POUR UNE EDUCATION A L'INFERENCE STATISTIQUE AU LYCEE

Philippe DUTARTE
Commission Inter-Irem
Lycées techniques

« Pour comprendre les pensées de Dieu, il faut étudier les statistiques, car elles constituent la mesure de ses desseins. »
Florence Nightingale (1820-1910).

« Quelquefois les phénomènes paraissent dépendre d'une cause régulière ; et cependant, ils ne sont que le résultat de ces causes irrégulières, variables et inconnues, auxquelles nous donnons le nom de hasard. C'est à l'analyse des probabilités à déterminer jusqu'à quel point une cause régulière est probable en vertu de ces phénomènes, et à l'indiquer aux philosophes, comme objet digne de leurs recherches. »
P. S. Laplace – Leçons à l'Ecole normale de l'An III (1795).

Introduction

Pour la première fois, les programmes de seconde, premières et terminales ES et S, entrés en application en 2000, 2001 et 2002, font une place aux méthodes statistiques dites « inférentielles », au travers de sujets comme, fluctuations d'échantillonnage, sondages, adéquation de données à un modèle équiréparti. L'intention des concepteurs de ces programmes, fort louable, était clairement affichée : *« Former les élèves en statistique, c'est leur donner les moyens de développer une forme de pensée critique sans laquelle ils seront exclus du débat social et scientifique »* (projet de programme de terminale ES et S). La presse, et pas seulement la presse scientifique, se fait en effet régulièrement l'écho de sondages, évoque les notions de risques, d'un illusoire risque zéro, de « preuves statistiques » (qui ne sont pas des preuves à 100%…), de modèles, de méthodes statistiques aidant à prendre des décisions dans un environnement incertain. Certains sociologues prétendent que nous sommes entrés dans la « société du risque », nécessitant une politique de « précaution », terrain qu'on ne saurait laisser aux seuls experts et groupes de pression. Cet article se propose de montrer, par quelques exemples, qu'une éducation, assez précoce, des futurs citoyens aux méthodes d'inférence statistique est sans doute aussi nécessaire que délicate.

Rappelons que la grande majorité des professeurs de l'enseignement secondaire n'a reçu aucune formation dans ce domaine bien spécifique des mathématiques, souvent éloigné des méthodes déductives habituelles. De plus, les a priori négatifs sur la statistique sont nombreux, y compris chez les professeurs de mathématiques. « S'agit-il d'ailleurs de mathématiques ? » pense-t-on souvent. Comment est traité le thème « Simulation et fluctuation d'échantillonnage » du programme de seconde ? souvent au mois de juin après le dernier conseil de classe... Nous exagérons à peine !

Les sections de techniciens supérieurs ont été confrontées à l'enseignement de la statistique inférentielle en mathématiques dès 1988. L'industrie française, avec quelque retard, notamment par rapport aux Etats-Unis, a en effet dû intégrer dans les années 1960/70 les méthodes statistiques de contrôle de qualité et de fiabilité. Face à la « déraisonnable » et spectaculaire efficacité de ces méthodes, appliquées à l'industrie japonaise et américaine, il en allait de la survie de l'industrie nationale. Dès 1988 sont mis en place des stages de formation pour les enseignants de mathématiques des sections de BTS concernées. A ce propos, Bernard Verlant, responsable de la Commission inter-IREM « Lycées technologiques », remarque : « *Je m'occupe de la formation continue depuis 1977 et la statistique inférentielle est le seul thème de formation qui ait nécessité durant cette période un dispositif aussi étalé dans le temps. Les stages n'ont pas désempli jusqu'en 2000… Quant à l'accueil réservé au premier exercice du bac ES Métropole 2003* [adéquation à une loi équirépartie]*, il ne fait que confirmer la difficulté d'introduction de cette notion* ».

Et pourtant, il serait souhaitable que la tentative des nouveaux programmes de seconde à terminale dans ce domaine ne soit pas un échec. Essayons de montrer pourquoi.

Variabilité et fluctuations naturelles

L'objectif de la statistique est l'étude de la variabilité (qui s'applique souvent à notre quotidien). Il s'agit de distinguer les variations naturelles, habituelles, imputables au seul hasard, des variations « significatives », signes d'autres causes. Remarquons que cela suppose que le hasard est régi par des lois, qu'il ne soit pas « n'importe quoi ». Ce qui ne va pas de soi, y compris pour les élèves.

L'un des premiers objectifs, lors de ce chapitre de seconde sur les fluctuations d'échantillonnage, est que l'élève dépasse le stade du simple « tout est possible, car c'est le hasard ». Cette attitude bloque toute analyse. Elle peut parfois correspondre à des a priori religieux. Aux débuts de la statistique, les études de la mortalité se sont opposées à l'idée que le destin humain est entre les mains de Dieu (mais les sociétés d'assurance tenaient à ces études). C'est ainsi, qu'actuellement, l'introduction des probabilités dans les programmes d'enseignement de certains pays provoque des débats avec des religieux... « On ne peut rien dire car tout est possible » est une attitude qui cependant se justifie lorsqu'on ne veut répondre qu'avec certitude, à 100% (ce que l'on recherche habituellement en mathématiques, mais que ne permet jamais la méthode statistique). Ce n'est pas une raison pour ne rien dire. Tout est possible (individuellement)… mais pas n'importe comment (sur un grand nombre de données). C'est ce que l'expérimentation, en particulier par simulation informatique, montrera.

Le classement des hôpitaux

Prenons l'exemple du classement des hôpitaux, paru il y a quelques années dans le magazine *Sciences et Avenir* (septembre 1998). Sur un total de 512 établissements, est établi, pour quatre spécialités, une liste des 50 meilleurs et des 50 plus dangereux. On trouve ainsi, pour la chirurgie digestive, l'hôpital de Briançon dans le groupe de tête, avec

1 décès sur 12 interventions au total (fréquence de décès $f_B = 1/12 \approx 8{,}3\%$), et dans le groupe de queue, l'hôpital de Méru (Oise), avec 3 décès sur 12 interventions au total (fréquence de décès $f_M = 3/12 = 25\%$). Or la différence observée, pour importante qu'elle paraisse (surtout si l'on doit subir une opération), n'est absolument pas « significative ». C'est-à-dire que le hasard seul suffit amplement à l'expliquer, sans que l'on doive invoquer une différence de qualité dans les soins prodigués dans les deux établissements. Pour prendre une image, disons que la même roue du destin peut avoir tourné dans les deux hôpitaux, mais que faire tourner la roue 12 fois ne suffit pas pour s'en faire une idée. Pour le comprendre, il faut avoir quelques notions d'échantillonnage, figurant au programme de seconde (un échantillon de taille 12, cela fluctue naturellement beaucoup).

La méthode statistique permettant d'évaluer si une différence est « significative » est celle des tests d'hypothèses évoquée dans la dernière partie de cet article. C'est cette procédure qui donne une définition rigoureuse de l'expression « différence significative ». Cependant une simple expérience peut montrer que les deux résultats observés, 1 sur 12 et 3 sur 12, peuvent fréquemment résulter de la même expérience aléatoire. Supposons que l'on face tourner 12 fois la roue de loterie de la figure 1, où le secteur noir représente $p = 16{,}7\%$ de la surface ($16{,}7\% \approx 2/12$ est la moyenne de f_B et f_M). Si la roue s'arrête sur le secteur noir, c'est un « décès ». Les élèves peuvent simuler la rotation de cette roue par l'instruction int(rand + 2/12) sur la calculatrice, ou ENT(ALEA( ) + 2/12) sur le tableur Excel. Il suffit de répéter 12 fois pour obtenir un échantillon de taille 12. Il est facile, sur tableur, de sélectionner l'échantillon et par recopie d'obtenir une centaine d'échantillons pour observer les fluctuations des résultats et constater que les cas 1 « décès » et 3 « décès » sont toujours très fréquents. Lorsque l'on connaît la loi binomiale, des calculs théoriques de probabilité viennent confirmer les observations sur les simulations. Si la variable aléatoire X suit la loi binomiale de paramètres $n = 12$ et $p = 2/12$, on a $P(X = 1) \approx 27\%$ et $P(X = 3) \approx 20\%$ (voir fig. 2).

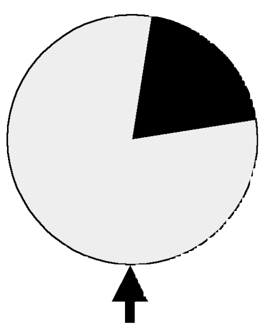

Fig.1 : La roue du destin

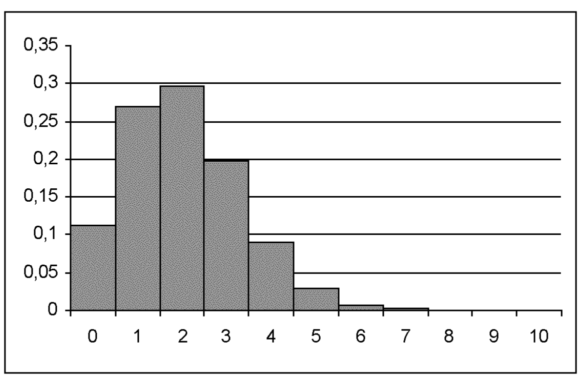

Fig. 2 : Distribution de la loi binomiale $n = 12$; $p = 2/12$

Qu'en serait-il si les statistiques avaient porté sur un échantillon de $n = 120$ opérations ? Les élèves peuvent faire l'expérience, en allongeant la taille des échantillons jusqu'à 120. On constatera alors qu'un taux de décès de 25% = 30/120 ou plus est extrêmement rare. La loi binomiale de paramètres $n = 120$ et $p = 2/12$ fournit $P(X \geq 30) \approx 1{,}3\%$. Alors que la différence entre 1 sur 12 et 3 sur 12 n'est pas significative, celle entre 10 sur 120 et 30 sur 120 l'est, alors qu'il s'agit des mêmes proportions respectives ! Pas si évident…

Les « clients » des hôpitaux américains peuvent accéder, sur le site www.healthgrades.com («*The single best source for healthcare quality ratings*» si on les croit), aux notes attribuées à chaque hôpital des U.S.A. selon l'acte chirurgical envisagé. Les notes vont de ***** quand l'établissement figure parmi les 10% les mieux notés, avec une « différence statistiquement significative » (selon la terminologie utilisée dans ce système de notation), à * lorsque l'établissement figure parmi les 10% plus mauvais, avec une « différence statistiquement significative ». Lorsque les résultats sont plus ou moins égaux à ceux attendus ou que la différence n'est pas statistiquement significative, la note attribuée est ***. Un établissement peut, malgré un bon score, n'avoir que la note ***. Pas sûr que le « consommateur » comprenne tout.

Comme l'affirme le document d'accompagnement des programmes de seconde, *« l'esprit statistique naît lorsque l'on prend conscience de l'existence de fluctuations d'échantillonnage »*. A la suite de nombreuses simulations, les élèves de seconde peuvent un peu quantifier les fluctuations d'échantillonnage dans le cas d'une fréquence : si l'on tire (avec remise) n boules dans une urne contenant une proportion p de boules rouges, la fréquence f des boules rouges dans les n tirages est, dans plus de 95% des cas, comprise dans l'intervalle $[\, p - \frac{1}{\sqrt{n}} \,,\ p + \frac{1}{\sqrt{n}}\,]$.

En revanche, une justification, qui serait fondée sur le théorème limite central, est hors de portée des élèves de lycée. En fait, ce n'est pas tant la justification mathématique du résultat qui pose problème (les simulations sont assez convaincantes), que sa compréhension approfondie, nécessaire pour les utilisations ultérieures. La connaissance des fluctuations d'échantillonnage permet en effet à la méthode statistique de se déployer dans deux directions, celle de l'estimation et celle des tests.

Sondages et estimation

Le citoyen est, depuis quelques décennies, inondé de sondages. L'utilité d'une éducation aux sondages est évidente. Dans le domaine politique, en France, cela a débuté en décembre 1965. Contre tous les observateurs, l'Ifop avait prévu le ballottage du général *De Gaulle*. *« C'est un triomphe pour l'Ifop, sinon pour le général »*, déclarait alors un ministre en privé. Cependant, lors des trois dernières grandes élections en France (présidentielles 1995, législatives 1997, présidentielles 2002), les instituts de sondage se sont en grande partie « trompés ».

En 1995, *Edouard Balladur* n'est pas au second tour. Réaction immédiate de *Nicolas Sarkozy* : *« C'est une formidable défaite pour les instituts de sondage »*.

En 2002, c'est la présence de *Jean-Marie Le Pen* au second tour qui n'était guère prévue. Le 30 avril 2002, l'agence *Reuter* rapporte la dépêche suivante : *« Le président du tribunal de Paris a rejeté une demande présentée par* *neuf journalistes et particuliers qui souhaitaient contraindre les quatre principaux instituts de sondage (Ipsos, CSA, Ifop et BVA) à mentionner obligatoirement les méthodes de correction utilisées, les modes de calcul et les marges d'erreurs.*

« Le matraquage de sondages faux au premier tour de l'élection présidentielle a eu pour résultat un abstentionnisme massif et les électeurs ont subi un préjudice », avait plaidé lors de l'audience l'avocat des demandeurs. Le juge a cependant estimé que « la responsabilité des manquements allégués ne saurait incomber aux instituts de sondage, dès lors que seuls les organes d'information assument la mission, et la responsabilité qui y est afférente, de publier les résultats. »

[…] *Les demandeurs devront payer 3000 euros de frais de procédure. »*

Les seules données que les instituts de sondage sont actuellement contraints de mentionner sont les dates de leur réalisation, la taille de l'échantillon et la méthode générale. En revanche, les instituts (et à plus forte raison les médias) ne précisent jamais que les résultats bruts après enquête sont systématiquement « redressés » par des méthodes de pondération, faisant appel aux résultats des scrutins antérieurs, puis « lissés » lorsque les résultats redressés paraissent trop fantaisistes.

Remarquons qu'à l'occasion des élections régionales de 2004, les fourchettes de sondages (même « sortis des urnes ») ont refait une apparition en force à la télévision (prudence enfin raisonnable).

Si les réactions sont parfois si vives à l'égard des sondages, c'est en effet souvent parce qu'on voudrait les voir affirmer ce qu'ils ne disent pas. Tout ce que l'on peut inférer d'un échantillon (très partiel) c'est un intervalle « de confiance » dans lequel on estime que se situe la réalité, avec un pourcentage de fiabilité (le coefficient de confiance) quant au taux de réussite de cette procédure.

L'élection présidentielle de 2002

Considérons de plus près l'exemple de l'élection présidentielle de 2002. Lors du premier tour, le dernier sondage publié par l'institut B.V.A. , effectué sur 1000 électeurs le vendredi 19/04/02, prévoyait :

Jacques Chirac	19 %
Lionel Jospin	18 %
Jean-Marie Le Pen	14 %

Les résultats du vote du 21/04/02 sont :

Jacques Chirac	19,88 %
Lionel Jospin	16,18 %
Jean-Marie Le Pen	16,86 %

Doit-on considérer que le sondage était « faux » ? S'il s'agissait de prévoir l'ordre des trois candidats, la réponse est affirmative, mais c'est prêter à la méthode statistique plus de pouvoir qu'elle n'en possède. Il convient, en revanche, tenant compte de la variabilité inhérente à la méthode du sondage, de définir sous forme de « fourchettes » la précision que l'on est en « droit » d'attendre du sondage.

Compte-tenu de la variabilité observée sur les échantillons de taille n, on convient, dans le thème « sondages » de seconde, que dans 95% des cas, la fréquence p inconnue se situe dans la fourchette $[\, f - \frac{1}{\sqrt{n}}\ ,\ f + \frac{1}{\sqrt{n}}\,]$ où f est la fréquence observée sur l'échantillon aléatoire de taille n. Calculons, à partir du sondage BVA, ces fourchettes à 95% de confiance. On obtient :

$$[\ 0{,}19 - \frac{1}{\sqrt{1000}}\ ;\ 0{,}19 + \frac{1}{\sqrt{1000}}\ ] \approx$$

[15,84% ; 22,16%] pour *Jacques Chirac*,

$$[\ 0{,}18 - \frac{1}{\sqrt{1000}}\ ;\ 0{,}18 + \frac{1}{\sqrt{1000}}\ ] \approx$$

[14,84% ; 21,16%] pour *Lionel Jospin*,

$$[\ 0{,}14 - \frac{1}{\sqrt{1000}}\ ;\ 0{,}14 + \frac{1}{\sqrt{1000}}\ ] \approx$$

[10,84% ; 17,16%] pour *Jean-Marie Le Pen*.

On constate que ces fourchettes ont une intersection commune non vide, donc que le sondage ne permet pas de prévoir l'ordre des candidats. De plus, chacune des valeurs réelles est effectivement contenue dans la fourchette à 95% qui lui correspond. Donc les fourchettes du sondage sont exactes. Compte-tenu de résultats somme toutes serrés, la taille $n = 1000$ est insuffisante. Bien entendu, une évolution de l'opinion dans les deux jours séparant le dernier sondage du vote n'est également pas à exclure. Toujours est-il que l'étude des fluctuations d'échantillonnage doit inciter le futur citoyen à la prudence et à l'esprit critique.

En devoir en classe (seconde), nous avons posé le petit exercice suivant :

> Un journaliste affirme à la télévision : « Ce mois-ci le président passe de 49% à 51% d'opinions favorables. Il franchit ainsi la barre fatidique des 50% d'opinions favorables. » Les résultats du sondage s'affichent à l'écran et on lit, en dessous, « sondage effectué sur un échantillon aléatoire de 1000 personnes ». Que pensez-vous du commentaire du journaliste (justifier) ?

Une bonne moitié des élèves ayant répondu à cette question fournit une réponse plus ou moins correcte (le « plus ou moins » provenant de la qualité de l'expression en français). Parmi ces réponses d'élève figure la suivante : *« Le journaliste ne peut pas tellement dire qu'il franchit la barre des 50% car la fourchette à* 95% *est* [0,49 ; 0,54] *».* Au moins un qui a compris et c'est un des plus mauvais en maths. C'est un des avantages de ce chapitre : peu de « prérequis » sont nécessaires. (Echantillon de $n = 27$ élèves, insuffisant pour une statistique, mais suffisant au quotidien).

Après l'introduction des probabilités en première, on peut aborder, en terminale ES ou S, l'utilisation de la méthode statistique pour un test d'adéquation à un modèle probabiliste.

Modélisation et tests

De la connaissance des fluctuations des échantillons aléatoires, la méthode statistique infère les qualités vraisemblables de la population dont ces échantillons sont issus, aidant ainsi à la prise de décision. Dans les procédures d'estimation, type fourchette de sondage, on n'a pas d'a priori sur les qualités que devrait posséder cette population. Bien souvent, au contraire, on a à l'esprit une norme que devrait vérifier la population (dans les contrôle de qualité par exemple). On procède alors à un test statistique. On formule une hypothèse (appelée « hypothèse nulle ») dont on suppose qu'elle est vérifiée par la population. Il s'agit de voir si les résultats observés sur l'échantillon sont, aux fluctuations près, compatibles avec cette hypothèse, ou si une différence significative avec les résultats attendus rend improbable l'hypothèse nulle. Le test retenu par les programmes de terminale ES et S est celui de l'adéquation à une loi équirépartie, dont l'exemple type est le test d'un dé. La question est de « savoir » si un dé est truqué d'après l'observation d'une centaine de lancers, par exemple. L'hypothèse nulle est celle d'un modèle équidistribué (chaque face apparaît avec une probabilité 1/6). La connaissance des fluctuations des résultats de 100 lancers selon ce modèle permet de déterminer des limites, au delà desquelles il est peu vraisemblable d'avoir affaire à un dé non truqué. Bien sûr, quand on dit « savoir », ce n'est pas une connaissance à 100% (pour cela il ne faut pas utiliser une méthode statistique mais passer le dé aux rayons X par exemple), mais l'intérêt de la méthode consiste à évaluer le risque d'erreur. Cependant si le risque de rejeter à tort l'hypothèse nulle est facile à quantifier, celui d'acceptation à tort de l'hypothèse nulle est moins évident, comme on le précise par la suite.

Encore une fois, tout n'est pas aussi simple qu'il y paraît à première vue. Raison de plus pour éduquer nos futurs citoyen à cette notion de risque lors d'une prise de décision selon une procédure statistique. Ces procédures sont en effet de plus en plus utilisées et la « gestion des risques » est un fait de société. Par ailleurs, la procédure du test statistique est souvent liée à celle de la validation d'un modèle et donc à la modélisation. Les scientifiques modélisent beaucoup, mais nombre de décisions sur des enjeux de société se prennent également sur la base de modélisations. Voici quelques exemples simples (voire simplifiés) pouvant en grande partie faire l'objet d'exercices en terminale (la fréquente interdisciplinarité des situations fait que la statistique inférentielle trouvera sa place dans nombre de projets type T.P.E.).

Un exemple pour comprendre : le Q.C.M.

L'expérience montre qu'un excellent exemple pour faire comprendre à nos élèves la notion de risque lors d'un test statistique est celui où le test en question est un examen scolaire (on s'appuie sur le vécu). La présentation suivante utilise la loi binomiale, qui est connue en terminale S.

Un professeur construit un Q.C.M. de 20 questions indépendantes, proposant 3 réponses possibles à chaque question, une seule réponse étant exacte. Il souhaite par cet examen éliminer environ 95% des élèves n'ayant rien appris et répondant au hasard. La détermination du nombre de bonnes réponses pour être admis à l'examen revient à la construction d'un test statistique. On désigne par p la probabilité qu'un élève donné a de bien répondre à une question portant sur le programme de ce Q.C.M.

L'hypothèse nulle H_0 est « $p = 1/3$ » (l'élève répond au hasard). L'hypothèse alternative sera « $p > 1/3$ » (lorsque l'élève ne répond pas au hasard).

Sous l'hypothèse nulle, le nombre de bonnes réponses correspond à la réalisation d'une variable aléatoire suivant la loi binomiale de paramètres $n = 20$ et $p = 1/3$ (répétition de 20 épreuves aléatoires indépendantes à deux issues possibles, succès ou échec, avec une probabilité de succès 1/3). Or la distribution de cette loi (cf. fig. 3 page suivante) montre que la probabilité d'avoir strictement moins de 11 réponses exactes est environ 96%. Le professeur décidera donc que pour être reçu à l'examen, on doit avoir au moins 11 bonnes réponses. Le risque de rejeter à tort l'hypothèse H_0 est environ 4%. Ce risque correspond au fait de recevoir à l'examen un élève qui répond au hasard. C'est un risque que les élèves sont prêts à prendre ! (en fait, c'est le risque du professeur).

Les élèves perçoivent bien qu'il existe un autre risque : celui d'échouer à l'examen alors

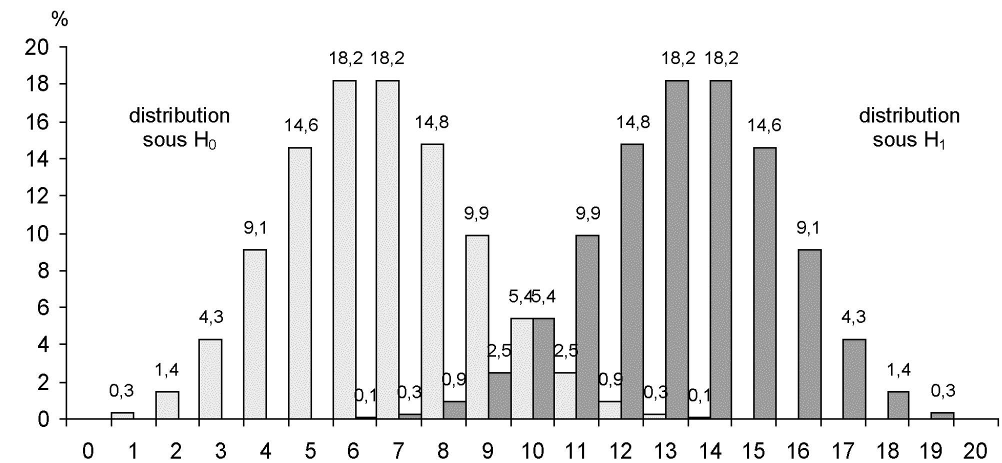

Fig. 3 : Distributions binomiales $n = 20$ $p = 1/3$ et $p = 2/3$

qu'on ne le mérite pas (la faute à pas de chance). C'est le risque consistant à accepter à tort l'hypothèse nulle H_0 .

Ce second risque est plus difficile à évaluer, puisqu'il dépend des connaissances du candidat. Pour mieux comprendre, calculons ce second risque dans le cas où l'hypothèse alternative H_1 correspond à une distribution de probabilité précise. Supposons que le candidat ait une probabilité p = 2/3 de répondre correctement à chaque question. Pour ce candidat, le second risque correspond à la probabilité d'avoir strictement moins de 11 succès selon la loi binomiale de paramètres n = 20 et p = 2/3, soit un risque d'être recalé à tort d'environ 9% (voir fig. 3).

On peut avoir le sentiment que dans le calcul de la « barre d'acceptation », on suspecte a priori le candidat d'être coupable de répondre au hasard. Cette impression est exacte. Dans la construction d'un test, l'hypothèse nulle est privilégiée dans la mesure où on ne la rejettera que si les observations sont vraiment trop peu compatibles.

Le réchauffement de la planète

Les données ci-contre, concernant la température moyenne du globe (un sujet « sensible »), sont celles du Groupe d'experts intergouvernemental sur l'évolution du climat, dépendant des Nations Unies (*Intergouvenmental Panel on Climate Change*, dont le site est www.ipcc.ch). Elles sont disponibles sur le site du *Climatic Research Unit* en Grande-Bretagne (www.cru.uea.ac.uk/cru/data/temperature).

Si l'on considère les 40 années les plus chaudes sur la période 1861-2003 , une seule apparaît avant 1930, en 1878 (sur la figure 4 l'échelle des ordonnées est en degrés Celsius, le 0 correspondant à une climatologie moyenne sur la période 1961-1990). Il n'est guère utile de procéder à un test statistique pour savoir si la répartition des 40 années les

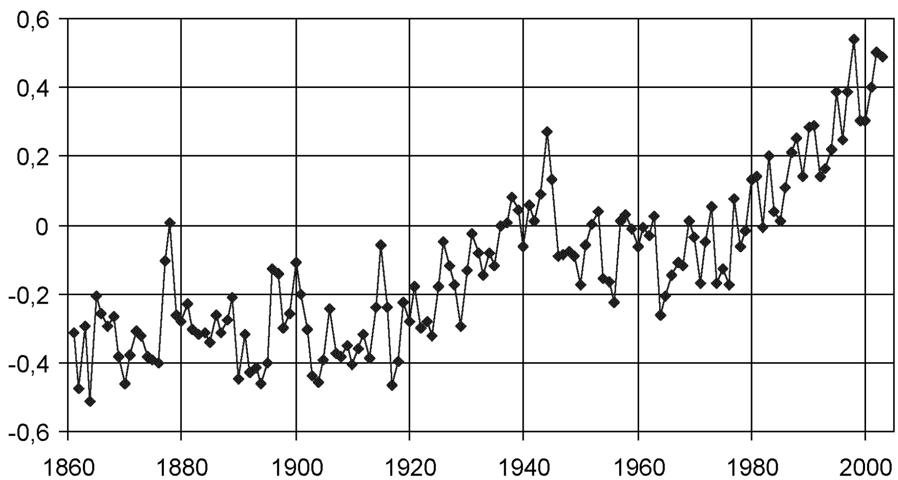

Fig. 4 : Températures annuelles globales de 1861 à 2003.
L'échelle est en degrés Celsius, le 0 correspondant à une climatologie moyenne sur la période 1961-1990.

plus chaudes peut être considérée comme totalement aléatoire sur la période 1861-2003. Cependant, comme la figure 4 suggère un changement qualitatif à partir des années 1920-1930, nous allons étudier s'il y a une accélération significative sur la deuxième période, en utilisant l'adéquation à un modèle équiréparti aux programmes de terminale ES et S (qui est un peu l'analogue du test du khi deux en utilisant la simulation au lieu de la loi de probabilité du khi deux.).

On choisit d'étudier la période [1924, 2003] et de la partager en quatre intervalles d'égale longueur (20 années) dans lesquels on considère la distribution des années les plus chaudes (il y en a 39).

Supposer que la répartition des 39 années les plus chaudes sur cette période de 80 ans se fait « au hasard » revient à dire que l'on observe ici le résultat d'un tirage simultané de 39 boules dans une urne contenant 80 boules

Classe n° i	Années	Eff. observés x_i	Fréq. observées f_i
1	[1924 , 1943]	6	0,154
2	[1944 , 1963]	7	0,179
3	[1964 , 1983]	6	0,154
4	[1984 , 2003]	20	0,513

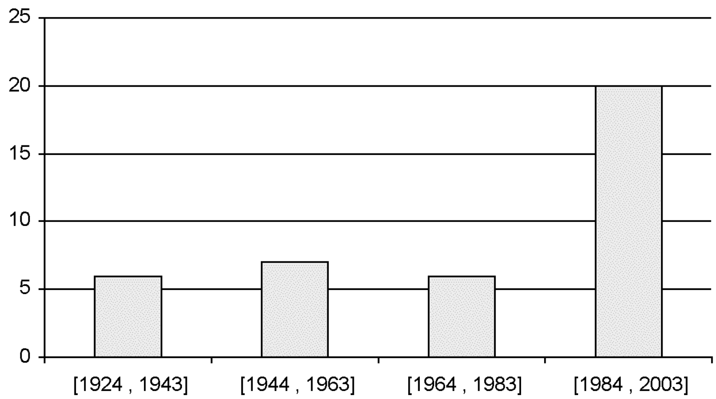

Fig. 5 : Répartition des 39 années les plus chaudes.

marquées du millésime de chaque année. De façon plus précise, utilisant quatre couleurs, on aurait peint chaque boule de la couleur associée à la période de 20 ans à laquelle elle correspond. Supposer que les années les plus chaudes sont le fruit du hasard implique que l'on prenne simultanément 39 boules dans cette urne quadricolore équirépartie. Etudions l'hypothèse selon laquelle la distribution de ces 39 années s'effectue au hasard selon un modèle équiréparti.

L'écart entre les fréquences observées et la fréquence 1/4 théorique peut être quantifié par $d_{obs}^2 = \sum_{i=1}^{4} (f_i - \frac{1}{4})^2$. On observe $100\times d_{obs}^2 \approx 9{,}3$.

On simule, sous l'hypothèse d'équirépartition, la distribution de 39 années entre 4 classes (tirages simultanés dans l'urne équirépartie) puis on calcule la valeur $100\times d_{obs}^2$ obtenue sur cette simulation. On peut effectuer 1000 simulations pour avoir un bon aperçu des fluctuations de $100\times d_{obs}^2$ sous l'hypothèse d'équirépartition (le tirage sans remise est simulé à l'aide d'une macro-commande). Les simulations montrent qu'environ 99% des valeurs de $100\times d_{obs}^2$ obtenues sous l'hypothèse d'équirépartition sont inférieures à 4.

Au risque de 1%, on rejette donc le modèle équiréparti pour expliquer la répartition des 39 années les plus chaudes dans les quatre classes. La différence observée est « significative ».

Telle est la présentation que l'on peut faire en terminale. Si au lieu d'un tirage simultané (sans remise) dans une urne équirépartie, on fait tourner une roue de loterie partagée en quatre quartiers égaux (ce qui équivaut à un tirage avec remise), on peut montrer (mais c'est hors programme de terminale) que la variable aléatoire correspondant aux réalisations de $4\times 39\times d_{obs}^2$ sous l'hypothèse d'équirépartition suit approximativement la loi du khi-deux à 3 degrés de liberté. On vérifie alors que la probabilité d'obtenir $100\times d_{obs}^2 \approx 9{,}3$ sous cette hypothèse est 0,002.

Cela signifie que la répartition (6, 7, 6, 20) observée avec les années les plus chaudes (fig. 5) est très peu vraisemblable comme issue d'un tirage avec remise dans une urne équirépartie, et encore moins vraisemblable comme provenant d'un tirage simultané (les écarts des tirages simultanés avec l'équirépartition ont tendance à être moindres que ceux des tirages avec remise). Cela laisse moins de 0,2% de chances à une explication totalement aléatoire de la répartition des années les plus chaudes sur la période 1924-2003. Il reste bien peu de place au hasard.

Un test fréquemment pratiqué est celui de comparaison de deux moyennes (« test de différence significative »). Celui-ci faisant intervenir la loi normale, il ne sera vu des élèves qu'après le bac (dans bien des domaines, en particulier la médecine). On constatera que le raisonnement est analogue. Effectuons ce test sur les moyennes des températures observées avant et après 1930 (cf. fig. 6).

Les observations fournissent les résultats suivants :

Années	moyenne	écart type	effectif
[1861,1929]	– 0,299	0,11	69
[1930,2003]	0,045	0,18	74

Faisons l'hypothèse nulle selon laquelle les deux échantillons de taille 69 et 74 sont extraits aléatoirement de populations ayant la même moyenne. Le théorème limite central montre que la variable aléatoire D faisant correspondre à deux tels tirages la différence d des moyennes observées suit approximativement une loi normale de moyenne nulle et d'écart type $\sqrt{\frac{0{,}11^2}{69}+\frac{0{,}18^2}{74}}$.

Or cette distribution de probabilité est telle que dans 99% des cas D prend une valeur comprise entre –0,065 et 0,065. La différence observée étant 0,344 on rejette l'hypothèse nulle

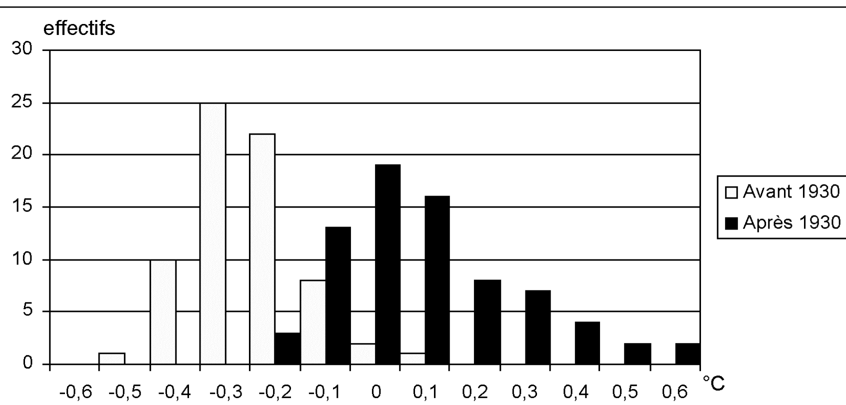

Fig. 6

au risque de 1% (en fait le risque est bien plus faible). La différence est bien significative.

Une modélisation de crues

Il est des domaines à risques où la modélisation est un enjeu décisionnel important. Celui des crues des fleuves fait fréquemment l'actualité. Les données de la figure 7 concernent l'Oise à Auvers (source : Ville d'Auvers/Oise et Direction Régionale de l'Environnement). Nous allons modéliser la hauteur du fleuve de deux façons, en introduisant une variable aléatoire à densité continue, notion figurant au programme de terminale S. Un test permettra d'évaluer la pertinence de ces modèles. Il est important de comprendre qu'il n'y a pas de vérité en la matière, un modèle unique, mais seulement des ajustements au mieux des observations, selon des critères choisis a priori.

Soit X la variable aléatoire qui, à une année prise au hasard, fait correspondre la hauteur maximale annuelle en mètres de l'Oise au pont d'Auvers. On suppose que X est une variable aléatoire continue de densité f, c'est-à-dire que la probabilité d'une hauteur maximale annuelle inférieure à x mètres est donnée par $P(X \leq x) = \int_0^x f(t)\,\mathrm{d}t$.

On dispose des observations suivantes :

Evénement x_i	Nb. d'observations sur l'historique de 86 années	Fréq. f_i observée
$X < 2{,}70$	68	0,791
$2{,}70 \leq X < 3{,}80$	16	0,186
$X \geq 3{,}80$	2	0,023

On cherche à coller au mieux à ces données, avec à notre disposition une liste de densités de probabilité.

Modèle 1 : Les élèves connaissent la loi exponentielle, vue dans le cadre de la désinté-

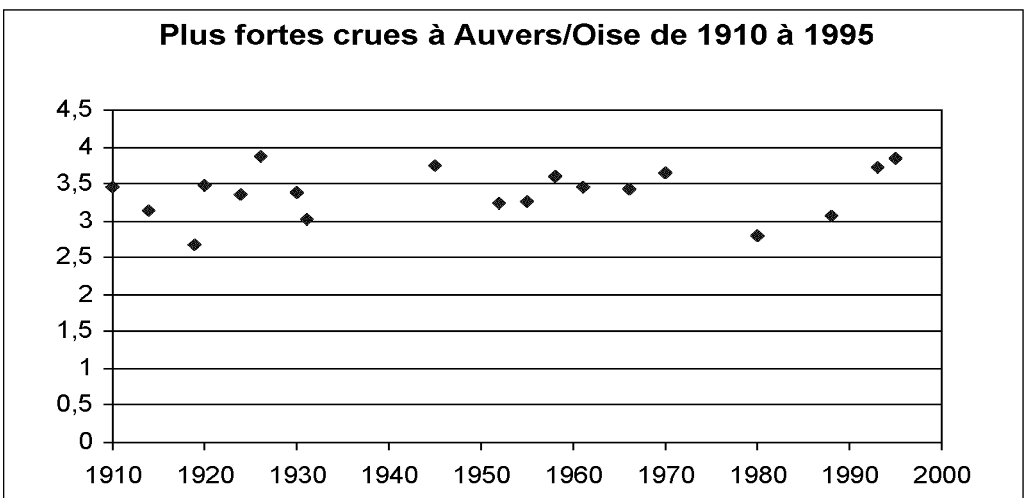

Fig. 7

gration nucléaire, correspondant à des phénomènes « sans mémoire » (où le futur est indépendant du passé) et qui s'applique assez bien à de telles situations, comme aux tremblements de terre, éruptions volcaniques (nos exemples ne sont guère réjouissants)…

On prend alors : $f(t) = 0{,}84\ \mathrm{e}^{-0{,}84\,t}$ et $P(X \le x) = 1 - \mathrm{e}^{-0{,}84\,t}$.

La valeur 0,84 du paramètre a été choisie de sorte à minimiser l'écart quadratique réduit $e = \sum_{i=1}^{3} \frac{(x_i - t_i)^2}{t_i}$ entre les effectifs observés x_i et les effectifs théoriques calculés t_i, ce qui peut se réaliser aisément sur un tableur (avec l'outil « Valeur cible » ou le « Solveur » d'Excel par exemple).

Modèle 2 : On peut rechercher une autre densité de probabilité, plus sophistiquée mais mieux à même de coller à la réalité. On prendra $f(t) = 2kt\ \mathrm{e}^{-k\,t^2}$ avec $k > 0$ (nommée loi de Rayleigh, mais peu importe).

Un petit calcul intégral (ce thème permet de fructueux changements de cadres…) convaincra les élèves qu'il s'agit bien d'une densité de probabilité, et permettra d'obtenir l'expression de $P(X \le x)$:

$$P(X \le x) = \int_0^x 2kt\ \mathrm{e}^{-k\,t^2}\ \mathrm{d}t = 1 - \mathrm{e}^{-k\,x^2}$$

et $\lim_{x \to +\infty} \int_0^x 2kt\ \mathrm{e}^{-k\,t^2}\ \mathrm{d}t = 1$.

On choisit $k = 0{,}23$ (même critère que pour le modèle 1) et $f(t) = 0{,}46\ t\ \mathrm{e}^{-0{,}23\ t^2}$. L'adéquation des observations à ces modèles est visualisée sur la figure 8. L'ajustement se mesure à partir du tableau situé en haut de la page suivante.

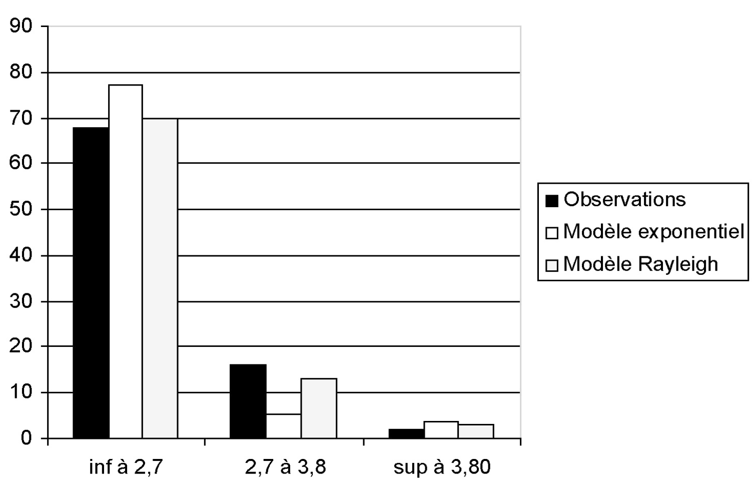

Fig. 8

i	Evènement i	Probabilité p_i selon le modèle 1	Effectif théorique $t_i = 86\, p_i$ selon le modèle 1	Probabilité p_i selon le modèle 2	Effectif théorique $t_i = 86\, p_i$ selon le modèle 2	Effectif x_i observé
1	$X < 2{,}70$	0,896	77	0,813	70	68
2	$2{,}7 \leq X < 3{,}8$	0,062	5	0,151	13	16
3	$X \geq 3{,}80$	0,041	4	0,036	3	2

A vue d'œil, il est immédiat que le second modèle est meilleur. Il existe cependant des différences entre les deux modèles et les observations. C'est peut être normal, voire même souhaitable, dans la mesure où les observations concernent un échantillon de taille n = 86 années. La mise en œuvre d'un test statistique a l'avantage de quantifier le risque d'erreur lorsque l'on rejette un modèle (mais pas quand on l'accepte). Procédons par simulation, un raisonnement a priori accessible aux élèves de terminale.

Faisons l'hypothèse que les crues se déroulent selon le modèle 1 (puis même chose avec le modèle 2). Simulons 86 réalisations selon ce modèle. Cela consiste à faire tourner 86 fois une roue de loterie comprenant trois secteurs de proportions 89,6% ; 6,2% et 4,1%. A défaut de tourner un nombre invraisemblable de fois une telle roue, on peut utiliser le tableur, ou fournir aux élèves les résultats de simulations.

Cette roue virtuelle peut se réaliser sur un tableur comme Excel à l'aide des instructions =ENT(1+86*ALEA()) et =SI(ref<78;1;SI(ref<83;2;3)) où ref correspond à la référence de la cellule du calcul précédent. On calcule l'écart entre les effectifs x_i obtenus par simulation et effectifs théoriques prévus par le modèle 1 sous la forme :

$$e = \frac{(x_1 - 77)^2}{77} + \frac{(x_2 - 5)^2}{5} + \frac{(x_3 - 4)^2}{4}.$$

On effectue ensuite 1000 simulations (par exemple) pour se faire une opinion des fluctuations naturelles de ces écarts, sous l'hypothèse du modèle 1.

Dans 95% des simulations sous l'hypothèse du modèle 1, l'écart e est inférieur à 6. Il en est de même pour les simulations avec la roue du modèle 2.

L'écart entre les observations de l'historique et le modèle 1 est :

$$e_{obs} = \frac{(68 - 77)^2}{77} + \frac{(16 - 5)^2}{5} + \frac{(2 - 4)^2}{4} \approx 26 > 6.$$

L'écart observé est très supérieur à celui fourni par le modèle 1 dans les simulations (dans 95 % des cas inférieur à 6). Le modèle 1 est donc rejeté au risque de 5% (et même beaucoup moins).

L'écart entre les observations de l'historique et le modèle 2 est :

$$e_{obs} = \frac{(68 - 70)^2}{70} + \frac{(16 - 13)^2}{13} + \frac{(2 - 3)^2}{3} \approx 1 < 6.$$

Le modèle 2 peut raisonnablement être retenu, au seuil de 5% (il ne s'agit pas ici du risque associé à cette décision). Les simulations montrent de plus que dans environ 50% des cas, la roue du modèle 2 (tournée 86 fois) donne un l'écart *e* supérieur à 1, qui est celui des observations. Reste que Vincent Van Gogh s'en moque, le cimetière d'Auvers est situé sur le coteau (ce n'est pas le cas des importantes installations électriques à proximité du pont).

Conclusion

Nous espérons, par les exemples précédents (libre à vous d'en exploiter d'autres, grâce en particulier aux ressources d'Internet), avoir montré l'intérêt de l'enseignement des méthodes de statistique inférentielle pour l'éducation de nos futurs citoyens. Vous aurez noté au passage qu'il s'agit de situations motivantes (il y a bien d'autres catastrophes possibles à envisager, mais aussi des exemples plus positifs, dans l'étude de traitements médicaux par exemple) où l'on exploite des outils mathématiques parfois variés, ainsi que les possibilités de l'informatique. Cependant, la compréhension de la méthode statistique (où l'on n'est sûr de rien) nécessiterait un apprentissage assez précoce (dès la seconde) et pas trop superficiel, ainsi qu'une formation conséquente des professeurs. Ce n'est pas le cas (réductions d'horaires, diminution des budgets de formation…), ce qui risque (non quantifié) de compromettre « la bonne intention », dans ce domaine, des nouveaux programmes de lycée.

Bibliographie

CALLON (Michel) – *Agir dans un monde incertain – Essai sur la démocratie technique* – Le Seuil 2001.

GIRARD (Jean-Claude) et HENRY (Michel) – *L'inférence statistique. Deux exemples d'application du calcul des probabilités : estimations et tests d'hypothèses – Enseigner les probabilités au lycée* Commission inter-IREM Statistique et Probabilité 1997.

HACKING (Ian) – *L'ouverture au probable, éléments de logique inductive* – Armand Colin 2004.

IREM PARIS-NORD – *Le nouveau programme de Statistique et Probabilité au lycée*– Villetaneuse 2003.

LOMBORG (Bjorn) – *L'écologiste sceptique* – Le Cherche Midi 2004..

PIEDNOIR (Jean-Louis) - DUTARTE (Philippe) – *Enseigner la Statistique au lycée : des enjeux aux méthodes* – IREM PARIS-NORD 2001.

REEVES (Hubert) – *Mal de Terre* – Seuil 2003.

ROBERT (Claudine) – *Contes et décomptes de la statistique. Une initiation par l'exemple* – Ed. Vuibert 2003.

SAPORTA (Gilbert) – *Probabilités, analyse des données et statistiques* –Ed. Technip.

VERLANT (Bernard) - SAINT-PIERRE (Geneviève) – *Statistique et probabilités - BTS* – Ed. Foucher.

WONNACOTT et WONNACOTT – *Statistique* – Ed. Economica.

Le Web pour les données statistiques.