation un peu rigoureuse sans la notion d'**intervalle de confiance** et de la probabilité-**risque** qui l'accompagne : $P(F_n - \varepsilon < p < F_n + \varepsilon) = 1 - \alpha$?

Cette remarque nous ramène à la nécessité de la notion de probabilité pour clarifier le statut des fourchettes de sondage et leur connexion avec la notion d'intervalle de confiance enseignée dans certains BTS ou DUT ainsi que dans les cours de Statistique de second cycle universitaire. Vu l'expérience acquise en post-bac, je reste réservé quant à la possibilité de faire bien comprendre ces questions aux élèves de Seconde (indifférenciée).

Les autres thèmes d'étude proposés (promenades aléatoires) semblent avoir provoqué de l'incompréhension chez les enseignants. Intéressants par leurs prolongements possibles, ils me semblent plutôt convenir pour l'illustration de modèles probabilistes à un niveau post-bac que pour un enseignement de la statistique.

e) Transposition didactique et formation des enseignants

Je conclurai ces quelques remarques d'ordre 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 qu'on a 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, en fonction de leur propre conception de l'enseignement des mathématiques ? Il y aura à vaincre des résistances au niveau 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 sceptiques la plupart des collègues qui par ailleurs vivent leur métier avec beaucoup d'investissement.

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 le message que je désire faire passer tout en assurant de la mobilisation des collègues de la commission inter-IREM "Statistique et Probabilités" pour contribuer à la réussite de cet enseignement : suivi d'expérimentations, préparation d'une brochure IREM-APMEP...

VI- Vers une proposition alternative : l'enseignement de la modélisation

La nécessité d'un véritable décroisement entre statistique et probabilité se dégage clairement des remarques qui précèdent. Il n'est pas, à mon sens, vraiment réalisé avec la progression proposée par les programmes de Seconde et de Première et les objectifs annoncés pour celui de Terminale. Ce décroisement suppose en fait une réforme globale, du Collège au Lycée puis à l'université. Intégrant les outils de la description statistique dans l'étude des grandes populations, cet enseignement introduirait les probabilités dans un processus de modélisation de situations de la réalité.

Cette approche expérimentale serait donc éclairée par l'introduction parallèle d'outils de description et de concepts théoriques simples :

- Descriptions de populations statistiques, caractères et opérations logiques,
- situations où le hasard intervient et descriptions d'expériences aléatoires,
- prélèvements au hasard dans une population et expériences de Bernoulli (succès-échec),
- répétition de telles expériences, stabilisation des fréquences. Urne de Bernoulli et pré-probabilité,
- modéliser le lancer d'une punaise : du modèle pseudo-concret au modèle mathématique.
- situations variées et modèles d'urnes, choix de modèles,
- schéma binomial, situations de sondages, principes de l'inférence statistique (nature probabiliste d'une décision statistique),
- distribution de fréquences et modèle probabiliste, la loi des grands nombres,
- générateurs aléatoires, distribution uniforme,
- caractères quantitatifs et variables aléatoires, lois de variables discrètes,
- résumés statistiques et moments d'une loi, théorème de la statistique,
- simulation informatique et modélisation. Probabilités géométriques (jeu du Franc-Carreau), modèles à lois continues,
- conflits de modèles et paradoxes (D'Alembert, les 3 bancs, Bertrand, ...),
- modéliser pour simuler. Résolution de problèmes concrets par simulation : files d'attente.

Ces objectifs n'ont pas de caractère exhaustif, ils ne sont pas opérationnalisés et sont de l'ordre de l'utopie. Ils montrent néanmoins quelles pourraient être les étapes d'un enseignement intégré de la statistique et des probabilités, vers lequel la réforme des programmes des années 2000 a tenté un premier pas.

Remarques sur le programme de Première S et sur le projet de programme de Terminale

Les programmes de la filière S, après ceux de la série L, ont été soumis au débat au premier trimestre de 1999-2000. Les programmes des classes de Première ont été adoptés par le CNP, à nouveau "sans réserve" et publiés au B.O. du 31 août. La CII Statistique et Probabilités avait fait part de son appréciation critique le 30 janvier. L'essentiel des remarques qui suivent reprennent cette appréciation.

La partie *Contenus* du programme de Première S est présentée en 2 parties que je présente d'abord succinctement, avec quelques premiers commentaires.

1 - La partie statistique complète celle du programme de Seconde par l'introduction de paramètres de dispersion : variance et écart type d'une part, intervalle inter quartile et « diagrammes en boîte » d'autre part. Dans la *mise en œuvre* le programme suggère de comparer par l'observation expérimentale les comportements respectifs de ces deux paramètres statistiques et leur robustesse, quand les données varient sensiblement. En les associant aux paramètres de position correspondants que sont respectivement la moyenne et la médiane, les *commentaires* limitent l'objectif de l'étude descriptive d'une série statistique à ces deux couples de résumés. L'interprétation des paramètres empiriques comme estimateurs possibles des paramètres correspondants dans une population échantillonnée est donc hors programme.

2 - La partie probabilités des *contenus* s'ouvre directement par la *Définition d'une loi de probabilité sur un ensemble fini* (lequel ?).

Cette définition ne fait pas l'objet d'une proposition dans la *mise en œuvre*. Elle paraît de nature pragmatique, rapportée à la notion de *distribution de fréquences* introduite en Seconde, via un *énoncé vulgarisé* (sic) de la loi des grands nombres. Cet énoncé est donné dans les *commentaires*. Il va bien au-delà de l'énoncé traditionnel du théorème de Bernoulli¹¹, car il englobe la convergence, repérée empiriquement et présentée sous la forme vague de « se rapprochent », des moyennes et des variances (pourquoi pas des médianes ?), en fait des lois, propriété connue par les spécialistes sous le nom de théorème de Glivenko-Cantelli¹², qu'il n'est bien évidemment pas question d'évoquer à cet endroit.

¹¹ Théorème de Bernoulli (non vulgarisé) : Dans une épreuve répétée de Bernoulli, la suite des fréquences de succès converge en probabilité vers la probabilité de ce succès, i.e. : $\forall \varepsilon > 0, P(|F_n - p| < \varepsilon) \rightarrow 1$ quand $n \rightarrow \infty$.
Conséquence : pour ε assez petit (précision) et n assez grand (taille de l'échantillon), $[F_n - \varepsilon, F_n + \varepsilon]$ est un intervalle de confiance pour p de niveau de confiance $P(|F_n - p| < \varepsilon) = 1 - \alpha$.

¹² Théorème de Glivenko-Cantelli (résumé) : La suite des fréquences cumulées tend en probabilité (loi faible) et uniformément presque sûrement (loi forte) vers la fonction de répartition de la loi limite.

Le programme indique ensuite : *espérance, variance, écart type d'une loi*, ce qui suppose pas mal d'implicites sur lesquels je n'insisterai pas, ces notions étant reprises ensuite à propos des *lois de variables aléatoires*, d'où une certaine redondance. Il n'y a pas d'étude systématique de loi élémentaires (Bernoulli et binomiales, par exemple) proposée en Première.

Les propriétés de base des probabilités sont ensuite introduites avec le concept d'« événement », sans indication sur les définitions à institutionnaliser, dans une progression qui semble difficile à gérer sur le plan didactique, allant de la compréhension globale de la notion de loi de probabilité à celle plus élémentaire de probabilité d'un événement ! L'équiprobabilité est ensuite présentée comme un cas particulier.

Le dernier paragraphe concerne la modélisation d'expériences aléatoires. Remarquons d'abord que la notion fondamentale d'« *expérience aléatoire* » est introduite à cet endroit sans faire l'objet d'indications particulières, limitée aux expériences de référence (les exemples donnés relèvent tous de l'équiprobabilité). Dans les commentaires, le programme précise : « *On indiquera que simuler une expérience consiste à simuler un modèle de cette expérience* », ouvrant ainsi la possibilité sur ce qu'il me semble essentiel : le travail sur la modélisation.

Les modalités de mise en œuvre proposent de la simulation d'une loi de probabilité image par une v.a. d'une loi équirépartie, ce qui suppose un sérieux recul théorique de la part des enseignants. La question du lien de cet enseignement avec l'aléatoire dans la réalité reste donc posée, avec celle de son articulation avec la partie statistique.

3 - Commentaires épistémologiques et didactiques de la CII Stat et Probas¹³

a) Sur la notion de probabilité

Une première constatation confirme la tendance relevée dans le programme de seconde : tout travail sur la notion de probabilité semble être exclu. L'introduction préconisée par le programme de 1991 est absente des objectifs affichés. Le statut théorique de la probabilité comme objet mathématique inscrit dans un modèle descriptif de la réalité sous certaines hypothèses, n'est pas non plus l'objet visé dans cet apprentissage. Les axiomes fondateurs de cet objet mathématique (très simplement la sous-additivité) ne sont pas proposés en tant que tels, ce qui invalide toute compréhension théorique de cette notion et sa transformation en concept abstrait, ce qui rend problématique son insertion dans des démonstrations mathématiques.

¹³ Je transcris ici, légèrement mise à jour, la partie concernant les commentaires de la Commission Inter-IREM « Statistique et probabilités » datés du 29 janvier 2000, à propos du projet soumis à discussion.

b) Sur la notion d'expérience aléatoire

De même la notion d'expérience aléatoire n'est abordée que d'un point de vue pragmatique, au travers d'exemples limités (jeux de hasard) et ne peut jouer le rôle de concept primitif autorisant une modélisation mathématique sous la forme d'un référentiel Ω , qu'il n'est pas question de spécifier dans chaque exemple, mais qui joue un rôle théorique irremplaçable dans la compréhension de la formalisation de la notion d'événement. Bref, les bases de la théorie probabiliste, et par là même d'une connaissance de type scientifique, sont absentes de ce programme.

Dans ces conditions, le processus de modélisation qui à juste titre est mis en avant explicitement, semble limité aux "expériences de référence amenant à une probabilité équirépartie", c'est à dire à des situations limitées et contextualisées qui pourraient être sources d'obstacles didactiques.

c) Sur l'équiprobabilité

Au delà de ce type d'obstacle, la limitation aux situations d'équiprobabilité porte en elle une question épistémologique de fond, déjà soulignée par J. Bernoulli dans *Ars Conjectandi* et indiquée par P. S. Laplace dans son deuxième principe. Les situations réelles relèvent-elles nécessairement de l'équiprobabilité quelque part ? Évidemment non. D'où la limitation aux jeux de hasard, "*définis par leurs inventeurs pour se ménager l'équité*" selon Bernoulli, du champ d'application des connaissances introduites dans ce programme. Cette limitation est-elle pertinente aujourd'hui où les problèmes de fiabilité, de files d'attentes, de modélisation d'une complexité (économique ou météorologique par exemple) sont l'objet même de cet apprentissage ?

Le programme de 1991 embrassait cette dimension sans difficulté, laissant aux hypothèses de modèle le soin de déterminer les probabilités en jeu. Les situations d'équiprobabilité étaient présentées comme des cas particuliers, même si cette hypothèse a l'avantage de permettre les calculs a priori dans un cadre combinatoire. La conséquence didactique prévisible est d'induire à nouveau chez les élèves l'idée que le calcul des probabilités n'est pas applicable à la réalité, se réduisant à d'aimables situations de dés ou urnes dans lesquelles on est conduit à des problèmes incontrôlables de dénombrement des cas.

d) Sur la notion de loi probabilité

La notion de loi de probabilité est proposée, sans que celle de probabilité fasse l'objet d'un travail préparatoire. On peut s'en étonner. Cette notion est présentée comme une structure formelle jouissant des propriétés des distributions de fréquences, lesquelles au travers d'un énoncé "vulgarisé" de la loi des grands nombres détermineraient cette loi limite.

Outre le fait qu'un tel énoncé, même vulgarisé, ne nous semble pas présentable en Première, sauf à prendre de grandes libertés avec la rigueur mathématique, il nous semble porter le risque d'assimiler ces probabilités comme des fréquences limites, ce qui revient à confondre la réalité observable avec le modèle théorique dans lequel on obtient des théorèmes.

Cette notion de loi de probabilité, qui nous semble essentielle effectivement, mérite plus qu'une simple description de la distribution des probabilités sur Ω . Comme pour la notion de fonction, introduite en Seconde, un tableau de valeurs ne suffit pas à rendre compte de la cohérence interne qui lie ces valeurs pour en faire des signifiants numériques des phénomènes modélisés. Pour donner du sens à cette notion, il convient de la rapporter à des situations concrètes. On pourrait par exemple opter pour l'introduction es qualité de la loi binomiale (qu'il faudrait anticiper par celle de Bernoulli), et peut-être l'accompagner d'autres exemples comme celui de la loi géométrique, très présente dans la pratique de tests industriels, malgré le cardinal théoriquement infini de son ensemble référentiel.

e) Sur le projet de programme de Terminale (extraits)

La notion d'indépendance nous semble devoir faire l'objet d'un réexamen. Une simple présentation intuitive n'est pas opératoire, et particulièrement source d'interprétations erronées. L'indépendance stochastique est plus précise que la non dépendance intuitive.

La notation en Terminale $P(A/B)$ ne nous semble pas préférable à $P_B(A)$ suggérée dans les programmes de 1991 et suivants, dans la mesure où elle induit l'idée d'événements conditionnés dont on a montré la non consistance, du moins à ce niveau. Plus profondément, elle induit l'idée subjectiviste d'une (distribution de) probabilité unique sur Ω , mais modifiable en fonction des renseignements disponibles par le sujet. D'un point de vue didactique, il est apparu plus clair de justifier un changement de probabilité par un changement d'expérience, le symbolisme est ici porteur de cette distinction. L'approche fréquentiste préconisée, que nous préférons pour sa généralité, ne nous semble pas pertinente dans le cadre de l'équiprobabilité imposé, où l'approche cardinaliste est plus directe.

Le projet d'introduire en Terminale S des lois continues nous semble ambitieux, mais pas inaccessible, modulo la formation des enseignants nécessaire. Mais comment peut-on expliquer qu'une variable aléatoire, transporteuse d'une telle loi de probabilité, puisse être définie sur un Ω fini (muni de plus d'une distribution d'équiprobabilité) et néanmoins prendre un ensemble continu de valeurs ($\mathbb{R}$ ou un intervalle $[a, b]$) ?

Les lois uniformes et gaussiennes (il nous semble préférable de les appeler "normales", dénomination de K. Pearson, ne serait-ce que pour rendre à leurs inventeurs Laplace, Moivre et peut-être Bernoulli ce qui leur revient en dimension 1) sont présentées par des définitions purement mathématiques en termes de calcul d'aires, sans que la relation

puisse être faite avec la notion de loi d'une variable introduite par ailleurs, et sans que leurs domaines d'applications concrètes puissent être explicités.

On peut s'étonner à cet endroit que l'on propose de recourir aux anciennes tables de valeurs de la fonction de répartition de la loi normale, alors que les calculatrices ou logiciels contemporains les donnent sans difficulté. La nouvelle colonne "informatique" aurait pu bénéficier de cette rubrique.

Ces propositions, ambitieuses pour les connaissances en jeu, mais insuffisantes quant aux savoirs probabilistes fondamentaux qu'une introduction intuitive et superficielle ne saurait rendre opératoires, sont séduisantes pour qui partage l'importance d'un enseignement modernisé de la statistique et des probabilités.