Les probabilités entrent au collège : quels enjeux, quelles activités ?

Rémy Coste

Rémy Coste a écrit cet article pour les « Chantiers de pédagogie mathématique », publication de la régionale Île De France dont il est un membre actif. Compte tenu de l'entrée des probabilités au programme de 3ème, il nous a semblé important de proposer cet article, riche en exemples, à un public plus large.

Rémy Coste est professeur de mathématiques au lycée Blaise Pascal à Orsay (91) et à l'IUFM de Versailles, site d'Étiolles, pour la formation des PE.

L'introduction de l'aléatoire dans les programmes de la classe de seconde en 2000, avait suscité beaucoup de perplexité et d'interrogations, voire de vives réactions. Peut-on enseigner conjointement des éléments de statistiques et de probabilités ? Quels savoirs et compétences sont visés ? Que recouvre précisément une « *approche fréquentiste des probabilités* » ? Quelle est la place et le rôle de la simulation ?

Passés les premières inquiétudes et l'inconfort à prendre en charge un enseignement pour lequel on n'a pas une vision claire, pas de recul, peu de ressources, et au mieux une formation universitaire qui est en net décalage avec les enjeux didactiques du secondaire, il est finalement apparu pour beaucoup d'enseignants, après une première année de pratique, que cette introduction à l'aléatoire aurait parfaitement sa place dès le collège. A ce sujet, Guy Brousseau, lors de l'élaboration de sa théorie des situations dans les années 70, a choisi d'expérimenter dans une classe du primaire, une situation relevant de l'aléatoire[1] (une bouteille opaque laisse voir dans le bouchon, quand on la retourne, une des billes blanches ou noires qu'elle contient. Question : quelle est la proportion de billes blanches et noires ?). Ce n'est donc pas une surprise de voir apparaître cette nouveauté dans les programmes de 3ème. Comme enseignant en seconde, je propose de faire le point sur ce que j'ai compris des enjeux de cette partie du programme après quelques années. Notons que la mouture du programme de troisième franchit un seuil par rapport à celui de seconde, en s'appelant « *Notions de probabilités* », en indiquant explicitement de « *calculer des probabilités* », et en demandant d'étudier « *des situations à une ou deux épreuves* ».

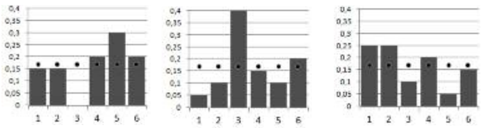
L'une des interrogations très souvent entendue est la suivante : quel est l'intérêt d'expérimenter des lancers de dés ou de pièces, tout ça pour « découvrir » que chaque face du dé a 1 chance sur 6 d'apparaître ? Comme scoop, ce n'est certes pas très excitant ! Il y a là un malentendu qu'il faut lever.

Une loi de probabilité est un modèle, que l'on choisit pour caractériser une expérience aléatoire. Trois cas de figure doivent être rencontrés avec les élèves, dans l'ordre suivant.

1er cas : on connaît la loi de probabilité. Il n'y a rigoureusement aucun doute sur la pertinence de la loi de probabilité choisie (la loi n'est pas et ne sera pas remise en question). C'est justement le cas des dés ou des pièces (éléments de symétrie attestés), des roulettes (secteurs parfaitement mesurables), des urnes (nombre de boules connu), etc. Le but de l'expérimentation est ici de faire constater

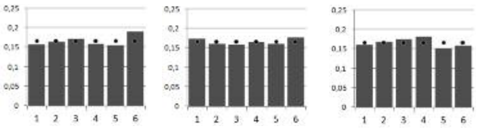
1 http://www.univ-orleans.fr/irem/cii/modules/news/article.php?storyid=68

que, alors que l'on connaît les probabilités de chaque éventualité (vues comme des « fréquences théoriques »), **on ne retrouve pas ces fréquences** en répétant l'expérience aléatoire un petit nombre de fois. De plus, en répétant des séries de même taille (réduite), on ne trouve pas des répartitions comparables. Ci-dessous, on a lancé un dé 20 fois et noté les fréquences observées des 6 occurrences (les points noirs ont pour ordonnée 1/6). On a recommencé ainsi 3 fois. Les 3 séries ne donnent pas la même répartition. Le 5 sort-il « trop souvent » (série 1) ou « trop peu » (série 3) ?



Fréquences observées sur 3 séries
de 20 lancers d'un même dé

L'enjeu est de casser les représentations intuitives fausses, souvent bien ancrées, qui induisent un comportement déterministe. Il existe des magazines tirant à plus de 100 000 exemplaires qui, constatant que la fréquence d'apparition de la boule 24 du loto lors des tirages de l'année est inférieure à celle des autres boules, prescrivent de la jouer : elle va rattraper son retard ! D'autres au contraire, face au même constat, reconnaissent là une preuve de plus de l'existence du grand complot, et recommandent de ne pas la jouer. A ce propos, lors de l'introduction du chapitre, je commence par un test du même type que celui proposé par le document d'accompagnement (page 16)[2]. Ce même test est rediscuté en fin de chapitre. La phase suivante consiste à augmenter le nombre de répétitions de l'expérience (i.e. la taille de l'échantillon). On constate alors que, d'une part il y a moins de disparités entre les séries d'expériences, d'autre part les fréquences observées ont tendance à se rapprocher des fréquences « théoriques ». C'est la fluctuation d'échantillonnage : elle existe, mais elle est d'autant plus faible que le nombre de répétitions est grand. Il y a donc bel et bien de l'aléatoire, mais il est quand même possible de faire des prévisions sur le résultat lorsque le nombre de répétitions est grand. C'est cette dualité aléatoire/prévision qui est en jeu.



Fréquences observées sur 3 séries
de 1000 lancers d'un dé

Évidemment, on en reste à des formulations qualitatives, les premiers calculs quantitatifs relevant de la classe de seconde (fourchettes de sondage) et des classes terminales (test d'adéquation avec une loi équirépartie). Pour éviter les confusions, avec des exemples de lois équiréparties, il faut impérativement rencontrer des exemples de loi non équirépartie. Ces dernières sont plus probantes pour distinguer ce qui est imputable à la fluctuation (ce qui est dû « au hasard ») et ce qui est imputable à la loi de probabilité.

2ème cas : on ne connaît pas *a priori* la loi de probabilité (en tout cas les élèves ne la connaissent pas). L'expérience, réalisée et répétée, donne des indications sur la loi adéquate à trouver et permet de formuler une conjecture ou de ruiner celle qu'on avait faite *a priori* (en probabilité l'intuition est souvent erronée, l'équirépartition naturellement adoptée par les élèves… souvent à tort). Puis, c'est par un raisonnement, un calcul, un arbre, un tableau, ou plus simplement en dévoilant le dispositif de l'expérience, que l'on va élucider *a posteriori* la situation et trouver la loi de probabilité

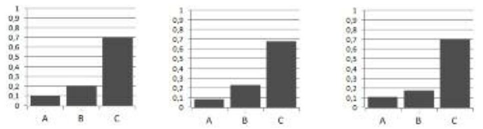
[2] http://eduscol.education.fr/D0015/LLPHAG00.htm#college

qui a prévalu. Les observations expérimentales ont pu éventuellement constituer une aide, dans la mesure où elles donnent des indications sur les valeurs à trouver.

Dans le document d'accompagnement, c'est l'exemple du « franc carreau », très riche par la consistance des considérations géométriques nécessaires pour aboutir à la loi.

3ème cas : on ne connaît pas la loi de probabilité et on est dans l'impossibilité de la déterminer par un raisonnement ou un calcul. Soit parce que les élèves ne sont pas encore outillés mathématiquement pour le faire, soit parce que ce n'est de toute façon pas possible. Ce sont alors les fréquences observées expérimentalement qui vont définir la loi de probabilité que l'on retiendra. C'est l'exemple du lancer de punaises évoqué dans le document d'accompagnement.

Mais il faut tenir compte de l'existence de la fluctuation d'échantillonnage. Pour cela, on recommence plusieurs séries de même taille. Si la répartition est à chaque fois sensiblement la même, c'est que la fluctuation est faible (nettement plus faible que les écarts constatés entre les fréquences). On pourra donc valider avec confiance la loi de probabilité définie par les fréquences observées expérimentalement.



Fréquences observées sur 3 séries de taille 200 d'une expérience à 3 éventualités (A, B et C) non équiprobables :
la fluctuation est faible, la loi de probabilité 0,1 / 0,2 / 0,7 peut être validée.

Ce qui tranche avec les pratiques habituelles en mathématiques, c'est qu'il ne s'agit pas d'une validation au sens preuve mathématique qui garantit à 100 % que le modèle est bon. Dans certains cas, les calculs probabilistes savants rencontrés dans les mathématiques universitaires ne le pourront pas plus ! Ils seront par contre capables d'évaluer le risque que l'on prend en décidant d'adopter tel ou tel modèle, uniquement sur la base des fréquences observées expérimentalement (tests d'hypothèses statistiques). Dans les applications scientifiques des statistiques et probabilités, l'évaluation de ce risque est nécessaire, mais très souvent suffisante pour mener efficacement les travaux. En particulier, et quoiqu'on dise, c'est le cas des sondages lorsqu'ils sont bien faits.

Quelle est la place de la simulation ? Elle est évidente : elle sert à faire très vite un grand nombre d'expériences, et d'être assistée par des outils numériques et graphiques puissants.

Mais il y a lieu de faire 2 remarques :

- Il ne faut pas faire l'économie de réaliser des expériences réelles. La dévolution des situations proposées est bien plus efficace. C'est tout à fait possible en mettant à contribution les élèves de la classe et en mutualisant au tableau (tableau de classe ou tableur) leurs données (pour une fois le nombre d'élèves est un atout !). On peut aussi leur demander de réaliser ce travail de collecte à la maison.
- En passant ensuite à la simulation sur calculatrice ou tableur, on utilise un générateur pseudo-aléatoire. Il ne me semble pas opportun d'évoquer avec les élèves les questions de fiabilité de ces générateurs. D'abord parce qu'ils le sont très largement pour les besoins de cet enseignement, ensuite parce que cela risquerait d'apporter de la confusion sur les conclusions que l'on cherche à leur faire identifier.

Pour mettre en œuvre concrètement cet enseignement, à côté du document d'accompagnement, plusieurs brochures de l'APMEP ou des IREM (celui de Paris Nord notamment), rédigées pour le lycée, sont en partie facilement recyclables pour le collège. Le site *Statistix*, dédié à cet enseignement, est également une ressource pour les enseignants.

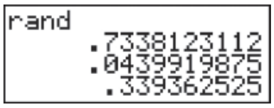
Comment simuler une expérience aléatoire ?

Avant de répondre à cette question, je précise qu'il est important selon moi de faire réaliser quelques expériences réelles aux élèves avant de faire des simulations (lancer de pièces ou de dés, prélèvement de billes dans une urne, tirage au sort, etc.). Cela permet une meilleure identification de la situation par les élèves et des questions que l'on se pose. La simulation permet ensuite de renouveler l'expérience un grand nombre de fois, mais aussi de réaliser virtuellement des expériences aléatoires que l'on ne peut pas réaliser réellement (c'est le cas de mon 1er exemple).

Je fais ici le choix d'utiliser une calculatrice. A titre d'exemple, je préciserai entre parenthèses comment accéder aux instructions sur la TI Collège Plus (on trouvera l'équivalent sur les autres modèles). On peut bien sûr utiliser un tableur, mais je veux montrer qu'une calculatrice est tout à fait opérationnelle et sans doute plus explicite pour les élèves, sans compter que la logistique en est simplifiée.

Le générateur aléatoire

Il est accessible par la commande ***Random***, notée en général ***rand*** sur les calculatrices (maths, RND, 1:rand). Par un algorithme interne très sophistiqué reposant sur des savoirs d'arithmétique, l'exécution de cette commande simule le tirage aléatoire équiprobable d'un nombre décimal à 14 chiffres pris dans [0;1[(à l'affichage, on ne voit que 10 chiffres). Ci-contre, l'exécution réitérée de ***rand*** (appuyer plusieurs fois sur Entrer).



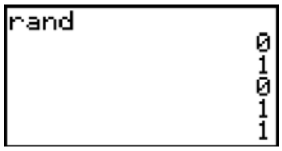
Utilisation de *rand* pour la simulation d'expériences aléatoires

1ère idée : On convient d'un codage des 10 chiffres de la partie décimale des nombres que renvoie ***rand***. Par exemple, pour le lancer de pièces, on convient que si un chiffre est pair c'est Pile, s'il est impair c'est Face. Ainsi, chaque exécution de rand donne une liste de 10 lancers. Dans la copie d'écran ci-dessus, on a ainsi : FFFPFPFFFP, puis PPFFFFFPFF, puis FFFFPPFPFP (le 10ème chiffre est 0, ce qui explique qu'il ne soit pas affiché).

Pour le lancer de dés cubiques, on ignore les chiffres 0, 7, 8 et 9. Les autres chiffres nous donnent les faces du dé obtenues. Dans l'exemple ci-dessus, on a : 3 3 1 2 3 1 1 2, puis 4 3 1 5, puis 3 3 3 6 2 5 2 5.

Avantage : c'est rapide car un seul nombre simule plusieurs lancers. Inconvénient : la lecture n'est pas très confortable.

2ème idée : Il est possible de régler la précision de l'affichage des nombres (Mode FLOTT n pour l'affichage de n chiffres après la virgule). Par exemple, pour le lancer d'une pièce, on règle l'affichage sur FLOTT 0, ce qui revient à arrondir les nombres à l'unité la plus proche. ***rand*** renvoie alors 0 si le nombre est inférieur à 0,5 (Pile) ou 1 si le nombre est supérieur à 0,5 (Face).



De même, si l'on veut obtenir un nombre aléatoire entre 0 et 99, on règle l'affichage sur FLOTT 2. Avantage : ce que l'on fait ressemble plus à l'expérience réelle, on lance une pièce à chaque appui sur **Entrer** , il y a deux issues, immédiates à décoder, ce qui permettra sans doute aux élèves de mieux être convaincus que cela est conforme au lancer d'une pièce. Inconvénient : à chaque appui de la touche, on ne simule qu'un seul lancer.

3ème idée : En multipliant ***rand*** par un entier n, on obtient un décimal dans [0 ; n[. Or il existe une commande qui renvoie la partie entière d'un nombre, en général ***int*** ou ***partEnt*** (maths, NUM, partEnt). L'instruction ***int***(***rand***xn) ou ***partEnt***(***rand***xn) renvoie donc, de façon équiprobable, un entier compris dans [0 ; n[, c'est à dire 0, 1, 2, …, ou n–1. Ainsi ***partEnt***(***rand***x2) renvoie 0 ou 1 et simule le lancer d'une pièce, ***partEnt***(***rand***x6)+1 renvoie 1, 2, 3, 4, 5, ou 6 et simule le lancer d'un dé cubique, etc... Avec les élèves, il suffit de leur faire expérimenter ***rand***xn, ensuite ***partEnt***(n), puis ***partEnt***(***rand***xn), et enfin ***partEnt***(***rand***xn)+1, en leur faisant formuler à chaque étape ce que l'on obtient.

Remarque : certains modèles de calculatrices disposent d'une commande, en général ***randn*** ou ***randInt*** (maths, RND, randn) qui permet d'obtenir directement la même chose. ***randn***(p;n) ou ***randInt***(p,n) renvoie aléatoirement un nombre entier compris entre p et n. (N.B. sur TI college Plus, le séparateur d'argument est le point-virgule, et le séparateur décimal est la virgule et non le point).

4ème idée : Pour simuler une expérience à deux issues mais non équiprobables (une pièce truquée par exemple), l'une des issues ayant une probabilité égale à p, et l'autre une probabilité égale à 1–p, il suffit d'utiliser l'instruction ***int***(***rand***+p) ou ***partEnt***(***rand***+p). En effet, ***rand***+p renvoie un nombre décimal dans [p;1+p[. Donc ***partEnt***(***rand*** + p) renvoie 0 si le nombre décimal est dans [p;1[et 1 s'il est dans [1;1+p[. Ainsi, 0 code l'issue dont la probabilité est 1–p, et 1 code celle dont la probabilité est p.

Trois exemples d'activités

Exemple 1 : Sondage pour un référendum

Objectif : Dans une situation où la loi de probabilité est parfaitement connue, constater l'existence d'une fluctuation dans les échantillonnages (ici les sondages) et l'ampleur de cette fluctuation selon la taille de l'échantillon (ici le nombre de personnes interrogées).

Situation : un référendum vient d'avoir lieu. 54 % des personnes ont voté OUI, et 46 % ont voté NON.
Question : un sondage à la sortie des urnes aurait-il permis de prévoir cette répartition avant la proclamation officielle ?

De la situation à la simulation : On est dans une situation où la **loi de probabilité est connue** (0,54 ; 0,46).
Pour expliquer la situation aux élèves, on peut expliquer que ce sondage sorti des urnes nous place exactement dans la même situation que si l'on tirait au hasard des cartes bleues ou blanches dans un jeu de 100 cartes dont 54 sont bleues et 46 sont blanches. On peut même réaliser vraiment ce jeu de 100 cartes. Les 54 cartes bleues sont numérotées de 00 à 53, les 46 cartes blanches sont numérotées de 54 à 99. Il y a bien concordance entre le numéro et la couleur, et c'est cette traduction numérique qui sera utilisée dans notre simulation. Quelques essais avec les cartes expliquent clairement le codage.

Réalisation de la simulation : Les élèves, regroupés en binômes, simulent chacun un sondage de taille 50 (par exemple) dans une population virtuelle contenant 54% de OUI et 46% de NON. Pour réaliser cette simulation, on peut utiliser avec les calculatrices la 2ème idée décrite ci-dessus : on règle l'affichage sur FLOTT 2 (2 chiffres après la virgule). A chaque exécution de *rand,* on regarde le nombre à 2 chiffres formé par la partie décimale. Si ce nombre est compris entre 00 et 53, c'est OUI, s'il est compris entre 54 et 99, c'est NON (cela fait bien 54 éventualités pour OUI et 46 pour NON).

Dans l'écran ci-contre, on a obtenu 2 OUI et 2 NON.

Mais on peut également utiliser la 4$^{\text{ème}}$ idée : l'instruction ***int(rand*** + 0,54) ou ***partEnt(rand*** + 0,54). OUI est alors codé par 1 et NON par 0, avec les probabilités voulues.

Une fois les simulations réalisées, les élèves calculent la fréquence observée de OUI sur leur sondage, ainsi que l'écart avec la fréquence réelle (54 %). Dans une classe de 28 élèves, cela fait 14 sondages de taille 50. Le relevé de ces résultats au tableau permet de constater la fluctuation des résultats et l'amplitude des écarts avec la valeur réelle. On fait l'hypothèse que cette amplitude est grande car la taille des échantillons est faible (50). On veut observer ce que l'on obtiendrait avec des échantillons de taille plus grande. Pour cela, il suffit de regrouper 4 sondages de 50 pour obtenir des sondages de taille 200 (on fait ainsi 3 sondages de taille 200 avec 12 sondages de taille 50). La fluctuation est moindre, les écarts avec la valeur réelle sont en général plus réduits. On regroupe finalement tous les sondages pour faire un seul échantillon (de taille 700 dans mon exemple, c'est-à-dire la taille couramment pratiquée dans les sondages d'opinion réels).

A l'aide d'un tableur vidéoprojeté, le professeur peut réitérer plusieurs simulations de sondages de taille 700, puis de taille encore plus grande. Sur tableur, ***rand*** correspond à ***ALEA()***. La formule à écrire est =ENT(ALEA()+0.54) que l'on recopiera sur une plage voulue (A1:N50 par exemple, pour avoir un sondage de taille 700). Pour comptabiliser le nombre d'occurrences DE 1 sur cette plage (c-à-d de OUI), c'est la formule =NB.SI(A1:N50;"=1").

On peut alors constater que la fluctuation est réduite, mais faire la remarque qu'elle ne l'est pas dans les mêmes proportions que la taille de l'échantillon : en passant d'échantillons de taille 50 à 700, la fluctuation n'est (en général) pas divisée par 14. On peut enfin répondre à la question initiale : la plupart des sondages prévoient l'issue du référendum. Mais pas tous ! La proportion étant proche de 50 %, il arrive que le hasard fasse mal les choses.

En conclusion, on peut à la fois montrer l'intérêt des sondages (interroger 700 personnes pour estimer l'opinion de millions de personnes a un sens), mais faire percevoir la prudence à observer quand on en donne le résultat. En particulier, on peut commenter les informations du type « *La cote de popularité du président de la République a perdu 2 points depuis le mois dernier* ». Toutefois, une erreur est à éviter : en insistant de façon excessive sur l'aspect incertain des sondages, on peut laisser penser aux élèves que les sondages sont à proscrire.

Exemple 2 : Le Vrai / Faux

Objectif : Dans une situation où la loi de probabilité n'est pas connue *a priori* (et ne se devine pas intuitivement), la simulation permet de mettre fortement en doute certaines conjectures intuitives, ou éventuellement d'en conforter d'autres, ou encore d'en formuler de nouvelles à la lumière de ce que l'on observe. Puis, la constitution organisée d'une liste des éventualités ou la réalisation d'un arbre de choix, permet d'élucider *a posteriori* la

situation et de trouver la loi de probabilité.

Situation : Dans un devoir, un exercice comporte 4 affirmations pour lesquelles il faut dire si elles sont vraies ou fausses. En répondant complètement au hasard, quelles sont les chances d'avoir toutes les réponses correctes ? Quelles sont les chances d'avoir exactement la moitié des réponses correctes ? Quelles sont les chances d'avoir au moins la moitié des réponses correctes ?

Ce scénario est un grand classique, relevant de la loi binomiale. Mais un élève « débutant » a très peu d'intuition pour conduire un calcul ou un raisonnement solide. Pour la 2ème question, par exemple, il est fréquent qu'il formule la conjecture : il y a 1 chance sur 2 que la moitié des réponses soient correctes. La simulation va mettre en doute cette conjecture, puis une démarche raisonnée va clarifier la situation.

De la situation à la simulation : ici, chaque éventualité est un quadruplet du type (V_1, F_2, F_3, V_4). Les indices, outre qu'ils rappellent le numéro de la question dans l'exercice V/F, permettent de distinguer l'événement « l'affirmation 1 est vraie » de l'événement « l'affirmation 4 est vraie ». En effet, considérer indistinctement l'événement « l'affirmation est vraie » pour les 4 affirmations, ce que l'on voit souvent dans les exercices de ce type, entraîne beaucoup de confusions chez les élèves.

Réalisation de la simulation

On fera remarquer également que l'expression « répondre au hasard » revient au même que s'il s'agissait d'un lancer de pièces répété 4 fois, ou de toute autre situation où l'on répète 4 fois une expérience à 2 issues équiprobables.

On règle la calculatrice pour obtenir un affichage des nombres arrondis à 10^{-4} près (mode FLOTT 4). Après chaque exécution de *rand,* on regarde la parité des 4 chiffres après la virgule. Un chiffre pair code Vrai et un chiffre impair code Faux. Ci contre, on obtient F1 F2 F3 F4 , V1 V2 F3 V4 , et F1 V2 F3 F4 . La séquence de 20 tirages est donc très rapide. Une fois terminé, ne pas oublier de remettre la calculatrice en affichage normal : (mode FLOTT).

V1	V2	F3	F4
F1	F2	F3	F4
F1	V2	V3	V4
…	…	…	…

Les élèves (en binôme) répètent la simulation 20 fois (par exemple) et recensent les résultats dans un tableau comme ci-contre. Le professeur dévoile alors quelle est la séquence correcte (on peut proposer aussi un exercice réel dans lequel les élèves peuvent valider les bonnes réponses). Chaque binôme compte combien de fois il observe le bon quadruplet dans son échantillon de 20. Puis on calcule le nombre d'occurrences du bon quadruplet pour l'ensemble de la classe (avec une classe de 32 élèves, on a ainsi un échantillon de 640 quadruplets aléatoires). La fréquence observée du bon quadruplet donne une estimation des chances d'obtenir un sans-faute en procédant au hasard.

Pour estimer les chances d'avoir exactement, ou au moins la moitié de bonnes réponses, les élèves reprennent leur tableau de 20 quadruplets. Pour chacun d'eux, ils barrent les mauvaises réponses. Par exemple, si le bon quadruplet est F1 V2 V3 V4, on barre comme ci-contre.

~~V1~~	V2	~~F3~~	~~F4~~
F1	~~F2~~	~~F3~~	F4
F1	V2	V3	V4
…	…	…	…

Cela permet ensuite de compter rapidement le nombre d'occurrences de l'événement « *il y a exactement 2 bonnes réponses* », puis de l'événement « *il y a au moins 2 bonnes réponses* ». Là encore, une mise en commun permet d'estimer les fréquences recherchées et d'éprouver les conjectures initiales.

Résolution du problème

Les élèves réalisent la liste exhaustive des 16 quadruplets possibles. Chaque V ou F ayant les mêmes chances d'être attribué pour constituer un quadruplet, on en conclut que les 16 quadruplets ont les mêmes chances d'être obtenus. Pour renforcer cette conviction, un arbre de choix pondéré par les fractions 1/2 permet de représenter le processus de fabrication et la liste des quadruplets. L'approche fréquentiste de la situation joue ici un rôle essentiel car elle permet de donner une interprétation de cette pondération : sur un très grand nombre de tirages aléatoires, **la moitié des cas** contiennent V_1, **parmi eux la moitié des cas** contiennent V_2, etc. Les élèves peuvent repasser en couleur le chemin de l'arbre correspondant au bon quadruplet, dans une autre couleur ceux qui contiennent exactement 2 bonnes réponses, etc. Pour la mise en œuvre, donner aux élèves des arbres vierges évite les « buissons » fantaisistes… Le calcul des probabilités par le quotient du nombre de cas recherchés sur le nombre de cas total permet de connaître les valeurs théoriques des fréquences que l'on compare à celles que l'on a obtenues expérimentalement, et bien sûr à celles que l'on avait conjecturées de façon intuitive au début de l'activité. La place de la simulation est ici centrale : outre qu'elle met en doute les conjectures intuitives, elle rend plus naturel et cohérent le recours à la liste des éventualités, ainsi que la constitution d'un arbre pondéré dont le sens est plus explicite.

Exemple 3 : Croix et pile

Objectif : rencontrer une situation où, contre toute attente, les issues ne sont pas équiprobables, et dans le cadre de ce que le programme a prescrit : « expériences aléatoires à une et *deux épreuves* » (celles à 2 épreuves ne sont pas dans le socle commun).

Situation : Il s'agit d'un jeu pratiqué sous l'ancien régime. Il se joue à 2. Les 2 joueurs déposent chacun une mise. L'un d'eux lance une pièce, si c'est CROIX, il remporte les mises. Si c'est PILE, il relance la pièce. Si c'est CROIX, il remporte les mises, sinon c'est l'autre joueur qui les gagne.
Quelles doivent être les mises de chacun pour que le jeu soit équitable ?[3]

De la situation à la simulation : Les élèves perçoivent que des mises égales ne seraient pas équitables, et admettent intuitivement qu'elles doivent être proportionnelles aux chances respectives des joueurs de gagner.
Les élèves ont pour 1ère tâche de déterminer les différentes issues de ce jeu. Il y en a trois que l'on peut coder par : C, PC, PP. Ils émettent ensuite des conjectures répondant à la question. Il serait surprenant que la répartition 2/3 1/3 ne soit pas envisagée. Si elle n'est pas proposée, on pourra évoquer la correspondance entre Diderot et D'Alembert, ce dernier soutenant la répartition 2/3 1/3. On réalise donc une simulation pour mettre cette conjecture à l'épreuve (ou d'autres s'il y en a).

Réalisation de la simulation :

Pour simuler une partie, il suffit d'exécuter l'instruction ***randn(0;1)*** ou ***partEnt(randx2)*** qui renvoie de façon équiprobable 0 (Croix) ou 1 (Pile) donc simule le lancer d'une pièce. Si c'est 1, il faut donc exécuter à nouveau l'instruction pour lancer la pièce une 2ème fois. Une partie est donc simulée par l'exécution de 1 ou 2 fois l'instruction selon les cas.

[3] Pour en savoir plus sur ce jeu, sur la controverse entre Diderot et D'Alembert, et pour réaliser des simulations en ligne, voir http://irem-fpb.univ-lyon1.fr/feuilles-probleme/feuille11/enonces/croixoupile.html

[4] Les élèves auront à comprendre que ce n'est pas équivalent à **2X*randn(0;1)*** ou **2X*partEnt(randX2)***.

Entre 2 parties, il vaut mieux effacer l'écran pour ne pas se perdre (pour les écrans à plus de 2 lignes d'écriture).

Chaque binôme réalise une série de 100 simulations. Les fréquences observées sur chaque série font apparaître une valeur très souvent au-dessus de 0,7, et ne sont pas réparties autour de 2/3, ce qui est contraire à l'intuition du novice. Pour réduire l'ampleur de la fluctuation d'échantillonnage, on met en commun toutes les séries des élèves pour obtenir une seule série de taille 1400. La fréquence observée reste bien éloignée de 2/3, alors que la taille de la série est conséquente (estimation qualitative qui n'a de sens que si l'on a précédemment rencontré des situations comme celle du sondage, dans lesquelles on a pu quantifier l'ampleur de la fluctuation selon la taille de l'échantillon).

Pour savoir si on peut quand même en imputer le résultat au hasard seul, on peut réaliser une simulation d'une autre expérience aléatoire, à 3 issues dont on est sûr qu'elles sont équiprobables. Les différentes séries de 100 obtenues par les élèves sont cette fois **nettement mieux réparties autour de 2/3**.

Résolution du problème

Pour élucider la situation, on peut proposer de reformuler le scénario de la façon suivante : on joue avec 2 pièces, l'une rouge, l'autre bleue, que l'on jette simultanément. Si la pièce rouge est Croix, le joueur gagne, que la pièce bleue soit Croix ou Pile. Si la pièce rouge est Pile, on regarde la pièce bleue. Si elle est Croix, le joueur gagne, sinon il perd.

Pour convaincre de l'équivalence des deux scénarios, on pourra faire l'arbre pondéré de ce nouveau scénario, puis effacer les deux branches « inutiles », ce qui donne l'arbre du 1^er^ scénario. La répartition 3/4 1/4 est établie, et elle est conforme aux estimations obtenues avec les premières séries de simulations.

Enfin, on peut réaliser une simulation de ce 2^nd^ scénario en exécutant l'instruction ***randn(0;1) + randn(0;1)*** ou ***partEnt(randX2) + partEnt(randX2)***[4]. Le tableau ci-dessous montre la conformité de l'instruction avec le scénario et renforce la compréhension de la répartition 3/4 1/4.

(R;B)	(0;0)	(0;1)	(1;0)	(1;1)
Somme	0	1	1	2
Issue obs.	0	1		2
Partie	Gagnée			Perdue

Les résultats des séries de simulations sont bien conformes à ceux obtenus dans la simulation du 1^er^ scénario (les fréquences sont proches de 0,75). Dans cette activité, les simulations ont eu à nouveau un rôle central puisqu'elles ont permis d'invalider la conjecture initiale (en établissant en particulier que la fluctuation ne pouvait expliquer le décalage observé), mais également de mieux comprendre le raisonnement qui a prévalu pour résoudre le problème.

Parmi les inquiétudes initiales des professeurs de seconde en 2000, il y avait celle de penser que les élèves ne prendraient pas au sérieux ces activités, estimant qu'elles manquent de consistance et les jugeant peu dignes du cours de mathématiques. L'expérience permet de faire un démenti formel. Les questions qui émergent chez les élèves sont souvent surprenantes et pertinentes. De plus, cet enseignement est particulièrement propice à une démarche sous forme de travail en groupe, de mise en commun, de débats, étayés par une expérimentation qui interroge sur la nature et le sens de la formalisation mathématique.

RÉFÉRENCES

ENSEIGNER LA STATISTIQUE AU LYCÉE : DES ENJEUX AUX MÉTHODES

par J.L. PIEDNOIR et P. DUTARTE.
Pour les professeurs : très clair, riche (applications, historique, exemples, ...). « Il est rare, dit P.L.Hennequin, de trouver un recueil aussi complet et aussi agréable à lire ».
Ed. IREM de Paris-Nord. Co-diffusion APMEP n° 820. Prix : 14 €

PROBABILITÉS AU LYCÉE

2ème édition fortement mise à jour par la commission Inter-IREM, d'une édition de 1979.
Ouvrage collectif sous la direction de Brigitte Chaput et Michel Henry.
Neuf articles aux signatures prestigieuses, sur l'évolution de l'enseignement des probabilités de 1965 à 2002, des analyses d'expérimentations, des modèles d'urnes pour les lois discrètes, les arbres, l'analyse en probas, modélisation et simulation, l'inférence statistique (principes, exemples).
Brochure APMEP n° 143 - Prix public : 13 € ; adhérent : 9 €.

STATISTIQUES AU LYCÉE (2 Volumes)

Commission Inter-IREM Statistique et Probabilités.
Coordination et réalisation : Brigitte CHAPUT et Michel HENRY.
Après Probabilités au lycée, la Commission Inter-IREM Statistique et Probabilités a entrepris un travail de fond sur l'enseignement de la statistique tel qu'il est conçu dans le cadre des programmes des années 2000 des lycées.
Cette publication Statistique au lycée est livrée en deux volumes.
Volume 1 : Les outils de la statistique
Ce volume est conçu comme une introduction, un débat et un élargissement autour des questions d'enseignement soulevées par les objectifs et la démarche adoptés dans l'ensemble des programmes de la seconde à la terminale.
Brochure APMEP n° 156. Prix public : 13 € ; adhérent : 9 €
Volume 2 : Activités statistiques pour la classe
Le second volume, paru en 2006, est plus centré sur les questions relatives à l'échantillonnage, aux situations de sondage, ainsi qu'à des exemples de simulation avec ou sans tableur.
Brochure APMEP n° 167. Prix public : 14 € ; adhérent : 10 €.

LES STATISTIQUES EN CLASSE DE SECONDE

par Pascale POMBOURCQ.
Cette brochure aborde aussi bien les aspects théoriques que pratiques de l'enseignement des statistiques dans cette classe. La réédition a permis à l'auteur d'adapter les chapitres traitant des fluctuations d'échantillonnage et des promenades aléatoires à une meilleure utilisation en classe. Un outil réconfortant.
Brochure APMEP n° 138 (2001). Prix public : 9 € ; adhérent : 6 €.

PRATIQUES DE LA STATISTIQUE. EXPÉRIMENTER, MODÉLISER ET SIMULER

Livre de l'Irem de Grenoble, avec onze intervenants, coordonné par Claudine SCHWARTZ.
Éditions Vuibert. Préface de J.-P. Raoult.
Co-diffusion APMEP n° 940. Prix public : 34 € ; adhérent 32,50 €.

CONTES ET DÉCOMPTES DE LA STATISTIQUE

par Claudine Robert
Une initiation par l'exemple avec, notamment, six leçons de statistique descriptive et trois de statistique inférentielle.
Ed. Vuibert. Réédition, revue, de « L'empereur et la girafe ».
Co-diffusion APMEP n° 930. Prix public : 20 €, prix adhérent : 19 €.

EN PASSANT PAR HASARD... LES PROBABILITES DE TOUS LES JOURS

par G. PAGES et C. BOUZITAT.
Des problèmes de la vie courante, en quatre parties indépendantes, et des développements récents.
Co-diffusion APMEP n° 907. Ed. Vuibert. Prix : 24,50 €.