

Statistiques inférentielles : un débat scientifique en classe de seconde

Jean-Marie Parnaudeau(*)

Initier à la méthode statistique n'est pas toujours facile. Évaluer ce qui en est retenu est probablement plus compliqué.

Les propos ci-dessous visent à décrire, non pas, comme c'est souvent le cas, une activité ou une réflexion sur les programmes, mais une évaluation en statistique. Évaluation qui a été suivie, non pas d'un corrigé, mais d'une analyse de quelques-unes des réponses, consistant, à partir des réponses fournies par les élèves, à engager une discussion et essayer d'élaborer une réponse raisonnable à la question posée. Cette analyse s'est déroulée sous la forme d'un débat, en salle d'informatique, les réponses étant vidéo-projetées. La règle étant : en sciences, lorsque l'on a une certaine expérience d'un sujet ou lorsque l'on dispose d'une théorie, on ne peut pas dire n'importe quoi.

L'évaluation en statistique sur tableur, proposée à une classe de seconde, comportait cinq questions. Les quatre premières questions étaient destinées à vérifier si l'élève était capable de mettre en place une simulation dans un cas suffisamment explicite, par exemple, lancer trois pièces équilibrées⁽¹⁾ (100 fois de suite) et compter le nombre de faces obtenus à chaque fois, puis résumer les résultats dans un tableau ; ou bien effectuer 50 lancers d'un dé régulier et visualiser, sur un graphique, le résultat de chaque lancer et la fréquence cumulée du 5. Dans ces quatre premières questions, il ne s'agissait pas, pour moi, d'évaluer quelque connaissance que ce soit en statistique, mais plutôt d'évaluer la maîtrise de quelques fonctionnalités du tableur utiles en statistique. Seule la cinquième question était une question de statistique, c'est seulement de cette cinquième question dont je veux parler ici.

Mais revenons en arrière. L'enseignement de statistique inférentielle, dans cette classe de seconde, s'est déroulé en deux phases : une phase expérimentale et une phase simulation.

La phase expérimentale me paraît indispensable pour comprendre la nature des problèmes qui relèvent de la statistique; le type de raisonnement à utiliser et le type de réponses à donner⁽²⁾. Par exemple, organiser un prélèvement aléatoire⁽³⁾, organiser

(*) jeanmarieparnaudeau@gmail.com

(1) La notion de « pièce équilibrée » est un modèle mathématique, que l'on devrait plutôt nommer « lancer de pièce équilibré » : certes la forme de la pièce intervient, mais aussi la façon dont elle est lancée et son environnement.

(2) « Comprendre ce qu'est une question de statistique et le type de réponse que l'on peut proposer », deuxième version du document d'accompagnement des programmes de seconde du 24/10/1999.

(3) L'expression « prélever au hasard » ou « prélèvement aléatoire » est préférable à « choisir au hasard » ; cette dernière expression est un oxymore.

une succession d'expériences identiques⁽⁴⁾, faire des comptages, estimer un paramètre, ... Mais aussi, et c'est peut être le plus important, sensibiliser à la variabilité des résultats.

La phase simulation (avec un tableur ou un algorithme) permet, comme l'indiquent les programmes, de faire de nombreuses fois une même expérience. L'objectif, à ce niveau, n'est pas de comprendre la théorie, mais de se forger une intuition sur le type de résultats que l'on peut obtenir, sur la variabilité des résultats, bref appréhender la fluctuation d'échantillonnage. Et donc, ensuite, de pouvoir porter un avis raisonné sur une question.

Dans les programmes il est indiqué : « *Faire réfléchir les élèves à la conception et la réalisation d'une simulation*⁽⁵⁾ ». De façon plus globale, il s'agit de concevoir une simulation d'expérience afin de répondre à une question⁽⁶⁾. Une fois cette simulation réalisée, il est nécessaire d'interpréter le résultat. Les deux activités intellectuellement intéressantes (et qui intéressent les élèves) sont la première (imaginer une expérience, un protocole permettant de répondre, en un certain sens, à la question), et la dernière, car c'est l'interprétation qui fait sens.

Sur une feuille du classeur tableur, la question posée, dans une zone de texte, était la suivante :

Une personne possède une pièce de monnaie. Elle pense que cette pièce n'est pas bien équilibrée.

Proposer une méthode (ou une expérience) pour répondre à cette question d'un point de vue statistique.

La réponse sera donnée dans cette zone de texte⁽⁷⁾.

Il n'y a pas de question explicite, c'est volontaire : on suggère que la personne a dû lancer la pièce plusieurs fois pour suspecter qu'elle n'est pas équilibrée.

Bien entendu, les élèves ne connaissent aucun cours qui permette une réponse dite « mathématique ».

Il y a deux objectifs à cette question, le premier est d'évaluer leur rapport à l'aléatoire, le deuxième apparaîtra en fin d'article.

Voici quelques réponses fournies par les élèves (ces réponses ont été mises en conformité avec la syntaxe et l'orthographe en vigueur, mais le sens de la réponse a été préservé).

(4) L'hypothèse de reproductibilité de l'expérience est fondamentale pour la suite (loi binomiale en 1S par exemple).

(5) Extrait du programme de la classe de seconde (BO n° 30 du 23 juillet 2009 page 8).

(6) Pour BACHELARD « Pour un esprit scientifique toute connaissance est une réponse à une question. S'il n'y a pas eu de question il ne peut pas y avoir connaissance scientifique. Rien ne va de soi. Rien n'est donné. Tout est construit. » (*La formation de l'esprit scientifique*). De même, en statistique, s'il n'y a pas de question, aucun élève ne fera de simulation.

(7) Les élèves ont été habitués à écrire les réponses à des questions dans une zone de texte ; cela permet d'éviter la séparation entre le fichier informatique et le cahier, mais aussi les forme à la « culture informatique » au sens où un fichier doit être documenté (comme on le ferait pour un algorithme).

1°) D'abord, certains élèves n'ont pas compris la question.

Mais, nous le savons très bien, quand il y a une question, il faut donner une réponse (les didacticiens diraient que cela fait partie du contrat !).

Par exemple :

Réponse : On pourrait lancer par exemple 10 fois la pièce de monnaie. Si elle tombe sur face 5 fois, on fait le pourcentage. Si elle tombe 5 fois sur pile, on fait de même.

On lance une pièce de monnaie 10 fois et on note le nombre de fois de pile et de face et si elle est équilibrée on a le même pourcentage d'avoir une face que d'avoir une pile (50 % pile et 50 % face).

Réponse : On effectue quatre lancers de cette pièce. Si la fréquence de pile est supérieure à celle des faces, la pièce est truquée de façon à ce que la fréquence de pile soit supérieure à 1/2. Si la fréquence des faces est supérieure, la pièce est truquée de façon à ce que la fréquence de face soit supérieure à 1/2.

Si aucun de ces deux cas n'est réalisé, la pièce est bien équilibrée.

Cette réponse (ainsi que la précédente) a été commentée et nous avons constaté (en utilisant le tableur) que si on lance quatre fois une pièce équilibrée, l'événement⁽⁸⁾ « ne pas avoir deux fois pile » est plus fréquent que l'événement « avoir exactement deux fois pile ». Ce qui en surprend plus d'un ! Ce résultat est d'autant plus flagrant que le nombre de lancers est important.

Les professionnels vont me dire que ce résultat est évident, mais pour un élève de seconde ce n'est pas aussi clair...

Il lui suffit de faire faire une expérience dans laquelle elle calculera le nombre de fois qu'une pièce retombe sur « Pile » et « Face ».

Si le nombre de pile ou de face est trop important, alors cette personne pourra affirmer que la pièce est truquée.

Après avoir constaté que l'expérience proposée est incomplète, le deuxième paragraphe a été commenté. En particulier nous avons retenu que plus l'écart entre le nombre « pile » et le nombre « face » est grand, plus il est probable que la pièce ne soit pas équilibrée (même si cela n'est pas statistiquement correct : en effet un écart de 10 pour 20 lancers n'a pas la même signification qu'un écart de 10 pour 1000 lancers).

2°) Certains élèves n'ont pas compris la question et pensent que, dans la réponse attendue, il faut utiliser le tableur.

Ils proposent donc des réponses de la forme :

Pour prouver que la pièce n'est pas équilibrée, on propose un pourcentage inégal pour pile et face.

Exemple: 20% pile ; 80% face.

*On utilise pour cela la fonction "Si(ALEA()*1>0,2;"face";"pile").*

(8) Même si les élèves ont fait des probabilités en troisième, le mot « événement » est pris ici au sens commun du terme.

Dans ce cas là, la pièce apparaîtra comme déséquilibrée puisqu'elle tombera plus de fois sur face que sur pile.

Ou bien

Pour voir qu'une pièce de monnaie est bien équilibrée, il faut faire une expérience. Grâce à EXCEL, nous pouvons voir si cette pièce est bel et bien équilibrée ou non. Nous procéderons alors, à faire un lancer de x fois de cette pièce, grâce à la fonction : =SI(ALEA() $<$ 0,50;"pile";"face").

Si cette pièce ne respecte pas cette fonction, c'est qu'elle est déséquilibrée, ou tout simplement truquée.

Bon, l'expérience proposée n'a rien à voir avec la question, mais restons positif. Ici l'idée est : si le résultat n'est pas conforme à la règle (« *cette pièce ne respecte pas la fonction* »), alors il y a un problème.

Même si cela ne permet pas de répondre à la question, l'élève a compris que si le résultat s'écarte de la situation « normale », alors il y a un problème. Cette réponse a suscité de nombreuses réactions : que signifie respecter la fonction, que vaut x ? Qu'est-ce qu'on a comme résultat (dans la description, il n'y a aucun comptage) ?

Une analogie : si on respecte le règlement intérieur, on n'a pas de problème ; si on fait des petits écarts, ça peut aller ; mais faut pas pousser le bouchon trop loin. « C'est normal qu'il ait pris trois jours, il a été trop loin... ».

En aparté, l'instruction tableur proposée aux élèves pour simuler un modèle de BERNOULLI de paramètre p ($p = 0,4$) est =Si(ALEA() $<$ 0,4;1;0) et non =ENT(ALEA()+0,4) comme indiqué dans le document ressource. Cette dernière instruction, répond, certes, à la question, mais, de mon point de vue, elle est difficile à comprendre pour un élève de seconde. Le point de vue de l'auteur est celui de BOREL⁽⁹⁾ : « *On sait que les questions de probabilités où interviennent des variables continues ne peuvent acquérir de sens qu'en vertu de **conventions** précises (...) La convention la plus commode, au moins dans le cas où l'ensemble des valeurs possibles de la ou des variables est **borné**, consiste à regarder la probabilité comme proportionnelle à l'**étendue** ».* C'est aussi le point de vue adopté naturellement par les élèves dans la mesure où dire que la probabilité d'obtenir pile est de 0,4, c'est dire que tout se passe comme si on procédait à des prélèvements dans une urne contenant 4 papiers marqués « pile » et 6 marqués « face » ; ce qui, dans le cas de l'utilisation de la fonction ALEA() qui renvoie un nombre « au hasard » entre 0 inclus et 1 exclu, consiste à décider que les 40% premiers correspondent à « pile » et les autres à « face ».

3°) Il faut sûrement utiliser la pièce et faire des lancers...

Réponse : On simule le lancer de la pièce 10 000 fois par exemple et on regarde la fréquence d'apparition de pile et de face et on regarde si on a environ 50% de pile et 50% de face. On peut approfondir et faire plus de lancers pour voir si l'équilibre

(9) *Remarques sur certaines questions de probabilités*, 1905, Bulletin de la Société Mathématique de France n° 33, disponible sur www.numdam.org par exemple. Les mots *conventions*, *borné* et *étendue* sont en italique dans le texte d'Émile BOREL.

se précise. On peut aussi comparer le résultat obtenu à celui d'une pièce non équilibrée.

En classe, nous avons fait des simulations de lancers de 1000, 10 000, 50 000 fois un dé⁽¹⁰⁾. L'élève s'en est inspiré et son idée est que si la pièce est équilibrée, alors « ça va se rapprocher de 50-50 », ce que nous appelons la stabilisation de la fréquence. Mais la réponse est incorrecte, en effet la réponse commence par « On simule le lancer de la pièce » ; or, et c'est pour moi fondamental, la simulation (que ce soit avec un tableur, une calculatrice ou un algorithme) ne peut se faire que si on connaît « l'état du monde », c'est à dire les caractéristiques du lancer de la pièce ; en termes plus savants, la simulation ne peut se faire que si l'on a choisi un modèle.

Réponse : Pour que la personne sache si sa pièce est bien équilibrée, il faudrait qu'il ou elle la lance un grand nombre de fois (1000 fois par exemple) et qu'il (elle) note à chaque fois le résultat (face ou pile). Il fera ensuite la somme des résultats pour chaque événement et si sa pièce est bien équilibrée chaque somme devrait approcher la moitié du nombre de lancers effectués (environ 500 pour 1000 lancers).

Cette réponse est plus structurée. D'une part, il y a description d'un protocole expérimental (d'une expérience aléatoire) destiné à donner des informations sur les caractéristiques de la pièce, mais on ne trouve pas de règle de décision⁽¹¹⁾ « chaque somme devrait approcher la moitié du nombre de lancers effectués ». L'élève en question ne savait pas trop à partir de combien il fallait se décider pour dire que la pièce n'est pas équilibrée.

4°) Quelques réponses plus conformes aux attentes...

Voici quelques réponses qui ont été lues et étudiées en cours. En effet, pour chacune d'entre elles, on trouve la mise en place d'une expérience destinée à réfuter l'hypothèse d'équiprobabilité⁽¹²⁾.

Réponse : Il faut faire l'expérience suivante:

- *On lance la pièce de cette personne une centaine de fois.*
- *Ensuite on compte le nombre de fois on l'on obtient pile et face.*
- *On construit un tableau dans lequel on note les résultats obtenus précédemment.*
- *Si les résultats d'obtention de pile et de face sont proches l'un de l'autre (environ 50% pour chacun) la pièce est donc bien équilibrée. Dans le cas contraire elle est truquée.*

(10) Afin de montrer la stabilisation de la fréquence, il faut faire beaucoup de lancers, l'organisation de simulation sur un tableur impose la disposition en lignes plutôt qu'en colonnes (256 colonnes pour beaucoup de lignes). Ceux qui, par simulation, souhaitent résoudre le problème du duc de Toscane savent qu'il faut faire beaucoup de tirages avant de « voir statistiquement » le résultat.

(11) L'expression « règle de décision » doit être prise au sens commun. On a des résultats, qu'est-ce qu'on fait ?

(12) On peut évoquer Karl POPPER, selon lequel la capacité d'une hypothèse à résister à l'expérience ne fait que nous renforcer dans son acceptabilité ; les expériences ne servent qu'à réfuter les hypothèses (ou à ne rien dire).

REMARQUE : Plus le nombre de lancers est conséquent, plus les résultats seront précis.

Pour cette réponse, l'idée est la mise en place d'un protocole expérimental ; par contre il y a confusion entre nombre de fois où l'on a pile et fréquence des piles. L'idée, exprimée par l'élève lors de cette séance, est que les fréquences de « pile » et celle de « face » doivent être proches de 50%. Cette partie de la réponse a été reformulée en classe.

De plus, et c'est un point important, la prise en compte du nombre de répétitions : sans connaître de résultats théoriques, cet élève affirme que « *Plus le nombre de lancers est conséquent, plus les résultats seront précis.* »

Ce qui signifie pour lui que plus on fera de lancers, plus on aura de chances que notre décision soit correcte.

Une réponse similaire :

Réponse : Pour cela, il faudrait effectuer une série de lancers où l'on noterait le résultat de chaque lancer. Par exemple 100 ou 1000 lancers de pièce. On dresserait ensuite un tableau avec le nombre d'apparitions de « face » et de « pile ».

Puis on comparerait les résultats en les mettant en pourcentage par rapport à l'effectif total de lancers.

Si les fréquences d'apparition des deux événements (avoir « pile » ; avoir « face ») ne s'équivalent pas (environ = 50% pour chacun des deux événements), on en conclurait que la pièce n'est pas bien équilibrée et donc qu'elle est truquée.

Dans ces deux réponses, comme dans certaines déjà citées, il y a la notion de « résultats proches », « qui s'équivalent », « écart pas trop important », ...

La question qui se pose est donc : à partir de quel écart on se décidera pour dire que la pièce n'est pas équilibrée (sachant que même si l'écart est important, il se peut que la pièce soit équilibrée) ? Qu'est-ce qu'un écart important ? Les élèves sont assez vite d'accord sur le fait que l'écart à prendre en compte est un écart « relatif ». De plus pour éviter des décisions trop personnelles, il faut définir des règles afin que pour un résultat donné, tous et toutes donnent la même réponse.

C'est ce besoin de « normalisation » qui a entraîné la notion d'intervalle de confiance et d'intervalle de fluctuation avec des seuils bien définis. La notion d'intervalle de fluctuation est au programme de seconde, la notion d'intervalle de confiance au programme de terminale.

Réponse : Du point de vue statistique, plus on effectuera de lancers de la pièce de monnaie en question, plus la valeur réelle du pourcentage de chance d'obtenir une face ou l'autre sera proche de la valeur de la série, il faut donc effectuer beaucoup de lancers pour vérifier (ou non) la pensée de la-dite personne.

Cependant, étant donné que faire des lancers à la main est long et laborieux, il lui faudra du temps pour pouvoir se faire une idée réelle de l'équilibrage de sa pièce ; on ne peut pas simuler cette expérience car on ne sait pas la probabilité qu'un événement « pile » ou qu'un événement « face » se produise.

Le point de vue de cet élève, exprimé lors de la discussion, est la stabilisation de la fréquence. Plus le nombre de lancers est important, plus la fréquence va « se rapprocher » de la probabilité. Ce raisonnement est discutable : est-ce que la probabilité existe effectivement, est-ce qu'elle est constante dans le temps, ... ? Restons modeste et tenons-nous au programme de seconde. Dans le document ressource concernant les probabilités et les statistiques⁽¹³⁾, on peut lire page 7 : « Les distributions de probabilités peuvent être estimées par observation de la stabilisation des fréquences sur de longues séries de fréquences... ». C'est le point de vue développé par Alfred RENYI : « Nous appellerons *probabilité* d'un événement le nombre autour duquel oscille la fréquence de l'événement considéré »⁽¹⁴⁾.

Le protocole expérimental est moins bien décrit que précédemment, mais la remarque intéressante est que cette question ne peut pas avec les connaissances des élèves de seconde être résolue en utilisant un tableur.

Cela a été l'occasion de leur parler d'une des expériences de WELDON.

WELDON⁽¹⁵⁾ a lancé (ou fait lancer) 49 152 fois un dé et il a obtenu 25 145 fois l'événement {4, 5, 6}.

On peut se poser la question : est-ce que, compte tenu de ces résultats, on peut admettre que le dé n'est pas régulier ?

Si le dé est régulier, la probabilité de l'événement considéré est 0,5, on doit observer en moyenne 24 576 fois cet événement. Soit un écart de 569.

On pourrait penser qu'un tel écart est faible (par rapport au nombre de lancers). En fait, un tel écart est extrêmement rare. On a donc de bonnes raisons de dire que le dé n'est pas régulier. On peut l'expliquer par le fait que sur les anciens dés où les numéros sont indiqués par des trous, les faces 4, 5 et 6 sont plus légères que les autres et donc apparaissent plus souvent que les autres.

Le deuxième objectif de cette évaluation, vous l'avez compris maintenant, c'est l'intervalle de fluctuation.

L'introduction aux élèves est semblable à celle proposée dans le document ressource déjà cité. Le 95% ne peut qu'être imposé, puisqu'il n'a aucun fondement de nature mathématique, sinon – c'est une boutade – qu'un type normal est à moins de deux écart types du type moyen.

Cette année, j'ai posé un problème identique à une classe de seconde⁽¹⁶⁾.

Suite à la discussion, nous sommes arrivés à la conclusion qu'il fallait étudier la différence entre le nombre de « pile » et le nombre de « face » et que plus cet écart serait grand, plus on serait en droit de penser que la pièce n'est pas équilibrée.

(13) http://cache.media.eduscol.education.fr/file/Programmes/17/9/Doc_ressource_proba-stats_109179.pdf

(14) Alfred RENYI *Calcul des probabilités* Dunod 1966 page 25.

(15) MG KENDALL et A STUART *The advanced theory of statistics* tome 1 Edition Griffin 1963 page 221.

(16) « Ma femme m'a dit : j'ai lancé 25 fois une pièce de 2 Euros, j'ai obtenu 15 fois pile, j'ai l'impression qu'elle n'est pas équilibrée ». Cette séance s'est déroulée dans une salle de classe « normale ».

Oui, mais l'écart proposé peut être positif ou négatif, j'ai imposé que l'on prenne l'écart sans le signe (sans parler pour autant de valeurs absolues). Les élèves (34) ont faits à la calculatrice des simulations (deux simulations de 25 lancers d'une pièce équilibrée), on a noté les résultats au tableau ; pourquoi n'y a-t-il que des nombres impairs ? Comment trouve-t-on l'écart si l'on ne connaît que le nombre de « pile » (dans le cas de l'utilisation d'un tableur) ? Pourquoi la forme de la distribution d'échantillonnage de la valeur absolue de la différence n'est-elle pas la même suivant que le nombre de lancers est pair⁽¹⁷⁾ ou impair ? C'est combien un écart grand ? Est-ce que l'on est sûr que la réponse est juste ?

Bref de vraies questions de mathématiques, de vraies questions de statistiques.

Lorsque j'ai donné la formule pour l'intervalle de fluctuation, j'ai eu deux réactions : « *vous auriez pu le dire avant* » et « *c'est normal qu'il y ait une règle simple, faut que tout le monde fasse pareil et on va pas à chaque fois prendre la calculatrice ou un tableur* ».

En conclusion, j'encourage les collègues à proposer des investigations, accepter les pistes proposées par les élèves, les étudier et les abandonner ou les garder, leurs laisser le temps de s'imprégner de la notion de fluctuation. Ce temps, soi-disant perdu pour le programme, me paraît indispensable pour la formation de l'esprit.

Comment organiser ce type de séance, comment les gérer, comment instaurer une synthèse, ... Bref, il reste encore du travail en didactique des statistiques et le Bulletin Vert peut en être une tribune.

ANNEXE : Quelques références à propos du débat scientifique en mathématiques

Sur le site de l'IREM de Grenoble ou bien dans Publimath, on trouvera de nombreuses entrées sur le débat scientifique en mathématiques, par exemple :

Une introduction parue dans PLOT :

www.apmep.asso.fr/IMG/pdf/Debat_Leroux.pdf

Deux références de Marc Legrand :

Un article paru dans Repères N° 10 :

www.univ-irem.fr/reperes/articles/10_article_68.pdf

Un document beaucoup plus conséquent :

http://www.cds-aub.be/www.cds-aub.be/uploads/file_/Le_principe_du_debat_scientifique_dans_l_enseignement-Legrand_M.pdf

(17) Pour avoir le résultat de 26 lancers, chaque élève a pris ses résultats pour 25 lancers et fait un lancer de plus.