Du bon usage d'un intervalle de fluctuation

Philippe Dutarte(*)

« L'éducation à l'aléatoire [...] devrait avoir pour but fondamental la prise de conscience que toute décision s'accompagne d'un risque, mais que ce risque peut être évalué. »
Norbert Meusnier – *Sur l'histoire de l'enseignement des probabilités et des statistiques* dans *Histoire de probabilités et de statistiques* – Ellipse 2004.

Introduction

Les programmes récents de mathématiques pour le lycée général et technologique font apparaître une notion « d'intervalle de fluctuation », dans le cas de l'observation d'une fréquence. Il s'agit d'une notion liée à des problématiques statistiques et, de ce fait, dépendante des contextes d'utilisation. Plusieurs « définitions » en sont possibles, mais, pour l'enseignement secondaire, il peut être souhaitable, d'une part d'en limiter le nombre, d'autre part d'assurer la cohérence avec les usages dans l'enseignement supérieur, personne n'imaginant que le point de vue adopté dans le programme de terminale puisse être assez différent de celui envisagé, le cas échéant, l'année suivante. Nous rejoignons l'avis porté ci-dessus par Norbert Meusnier, selon qui l'essentiel est d'initier à l'évaluation du risque dans la prise de décision.
Trois formes d'intervalles de fluctuation apparaissent dans les programmes (on note p la proportion d'individus présentant le caractère auquel on s'intéresse dans la population et n la taille de l'échantillon aléatoire prélevé dans cette population) :
– en seconde générale et technologique (2009) :

$$\left[p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}}\right],$$

au seuil de 95% sous certaines conditions ;
– en première, notamment S et ES (2011) : « déterminé à l'aide de la loi binomiale, pour rejeter ou non une hypothèse sur une proportion » (le programme ne précise pas la forme de l'intervalle mais limite explicitement les capacités attendues au contexte de la prise de décision) ;
– en terminale, notamment S et ES (2012) :

$$\left[p-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}},p+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right],$$

intervalle de fluctuation « asymptotique », ici au seuil de 95%.

Nous ferons un assez long détour, sans doute instructif, par l'histoire de ces notions, puis envisagerons des situations d'utilisation au lycée de ce concept d'intervalle de

(*) IA-IPR de mathématiques, philippe.dutarte@ac-creteil.fr

fluctuation, dont l'introduction en classe doit se faire d'abord par des exemples (nous verrons pourquoi), avant d'être ensuite formalisé.

1. Ce que nous apprend l'histoire

Dans les situations nous occupant ici, on s'intéresse à la région où se manifeste « habituellement » la variabilité de la fréquence observée sur un échantillon, dans le but d'inférer à la population entière des informations recueillies sur un échantillon. Formellement, cela peut s'effectuer de deux façons, soit en utilisant « directement » un intervalle de fluctuation dans le cadre d'un « test d'hypothèse » (on procède ainsi lorsque l'on a de bonnes raisons de raisonner à partir d'une hypothèse sur la valeur de p, l'intervalle de fluctuation apparaissant alors comme la région d'acceptation de cette hypothèse à un certain seuil), soit en utilisant un processus consistant à « retourner » l'intervalle de fluctuation, pour obtenir un « intervalle de confiance[(1)] » pour p, calculé à partir de la fréquence f observée (on procède ainsi lorsque l'on n'a pas d'*a priori* sur la valeur de p). Ces notions n'ont été formalisées et comprises, sans toutefois faire consensus, qu'à partir des années trente … du vingtième siècle. Il n'est pas certain que plonger les élèves de lycée dans les tâtonnements historiques précédant cette époque soit directement productif, mais, pour le professeur du moins, c'est très instructif des difficultés que présentent ces notions dans leur utilisation à bon escient[(2)]. On pourra à ce sujet se reporter à l'article de B. Parzysz dans le dossier du Bulletin n° 502.

L'une des premières manifestations du raisonnement « par l'absurde » présent dans les tests d'hypothèse, lors du rejet de celle-ci, se rencontre dans l'œuvre du médecin écossais John Arbuthnot[(3)] : « *An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes* » (1710). Sous une hypothèse de pur hasard, il est quasiment impossible d'atteindre et de dépasser l'écart observé. Ce qui discrédite l'hypothèse initiale. Arbuthnot constate qu'à Londres, sur 82 années consécutives, il naît chaque année davantage de garçons que de filles. Sous l'hypothèse de l'équiprobabilité des sexes, la probabilité d'observer un tel événement, $\left(\frac{1}{2}\right)^{82} \simeq 2.10^{-25}$, est si faible que cette hypothèse doit être rejetée. Arbuthnot en concluait :

> « cette inégalité des mâles et des femelles n'est pas l'effet du Hasard mais celui de la Providence Divine. »

Michel Armatte souligne[(4)] qu'au XVIII^{e} siècle, des « raisonnements en forme de syllogisme aléatoire » se retrouvent dans les discussions sur l'efficacité de l'inoculation, sur des contrats aléatoires juridiques ou économiques, ou sur des conjectures scientifiques. Il affirme que « la systématisation de ces raisonnements,

(1) Le point de vue de l'intervalle de confiance dépasse le cadre de cet article.
(2) À propos du rôle de l'histoire dans l'enseignement de la statistique, on pourra consulter avec profit l'article de Michel ARMATTE [2].
(3) Cité par Michel ARMATTE dans [1] p. 124.
(4) Voir [1].

armée du calcul analytique des probabilités, triomphe avec les travaux de Pierre Simon Laplace. ». Un exemple remarquable est celui où, en 1812, Laplace se demande si les comètes sont des éléments du système solaire ou viennent d'un autre système(5). Le critère observé est celui de l'angle que font les planètes du système solaire et les comètes avec l'écliptique (le plan dans lequel la Terre tourne autour du Soleil). L'hypothèse « testée », si l'on peut s'exprimer ainsi, est celle d'une inclinaison pour chaque objet résultant d'une loi uniforme(6) entre 0 et 100 : « si toutes les inclinaisons étaient également possibles » selon l'expression de Laplace. Laplace considère, d'une part, l'inclinaison moyenne des 10 planètes du système solaire au commencement de 1801 qui vaut environ 9,14 grades, et, d'autre part, celle des 100 comètes identifiées jusqu'en 1811, qui vaut 51,7 grades environ, à comparer à une moyenne de 50 grades selon la loi uniforme. Pour le cas des 10 planètes, Laplace calcule la probabilité d'observer, selon l'hypothèse d'uniformité, un résultat inférieur ou égal à celui obtenu et conclut que l'événement complémentaire :

> « s'approche tellement de la certitude que le résultat observé devient invraisemblable dans cette hypothèse ; ce résultat indique donc, avec une très grande probabilité l'existence d'une cause primitive qui a déterminé les mouvements des planètes à se rapprocher du plan de l'écliptique [...](7) ».

Dans le cas des 100 comètes, sous l'hypothèse d'une loi uniforme, « la probabilité que l'inclinaison moyenne [soit] inférieure à l'inclinaison observée » est, selon Laplace, 0,737. Ce résultat n'est pas jugé assez grand pour rejeter l'hypothèse de loi uniforme.

Des fondateurs : R.A. Fisher, E. Pearson, J. Neyman

Les tests statistiques couramment utilisés de nos jours reposent essentiellement sur deux théories : celle des tests de « signification » de Ronald Aymler Fisher (1890-1962) et celle des « tests d'hypothèse », intégrant une théorie de la décision, d'Egon Pearson (1895-1980) et de Jerzy Neyman (1894-1981). Il s'agit de points de vues relativement différents et de vives controverses ont opposé leurs auteurs (et se poursuivent en partie actuellement).

Fisher s'adresse essentiellement aux chercheurs et voit les tests comme un moyen d'apprendre à partir des données expérimentales. Ses ouvrages, *Statistical Methods for Research Workers* (première édition 1925) et *The Design of Experiments* (première édition 1935) connurent un succès considérable. La procédure de test utilisée par Fisher est la suivante :

– fixer une hypothèse nulle de type « sans effet » ;
– calculer, sous cette hypothèse, la probabilité P d'obtenir un résultat au moins aussi extrême que celui observé (cette probabilité conditionnelle est nommée

(5) Jean-Jacques Droesbeke et Philippe Tassi signalent dans [4], que Jerzy Neyman mentionne, qu'à sa connaissance, la première approche d'un test d'hypothèse est due à Laplace, à propos du cas des comètes.
(6) Laplace utilise le grade (centième partie du quart de cercle) comme unité de mesure d'angle.
(7) Cité par Michel Armatte dans [1]. Comme on le sait, Laplace s'affranchit de l'hypothèse (alternative) d'une Divine providence.

« niveau de signification », « valeur critique » ou « *p*-valeur » correspondant à l'échantillon) ;

– considérer que plus P est faible, moins l'hypothèse nulle est crédible.

Dans cette procédure, l'hypothèse nulle est l'hypothèse à réfuter (celle qui n'intéresse pas le chercheur). Le résultat du test est soit le rejet de l'hypothèse nulle (si P est assez faible), soit la suspension du jugement. Il n'y a donc qu'un seul type d'erreur : celle de rejeter l'hypothèse nulle alors qu'elle est vraie.

En 1926, Egon Pearson commence à interroger Neyman sur la problématique des tests. Son interrogation porte sur ce que l'on peut dire lorsque les données sont compatibles avec la valeur critique de Fisher (lorsqu'on ne rejette pas l'hypothèse nulle). Il s'agit alors du cas particulier de l'adéquation de données observées avec une répartition normale. Que dire si la différence avec le modèle normal n'est pas « significative » ? Il n'y a pas de raison de rejeter le modèle normal, mais quelles sont les raisons de l'accepter ? Cette question fondamentale suscite une très riche collaboration entre Neyman et Pearson jusqu'en 1933, l'un en Pologne, l'autre à Londres. Certaines idées initiales viennent de Pearson : la correspondance Pearson-Gosset, six mois avant la rencontre de Neyman, montre que Pearson avait déjà en tête les idées d'hypothèse alternative et des différents risques d'erreur, peut-être en partie suggérées par Gosset. En revanche, Pearson reconnaîtra que Neyman apporta des « fondations mathématiques à ses vagues idées ».

À la différence de la théorie de Fisher, où l'on détermine la *p*-valeur à partir de l'échantillon, celle des tests d'hypothèse d'Egon Pearson et Neyman ne part pas de l'observation des données et est davantage élaborée à partir de la vision fréquentiste de la probabilité. Dans les tests de Pearson et Neyman, on affecte une valeur *a priori* au paramètre (c'est l'hypothèse « nulle ») et on évalue les risques de rejet d'une hypothèse nulle vraie et, c'est plus délicat, d'acceptation d'une hypothèse nulle fausse. Cette valeur *a priori* du paramètre est, par exemple, la norme d'une production industrielle ou est fournie par un modèle. Selon les valeurs observées, on considèrera que l'hypothèse selon laquelle cette norme ou ce modèle est respecté, est, ou non, acceptée.

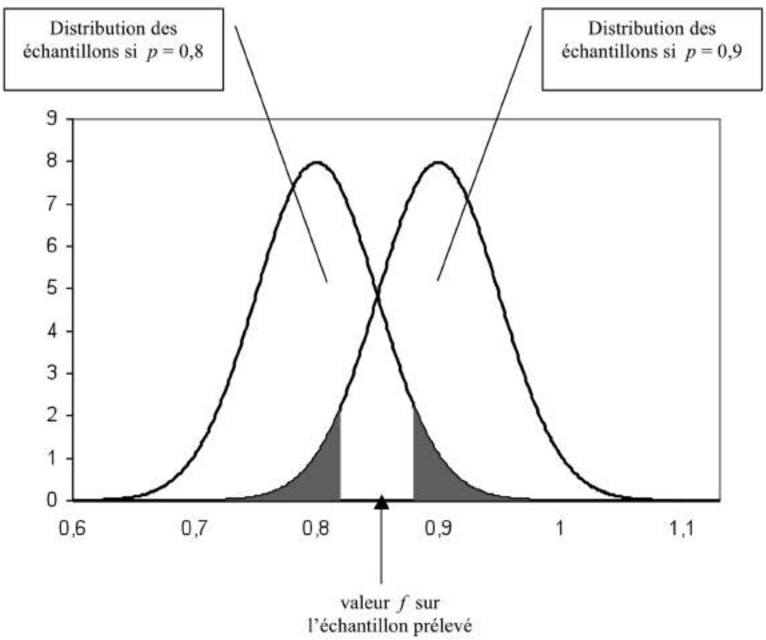
De façon générale, E. Pearson et Neyman mettent en évidence en 1928 le rôle dissymétrique joué par l'hypothèse nulle H_0, que l'on privilégie en quelque sorte, et par l'hypothèse H_1 qu'on lui oppose. La décision prise, rejet ou acceptation de l'hypothèse, dépend des choix faits *a priori* par l'utilisateur ou plutôt les utilisateurs. C'est-à-dire que pour la même situation, deux tests différents peuvent bien entendu conduire à deux décisions différentes. Prenons un exemple : un laboratoire affirme, qu'à la différence de ses concurrents, dont le produit est efficace à 80 %, le sien l'est à 90 %. Deux tests sont envisageables selon le point de vue.

– Test fabricant :

On choisit comme hypothèse nulle $p = 0{,}9$ et l'on construit un test dit « unilatéral à gauche », c'est-à-dire que l'on ne rejette l'hypothèse nulle selon laquelle le nouveau produit est meilleur ($p = 0{,}9$) que si le résultat de l'échantillon est « significativement » inférieur à 0,9. Ce test privilégie l'hypothèse $p = 0{,}9$ que l'on ne rejette que si elle est vraiment intenable. C'est l'intérêt du fabricant.

– Test client :
On choisit comme hypothèse nulle $p = 0{,}8$ et l'on construit un test dit « unilatéral à droite », c'est-à-dire que l'on ne rejette l'hypothèse nulle $p = 0{,}8$ que si le résultat de l'échantillon est « significativement » supérieur à 0,8.
Le client privilégie l'hypothèse $p = 0{,}8$. Il n'est prêt à changer d'avis que si les résultats de l'échantillon sont significativement meilleurs.

Supposons que l'on ait représenté sur le graphique suivant les distributions d'échantillonnage selon que $p = 0{,}8$ ou que $p = 0{,}9$ ainsi que les zones de rejet de l'hypothèse nulle pour chacun des tests, à gauche pour le test « fabricant » et à droite pour le test « client ».



Supposons que l'échantillon prélevé amène la valeur $f = 0{,}85$. La décision de chaque test est de conserver son hypothèse nulle. D'où des décisions contraires, le fabriquant considère que $p = 0{,}9$ et le client que $p = 0{,}8$.
Ceci peut donner une impression de désordre ou faire penser que l'on peut faire dire ce que l'on veut aux chiffres. En fait, cela signifie simplement que la procédure de test doit faire l'objet d'un consensus préalable à la prise d'échantillon, en connaissance de cause, c'est à dire en sachant quels sont les risques d'erreurs de décision[8].

(8) La situation est analogue lorsque la zone d'acceptation de l'hypothèse correspond à un intervalle de fluctuation bilatéral au programme des classes de lycée. On voit là le rôle formateur que le professeur de mathématiques a à jouer par rapport à une utilisation « presse-bouton » des logiciels.

En 1933, E. Pearson et Neyman introduisent la distinction entre erreur de « première espèce » (rejet à tort de H_0) et de « seconde espèce » (acceptation à tort de H_0). Tous ces concepts ne vont pas de soi et n'ont d'ailleurs pas fait immédiatement l'unanimité, mais sont essentiels lors d'une prise de décision et, semble-t-il, incontournables. Fisher critiqua la formulation de Neyman-Pearson dès sa conception et jusqu'à sa mort. Il écrira par exemple(9) :

> « Neyman, pensant qu'il corrigeait et améliorait mon propre premier travail sur les tests de signification, comme un moyen de " l'amélioration de la connaissance naturelle ", le réinterprétait en fait en termes de cet appareil technologique et commercial connu comme étant une procédure d'acceptation. Maintenant, les procédures d'acceptation sont d'une grande importance dans le monde moderne. Quand une grande maison comme la Royal Navy reçoit du matériel d'une entreprise industrielle, il est, je suppose, soumis à une inspection suffisamment minutieuse qui teste afin de réduire la fréquence d'acceptation des pièces mauvaises ou défectueuses… Mais, la différence logique entre une telle opération et le travail de découverte scientifique par expérimentation physique ou biologique me semble si large que l'analogie avec elle n'est pas utile, et l'identification des deux sortes d'opération résolument trompeuse. »

2. Prise de décision dans un contexte bilatéral

Un exemple

L'exercice suivant, où il est question de variable aléatoire, peut être proposé au niveau première ou terminale.

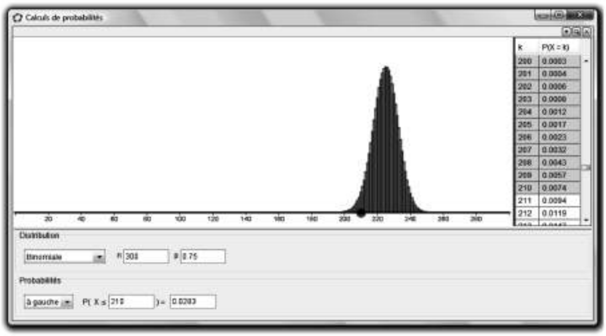
> « On considère la couleur des yeux des mouches drosophiles. Par croisement de mouches homozygotes « yeux rouges » de gènes AA et de mouches homozygotes « yeux bruns » aa, on obtient des mouches hétérozygotes Aa. Si l'on croise ces hétérozygotes entre eux, on doit obtenir, selon les lois de Mendel, dans cette seconde génération, 75% de « yeux rouges » (AA et Aa) et 25% de « yeux bruns » (aa).
> On souhaite tester l'hypothèse selon laquelle la proportion de drosophiles « yeux rouges » de seconde génération est $p = 0{,}75$ en mettant en place une expérimentation permettant d'observer 300 drosophiles de seconde génération (considérées comme un échantillon aléatoire).
> 1. Sous l'hypothèse $p = 0{,}75$, déterminer l'intervalle de fluctuation au seuil de 95% de la variable aléatoire correspondant à la fréquence du caractère « yeux rouges » sur un échantillon aléatoire de taille 300.
> 2. Énoncer la règle de décision permettant de rejeter, ou non, l'hypothèse $p = 0{,}75$, au seuil de 5%, sur un échantillon aléatoire de taille 300.
> 3. L'expérience permet d'observer 237 « yeux rouges » et 63 « yeux bruns ». Cette répartition est-elle conforme à la loi de Mendel, au seuil de décision de 5% ?

Détaillons le raisonnement que l'on peut suivre(10).

1. On considère que les 300 drosophiles qui seront observées résultent d'un tirage au hasard et avec remise dans la population des drosophiles de seconde génération. Dans ce cadre, la variable aléatoire X correspondant au nombre de drosophiles présentant le caractère « yeux rouges » dans un tel échantillon suit, sous l'hypothèse $p = 0{,}75$, la loi binomiale de paramètres $n = 300$ et $p = 0{,}75$. Il est donc possible, sous

(9) Cité par David SALSBURG dans [9].
(10) Il ne s'agit pas de la rédaction attendue d'un élève de première ou de terminale.

l'hypothèse $p = 0{,}75$, de prévoir la « variabilité » de la variable aléatoire $F = \dfrac{X}{300}$ correspondant à la fréquence du caractère « yeux rouges » sur un tel échantillon. Compte tenu de la nature du problème (on accepte l'hypothèse $p = 0{,}75$ à condition que la fréquence observée ne s'en éloigne pas trop, ni à gauche, ni à droite), on recherche un intervalle de fluctuation « bilatéral » (c'est-à-dire avec une zone de rejet à gauche et une zone de rejet à droite) correspondant à une probabilité d'au moins 95%. Ceci de sorte que le risque(11) de rejeter à tort cette hypothèse soit d'au plus 5% et même, ici, d'au plus 2,5% de chaque côté. On est donc amené à rechercher les entiers a et b tels que : $P(X < a) \leq 0{,}025$; $P(a \leq X \leq b) \geq 0{,}95$ et $P(X > b) \leq 0{,}025$(12). Le logiciel GeoGebra, dans sa version 4, permet de déterminer aisément (et de manière visuelle) ces valeurs a et b pour lesquelles les probabilités « à gauche » et « à droite » dépassent le seuil de 0,025. On trouve $a = 210$ et $b = 239$.



En divisant par 300, on obtient l'intervalle de fluctuation au seuil de 95% de F : $I = [0{,}7 \,;\, 0{,}797]$.

En classe de terminale, on peut préférer utiliser l'intervalle de fluctuation asymptotique, déterminé à partir de la loi normale, et qui fournit un résultat extrêmement proche : $[0{,}701 \,;\, 0{,}799]$. L'intervalle de fluctuation donné en seconde, majoration du précédent, est $[0{,}69 \,;\, 0{,}81]$.

2. Après prélèvement d'un échantillon aléatoire de taille 300 et calcul de la fréquence f du caractère « yeux rouges » sur cet échantillon, la règle de décision, au seuil de 5%, est la suivante : si f n'appartient pas à l'intervalle I de fluctuation à 95%

(11) L'apport de Pearson et Neyman nous invite à porter le regard sur le risque (ici de première espèce) de sorte que l'intervalle de fluctuation se détermine surtout par considération de son complémentaire. N'importe quel intervalle de fluctuation ne fait pas l'affaire et, d'une certaine façon, il y a ici unicité de la réponse à apporter dans la recherche de l'intervalle de fluctuation adapté au problème.

(12) C'est l'intervalle de fluctuation ainsi défini qui permet d'obtenir la notion d'intervalle de confiance d'une proportion, en quelque sorte « exact », obtenu à l'aide de la loi binomiale et que l'on pourrait définir de la façon suivante (cette notion est bien entendu hors programme au lycée) :

On considère une variable aléatoire X de loi binomiale $\mathcal{B}(n,p)$ et $F = X/n$. À partir d'une réalisation f de F, l'intervalle de confiance de p au niveau de confiance de 95% est l'intervalle $[p_1, p_2]$ tel que, si $p < p_1$ alors $P(F \geq f) > 0{,}025$, et si $p > p_2$ alors $P(F \leq f) > 0{,}025$.

précédent, on rejette l'hypothèse $p = 0{,}75$, si f appartient à I, on ne rejette pas cette hypothèse(13).

Il est important de savoir que les 5%(14) correspondent à la probabilité de rejeter à tort l'hypothèse $p = 0{,}75$ (c'est comme cela qu'a été construite la procédure de décision).

3. On observe $f = 0{,}79$. Cette valeur appartient à l'intervalle I de fluctuation à 95%. Cette observation est donc « conforme à » (ou « compatible avec ») l'hypothèse $p = 0{,}75$ au seuil de décision de 5%. On s'exprime ainsi car, dans cette situation, il serait un peu étrange d'affirmer qu'on « ne rejette pas la loi de Mendel ». En cas de rejet de l'hypothèse, on serait sans doute amené à étudier les conditions du protocole expérimental.

Les trois types d'intervalles de fluctuation présents dans les programmes des classes du lycée général et technologique et leur mise en œuvre

Les concepteurs des programmes ont fait le choix de restreindre le domaine d'investigation de la statistique inférentielle au lycée à celui de l'étude de la proportion d'un caractère dans une population. L'objectif de cet enseignement dans un cours de mathématiques n'est pas, en effet, d'accumuler les techniques, mais de comprendre le sens des procédures mises en œuvre. Pour ce qui concerne la prise de décision et l'estimation par intervalle de confiance, on limite les attendus de la part des élèves, aux situations bilatérales. Dans cet esprit (mieux vaut en faire peu mais le mieux possible), trois formes d'« intervalles de fluctuation » sont aux programmes du lycée. Il est sage de s'y tenir.

D'une certaine façon, le modèle adapté à la situation (en supposant la population suffisamment grande pour considérer l'échantillonnage comme résultant d'un tirage avec remise) est celui de la loi binomiale, accessible en classes de premières S, ES, STI2D-STL et STMG. Dans le cadre bilatéral de la prise de décision, l'intervalle de fluctuation (binomial) à 95% doit être défini de la façon suivante(15) (après avoir considéré quelques exemples « concrets » justifiant cette définition(16)) :

> On considère une variable aléatoire X de loi binomiale de paramètres n et p et $F = X/n$. Soit a le plus grand entier tel que $P(X < a) \leq 0{,}025$ et b le plus petit entier tel que $P(X > b) \leq 0{,}025$. L'intervalle $[a/n, b/n]$ est nommé intervalle de fluctuation à 95% de F.(17)

(13) On peut faire une distinction entre « on ne rejette pas l'hypothèse » et « on accepte l'hypothèse ». Dans le cadre d'une prise de décision, le non rejet revient, cependant, à une acceptation de l'hypothèse. En revanche, le risque d'erreur en cas d'acceptation (probabilité – conditionnelle – d'accepter l'hypothèse alors qu'elle est fausse) est plus difficile à estimer puisqu'il dépend de la valeur du paramètre au sein de l'hypothèse alternative (risque de seconde espèce).

(14) Avec la loi binomiale, ce n'est pas exactement 5% (on peut déterminer la probabilité correspondant à la zone de rejet).

(15) Ou de toute autre manière équivalente. Le document ressources [7] donne ainsi une définition « algorithmique » très opérationnelle : on recherche les plus petits entiers a et b tels que $P(X \leq a) > 0{,}025$ et $P(X \geq b) \geq 0{,}975$.

(16) On notera dans cette définition le rôle des zones de rejet et du risque de 5% (maximum) qui leur correspond.

(17) On constate immédiatement que F appartient à cet intervalle avec une probabilité au moins égale à 0,95.

En terminale, on dispose de la loi normale, permettant de remplacer dans certains cas (on donne généralement $n \geq 30$, $np \geq 5$ et $n(1 - p) \geq 5$) l'intervalle précédent par l'intervalle de fluctuation asymptotique qui lui est proche. À une époque, pas si lointaine, cela permettait d'éviter des calculs inutilement pénibles avec la loi binomiale (ce qui est beaucoup moins vrai aujourd'hui). Reste que l'expression obtenue, rappelée au début de cet article, a l'avantage de fournir une formulation explicite des bornes de l'intervalle en fonction de la taille n de l'échantillon. Par ailleurs, l'obtention d'un intervalle de confiance est moins difficile à expliciter à partir de l'intervalle de fluctuation asymptotique, qu'elle ne l'est à partir de la définition de l'intervalle de fluctuation binomial.
En seconde générale et technologique, ne disposant pas de l'outil de la loi binomiale, une majoration[(18)] de l'intervalle de fluctuation asymptotique (bilatéral) à 95% est admise et expérimentée par simulation.
La mise en œuvre d'un intervalle de fluctuation, pour la prise de décision, devrait répondre, au lycée, à un certain nombre de critères, garantissant le respect de l'esprit de la démarche statistique, susceptible, selon les orientations choisies, d'être poursuivie dans l'enseignement supérieur[(19)].
– Le premier critère nous semble devoir être de poser un problème, une situation, ayant quelque écho avec le monde réel. Cette situation doit être telle que la proportion p dans la population est supposée connue : c'est l'hypothèse de départ, avec laquelle on construit la règle de décision. Cela exclut les situations où l'on n'a aucun *a priori* sur la valeur de p (type « sondage »), pour lesquelles on aura recourt à un intervalle de confiance[(20)].
– Il faut ensuite que cette situation soit de type « bilatérale » : par rapport à la valeur de p « visée » (celle de l'hypothèse) on s'intéresse à un écart trop important à gauche et à droite.
– La règle de décision doit être élaborée, autant que possible[(21)], avant la prise d'échantillon. Il est plus honnête de décider d'une règle avant de jouer, qu'après la partie. Par ailleurs, l'approche de Neyman-Pearson est fréquentiste. Il ne s'agit pas d'élaborer une règle de décision à partir d'un échantillon, mais d'élaborer une règle de décision en amont dont on sait évaluer les risques sur un grand nombre d'échantillons.
– L'échantillon doit être prélevé par tirage au hasard avec remise (équiprobabilité garantie par randomisation).

(18) $1{,}96\sqrt{p(1-p)}$ est majoré par 1.

(19) Satisfaire tous ces critères dans un exercice scolaire de lycée n'est pas évident. Le premier (une situation-problème) est cependant incontournable : le choix d'une méthode statistique dépend de la situation.
(20) L'estimation par intervalle de confiance est une démarche inductive à l'état d'esprit très différent de celui d'un test. On n'y retrouve pas la démarche déductive de type « raisonnement par l'absurde » à l'œuvre dans la prise de décision correspondant aux tests, en cas de rejet de l'hypothèse. Il faut absolument éviter de mener les deux démarches pour une même situation.
(21) Dans certains cas, voir par exemple les cas de leucémies à Woburn (document ressources pour le lycée professionnel), on ne peut procéder ainsi. Le raisonnement consiste à supposer que l'échantillon étudié résulte d'un échantillonnage aléatoire sous une certaine hypothèse.

– En cas de rejet de l'hypothèse, il faut savoir que l'on peut interpréter le 5% (ou le 1% en terminale S dans le cas d'un intervalle de fluctuation asymptotique à 99%[22]) comme la probabilité de commettre une erreur de décision (probabilité de rejet de l'hypothèse sachant qu'elle est vraie). En cas d'acceptation de l'hypothèse, il faut savoir que l'estimation de la probabilité d'erreur de décision est plus compliquée[23].

3. Représentativité d'un échantillon

Le document ressources de terminale[24] indique un champ d'application intéressant des intervalles de fluctuation, correspondant à la notion d'échantillon représentatif. Il y est indiqué que cette notion « ne constitue en aucun cas un objectif du programme » (aucun attendu de la part des élèves sur ce point) mais permet de « contextualiser » utilement les activités.

L'exemple présenté est celui de l'estimation de la proportion de personnes en surpoids dans une ville V. Pour cette étude, on prélève un échantillon aléatoire de 460 personnes. On sait par ailleurs que certains critères peuvent influer sur le surpoids et, notamment, l'âge. On souhaite ainsi que l'échantillon prélevé soit « représentatif » de la proportion $p = 0{,}2$ des personnes âgées de plus de 60 ans dans la ville V au sens que la fréquence f correspondante sur l'échantillon devra appartenir à l'intervalle de fluctuation à 95% (condition nécessaire pour retenir l'échantillon).

En utilisant l'intervalle de fluctuation asymptotique à 95%, on obtient [0,16 ; 0,24]. Sur l'échantillon prélevé, on constate que $f = 0{,}234$. On conserve donc cet échantillon, considéré comme « représentatif » pour ce critère d'âge.

Le contexte est intéressant, correspond à une pratique statistique réellement mise en œuvre, dans une situation bilatérale avec un échantillonnage aléatoire. En revanche, il ne s'agit pas de la situation « habituelle » d'une prise de décision de type « test statistique ». On ne fait pas d'hypothèse sur p puisque l'on connaît (ici par le recensement) la valeur de p. On ne commet donc pas d'erreur d'aucune espèce. On sait que dans 5% des cas, on rejettera l'échantillon, parce qu'on n'a pas eu de chance et que l'on craint que la sous ou surreprésentation de ce critère n'influe sur l'étude menée par la suite.

4. Prise de décision dans un contexte unilatéral

Concernant une proportion, nombre de situations de prise de décision sont, de

(22) Le choix du seuil de décision (1% ou 5%) doit, bien entendu, se faire avant la prise d'échantillon. Il dépend du risque (de première espèce) consenti (si l'on diminue le risque de première espèce, on augmente le risque de seconde espèce, à taille d'échantillon égale). Ce choix dépend des utilisateurs et non du statisticien.

(23) Les notions d'erreur de première et de seconde espèces ne sont pas au programme au lycée (aucune autonomie des élèves à ce sujet n'est attendue, ni même la connaissance de ce vocabulaire). Pour autant, il nous semble inconcevable de passer cette problématique (il y a deux types d'erreurs et donc de risques) complètement sous silence, tant elle est essentielle. Une analogie simple suffit à faire comprendre la situation. Une prise de décision est comme un jugement au tribunal. L'hypothèse est que le prévenu est présumé innocent. Il y a deux risques au jugement : celui de condamner un innocent (rejet à tort de l'hypothèse, première espèce), ou d'absoudre un coupable (acceptation à tort de l'hypothèse, seconde espèce).

(24) Voir [8] page 31.

manière plus « naturelle », unilatérales plutôt que bilatérales. Il nous semble dommage de s'interdire l'accès à ces situations, tout en étant conscient qu'aucune autonomie des élèves n'est attendue à cet égard. Il ne s'agit pas pour autant de taire le problème et de faire, dans une situation annoncée comme unilatérale, « comme si de rien n'était ».

Dans le cadre de la loi binomiale, un exemple est fourni dans le document ressources pour la classe de première (seuil de radioactivité(25)).

Nous reprenons ici un exemple présenté dans le document ressources pour la classe de terminale(26), en insérant la question 2, essentielle à la compréhension du « 95% ».

> On admet que dans la population d'enfants de 11 à 14 ans d'un département français le pourcentage d'enfants ayant déjà eu une crise d'asthme dans leur vie est de 13%. Un médecin d'une ville de ce département est surpris du nombre important d'enfants le consultant ayant des crises d'asthme et en informe les services sanitaires. Ceux–ci décident d'entreprendre une étude et d'évaluer la proportion d'enfants de 11 à 14 ans ayant déjà eu des crises d'asthme.
>
> Ils sélectionnent de manière aléatoire 100 jeunes de 11 à 14 ans de la ville.
>
> La règle de décision prise est la suivante : si la proportion observée est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 95% alors une investigation plus complète sera mise en place afin de rechercher les facteurs de risque pouvant expliquer cette proportion élevée.
>
> 1. Déterminer l'intervalle de fluctuation asymptotique au seuil de 95% de la proportion de jeunes de 11 à 14 ans ayant eu une crise d'asthme dans un échantillon de taille 100.
> 2. Indiquer une valeur approchée de la probabilité de mener une investigation supplémentaire à tort (c'est-à-dire alors que la proportion d'enfants de 11 à 14 ans de la ville V ayant déjà eu une crise d'asthme dans leur vie est de 13%).
> 3. L'étude réalisée auprès des 100 personnes a dénombré 19 jeunes ayant déjà eu des crises d'asthme. Que pouvez-vous conclure ?

1. On trouve [0,06 ; 0,20]. Remarquons qu'il faut lire attentivement l'énoncé pour comprendre qu'il y a deux populations. Celle du département, où $p = 0{,}13$ est connu, et celle de la ville V, où l'on fait en quelque sorte, l'hypothèse que $p = 0{,}13$.

2. La problématique de cet exercice est unilatérale, ce qui est clairement affirmé dans la règle de décision. En revanche, l'intervalle de fluctuation asymptotique au seuil de 95% est bilatéral (c'est le seul au programme de terminale). On ne peut donc répondre « 5% » à la question posée. En revanche, un élève de terminale connaît la propriété de symétrie de la courbe de Gauss et sait donc que ces 5% (correspondant aux fréquences observées situées en dehors de l'intervalle de fluctuation lorsque $p = 0{,}13$) se partagent en 2,5% de chaque côté de l'intervalle de fluctuation. La probabilité demandée est donc d'environ(27) 2,5%.

Il est essentiel, et simple, de comprendre que le risque consenti de mener une investigation supplémentaire inutile (et éventuellement coûteuse) est de 2,5%.

(25) Voir [7] page 71. Il s'agit bien d'une situation unilatérale : le « risque » se situe pour un seuil de radioactivité jugé élevé (il est rare que l'on s'inquiète d'une radioactivité anormalement basse).

(26) Voir [8] page 22.

(27) Le « environ » provient de l'aspect asymptotique de l'intervalle de fluctuation. Pour un résultat « exact », travailler avec la loi binomiale.

3. La valeur 0,19 est à l'intérieur de l'intervalle de fluctuation asymptotique au seuil de 95%, on en conclut que la règle de décision choisie ne prévoit pas de réaliser une enquête supplémentaire(28).

Conclusion

L'enseignement au lycée de la notion d'intervalle de fluctuation est essentiel, tout d'abord dans le cadre de la prise de décision, ensuite comme préliminaire à l'investigation de la notion d'intervalle de confiance. Il faut faire comprendre « que toute décision s'accompagne d'un risque, mais que ce risque peut être évalué ». Dans cet enseignement, l'apport principal du professeur de mathématiques doit être de dégager la logique qui préside à la démarche statistique, les raisonnements mis en œuvre. On pourrait confier les détails des calculs à un logiciel (ou programmer un algorithme), là n'est pas l'essentiel. Partir de situations-problèmes « concrètes » est indispensable pour mettre en perspective les principes introduits, comprendre les définitions et les précautions d'utilisation (rôle de la zone de rejet, distinction entre la règle de décision et la prise d'échantillon, signification des pourcentages 95%, 5%, …). Il ne s'agit pas d'élaborer des théories mathématiques « à vide », dans un cadre purement probabiliste : à l'arbitraire de cette introduction s'ajouterait l'incapacité d'appliquer à bon escient ces concepts. Ce n'est pas là l'objectif de l'enseignement de la statistique inférentielle en cours de mathématiques. Ces notions ont été forgées pour répondre à des problèmes concrets. Connaître quelques repères historiques, particulièrement ceux de leur élaboration aux débuts du XXe siècle, permet à l'enseignant de mieux comprendre les difficultés d'utilisation, les différents points de vue et l'importance des choix didactiques à effectuer.

Références

[1] ARMATTE, Michel, *Contribution à l'histoire des tests Laplaciens* dans *Mathématiques et sciences humaines* nº 176 – 2006, http://www.ehess.fr/revue-msh/pdf/N176R1254.pdf

[2] ARMATTE, Michel, *Le rôle de l'histoire dans l'enseignement de la statistique*, Revue électronique *Statistique et enseignement*, SFdS 2010, www.statistique-et-enseignement.fr.

[3] BERNOULLI, Jakob (1713), *Ars Conjectandi*, 4ème partie. Traduit du latin par Norbert MEUSNIER dans *Jacques Bernoulli et l'Ars conjectandi*, IREM de Rouen, 1987.

(28) On peut faire remarquer que dans cette situation d'acceptation de l'hypothèse, on ne sait pas, en l'état, quantifier le risque. Cela peut être prétexte à prolongement de l'exercice. Supposons par exemple que la proportion d'enfants de cet âge asthmatiques dans la ville V (connue comme souffrant de pollution de l'air) soit en fait de 25%, quelle est la probabilité que la règle de décision précédente conduise à accepter (à tort) l'hypothèse $p = 0{,}13$? En terminale, il s'agit d'une question « ouverte » : on peut effectuer des simulations, utiliser la loi binomiale ou calculer $P(F \leq 0{,}20)$, où F suit la loi normale de moyenne 0,25 et d'écart type $\sqrt{\frac{0{,}25 \times 0{,}75}{100}}$.

[4] Droesbeke, Jean-Jacques, Tassi, Philippe, *Histoire de la statistique* – Que-sais-je ?, PUF 1997.

[5] Henry, Michel, *La démonstration par Jacques Bernoulli de son théorème dans Histoire de probabilités et de statistiques*, Ellipse, 2004.
Voir aussi : http://www.apmep.asso.fr/IMG/pdf/Asm10_Atelier_M_Henry.pdf

[6] Meusnier, Norbert, *Sur l'histoire de l'enseignement des probabilités et des statistiques* dans *Histoire de probabilités et de statistiques*, Ellipse, 2004.

[7] Ressources pour la classe de première générale et technologique – *Statistiques et probabilités*, DGESCO 2011.

[8] Ressources pour la classe terminale générale et technologique – *Probabilités et statistique*, DGESCO 2012.
Les documents ressources sont accessibles par :
http://eduscol.education.fr/cid45766/mathematiques-pour-le-college-et-le-lycee.html

[9] Salsbourg, David, *The lady tasting tea, How statistics revolutionized science in the twentieth century*, First Owl Books New York 2002.