

ÉVOLUTIONS DE LA PRATIQUE STATISTIQUE DANS L'ENSEIGNEMENT DU SECOND DEGRÉ EN FRANCE : DU CALCUL STATISTIQUE AU DÉVELOPPEMENT DE LA PENSÉE STATISTIQUE¹

Philippe DUTARTE²

TITLE

Evolution of statistics education in secondary education in France: from statistical calculation to statistical thinking

RÉSUMÉ

L'enseignement de la statistique dans le second degré en France a connu une évolution de fond depuis les années 1990, passant progressivement de la pratique du calcul statistique à l'exercice de la pensée statistique. Ce mouvement, initié dans les programmes des sections de techniciens supérieurs, s'est consolidé dans ceux de lycée en 2000-2003, puis affirmé par l'introduction de l'aléatoire dans les programmes de collège en 2008, la réforme de la voie professionnelle en 2009, puis les nouveaux programmes du lycée général, encore en cours d'élaboration. On peut suivre, dans cette évolution, le rôle de l'outil informatique et de la démarche d'investigation, la prise en compte de l'aléatoire et de l'induction, un certain glissement des objectifs vers des enjeux citoyens et professionnels, et une évolution de l'évaluation vers davantage d'expérimentation et de réflexion. Bien sûr, beaucoup reste à faire, mais du chemin a cependant été parcouru.

ABSTRACT

In France, the teaching of statistics in secondary education has undergone a major transformation since the 1990s, the emphasis gradually shifting from statistical calculation to statistical thinking. This transformation, which first appeared in technical college curricula, later spread to high school curricula between 2000 and 2003. It gained prominence with the introduction of ideas of randomness in the curricula of lower secondary schools in 2008, and in the reform of vocational colleges in 2009; it is also expected to become part of the upper secondary school curricula currently under revision. One can trace, in this transformation, the increasing influence of computing and inquiry-based learning, as well as the emergence of randomness and inductive reasoning, and a greater emphasis both on social and professional issues in programs, and on experimentation and reflection in student evaluations. While statistical education can no doubt still be improved, it has already come a long way.

1 La statistique dans les programmes d'enseignement français depuis une vingtaine d'années

Une impulsion importante à l'enseignement des méthodes statistiques d'inférence par les professeurs de mathématiques de l'enseignement secondaire est due, en France, aux programmes des sections de techniciens supérieurs. Bien que relevant de l'enseignement supérieur, ces sections sont implantées dans les lycées et les intervenants sont des professeurs

¹ Cet article a été rédigé en novembre 2010.

² Inspecteur pédagogique régional de mathématiques, philippe.dutarte@ac-creteil.fr

du secondaire. Les programmes de 1989³ prévoient un enseignement de statistique inférentielle en précisant « qu'à l'aide des modèles théoriques que le calcul des probabilités permet de dégager, les notions de statistique inférentielle figurant au programme ont pour but la prise de décision concernant la conformité d'une production à un cahier des charges, soit en cours de fabrication, soit lors de la réception de la marchandise, et l'estimation de la durée de vie d'un équipement ». L'objectif de ces programmes est donc ciblé sur des applications très particulières. Une rédaction ultérieure élargira le champ d'application, notamment par l'introduction des plans d'expérience, tout en explicitant le contexte de cet enseignement de la statistique. On lit⁴ ainsi que « sous l'impulsion notamment du mouvement de la qualité, les méthodes statistiques sont aujourd'hui largement utilisées dans les milieux économique, social ou professionnel » et « qu'il s'agit de faire percevoir [...] ce que sont les procédures de décision en univers aléatoire, ainsi que leur pertinence ». Ces méthodes ont ainsi d'abord trouvé leur légitimité dans l'enseignement technologique. L'effort de formation des enseignants, dans lequel les IREM⁵ ont joué un rôle déterminant, avec une pédagogie adaptée aux élèves de lycée, a contribué, dans bien des établissements, à faciliter l'introduction de la statistique inférentielle en classe de seconde par les programmes de 2000.

Les programmes de lycées, mis en place à partir de 2000, de façon extrêmement volontariste sous l'égide du « Conseil national des programmes » dont la section mathématiques était présidée par Claudine Schwartz, elle-même chercheur en statistique, ont marqué un tournant dans l'enseignement de la statistique à ce niveau d'étude, en France.

Le programme de seconde⁶, mis en place en 2000, met en avant l'apport fondamental de l'outil informatique et la possibilité d'expérimentation qu'il présente. L'étude de la fluctuation d'échantillonnage et la simulation des phénomènes aléatoires constitueront un élément essentiel de la démarche d'expérimentation rendue possible par l'ordinateur. L'introduction du programme précise que : « L'informatique, devenue aujourd'hui absolument incontournable, permet de rechercher et d'observer des lois expérimentales. Cette possibilité d'expérimenter, classiquement davantage réservée aux autres disciplines, doit ouvrir largement la dialectique entre l'observation et la démonstration, et, sans doute à terme, changer profondément la nature de l'enseignement. [...] Un des apports majeurs de l'informatique réside aussi dans la puissance de simulation des ordinateurs ; la simulation est ainsi devenue une pratique scientifique majeure : une approche en est proposée dans le chapitre statistique ». Dans le domaine de la statistique descriptive, le programme de 2000 insiste sur la pensée statistique qui doit être mise en œuvre. Il ne s'agit plus simplement de savoir calculer, mais de réfléchir : « L'objectif est de faire réfléchir les élèves sur la nature des données traitées, et de s'appuyer sur des représentations graphiques pour justifier des choix de résumé ». L'enseignement de la statistique s'ouvre sur le monde. Le programme oblige ainsi à traiter des données en nombre suffisant, sur des thèmes motivants : « L'enseignant traitera des données en nombre suffisant pour que cela justifie une étude statistique ; il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts ». Toutes ces nouveautés n'ont trouvé que très difficilement leur place dans les classes. Dix ans après, nous verrons que les nouveaux manuels de seconde ne prennent encore en

³ BOEN n° 21 du 21 mai 1989.

⁴ Annexe de l'arrêté paru au BO hors série du 13 septembre 2001.

⁵ Instituts de Recherche sur l'Enseignement des Mathématiques.

⁶ Le lycée, en France, accueille les élèves pour 3 ans à l'issue de leur scolarité commune de 9 ans (école élémentaire puis collège) ; les classes y sont dénommées, en ordre chronologique, seconde, première et terminale ; on distingue les lycées généraux, les lycées technologiques et les lycées professionnels.

Ph. Dutarte

compte que de façon inégale les recommandations précédentes. On peut cependant affirmer que le chemin parcouru depuis 2000 est significatif et que, rétrospectivement, l'impulsion apportée par ces programmes a été essentielle et commence réellement à porter ses fruits.

Les programmes de première (2001) et terminale (2002) scientifiques s'inscrivent dans la continuité de celui de seconde. Pour la statistique descriptive, celui de première insiste sur la nécessité de ne pas faire de longs calculs à vide et sur l'articulation entre statistique et probabilités : « Cette partie descriptive ne doit pas faire l'objet de longs développements numériques, ni être déconnectée du reste du programme de probabilités et statistique ». L'articulation statistique-probabilités est éclairée par un « énoncé vulgarisé de la loi des grands nombres » et la notion de modélisation est mise en avant. En terminale scientifique sera abordée la notion de « l'adéquation d'une loi de probabilité à des données expérimentales » à travers l'exemple de l'adéquation à une loi équirépartie. Pour la première fois, le programme de terminale scientifique quantifie le temps consacré à chacun des trois domaines des mathématiques : « À titre indicatif, la répartition horaire des différents chapitres peut être : analyse 45 % (environ 14 semaines), géométrie 35 % (environ 11 semaines), probabilités et statistique 20 % (environ 6 semaines) ». Bien sûr, cette formulation n'est qu'à « titre indicatif », mais elle a le mérite de la clarté, d'autant qu'en 2002, pour la plupart des classes de terminale scientifique, la part de l'enseignement de la statistique et des probabilités est sans doute plus proche de 10 % que de 20 %. Les programmes actuellement rédigés pour 2012 ne quantifient pas ce partage, mais la lecture des contenus montre, hors spécialité, une répartition du temps sans doute proche de 50 % pour l'analyse et l'algèbre, 25 % pour la géométrie et 25 % pour la statistique et les probabilités.

Les rentrées 2007 et 2008 voient la mise en place de nouveaux programmes en première et terminale des filières technologiques tertiaires et socio-médicales. L'aspect pluridisciplinaire des activités statistiques, notamment en liaison avec les enseignements technologiques, est mis en avant, mais aussi le rôle fondamental de ce domaine pour le développement de l'esprit critique du citoyen : « On privilégiera les données issues des autres disciplines ou des médias, afin de développer une attitude critique vis-à-vis des informations chiffrées. L'emploi du tableur et de la calculatrice facilite l'exploitation des données ». De nouveau, l'enseignement de la statistique doit se faire une part dans l'emploi du temps, qui n'existait pas ainsi précédemment : « L'étude de ce chapitre doit constituer un moment important de la formation des élèves (développement de l'esprit critique, capacité à analyser les résultats d'une enquête...). Il est donc nécessaire que les élèves disposent d'un temps suffisant pour se familiariser avec ces notions ». Enfin, on retrouve des indications favorisant l'esprit statistique plutôt que l'aspect calculatoire : « Le tableur permet d'aborder des populations de grand effectif [...] ; le calcul d'un écart-type à la main à l'aide d'un tableau n'est pas un objectif et ne peut être exigible [...] ; rédiger l'interprétation d'un résultat ou l'analyse d'un graphique [est une] capacité attendue pour l'intégralité de l'étude des séries de données statistiques à une variable ».

Parallèlement, notamment pour répondre aux manques mis en évidence par les enquêtes internationales PISA, l'introduction de l'aléatoire dans les programmes de collège s'effectue en classe de troisième à la rentrée 2008 : « La notion de probabilité est abordée à partir d'expérimentations qui permettent d'observer les fréquences des issues dans des situations familières (pièces de monnaie, dés, roues de loteries, urnes, etc.). La notion de probabilité est utilisée pour modéliser des situations simples de la vie courante. Les situations étudiées concernent les expériences aléatoires à une ou à deux épreuves ». Cette initiation en classe de troisième sera prise en compte dans les nouveaux programmes de seconde.

À la rentrée 2009, la filière du lycée professionnel (baccalauréat professionnel en trois ans) est la première du lycée à être globalement réformée. C'est l'occasion d'y revoir complètement l'enseignement de la statistique et des probabilités, qui y était moribond. L'accent est fortement mis sur la formation du citoyen : « Ce domaine [la statistique] constitue un enjeu essentiel de formation du citoyen. Il s'agit de fournir des outils pour comprendre le monde, décider et agir dans la vie quotidienne ». Dans la continuité du programme de troisième, l'étude de la fluctuation d'échantillonnage est introduite sur la base de l'expérimentation à l'aide des moyens informatiques : « La notion de fluctuation d'échantillonnage, essentielle en statistique, est abordée dans cette partie du programme en étudiant la variabilité d'observation d'une fréquence. Elle favorise une expérimentation de l'aléatoire. L'objectif est de faire comprendre que le hasard suit des lois et de préciser l'approche par les fréquences de la notion de probabilité initiée en classe de troisième ». Le programme de première professionnelle, mis en place à la rentrée 2010, prévoit une initiation à la prise de décision : « La connaissance de la « variabilité naturelle » des fréquences d'échantillons (la probabilité qu'un échantillon aléatoire de taille n fournisse une fréquence dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ est supérieure à 0,95) permet de juger de la pertinence de certaines observations, où p désigne la proportion du caractère étudié dans la population de laquelle on prélève cet échantillon. Cette propriété se présente ainsi comme une première approche de la loi des grands nombres ».

La rénovation des programmes de mathématiques du lycée général et technologique a commencé par la mise en place d'un nouveau programme de seconde à la rentrée 2009, confirmé à la rentrée 2010. En statistique descriptive, ce programme demande de traiter des données réelles et met en avant la réflexion statistique : « L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, d'un fichier mis à disposition par l'INSEE), synthétiser l'information et proposer des représentations pertinentes ». Ce programme est aussi plus précis et plus ambitieux que celui de 2000 sur l'exploitation de la fluctuation d'échantillonnage : « L'objectif est d'amener les élèves à un questionnement lors des activités suivantes :

- l'estimation d'une proportion inconnue à partir d'un échantillon ;
- la prise de décision à partir d'un échantillon ».

Une initiation aux problématiques d'un intervalle de confiance ou d'un test, dans le cadre de l'étude d'une proportion, est donc mise en place.

Les programmes de première scientifique⁷, économique et sociale, puis des sciences technologiques et industrielles et du laboratoire, qui seront en application à la rentrée 2011, s'inscrivent en continuité avec le programme de seconde. Dans le domaine de la statistique inférentielle, on approfondit la problématique du test d'hypothèse à propos d'une proportion, dans le cadre de la loi binomiale : l'élève doit être capable « d'exploiter l'intervalle de fluctuation à un seuil donné, déterminé à l'aide de la loi binomiale, pour rejeter ou non une hypothèse sur une proportion. Et le programme explicite son objectif : « amener les élèves à expérimenter la notion de « différence significative » par rapport à une valeur attendue et à remarquer que, pour une taille de l'échantillon importante, on conforte les résultats vus en classe de seconde ». On peut penser qu'en terminale⁸, l'introduction de quelques lois

⁷ BO spécial n° 9 du 30 septembre 2010.

⁸ Il ne s'agit que d'hypothèses, ces programmes sont en cours d'écriture et non encore soumis à la consultation des enseignants.

Ph. Dutarte

continues, dont la loi normale, permettra de reprendre cette problématique dans le cadre gaussien et de revenir sur la notion d'intervalle de confiance. Les élèves de classe terminale pourront alors justifier, en admettant l'approximation d'une loi binomiale par une loi normale, l'expression $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ expérimentée en classe de seconde. Des échanges fructueux avec les autres disciplines scientifiques sont également envisageables à propos de certaines pratiques statistiques présentes en physique ou en sciences de la vie et de la terre, erreurs de mesures, barres d'erreurs, comparaison de deux populations à l'aide d'intervalles de confiance ou de barres d'erreurs.

2 Une analyse succincte des manuels de la classe de seconde

Bien entendu, les « injonctions » des programmes ne se traduisent pas immédiatement dans les pratiques pédagogiques. Il y a plusieurs raisons à cela, notamment des questions de formation des enseignants, de ressources disponibles et la pression de l'examen. Une première mesure de la distorsion éventuelle entre les programmes et les pratiques en classes peut s'effectuer par l'examen des manuels disponibles. Nous avons examiné 9 des manuels à la rentrée 2010, correspondant au nouveau programme de 2009⁹. L'étude succincte qui a été menée porte sur la partie « Statistique descriptive - analyse des données » du programme qui généralement constitue un chapitre, mais est parfois jointe à la partie « étude des fluctuations d'échantillonnage ». Cette partie a été choisie parce qu'elle est, a priori, la plus connue des enseignants et est pratiquée souvent de longue date.

La première chose examinée a été la proportion d'exercices purement « calculatoires » ou du moins sans aucune utilisation des calculs effectués ou des graphiques réalisés. Les résultats sont très variables, ce qui est déjà un bon signe car il y a quelques années la part de ces exercices « gratuits » était massive, de façon générale. Un exemple particulièrement frappant est l'exercice de « statistique » suivant¹⁰ :

27 * On donne la série statistique suivante :

Valeurs	1	a	5	2a	9
Effectifs	28	22	14	20	16

a) Déterminez a sachant que la moyenne est égale à 4,59.
b) Calculez la médiane de cette série.

Dans deux manuels, les exercices calculatoires représentent encore 80 % et 76 % des exercices proposés. La place laissée à la « pensée statistique » est donc de l'ordre de 20 %, si les élèves y parviennent. À l'opposé, dans trois manuels, la part de ces exercices est de l'ordre de 20 %. Ce type d'exercice est en effet sans doute nécessaire pour installer certains automatismes, le bon dosage n'étant pas 80 %. On pourra voir « le verre à moitié plein » en constatant que les deux tiers des manuels comportent moins de 50 % d'exercices uniquement calculatoires. Il y a un réel progrès dans ce domaine.

⁹ Il s'agit des manuels suivants : Bordas « Indice », Bordas « Pixel », Didier « Math'x », Hachette « Déclic », Hachette « Repères », Hatier « Odyssée », Nathan « Collection Antibi », Nathan « Hyperbole », Nathan « Transmath ». Nous ne disposons pas de l'ouvrage Belin « Symbole ».

¹⁰ Manuel de seconde Nathan « Collection Antibi », page 237.

Le deuxième aspect examiné est la part des exercices traitant des données « réelles », qui est un attendu du programme. Il s'agit de données dont on donne généralement la source ou, si ce n'est pas le cas, dont on conçoit qu'elles correspondent à une réalité crédible. Là encore les résultats sont très inégaux, allant de 7 % à 40 %, mais l'ensemble est assez positif puisqu'une majorité d'ouvrages (5 sur 9) consacrent plus de 20 % des exercices à l'étude de données réelles. Certains manuels proposent, de plus, des thèmes de recherche ou des sujets d'exposés, notamment sur des thèmes statistiques de société.

Nous avons ensuite considéré le nombre d'activités fondées sur l'analyse d'un grand nombre de données, ainsi que le préconisent les programmes, depuis 2000. La moisson est assez maigre. Trois manuels n'en proposent aucune et deux manuels n'en proposent qu'une seule. Autant dire que l'apprentissage de la statistique en classe de seconde, vu au travers de ces manuels, ne passe pas par la problématique d'avoir à résumer un grand nombre de données. Les quatre autres manuels examinés prennent en compte cet enjeu en proposant de 2 à 6 activités, en donnant l'adresse de sites Internet fournissant un fichier de données à un format tableur (l'INSEE par exemple) ou en fournissant de tels fichiers, par le biais notamment d'un site dédié au manuel. Il est à souhaiter que le développement des manuels numériques permette de faciliter l'accès des enseignants et de leurs élèves à de telles ressources.

Enfin, l'enjeu de la place de l'informatique dans l'élaboration des connaissances statistiques étant fondamental (voir le paragraphe suivant), nous avons considéré dans ces manuels, pour la partie « statistique et analyse des données », le nombre d'activités d'expérimentation proposées à l'aide des TIC¹¹ (nous ne faisons pas entrer dans cette catégorie les activités « mode d'emploi » de découverte du tableur). Les résultats sont inégaux, mais c'est mieux que pour le traitement d'un grand nombre de données. Seuls deux manuels n'en proposent aucune et deux autres n'en proposent qu'une seule. La majorité des manuels étudiés (5 sur 9) proposent au moins trois travaux pratiques d'investigation avec l'ordinateur. Il s'agit d'activités conséquentes demandant généralement une séance en salle informatique chacune. L'offre est donc, dans ce cas, suffisante.

Comme on peut le constater par cette petite enquête, la pensée statistique dans les manuels est diversement avancée, mais les progrès depuis 10 ans sont cependant significatifs. L'ouverture vers les données réelles est grandement réalisée et, sauf exceptions, la part des exercices demandant de simples exécutions, sans interprétation ni analyse, a grandement diminué. Restent des progrès à réaliser, parfois dans la prise en compte des outils informatiques, souvent pour fournir des ressources d'analyse d'un grand nombre de données.

3 Le rôle de l'outil informatique et de la démarche d'investigation

La place de l'informatique, comme outil favorisant une démarche d'investigation, est de plus en plus grande dans l'enseignement des mathématiques et de la statistique en particulier. Le choix, en 2000, d'inscrire au programme de seconde la simulation informatique pour expérimenter la fluctuation d'échantillonnage, était d'ailleurs une façon de répondre à la demande du ministère de mieux intégrer, en mathématiques, une démarche liée à la dimension informatique.

¹¹ Technologies de l'information et de la communication.

Ph. Dutarte

L'impulsion donnée à la statistique dans les programmes mis en place à partir de 2000 est donc en partie due au développement de l'utilisation de l'outil informatique et la première entrée à été celle de la simulation. Dans le cadre de ce programme, la simulation favorise une approche fréquentiste de la notion de probabilité, en prolongeant les expériences physiques, pour expérimenter la loi des grands nombres, et permet, par ailleurs, une compréhension de la fluctuation d'échantillonnage, en multipliant les échantillons d'une taille donnée. Nous donnons ci-dessous un exemple d'activité développée en classe de seconde, suite au programme de 2000, et toujours d'actualité.

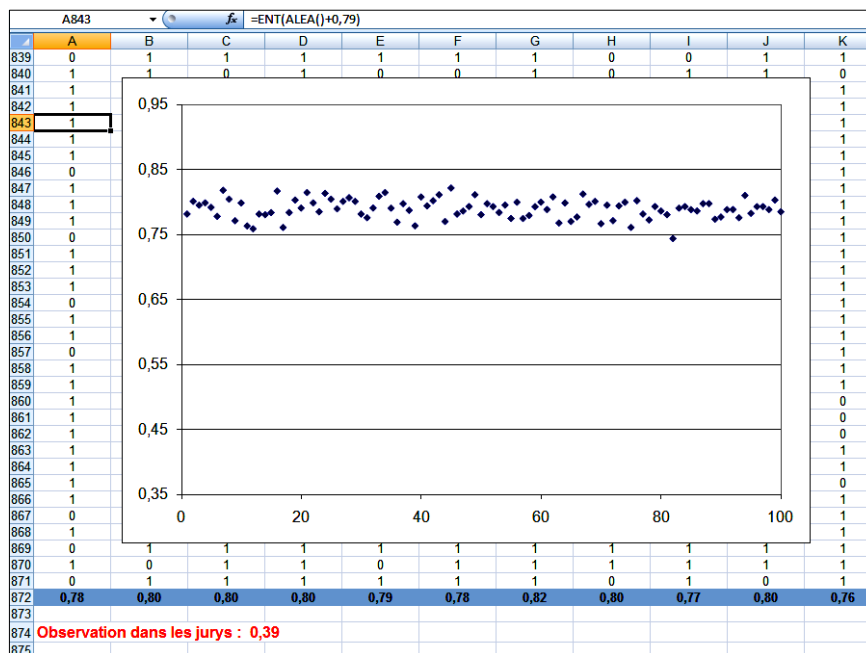
Exemple 1 : l'affaire Castaneda contre Partida

L'énoncé suivant a été donné dans des classes. On demande aux élèves de se placer dans le rôle d'un expert en statistique, fournissant des arguments scientifiques à la cour.

En novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79 % de la population du comté était d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors des 11 années précédentes, il n'y eut que 339 personnes d'origine mexicaine.

Devant la Cour Suprême, un expert statisticien produisit des arguments pour convaincre du bien-fondé de la requête de l'accusé. En vous situant dans le rôle de cet expert, produisez à votre tour des calculs, des raisonnements, des graphiques... pour montrer que le hasard ne peut pas « raisonnablement » expliquer à lui seul la sous-représentation des américains d'origine mexicaine dans les jurys de ce comté.

- Vous commencez ce travail en binômes en utilisant les documents disponibles, la calculatrice, le tableur.
- Vous terminez la rédaction (arguments en français, calculs, graphiques...) en devoir individuel, à la maison.



Evolutions de la pratique statistique dans l'enseignement du second degré en France

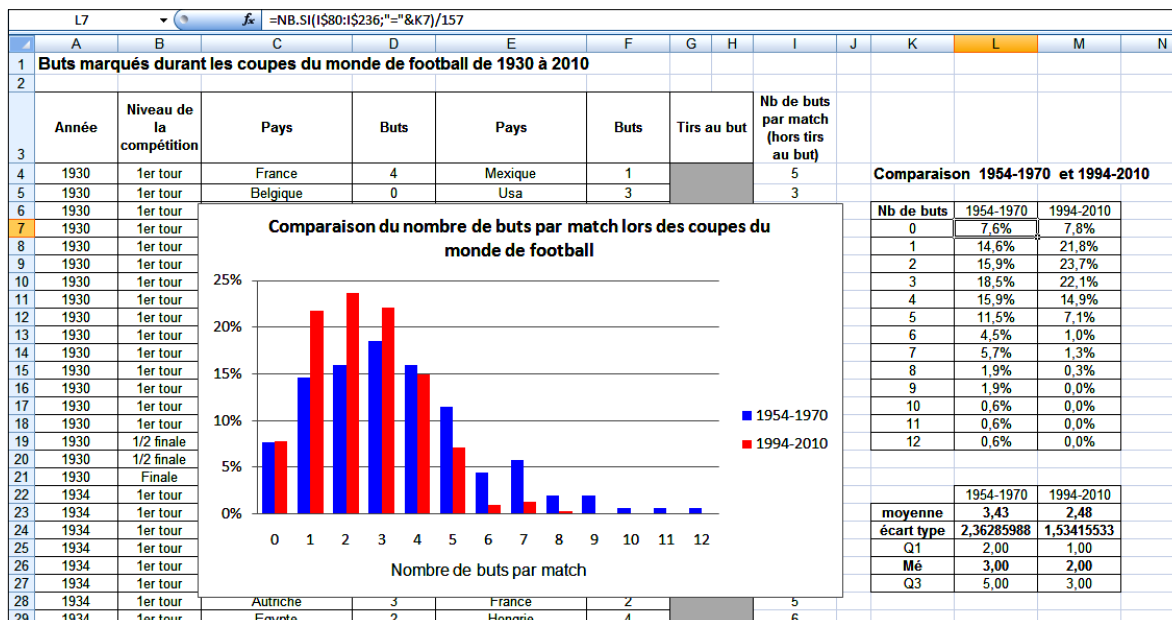
En simulant un grand nombre d'échantillons aléatoires de taille $n = 870$ dans une population où la proportion du caractère considéré est $p = 0,79$, les élèves constatent aisément que la proportion $f = \frac{339}{870} \approx 0,39$ est très éloignée de la zone correspondant aux fluctuations « naturelles » d'échantillonnage. On a là une « preuve statistique » du fait que la constitution des jurys ne peut pas être considérée comme un tirage aléatoire.

Le deuxième rôle de l'outil informatique dans l'enseignement de la statistique est de permettre l'investigation d'un grand nombre de données. Internet, et certains sites spécialisés¹², donnent accès à des « fichiers tableurs » importants. Le tableur libère l'élève des difficultés techniques de calcul ou de représentation graphique et ouvre la porte à la réflexion et aux choix pertinents.

Exemple 2 : les scores des matchs des coupes du monde de football

Les statistiques sportives abondent, notamment sur Internet, et l'intérêt des élèves dans ce domaine est assez général. Il est ainsi plutôt séduisant de leur faire mettre en œuvre leurs connaissances statistiques pour explorer des données sportives. Nous prenons ici les scores de tous les matchs des coupes du monde de football depuis 1930 (ces données sont aisément accessibles sur Internet).

On peut par exemple comparer le nombre de buts marqués par match durant la période 1954-1970 et durant la période 1994-2010. On constate que les distributions sont assez différentes. On avait tendance à marquer davantage de buts « jadis », avec une dispersion plus grande.



¹² Le site de l'INSEE, de l'OCDE, de l'INED, le site Statistix fournissent de nombreuses ressources.

Ph. Dutarte

Un troisième aspect important de l'intérêt des TIC dans l'enseignement de la statistique est celui d'aider à la compréhension des notions. Nous avons déjà évoqué le rôle de la simulation pour mieux appréhender la loi des grands nombres ; prenons ici deux exemples illustrant le nouveau programme des classes de premières scientifiques et technologiques industrielles (rentrée 2011).

Exemple 3 : algorithme de recherche d'un intervalle de fluctuation dans le cadre d'une prise de décision à l'aide de la loi binomiale

Le programme de première scientifique, mis en œuvre à la rentrée 2011, prévoit « d'exploiter l'intervalle de fluctuation à un seuil donné, déterminé à l'aide de la loi binomiale, pour rejeter ou non une hypothèse sur une proportion » et de « remarquer que, pour une taille de l'échantillon importante, on conforte les résultats vus en classe de seconde » ; il stipule que « l'intervalle de fluctuation peut être déterminé à l'aide d'un tableur ou d'un algorithme ». De quoi s'agit-il ? Il a été admis, en classe de seconde, que, sous certaines conditions, plus de 95 % des fréquences observées sur des échantillons aléatoires de taille n se situent dans l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$, où p est la proportion du caractère étudié dans la population. En classe de première, on dispose de la loi binomiale et l'on peut considérer la variable aléatoire X correspondant au nombre de fois où le caractère étudié apparaît dans l'échantillon.

Dans ce cadre, sous l'hypothèse que la proportion du caractère vaut p dans la population, X suit la loi binomiale de paramètres n et p . On cherche à partager l'intervalle $[0, n]$, où X prend ses valeurs, en trois intervalles $[0, a - 1]$, $[a, b]$ et $[b + 1, n]$ de sorte que X prenne ses valeurs dans chacun des intervalles extrêmes avec une probabilité aussi proche que possible de 0,025, sans dépasser strictement cette valeur¹³. On recherche donc le plus grand entier a tel que $P(X < a) \leq 0,025$ et le plus petit entier b tel que $P(X > b) \leq 0,025$. D'un point de vue algorithmique, il est plus efficace de travailler avec les probabilités cumulées croissantes, que la calculatrice ou le tableur fournissent facilement. En tabulant les probabilités cumulées $P(X \leq k)$, pour k allant de 0 à n , il suffit de déterminer le plus petit entier a tel que $P(X \leq a) > 0,025$ et le plus petit entier b tel que $P(X \leq b) \geq 0,975$.

Cet algorithme est aisé à réaliser sur un tableur¹⁴. La modification des valeurs des paramètres n et p permet alors de mieux comprendre sa détermination et de le comparer à celui utilisé en classe de seconde.

¹³ Cette détermination du segment d'extrémités a et b prend bien sûr tout son sens dans le cas où l'on cherche à prendre une décision sur l'hypothèse selon laquelle la proportion du caractère vaut p dans la population. Il s'agit là d'une problématique de statistique inférentielle qui ne relève pas du programme de la classe de première scientifique pour laquelle cette activité est proposée. Mais certains enseignants peuvent vouloir déjà en profiter pour donner un « coup de projecteur » dans cette direction.

¹⁴ Voir l'image d'écran. La recherche de a , exprimé en fréquence, s'effectue en recopiant vers le bas la formule =SI(B6>0,025;A6/B\$3;"") introduite en C6. Celle de b , exprimé en fréquence, en recopiant la formule =SI(B6>=0,975;A6/B\$3;"") introduite en D6. Les valeurs a et b sont données par les formules =MIN(C6:C1006) et =MIN(D6:D1006).

Evolutions de la pratique statistique dans l'enseignement du second degré en France

D6													
=SI(B6>=0,975;A6/B\$3;"")													
A	B	C	D	E	F	G	H	I	J	K	L	M	
Intervalle de fluctuation à 95 % d'une fréquence donné par la loi binomiale													
n = taille de l'échantillon		p = proportion supposée dans la population											
n= 50		p= 0,35		Intervalle classe de seconde :					p-1/rac(n)= 0,209		p+1/rac(n)= 0,491		
k	P(X<=k)	recherche a	recherche b	Intervalle de fluctuation à environ 95 % (selon la loi binomiale):								0,22	0,48
0	4,4225E-10												
1	1,2349E-08												
2	1,69426E-07												
3	1,52271E-06												
4	1,00848E-05												
5	5,25002E-05												
6	0,000223793												
7	0,000803553												
8	0,002481512												
9	0,006697923												
10	0,016006462												
11	0,03423297	0,22											
12	0,066129359	0,24											
13	0,116333143	0,26											
14	0,18777699	0,28											
15	0,280104423	0,3											
16	0,388855486	0,32											
17	0,505972015	0,34											
18	0,621587051	0,36											
19	0,726436313	0,38											
20	0,813945121	0,4											
21	0,881259589	0,42											
22	0,929038739	0,44											
23	0,96035885	0,46											
24	0,97933161	0,48	0,48										
25	0,989956356	0,5	0,5										
26	0,995457334	0,52	0,52										

Exemple 4 : notion de seuil de signification lors d'une prise de décision

L'activité suivante est inspirée¹⁵ de présentations pédagogiques effectuées, pour des physiciens, par Alain Vivier, responsable pédagogique à l'INSTN¹⁶, institut dépendant du Commissariat à l'énergie atomique à Saclay. Convaincu de l'intérêt pédagogique des expérimentations sur tableur, Monsieur Vivier déclare : « Pour ma part je n'aurais jamais approfondi ces aspects [de statistique et probabilités], indispensables en physique, sans cet outil [le tableur]. Cela m'a été utile non seulement pour des aspects enseignements, mais aussi de recherche (notamment dans le domaine de la problématique du seuil de décision, qui peut s'avérer parfois difficile). ». Il est extrêmement intéressant de pouvoir travailler, en classe de mathématiques, avec des méthodes statistiques pratiquées dans les autres disciplines scientifiques.

Une activité sur le thème de la radioactivité, consistant à rechercher un seuil de différence significative à l'aide de la loi binomiale¹⁷, peut se présenter comme suit.

On mesure en laboratoire, avec un compteur Geiger, un objet pouvant être « radioactif ». Le compteur est réglé selon une certaine sensibilité et on effectue une mesure à un mètre de l'objet, pendant dix secondes. L'instrument compte 37 désintégrations ou « coups ». Cependant, avec ce réglage et dans ces conditions, une mesure de « bruit de fond » (correspondant à l'environnement du laboratoire) donne en moyenne un comptage de 30 coups. La question qui se pose est de savoir si la différence observée est assez importante pour considérer l'objet

¹⁵ Avec l'aimable autorisation de Monsieur Alain Vivier.

¹⁶ Institut national des sciences et techniques nucléaires.

¹⁷ Le comptage du nombre de désintégrations durant un intervalle de temps donné suit, a priori, une loi de Poisson de paramètre λ . Cependant, la loi binomiale de paramètres n et p (facile à simuler et connue des élèves, ce qui n'est pas le cas de la loi de Poisson) est, sous certaines conditions, proche de la loi de Poisson de paramètre $\lambda = n \times p$. C'est le cas ici, où la loi binomiale correspond à un comptage à temps discret.

Ph. Dutarte

comme « radioactif ».

On suppose que dans le laboratoire, durant chaque centième de seconde, le compteur est aléatoirement susceptible de compter un coup de bruit de fond avec une probabilité 0,03, ce que l'on simulera à l'aide d'un tableur avec l'instruction =ENT(ALEA()*0,03).

1. Simuler en colonne A un comptage de bruit de fond pendant 10 secondes, puis recopier vers la droite pour obtenir la simulation de 100 comptages.

2. Calculer la moyenne des 100 comptages simulés. Est-elle proche de 30 coups ? (Faire F9 pour obtenir d'autres simulations.)

3. Un comptage supérieur ou égal à 37 coups vous semble-t-il exceptionnel ?

4. Déterminer les paramètres n et p de la loi binomiale suivie par la variable aléatoire X modélisant un comptage de bruit de fond.

5. Sur une nouvelle feuille, calculer une table fournissant $P(X \leq k)$ pour k allant de 0 à 1 000.

6. On considérera qu'il y a radioactivité (comptage « significatif ») à partir d'un comptage de $N + 1$ coups, où N est le plus petit entier tel que $P(X \leq N) > 0,95$.

a. Déterminer la valeur de N .

b. Contrôler la valeur de N obtenue en vérifiant qu'environ 95 % des comptages simulés (en moyenne) sont inférieurs ou égaux à N (utiliser la fonction NB.SI et faire plusieurs fois F9).

c. Un comptage de 37 coups est-il « significatif » ? Quel est le comptage minimal à partir duquel on considère qu'il y a une radioactivité significative ?

d. Quelle est la probabilité de commettre l'erreur suivante : on considère que le comptage est « significatif » alors que c'est un bruit de fond (il n'y a pas de radioactivité) ?

4 La place de l'évaluation

L'évaluation joue un rôle très important dans l'évolution, ou non, des pratiques pédagogiques. Le « pilotage par l'examen » demeure en France assez essentiel. Le premier objectif, qui devrait être celui de la formation, est davantage celui de la préparation à l'examen et, pour le lycée, au baccalauréat. L'évaluation en mathématiques dans l'enseignement secondaire français demeure globalement très « classique » : évaluation sommative de connaissances, à l'écrit, chiffrée, résumée par une moyenne. Le prototype de cette évaluation peut être le baccalauréat. Or ce point de vue est assez peu compatible avec nombre des compétences correspondant à la « pensée statistique » et développées par le traitement d'un grand nombre de données ou l'expérimentation à l'aide des TIC.

La situation évolue cependant et de manière favorable à l'enseignement de la statistique. Les performances décevantes des élèves français aux évaluations PISA¹⁸ ont remis en question quelques certitudes. La notion de « culture mathématique » au sens de PISA¹⁹ trouve un écho dans la rédaction des programmes. L'ouverture des mathématiques au monde est

¹⁸ Programme International du Suivi des Acquis des élèves, à l'initiative de l'OCDE.

¹⁹ « La culture mathématique est l'aptitude d'un individu à identifier et à comprendre les divers rôles joués par les mathématiques dans le monde, à porter des jugements fondés à leur propos, et à s'engager dans des activités mathématiques en fonction des exigences de la vie présente et future en tant que citoyen constructif et réfléchi. »

favorable à un apprentissage plus conséquent de la statistique et des probabilités. Le sens critique et la capacité à prendre des initiatives doivent être développés et évalués. La prise en compte, dans l'évaluation des élèves, des compétences développées grâce à l'outil informatique se met progressivement en place. L'évaluation de travaux pratiques sur informatique est demandée aux professeurs par les corps d'inspection. Le baccalauréat professionnel sera désormais évalué, pour les mathématiques, en deux contrôles en cours de formation, comprenant chacun une partie sur informatique évaluée durant la passation. Une épreuve pratique de mathématiques au baccalauréat général est envisagée. Une meilleure préparation aux études supérieures exige en effet que l'évaluation au lycée soit en meilleure adéquation avec les objectifs de formation.

5 Des résistances et des perspectives

Des résistances et des obstacles demeurent, s'opposant à un enseignement de la statistique plus performant, c'est-à-dire développant davantage les compétences liées à la pensée statistique, plus qu'à la simple exécution de tâches techniques. On peut citer, comme premier obstacle, celui de la formation des enseignants. La statistique a rarement fait partie de la formation initiale des enseignants de mathématiques de l'enseignement secondaire et les besoins de formation sont importants. Ceci est clairement remonté de la consultation des enseignants menée à propos des nouveaux programmes de première. En second lieu, l'image de la statistique chez certains enseignants de mathématiques est encore assez négative, notamment par manque d'information ou de mise en perspective. Il est vrai aussi que la démarche inférentielle est inhabituelle et trouve « au pays de Descartes » un terrain qui n'est pas toujours favorable, ce qui renvoie à l'historique de la place de la statistique dans l'Université française et dans l'enseignement. Troisième difficulté, les problématiques de la statistique trouvent souvent leur sens dans des contextes pluridisciplinaires. Les mathématiques sortent d'une tour d'ivoire confortable. Outre les difficultés liées au dialogue entre disciplines, les professeurs de mathématiques se trouvent confrontés, en statistique, à des situations inhabituelles, touchant à des thèmes de société, portant à débat et pouvant soulever des questions éthiques. Enfin, nous l'avons dit, l'omniprésence du devoir surveillé classique et la difficile évolution du baccalauréat freinent le changement des pratiques pédagogiques.

Pour autant, les signes positifs sont « significatifs ». Le rapide historique, effectué au début de cet article, montre que, depuis une vingtaine d'années, les programmes du secondaire ont constamment évolué dans les mêmes directions, celle d'une part de plus en plus importante accordée à la partie statistique et probabilités, occupant à présent de l'ordre de 20 % à 25 % du temps d'enseignement et celle d'un développement de la pensée statistique, intégrant une dimension aléatoire dès la fin du collège et une initiation à l'estimation et à la prise de décision se déployant de façon cohérente de la seconde à la terminale. L'évolution de la forme des examens, prenant en compte la démarche d'investigation fondée sur l'outil informatique, est engagée dans la filière professionnelle. Enfin, l'équipement informatique, encore inégal dans les collèges, devient performant dans les lycées français et s'accompagne de l'amélioration de la formation des enseignants à l'usage de cet outil. Au total, il nous semble que ces indices plaident en faveur d'un optimisme raisonnable. La pratique de la statistique dans l'enseignement secondaire français évolue résolument dans le bon sens.