

SIMULATION STATISTIQUE ET ENSEIGNEMENT

Jean-Pierre RAOULT

Laboratoire d'Analyse et de Mathématiques Appliquées
CNRS, UMR 8050
Université de Marne-la-Vallée
jean-pierre.raoult@univ-mlv.fr

**Université d'Eté d'ANIMATHS
Saint-Flour, le 26 août 2004**

Quatre années après l'apparition de la simulation statistique dans les programmes d'enseignement des Mathématiques en classe de seconde dans les lycées en France, ce volet des programmes, s'il donne moins lieu à polémique que lors de son introduction, continue cependant à susciter dans le corps enseignant interrogations et inquiétudes et de nombreux professeurs avouent n'aborder cette tranche du programme qu'en fin d'année, au risque souvent de la laisser finalement de côté.

A mon avis, la réflexion sur :

- ce qu'est
- à quoi sert,
- comment peut se traiter

la SIMULATION STATISTIQUE, au regard de l'enseignement en Collège et Lycée, ne peut se mener sans une mise en perspective plus générale de l'ENSEIGNEMENT DE LA STATISTIQUE à ces niveaux d'études.

Ma conviction est qu'enseigner la statistique est :

- NECESSAIRE,
- DIFFICILE,
- POSSIBLE,

ces trois termes ne prenant leur sens qu'en fonction des buts qu'on assigne à cet enseignement.

1. Enseigner la statistique est NECESSAIRE

Personnellement, si je souhaite que la STATISTIQUE et le CALCUL DES PROBABILITES fassent partie du bagage des bacheliers, selon des modalités diverses selon les filières, ce n'est pas essentiellement en tant que **composantes de la culture mathématique** (ce qu'ils sont pourtant), mais d'abord en tant qu'**éléments de la culture du citoyen** (ou plus généralement de l'être humain en tant que sujet doté de SENS CRITIQUE).

A ce titre, Condorcet reste pour moi d'actualité, qui écrivait dans son texte sur l'enseignement des "Mathématiques Sociales" :

Cette exposition montrera toute l'utilité de cette science : on verra qu'aucun de nos intérêts individuels ou publics ne lui est étranger, qu'il n'en est aucun sur lequel elle ne nous donne des idées plus précises, des connaissances plus certaines ; on verra combien, si cette science était plus répandue, plus cultivée, elle contribuerait au bonheur et au perfectionnement de l'espèce humaine.

En d'autres termes, reprenant une belle citation faite par Yves CHEVALLARD dans sa conférence à cette même université d'été d'Animaths le 25 août 2004, je trouve la Statistique digne d'être incluse dans la formulation de Paul GAGNON dans son article *What Should Children Learn* (in *The Atlantic Monthly*, décembre 1995, p. 71-72) :

the essential core of learning that all students in a modern democracy have the right not to be allowed to avoid.

MAIS ON PEUT OBJECTER (ET D'AUCUNS LE FONT, DE L'INTERIEUR COMME DE L'EXTERIEUR DE LA COLLECTIVITE MATHEMATIQUE) :

Si l'enseignement de la Statistique est nécessaire, est-ce nécessairement le cours de Mathématiques qui en fournit le cadre optimum ?

De nombreuses raisons sont avancées pour situer l'enseignement de la Statistique au sein de celui des Mathématiques. Certaines ne sont à mon sens que partiellement valables ; passons les en revue.

a. Le bon usage de la Statistique (ou des statistiques) ne va pas sans raisonnement ; or le cours de Mathématiques est le lieu privilégié d'apprentissage du raisonnement.

CERTES, MAIS, en matière de regard raisonné sur la Statistique, il ne s'agit pas uniquement du raisonnement hypothético-déductif qui fonde les Mathématiques.

b. On fait grand usage en Statistique, comme en Mathématiques, d'outils de calcul numérique et de représentations géométriques.

CERTES, MAIS d'autres disciplines scolaires (Physique, Biologie, Economie ...) font aussi usage de tels outils et le font parfois de manière plus accessible aux élèves, car plus proche des besoins immédiats.

c. L'usage des calculatrices, des ordinateurs, est de plus en plus partie intégrante des enseignements de Mathématiques ; or la Statistique en constitue un domaine d'emploi privilégié.

CERTES, MAIS les employer dans des secteurs plus traditionnels de l'enseignement des Mathématiques (tels les calculs approchés de fonctions, d'intégrales, de solutions d'équations numériques ou différentielles ...) fournit déjà un vaste champ de mise en œuvre et on pourrait envisager de laisser un usage "routinier" des fonctions statistiques aux autres disciplines.

A MON AVIS, LA RAISON PRINCIPALE POUR FAIRE FIGURER LA STATISTIQUE AU SEIN DU COURS DE MATHÉMATIQUES EST LA SUIVANTE :

Le caractère “universel” des Mathématiques est indispensable pour donner à l’élève les clefs pour un usage lui aussi universel de la Statistique.

Pour mettre ici en avant la danger de la spécialisation dans tel ou tel type d’emploi, que seul le cadre du cours de Mathématiques permet d’éviter, je m’appuie sur l’avis de “grands praticiens” de la Statistique dans des domaines déterminés, tels Daniel SCHWARTZ (en Biologie et Médecine) ou Edmond MALINVAUD (en Economie).

Il faut en effet pouvoir à la fois :

- **donner vie aux concepts statistiques**, indépendamment d’un domaine d’application particulier (et cela seul l’enseignant de Mathématiques peut le faire) ;
- **motiver et concrétiser ces concepts** par des applications diversifiées, ce qui suppose de fournir au professeur de Mathématiques de la documentation appropriée et des occasions de contact avec les collègues d’autres disciplines ;
- **mettre en garde** contre des “illusions mathématiques”, tel l’usage abusif du “linéaire” (et ceci encore est l’apanage du professeur de Mathématiques).

2. Enseigner la statistique est DIFFICILE

Il serait illusoire de nier cette difficulté.

Cette difficulté est d’abord attestée par l’expérience (presque tous les professeurs le disent !).

Personnellement, en plus de 40 ans où j’ai pratiqué des maquettes diverses pour l’enseignement de la Statistique dans différents contextes (ou aussi participé à leur élaboration), je n’ai jamais été pleinement satisfait : les voies possibles sont multiples, parfois fort éloignées les unes des autres, et toutes comportent des obstacles.

Il en résulte un premier objectif que j’assigne à mon présent exposé : **je voudrais pouvoir ici reconforter ceux de mes collègues qui “complexent” face à l’enseignement de la Statistique** ; j’aimerais leur dire, avec mon “autorité” de spécialiste et l’ancienneté de ma pratique :

Oui, c’est normal que vous trouviez cela plutôt dur !

Alors, que faire ?

Nous détaillerons un peu plus ci-dessous pourquoi et comment enseigner la Statistique nous paraît possible. Pour l’instant, nous affirmerons seulement qu’il est utile de s’appuyer sur une documentation appropriée (par manuels, fascicules ou internet) de plus en plus élaborée et que, en tout cas, il importe de se convaincre que le succès de cet enseignement exige chez l’élève une maturation progressive qui impose de le traiter, à une place modeste certes (celle que lui attribuent les programmes) mais **en continu sur l’année scolaire**.

Essayons d’analyser des raisons intrinsèques à cette difficulté (sous-jacentes aux arguments de “nécessité” que j’ai exposés plus haut) et ce sans nous limiter pour l’instant au cas de l’enseignement secondaire qui sera ensuite l’essentiel de notre propos

L'apprentissage de la Statistique conduit à ARTICULER et HIERARCHISER :

- applications "concrètes" et outils "généralisants",
- aspect descriptif et initiation à l'aléatoire,
- usage d'outils logiciels ("boîtes noires") et compréhension de ce qu'il y a "derrière" et de ce qu'ils produisent "vraiment",
- résultats précis et approximations (souci de l'ordre de grandeur, que les enseignants de mathématiques avaient parfois tendance à évacuer vers d'autres disciplines plus expérimentales),
- éléments de théorie mathématique (par exemple la loi des grands nombres) et raisonnement de type inductif ("l'inférence statistique" où tout ce que l'on énonce a une "probabilité d'erreur").

C'est pourquoi dans ce domaine toute mise en place d'un cursus d'enseignement suppose de faire un CHOIX DANS CES ARTICULATIONS ET CES HIERARCHIES. Il n'est donc pas étonnant que les programmes d'enseignement, en France, aient présenté plusieurs tels choix au cours des années ; cette variété de points de vue était naturelle mais a pu accroître parfois le désarroi des enseignants.

La meilleure chose à faire face à un désarroi est de tenter de l'analyser ; ceci me semble en assez bonne voie maintenant en France et à cet égard j'ai choisi, pour la suite de cet exposé, de faire référence non à la lettre des programmes ou à des documents d'accompagnement de ceux-ci (pourtant très riches) mais à des contributions datées de cette année (2004) :

1. Actes des sessions *Enseignement de la Statistique* lors des journées de la SFDS (Société Française De Statistique) tenues à Montpellier en mai 2004 (sessions dont le thème était cette année "enseignement de la simulation") ; on y fera référence par la suite avec le sigle SFDS.

2. Manuscrit d'un ouvrage, à paraître début 2005 sous l'égide de l'APMEP (Association des Professeurs de Mathématiques de l'Enseignement Public), émanant de la Commission Inter IREM (Instituts de Recherche sur l'Enseignement des Mathématiques) *Statistique et Probabilités* ; il s'agit du premier tome (*Les outils de la Statistique*) d'un projet en deux tomes dont le titre global sera *Statistique au Lycée* ; on y fera référence par la suite avec le sigle CII.

3. Enseigner la statistique est POSSIBLE

Concentrons nous maintenant sur le choix actuel des programmes en France (hors sections technologiques ou professionnelles qui ont leur contraintes propres). Rappelons-en brièvement les caractéristiques :

AU COLLEGE

Présentation très simple de caractéristiques DESCRIPTIVES des séries numériques.

EN SECONDE

Prise de conscience de la VARIABILITE DES DONNEES ALEATOIRES :

- par la biais de la **simulation** (recours à l'ordre "RANDOM" des ordinateurs),
- sans théoriser les variables aléatoires (usage d'un vocabulaire volontairement NAIF en termes de "chances").

EN PREMIERE ET TERMINALE

Introduction mathématisée des PROBABILITES (permettant la STATISTIQUE INFERENTIELLE)

- Mon avis personnel est que ce choix, qui n'était certes pas le seul envisageable, est :
- COHERENT,
 - PORTEUR DE SENS,
 - REALISABLE.

Bien sûr, cette progression est à respecter sans dogmatisme (comme toujours dans l'enseignement) et en particulier je considère pour ma part que, en classe de seconde, l'enseignant peut s'autoriser éventuellement un peu de vocabulaire probabiliste si c'est ainsi qu'il se sent plus à l'aise.

Mais il est essentiel de faire travailler les élèves sur des données et pour l'obtention de celle-ci d'articuler, là encore dès la seconde, simulations et exemples. Les documents d'accompagnement y aident et par ailleurs je recommande très vivement l'usage du **logiciel français gratuit SMEL**, qui a été très précisément conçu pour cela, à l'UFR de Mathématiques et Informatique de l'Université René Descartes (Paris V), avec le soutien du Ministère de l'Education Nationale et de l'INRIA (Institut de Recherches en Informatique et Automatique). Le sigle SMEL signifie Statistique Médicale En Ligne, car il s'agit de l'adaptation, pour les besoins de l'enseignement secondaire, d'un outil conçu au départ pour être convivial à l'usage de médecins plutôt ignares en Statistique ; il garde de cette histoire un riche corpus d'exemples réalistes mais accessibles (d'où, retombée intéressante, des possibilités de dialogue avec les enseignants de biologie). On y accède sur

<http://www.math-info.univ-paris5.fr/smel>

Limitons désormais nos considérations à la simulation, en nous focalisant sur les problèmes :

- de conception,
- de mise en œuvre,

liées à son introduction comme AXE DU PROGRAMME DE STATISTIQUE EN SECONDE depuis 4 ans.

Je conseille en particulier ici la lecture des travaux de Claudine VERGNE qui effectue une thèse de didactique de la Statistique à l'université Montpellier II (s'adresser à l'IREM de Montpellier : irem@math.univ-montp2.fr) ; je reproduis le résumé de sa contribution aux Journées de la SFDS.

LA NOTION DE VARIABILITE DANS LES PROGRAMMES ACTUELS DE LA CLASSE DE SECONDE

La variabilité, voilà le mot clef

Daniel SCHWARTZ, Le jeu de la Science et du Hasard

Les nouveaux programmes de Mathématiques au Lycée (2000) ont mis l'accent sur la Statistique et tout particulièrement sur l'introduction de la notion de VARIABILITE. Il semble que

cette partie des nouveaux programmes ait du mal à vivre dans les classes. Notre étude concerne la classe de seconde; elle porte sur des éléments du curriculum officiel (programmes, document d'accompagnement, texte de la CREM (Commission de Réflexion sur l'Enseignement des Mathématiques)...) et des éléments du curriculum réel (observation des professeurs dans la classe, cahiers d'élèves, manuels, questionnaires...). Elle tente de dégager la praxéologie en construction et d'envisager la question de sa viabilité, dans le but de mieux comprendre les raisons de ces difficultés. Ceci pourrait éventuellement servir d'appui à des propositions visant à situer l'enseignement de la Statistique dans les classes ordinaires.

Il s'agit d'une étude en cours dont nous commenterons les premiers résultats.

Pour dégager les possibilités qui s'offrent pour l'enseignement de la simulation, il fait d'abord éliminer les ambiguïtés qui entourent cette notion.

Lors de la *table ronde* sur ce thème aux journées de la SFDS déjà citées, un éminent pédagogue de la Statistique, Pierre DAGNELIE (Professeur de Statistique en université d'Agronomie à Gembloux (Belgique), auteur de nombreux manuels réputés) a posé aux enseignants français la question suivante :

Que prétendez-vous simuler :

- *une loi de probabilité ?*
- *un phénomène ?*
- *des mesures expérimentales ?*

En réponse à cette question, remarquons tout d'abord que les acceptions les plus courantes du mot renvoient clairement à la SIMULATION D'UN PHENOMENE. Donnons en deux preuves.

La première preuve dépasse le cadre de la Statistique; c'est un extrait de la définition qu'on trouve dans l'Encyclopædia Universalis :

La simulation est l'expérimentation sur un modèle. C'est une procédure de recherche scientifique qui consiste à réaliser une reproduction artificielle (modèle) du phénomène que l'on désire étudier, à observer le comportement de cette reproduction...

La deuxième preuve relève, elle, directement du champ de notre exposé; on la trouve à l'intérieur de l'article *Statistique*, rédigé par Yadolah DODGE, professeur de Statistique à l'Université de Neuchâtel (Suisse), pour le Dictionnaire Encyclopédique (Dunod, 1993) :

La simulation est la méthode statistique permettant la reconstitution fictive de l'évolution d'un phénomène. C'est une expérimentation qui suppose la constitution d'un modèle théorique présentant une similitude de propriétés ou de relations avec le phénomène faisant l'objet de l'étude.

A partir de ces références, il pourrait sembler difficile de faire pratiquer à des élèves de seconde LA SIMULATION tout en faisant l'économie de la MODELISATION. C'est pourtant à cette économie qu'invitent les programmes français actuels, car ce n'est que pour la classe de première qu'on y lit :

On clarifiera brièvement les positions respectives de la modélisation et de la simulation...; on parlera de simulation de la loi de probabilité P .

mais aussi, après présentation de quelques situations vues comme “simples” :

En dehors de tels cas, le choix d'un modèle à partir de données expérimentales est beaucoup plus délicat et ne sera pas abordé dans l'enseignement secondaire.

On remarquera au passage que, parmi les interprétations proposées par Pierre Dagnélie, les auteurs du programme de première ont retenu : SIMULATION D'UNE LOI DE PROBABILITE.

Des enseignants qui ont réfléchi à cette question relèvent la difficulté qu'il y aurait à dissocier Modélisation et Simulation ; on lira avec profit, dès la parution du volume de la CII, la contribution de Jean-Claude GIRARD et Michel HENRY titrée : *Modélisation et Simulation en classe : quel statut didactique ?*

En fait, la ligne directrice du programme actuel de seconde me paraît tenir en les trois propositions suivantes :

- la notion d'observations “équiprobables et indépendantes” est intuitive à partir d'un vécu concret antérieur (dés, pile ou face, tirages dans urnes avec remise et mélange), et peut donc être exploitée sans qu'il soit déjà besoin de la théoriser ;
- la sortie de “nombres au hasard” par une machine peut être naturellement admise par l'élève comme reproduisant le même type de phénomène ;
- la puissance et la rapidité de la machine permettent de mieux observer (et donc de mieux comprendre) de tels phénomènes.

Je tiens cette conception pour parfaitement viable. Mais je comprends les inquiétudes qu'elle soulève et dont je vais retracer ici certaines.

a. Lors de la table ronde de la SFDS, Gilbert SAPORTA, Professeur de Statistique au CNAM (Conservatoire National des Arts et Métiers) à Paris, a déclaré dans son intervention :

Il y a comme un cercle vicieux :

- *on veut faire comprendre les variables aléatoires via la simulation,*
- *pour comprendre la simulation, il faut savoir ce qu'est une variable aléatoire (et aussi des variables aléatoires indépendantes).*

b. De nombreux professeurs de Mathématiques dans l'enseignement secondaire m'ont déclaré en substance :

“On sait bien” que les générateurs de nombres au hasard dans les ordinateurs sont en fait déterministes.

En d'autres termes, il y a bien “simulation” dans l'appel au RANDOM, mais avec une autre signification que le sens “savant” développé jusqu'ici : il s'agit du sens courant (et un peu dévalorisant) de “faire semblant”, comme on “simule une maladie” pour se faire accorder un congé ! ainsi l'ordinateur “ferait semblant” de sortir des réalisations de variables aléatoires équiréparties (sur l'ensemble des 10 chiffres) et indépendantes.

c. La tentation est alors forte, pour l'enseignant qui se veut honnête vis-à-vis des élèves, de leur faire observer des suites sorties par RANDOM et les amener à **constater des régularités**. Mais, toujours dans l'article de CII déjà cité, Jean-Claude GIRARD et Michel HENRY écrivent :

Souhaitant simuler un sondage . . . , avons nous réellement simulé quelque chose ? Ou avons nous seulement vérifié que la fréquence des réponses favorables est voisine de p , ce qui ne fait que confirmer que le fabricant de l'ordinateur et le concepteur du logiciel ont bien rempli leur cahier des charges ?

d. L'aspect "boîte noire" du recours au RANDOM semble à certains professeurs contraire à la déontologie du mathématicien. C'est ainsi que Bernard PARISZ écrit, dans le volume CII :

Ce qu'il faut, c'est tempérer chez l'élève "l'effet boîte noire" inhérent à l'utilisation d'une machine.

Tous ces scrupules sont compréhensibles, mais ils me paraissent plus liés à la référence à des pratiques de l'enseignement mathématique qu'à des objections proprement scientifiques. En particulier, "BOITE NOIRE" ne signifie pas "ESCROQUERIE". On peut considérer que toute la science progresse en admettant ici ou là des zones d'ombre autour desquelles on fait la lumière.

Précisons en ce qui nous concerne cette dialectique des zones d'ombre et de la lumière.

Nous pouvons, il me semble, faire franchir un pas considérable à l'élève de seconde en **admettant** avec lui que l'on se trouve face à des phénomènes fondamentalement de même nature quand on enregistre des sorties successives de l'ordre RANDOM ou que, par exemple, on observe des jets successifs d'un dé **considéré** non truqué (j'insiste sur ma mise en gras des mots **admettant** et **considéré**). Dans un cas comme dans l'autre, nous ne savons pas tout sur le phénomène qui se produit ; nous sommes tout aussi ignorants (et même plus !) des paramètres "physiques" qui génèrent la trajectoire du dé que des règles qui, dans le "comportement interne" de l'ordinateur, génèrent les sorties du RANDOM. Est-ce fondamentalement plus gênant de laisser dans l'ombre les secondes que les premiers (qui sont totalement inaccessibles) ? Je pense que NON.

En revanche, ce que nous devons éclairer pour l'élève, c'est, à partir de telles observations :

- QUOI REGARDER ?

- COMMENT REGARDER ?

et ainsi le faire progresser dans la familiarité avec les notions intuitives évoquées ci-dessus (ce qui plus tard servira de fondement pour élaborer des modèles et mettre ceux-ci à l'épreuve).

Je ne reprendrai pas ici des exemples qui ont pu être évoqués dans mon exposé devant les participants à cette Université d'Été ou dans la discussion qui l'a suivi ; ils abondent aussi dans différents documents d'accompagnement des programmes, travaux des IREM . . . Je voudrais seulement dire que l'archétype de la "familiarisation" que j'appelais ci-dessus de mes vœux est, pour les concepteurs du programme de seconde comme pour moi-même, l'approche raisonnée de la loi des grands nombres, d'autant plus nécessaire que cette expression "loi des grands nombres" est tellement tombée dans le domaine courant qu'elle ne veut plus y dire grand chose.

Un entraînement type à cet égard consiste en faire effectuer plusieurs tracés, en fonction de n , de la proportion de succès lors de n expériences indépendantes résultant soit en un succès

avec probabilité p soit en un échec avec probabilité $1 - p$, et ceci pour $p = 0,5$ (dé non truqué), ou $p = 0,51$ ou $p = 0,501$; dans ces derniers cas de dés truqués, à partir de quelle valeur de n “voit-on”, dans toutes les séries réalisées, la proportion rester distincte de $0,5$?

Certes, il s’agit ici de constats et non de démonstrations mathématiques, et certains collègues peuvent se sentir gênés par ce type de pratique enseignante; mais il y a là un excellent entraînement à la logique des tests d’hypothèse ou des intervalles de confiance, qu’il est évidemment trop tôt pour théoriser en classe de seconde, mais dont on sait combien elle est difficile d’accès pour des esprits non préparés.

Une autre réticence des enseignants devant ce type d’activité tient à ce qu’elle se prête mal à l’évaluation chiffrée des élèves (mais est-ce si grave en seconde, s’agissant d’une fraction somme toute bien minoritaire du programme?).

Enfin certains professeurs voient mal comment organiser une trace écrite bien structurée du travail effectué, ce qui explique sans doute que soient peu pratiqués les “cahiers de statistiques” qui avaient été recommandés lors de la publication de ces programmes; ils seraient pourtant bien utiles: l’élève y collerait des sorties sur imprimantes des graphiques réalisés et le professeur veillerait à la précision et à la correction grammaticale des commentaires.

C’est sur cet exemple et ce vœu, somme toute assez modestes, que je clos cet exposé, espérant, pour reprendre l’articulation de mon propos, avoir convaincu mes collègues de la nécessité de l’enseignement de la Statistique, les avoir apaisés quant à sa difficulté et les avoir stimulés quant à sa possibilité.

Achévé de rédiger à Marne-la-Vallée
Le 4 janvier 2005