Aux urnes, collégiens !

Louis-Marie Bonneval(*)

Depuis la rentrée 2008, les probabilités sont au programme de la classe de Troisième, ce qui à mon avis est une bonne chose. Mais je voudrais dire mes inquiétudes quant à la forme que semble prendre ce premier contact.

Le libellé du programme est le suivant :

Connaissances	Capacités	Commentaires
1.4. Notion de probabilité	– Comprendre et utiliser des notions élémentaires de probabilité. – Calculer des probabilités dans des contextes familiers.	La notion de probabilité est abordée à partir d'expérimentations qui permettent d'observer les fréquences des issues dans des situations familières (pièces de monnaie, dés, roues de loteries, urnes, etc.). La notion de probabilité est utilisée pour modéliser des situations simples de la vie courante. Les situations étudiées concernent les expériences aléatoires à une ou à deux épreuves.

Pendant toute ma carrière d'enseignant de lycée, j'ai enseigné les bases des probabilités dans toutes les sections, scientifiques ou non, et cette expérience m'a appris quels étaient les principaux obstacles rencontrés par les élèves. Ces obstacles me semblent sous-estimés par les auteurs du programme, si j'en crois le document d'accompagnement [1].

Approche fréquentiste ?

Je suis très inquiet qu'on prétende démarrer au collège par une approche fréquentiste. Je ne méconnais pas l'intérêt de cette approche, mais elle me semble hors de portée d'un élève de Troisième.
L'annexe 1 du document d'accompagnement rappelle les deux grandes interprétations du concept de probabilité :
Dans l'approche traditionnelle(1), on définit la probabilité d'un événement comme le rapport du nombre de cas favorables au nombre de cas possibles, supposés également possibles (on parle de probabilité *a priori*, ou *subjective*). Cette démarche a le défaut d'être inopérante pour les univers infinis, et même pour les univers finis quand on ne peut pas se ramener à l'équiprobabilité.
Dans l'approche fréquentiste(2), on définit la probabilité d'un événement comme la limite presque sûre de sa fréquence d'apparition quand le nombre de répétitions de

(*) louis-marie.bonneval@libertysurf.fr
(1) dite *laplacienne* ou *épistémique*. Elle fonde la « géométrie du hasard » initiée par Pascal en 1654 et exposée par Laplace dans l'*Essai philosophique sur les probabilités* (1814).
(2) initiée par Jakob Bernoulli (*l'Art de conjecturer*, 1713) et développée notamment par Laplace, Gauss, Fisher, Student, Neyman, Pearson, …

l'épreuve tend vers l'infini (on parle de probabilité *a posteriori*, ou *objective*). Cette démarche a le grand mérite de s'appliquer à tout type d'épreuve. Mais elle présente des difficultés(3) :
1) Parler de limite presque sûre fait référence à la probabilité pour définir la probabilité, ce qui ressemble à un cercle vicieux(4).
2) En pratique, le nombre de répétitions est nécessairement fini, ce qui fait de la fréquence observée une valeur approchée de la probabilité. Pour mesurer la précision de cette approximation, il faut étudier la fluctuation d'échantillonnage. Cela nécessite soit l'appel à une théorie de niveau universitaire (loi normale), soit un travail de simulation un peu délicat (qui est actuellement au programme de Seconde).
3) Il faut choisir (arbitrairement ?) une valeur dans l'intervalle obtenu. Cela laisse planer un doute quant à l'unicité de la probabilité d'un événement.
À ces obstacles épistémologiques s'ajoutent des obstacles didactiques : il n'est pas question de parler de limite à des élèves de collège, qui ont souvent des difficultés avec les inégalités et les intervalles, voire avec les fractions et les proportions…
Or il me semble qu'on peut surmonter ces difficultés, et même faire une synthèse des deux approches, en pratiquant en Troisième une démarche traditionnelle et en Seconde une démarche fréquentiste.

Distinguer réalité et modèle

Un préalable indispensable est de bien distinguer réalité et modèle. Cela permet de répondre à l'objection 3 ci-dessus, et surtout cela fournit aux élèves une règle du jeu mathématique claire, en distinguant ce qui relève de l'observation et ce qui relève du raisonnement.
Or le document d'accompagnement, loin de faire cette distinction, me semble entretenir une confusion regrettable. Le mot « modèle » n'y apparaît d'ailleurs jamais, alors que le terme « modéliser » figure dans le libellé du programme.

Quelques exemples :

– page 2, il est question de lancer une pièce « *équilibrée* » : pourquoi des guillemets autour de ce mot ? Il faut dire qu'il s'agit d'une définition : dire qu'une pièce est équilibrée signifie que, lors d'un jet, pile et face sont équiprobables. De même, dire qu'un dé est régulier signifie que, lors d'un jet, ses faces sont équiprobables(5). Il s'agit d'objets mathématiques, comme le triangle équilatéral ou l'hexagone régulier.

– page 4 : *un tireur tire parfaitement au hasard sur la cible, sans jamais la rater*(6) : il y a dans l'expression « parfaitement au hasard » une confusion entre la réalité et le modèle : dire « au hasard » c'est choisir comme modèle une loi uniforme (pour laquelle par définition *la probabilité d' une région est proportionnelle à son aire*) ; or un modèle mathématique ne peut pas rendre compte « parfaitement » d'une réalité

(3) Voir par exemple [2], pages 30-31, et [3], chapitres I et II.
(4) Objection discutée par exemple par Renyi (*Calcul des probabilités*, 1966).
(5) Pour être précis, c'est le *lancer de la pièce* qui est équilibré, et le *lancer du dé* qui est régulier.
(6) L'auteur est conscient du cocasse de cette hypothèse, puisqu'il met un point d'exclamation entre parenthèses.

complexe. Pour faire une analogie avec la géométrie, c'est comme si on parlait d'une chambre « parfaitement carrée » : la chambre réelle ne peut être considérée comme carrée qu'à un certain niveau d'approximation. De même ici, la répartition des impacts ne peut être considérée comme uniforme qu'à un certain niveau de confiance.

– page 6 (choix de deux points au hasard sur un segment), on évoque la fonction ALEA() sans préciser qu'elle simule une loi uniforme (il faudrait rappeler ce que ça veut dire : la probabilité que l'abscisse soit dans un segment est proportionnelle à la longueur de ce segment). Ce qui est capital, c'est d'expliquer qu'*on simule toujours un modèle*. Dans l'exemple choisi, ce modèle est constitué de deux hypothèses : chaque abscisse suit la loi uniforme sur [0 ; 1] ; les deux abscisses sont indépendantes. C'est avec ce modèle là qu'on peut – comme c'est fait page 7 – justifier que 0,25 est bien la probabilité cherchée. Mais la simulation à elle seule ne conduit pas nécessairement à cette valeur : pourquoi pas 0,24 ou 0,245 ? La bonne problématique, qui n'apparaît guère dans le document, est la suivante : comment choisir un modèle (estimation) ? Comment juger si un modèle est valable (tests) ?
Sur le même exemple, le commentaire en haut de la page 8 parle de difficultés liées au régionnement : à mon sens elles sont plutôt liées à la notion de probabilité-produit, inaccessible au collège : comment justifier que la répartition de probabilité est uniforme sur le carré ? *On peut se contenter d'une illustration au tableur* : pourquoi pas, mais quel est l'objectif pédagogique ? Sur le fond, je pense que les univers infinis sont prématurés au collège (et même au lycée).

– en annexe 1, en bas de la page 18, à propos d'une loterie de 1 000 boules numérotées : « *la probabilité comme fréquence limite : lors de tirages répétés avec remise, chaque boule est tirée aussi souvent qu'une autre* ». Ce n'est pas vrai : même s'il y a équiprobabilité, lors de tirages répétés les fréquences sont en général différentes ! Il s'agit de savoir si elles sont *significativement* différentes, ce qui mériterait d'autres commentaires (test du khi-deux).

Que faire au collège ?

L'urne de Bernoulli[7] (contenant des boules – ou des jetons – indiscernables) est évoquée dans le programme comme un exemple de « situation familière ». Je pense qu'elle peut être beaucoup plus que cela : une *épreuve de référence* incontournable, au même titre que les figures-clés de la géométrie ou les fonctions de référence de l'analyse.
Sur le plan didactique, elle demeure selon moi le meilleur outil pour fournir une bonne image mentale, fixer le vocabulaire de base (épreuve, issues, événements, probabilités) et établir les premières propriétés [3].
Par rapport aux autres « situations familières » (pièces, dés, cartes, roues, …), elle présente plusieurs avantages :
– par l'adjonction de couleurs et de numéros sur les boules, elle permet de traiter facilement de la réunion et de l'intersection d'événements ;
– elle établit un lien naturel avec les statistiques, puisqu'elle peut représenter une

(7) « *Je suppose que, dans une urne, à ton insu soient placées 3 000 pierres blanches et 2 000 pierres noires …* » (Jakob Bernoulli, *l'Art de conjecturer*).

population dans laquelle on s'intéresse à diverses sous-populations.
– elle peut facilement être matérialisée en classe (du moins pour un nombre de boules pas trop grand) par un tas de bouts de papier identiques (qu'on retourne et mélange, après y avoir inscrit les informations voulues).
Mais même si on peut en faire une maquette, il doit être très clair qu'il s'agit d'un objet mathématique, et que par définition, lors d'un tirage dans une urne de Bernoulli, chaque boule a autant de chances d'être tirée qu'une autre. Autrement dit, s'il y a n boules, chacune a une chance sur n (une probabilité $\frac{1}{n}$) d'être choisie.
Bien sûr, cela ne permet pas d'envisager des univers infinis. En revanche, toute épreuve concrète d'univers fini peut être modélisée par un tirage dans une urne. En effet, pour une telle épreuve, un modèle est la donnée des issues $(\omega_1, \omega_2, \ldots, \omega_k)$ et de leurs probabilités $(p_1, p_2, \ldots, p_k)$. Or, tout modèle étant une approximation de la réalité, et tout réel pouvant être approché aussi près qu'on veut par un rationnel, on peut toujours supposer les p_i rationnels[8]. On peut alors représenter l'épreuve par une urne contenant n boules dont n_i sont marquées ω_i, les proportions étant $(p_1, p_2, \ldots, p_k)$.
Par exemple,
– un dé régulier peut être représenté par une urne contenant 6 boules numérotées 1, 2, 3, 4, 5, 6 ;
– un dé truqué où le 6 a trois fois plus de chances d'apparaître que chacun des autres numéros peut être représenté par une urne contenant cinq boules numérotées 1, 2, 3, 4, 5 et trois boules numérotées 6.
À propos de ces deux exemples, il se peut qu'un élève particulièrement futé pose la question : « Comment peut-on savoir qu'un dé est régulier ? Comment peut-on savoir que le 6 a trois fois plus de chances d'apparaître que chacun des autres numéros ? ». S'il est capable de poser cette question, il sera capable d'en deviner la réponse : « en le lançant un grand nombre de fois… ». C'est ce qui motive la démarche fréquentiste, qui, selon moi, devrait être reportée en Seconde (cf. plus loin).

L'urne de Bernoulli permet de poser les bases des probabilités, telles qu'elles sont décrites au paragraphe 4 du document d'accompagnement :
1) Épreuve aléatoire, issues, répartition de probabilité (équiprobabilité ou pas équiprobabilité, voilà la question…).
2) Notion d'événement (condition sur les issues / sous-ensemble de l'univers : passage d'un langage à l'autre).
3) Probabilité d'un événement (somme des probabilités de ses issues).
4) Événement impossible, événement certain.

(8) Je parle d'épreuve *concrète* (pour laquelle les probabilités des issues doivent être *estimées*) et non pas d'épreuve conçue selon un modèle mathématique (cas de l'aiguille de Buffon par exemple). Pour faire une analogie avec la géométrie, on peut bien entendu concevoir un rectangle à côtés irrationnels, mais s'il s'agit de *mesurer* une salle rectangulaire, on obtiendra des valeurs rationnelles (et même en général décimales).

5) Intersection, réunion (travail sur le « et » et le « ou »).
6) Événements incompatibles (la probabilité de l'un ou l'autre est la somme des probabilités de chacun).
7) Événement contraire d'un autre (sa probabilité est le complément à 1 de celle du premier).

On a largement de quoi s'occuper avec ça. C'est peut-être même déjà trop : quel est l'horaire à consacrer aux probabilités ?
Comprendre et acquérir le vocabulaire et les notations n'est pas immédiat pour les élèves de lycée : cela demandera du temps aux élèves de collège. Par exemple, la confusion entre un événement et sa probabilité est courante chez les débutants.
Les propriétés ci-dessus sont à rapprocher des propriétés des proportions (ou fréquences) vues en statistique, les unes confortant les autres. Par exemple la propriété 6 n'est autre que la mal-nommée « somme de pourcentages ». Le passage du langage des statistiques (« 75 % des vieillards sont des femmes ») à celui des probabilités (« la probabilité qu'un vieillard soit une femme est 0,75 »), et inversement, est une compétence à développer.
On peut aussi en profiter pour conforter quelques bases de logique : j'ai parlé du « et » et du « ou », on peut parler de leur négation, ou encore du lien entre implication et inclusion.

Le programme parle d'*expériences à deux épreuves*[9] : cela me paraît ambitieux en Troisième, mais si on trouve le temps de le traiter, cela se fait très bien avec une urne dans laquelle on tire deux fois de suite. Comparer tirage avec remise et sans remise introduit très naturellement les notions de probabilité conditionnelle et d'indépendance, ainsi que l'usage des arbres.
L'approche traditionnelle est alors bien plus simple et satisfaisante que l'approche fréquentiste pour établir les propriétés. Il suffit pour s'en convaincre de lire le document d'accompagnement (page 11) qui prétend utiliser des égalités approchées pour établir une égalité exacte ! C'est comme si en géométrie, on disait : « Ayant tracé un carré de côté 1 dm et mesuré sa diagonale, j'ai trouvé 1,41 dm, donc j'affirme que la diagonale mesure $\sqrt{2}$ » ! Ce paragraphe est un véritable « pousse-au-crime » : si les professeurs le mettent en pratique, il ne faudra pas s'étonner de trouver dans les copies des raisonnements approximatifs !

En bref, l'essentiel au collège, me semble-t-il, tient en deux objectifs :
– Savoir modéliser une « situation familière » par un tirage dans une urne (éventuellement deux tirages successifs).
– Savoir traiter un tirage dans une urne (éventuellement deux tirages successifs).

(9) Cette formulation m'étonne, car pour moi les expressions « expérience aléatoire » et « épreuve aléatoire » sont synonymes. Si e et e' désignent deux épreuves, je suggère de noter ee' l'épreuve qui consiste à effectuer e puis e' de façon indépendante. Avec cette notation, on noterait e^n l'épreuve qui consiste à effectuer n fois l'épreuve e.

Continuité avec le lycée

L'approche fréquentiste pourrait fort bien attendre le lycée, sans même qu'il soit nécessaire d'en modifier les programmes.
En effet cette démarche est actuellement abordée en Seconde : on répète des tirages (matériellement, puis par simulation), et on découvre expérimentalement qu'il y a un lien entre la proportion de boules rouges dans l'urne et la fréquence d'apparition d'une boule rouge lors de tirages répétés, et que ce lien est d'autant plus étroit que le nombre de répétitions est élevé. On en tire l'idée que la répétition d'une épreuve permet d'estimer une probabilité inconnue, et donc de choisir un modèle, à savoir l'urne qui peut représenter l'épreuve. À partir de ce modèle on peut raisonner, puisqu'on a appris à le faire en Troisième.
Exemple classique (évoqué en page 5 du document d'accompagnement) : on jette une punaise, qui peut tomber sur la tête ou sur la pointe. Aucune propriété de symétrie ne pouvant suggérer un modèle, la répétition de l'épreuve est incontournable pour estimer la répartition de probabilité.
Autre exemple : en lançant une pièce un grand nombre de fois, on découvrira peut-être qu'elle n'est pas équilibrée, et on pourra envisager un autre modèle mieux adapté.

En Première il sera bien temps de parler de probabilités conditionnelles, d'indépendance, d'arbres pondérés, et peut-être de variables aléatoires.

Et en Terminale on aura ainsi les outils théoriques pour revenir sur les épreuves répétées, qu'on pourra aborder cette fois par le raisonnement.

Quant aux univers infinis, je ne vois pas l'urgence. On pourrait certes parler du jeu de franc-carreau ou du tir sur une cible, mais cela ne me paraît pas prioritaire, vu le peu de temps dont on dispose en Troisième.

Il serait tellement dommage que les difficultés conceptuelles et techniques dégoûtent des probabilités les élèves de collège !

Bibliographie succincte

[1] Probabilités au collège (mars 2008) :
http://eduscol.education.fr/D0015/doc_acc_clg_probabilites.pdf

[2] *Enseigner les probabilités au lycée*, par la commission inter-IREM Statistiques et probabilités, 1997.

[3] *À propos de l'enseignement du calcul des probabilités*, par Jean-Claude Thiénard (IREM de Poitiers, 1993).