DE L'EXPERIENCE ALEATOIRE AU MODELE, *via* LA SIMULATION

Bernard PARZYSZ
Irem Paris-Diderot

Résumé. Dès la classe de Troisième, les programmes de statistique mettent l'accent sur la notion d'expérience aléatoire de référence et sur la simulation d'expériences à l'aide de la calculatrice et de l'ordinateur. Mais simuler suppose en principe de se référer à un modèle probabiliste, concept auquel les élèves n'accèderont qu'à partir de la classe de Première. Cette difficulté de fond peut cependant être contournée en travaillant plus particulièrement, dans les premiers temps, deux points de la démarche de simulation : la mise en œuvre de procédures de simulation qui soient aussi « isomorphes » que possible à l'expérience simulée et l'explicitation des hypothèses qui sous-tendent ces procédures. Le repérage des analogies présentées par des expériences diverses grâce à l'étude des procédures mises en œuvre pour les simuler — notamment à l'aide du tableur qui se prête bien aux comparaisons — amène à la notion de schéma d'expérience, préparant l'accès à celle de modèle probabiliste.

A propos de la notion de probabilité, le programme de troisième parle de « *situations de la vie courante pouvant être modélisées simplement…* ». Il s'agit donc de modéliser des expériences aléatoires, et par conséquent de faire prendre conscience aux élèves — et de les convaincre — de ce que peuvent avoir en commun des expériences fort diverses, réalisées à partir de lancers de pièces de monnaie ou de dés, de tirages de boules d'un sac, etc., c'est-à-dire de prendre conscience de l'existence d'un même modèle d'expérience qui leur est associé. Cette démarche ne va pas de soi, et pour la mettre en œuvre je propose ici, dans l'esprit des programmes de statistique du lycée, le recours à la simulation à l'aide d'un tableur, qui selon moi peut constituer un bon intermédiaire sur la voie de la modélisation, en faisant apparaître l'analogie entre diverses expériences lors de l'élaboration de la procédure de simulation. Le sujet concernant en fait également le lycée, je ne me cantonnerai pas à la seule classe de troisième, et pour introduire et illustrer mon propos je m'appuierai sur quelques exemples « classiques » d'expériences aléatoires.

1. Un premier exemple historique

Il s'agit du jeu de Croix ou Pile, auquel d'Alembert consacra une entrée de l'Encyclopédie, et qui m'avait servi de support pour une expérimentation sur l'approche fréquentiste en classe de Première ES [Parzysz 2007].

Ce jeu se pratique à deux joueurs, A et B, et ne nécessite qu'une pièce de monnaie, dont les deux faces sont désignées sous le nom de « croix » et « pile ». Une partie se joue en un ou deux lancers de la pièce, au cours desquels on observe quelle face apparaît. Le joueur A jette la pièce, et si c'est « croix », il a gagné ; par contre, si c'est « pile » A relance la pièce, et si cette fois il fait « croix » il a gagné ; sinon, c'est B qui a gagné. La question que pose d'Alembert est : Quelle chance chacun des deux joueurs a-t-il de gagner la partie ? Il propose alors deux solutions :

— la première, celle « *qu'on trouvera dans tous les auteurs* », consiste à dire qu'il y a quatre combinaisons (croix au premier coup et croix au second, pile au premier coup et croix au second, croix au premier coup et pile au second, pile au premier coup et pile au second), et que « *de ces quatre combinaisons une seule fait perdre et trois font gagner* », et donc que A possède 3 chances sur 4 de gagner la partie.

— la seconde, qui semble avoir la préférence de d'Alembert, part de la remarque que « *dès qu'une fois croix est venue, le jeu est fini, et le second coup est compté pour rien* », ce qui conduit à ne considérer que trois combinaisons : croix au premier coup, pile au premier coup et croix au second, pile au premier coup et pile au second. Il en résulte que A possède 2 chances sur 3 de gagner la partie.

L'intérêt de proposer une telle situation à la réflexion d'une classe est que la plupart des élèves pensent certes que le joueur A a plus de chances de gagner que le joueur B, mais qu'ils n'ont pas beaucoup d'idées pour évaluer et comparer les chances de l'un et de l'autre. Chacune des deux solutions proposées par d'Alembert est donc susceptible d'emporter l'adhésion d'une partie de la classe, et c'est effectivement ce qui se passe [1] [Parzysz, *op. cit.*]. En outre, l'idée de jouer un certain nombre de parties apparaît spontanément. Cette démarche peut être « officialisée » par l'enseignant et conduire, par cumul des résultats, à discréditer par l'expérience certaines des réponses envisagées. Il s'agit bien de rejeter comme non plausibles certaines solutions, et non pas d'en valider une, car le fait d'obtenir par exemple une fréquence de succès de 0,76 sur 520 parties jouées permet de dire que ce résultat rend la réponse « 2 chances sur 3 » beaucoup moins vraisemblable que la réponse « 3 chances sur 4 », mais ne permet pas pour autant d'affirmer que celle-ci est exacte, étant donné qu'on ne sait rien de la pièce et qu'une autre série de parties fournirait sans doute un résultat différent.

2. Expérience et simulation

Mais la réalisation effective d'un grand nombre de parties est très longue, et c'est pourquoi on peut se tourner vers la simulation. Notons au passage que, même si c'est presque toujours le cas aujourd'hui, la simulation ne nécessite pas le recours aux technologies, et qu'on a simulé des expériences aléatoires bien avant l'invention de la calculatrice et de l'ordinateur. Ce fut par exemple le cas au 19ème siècle pour Walter Weldon [2], qui s'intéressa à la fréquence d'apparition des diverses faces d'un dé lors de lancers successifs. Dans

1 Il apparaît même parfois une troisième solution, selon laquelle chaque joueur a une chance sur deux de gagner, solution sans doute basée sur le même principe que celui qui consiste, dans une situation d'incertitude, à attribuer par défaut la même probabilité à toutes les éventualités (« *Demain, il y a une chance sur deux pour qu'il pleuve* »).

2 Walter Weldon (1860 – 1906), zoologiste, créateur de la biométrie. Il utilisa les techniques que Francis Galton venait de développer à l'occasion de travaux sur la morphologie de la crevette et travailla également avec Karl Pearson.

la littérature qui évoque la désormais fameuse « expérience de Weldon », on trouve tantôt la mention de 315 672 lancers d'un dé, et tantôt celle de 26 306 lancers de 12 dés ($26\,306 \times 12 = 315\,672$). De là à conclure que ce n'est pas un seul dé que lançait Weldon, mais plutôt des poignées de 12 dés, il n'y a qu'un pas qu'on peut à mon avis franchir sans ambages…

Mais qu'est-ce que simuler ? Une recherche dans les manuels de Seconde permet d'y distinguer quatre groupes de réponses à cette question :

— un premier groupe se borne à relever l'idée de *substitution* : « *la simulation remplace l'expérience* » [3] ;

— un deuxième groupe met en avant l'*analogie* qui doit exister entre l'expérience initiale et la simulation : « *simuler une expérience aléatoire, c'est la remplacer par une autre expérience aléatoire qui permet de produire des résultats que l'on obtiendrait par la réalisation pratique de la première* » [4] (ce qui ne dit hélas pas comment on peut s'assurer de cette similitude de résultats) ;

— un troisième groupe souligne l'idée *d'économie de moyens* réalisée avec la simulation : « *simuler une expérience, c'est remplacer cette expérience par une autre, plus économe, plus rapide et qui permet d'obtenir des résultats identiques* » [5] (même remarque) ;

— enfin, un dernier groupe convoque l'idée de *modèle théorique* (« *simuler une expérience, c'est choisir un modèle pour cette expérience* » [6]).

3 Collection Repères, Hachette 2004 (p. 210).
4 Collection Pyramide, Hachette 2000 (p. 170).
5 Belin 2000 (p. 170).
6 Collection Modulo, Didier 2004 (p. 170).

Pour synthétiser ces différents aspects de la simulation, on peut dire en définitive qu'elle consiste à remplacer une expérience aléatoire qu'on se propose d'étudier par une autre expérience, plus facile et/ou rapide à mettre en œuvre, mais qu'il faut pour cela s'assurer que cette simulation reflète statistiquement l'expérience initiale, l'adéquation entre les deux étant en quelque sorte assurée par la qualité du modèle probabiliste déterminant la simulation. Ce que confirme le statisticien Yadolah Dodge, qui donne la définition suivante :

> « *La simulation est la méthode statistique permettant la reconstitution fictive de l'évolution d'un phénomène. C'est une expérimentation qui suppose la construction d'un modèle théorique présentant une similitude de propriétés ou de relations avec le phénomène faisant l'objet de l'étude* ». [Dodge 1993]

Ceci ne veut cependant pas dire que pour simuler une expérience on doit connaître son modèle probabiliste (alors il n'y aurait plus de problème, sauf de calcul). La démarche est en fait la suivante : vu le protocole expérimental et en fonction des outils que j'ai à ma disposition, je propose un modèle pour cette expérience. Les données expérimentales, confrontées aux simulations de ce modèle, me permettent ensuite (ou non) d'accepter ce modèle comme adéquat ou de le rejeter. Le document d'accompagnement du programme de Seconde précise cette position par une formule lapidaire : « *simuler une expérience c'est choisir un modèle de cette expérience puis simuler ce modèle* ». La référence à un modèle probabiliste est donc — implicitement ou explicitement — sous-jacente à l'idée de simulation. Ainsi, Weldon admet implicitement que la probabilité P(n) d'obtenir la face

n (avec $n = 1, 2,,\ldots,6$) est la même pour chacun des 12 dés, et que lancer 12 dés simultanément revient au même que lancer 12 fois le même dé. Je reviendrai plus loin plus en détail sur les rapports entre simulation et modèle, mais je dirai d'abord quelques mots sur la simulation « technologique », c'est-à-dire celle qui fait intervenir la fonction *random* de la calculatrice ou la fonction *aléa* du tableur de l'ordinateur.

Elle contredit en fait l'une des idées indiquées ci-dessus, puisque dans ce cas la simulation *n'est pas* une expérience aléatoire [7] ; en effet, les nombres produits par la machine sont les termes successifs d'une suite récurrente, et sont donc parfaitement déterminés [Parzysz 2005]. Ils sont cependant supposés être « équirépartis », ce qui signifie qu'à chaque appui sur la touche chacun des 10^n nombres possibles a une chance sur 10^n d'apparaître [8].

Mais certains manuels proposent, à l'aide du seul nombre affiché, de simuler n lancers successifs d'une pièce de monnaie équilibrée en considérant séparément chaque chiffre et en décidant par exemple que 0, 1, 2, 3, 4 correspondront à « pile » et 5, 6, 7, 8, 9 correspondront à « face », ce qui n'a de sens que si la répartition de chacun des chiffres du nombre est elle-même équirépartie sur $\{0, 1, \ldots,9\}$. Ceci pose la question de l'utilisation des nombres produits par la machine : si on peut légitimement considérer chacun des nombres obtenus comme résultant d'un tirage au hasard parmi les 10^n nombres possibles, peut-on pour autant considérer ses chiffres successifs comme résultant de tirages au hasard parmi l'ensemble $\{0, 1, 2, \ldots, 9\}$? De même, peut-on considérer les « tranches » successives de deux chiffres comme résultant de tirages au hasard parmi l'ensemble $\{00, 01, \ldots, 98, 99\}$? Il s'agit là d'une hypothèse qu'il est nécessaire de faire (et qui peut être testée). En définitive, le recours à la simulation technologique nécessite donc un *acte de foi* ou (selon les positions idéologiques) une certaine confiance dans les garanties fournies par les concepteurs des machines et des logiciels : l'acceptation, non seulement de l'hypothèse que la suite de nombres produite est équirépartie, mais aussi de l'hypothèse qu'il en va de même pour la suite des chiffres obtenue en concaténant dans l'ordre les décimales de ces nombres (ainsi que pour la suite des couples de chiffres, etc.). L'« acte de foi » comporte en outre un aspect moins visible : on suppose l'« absence de mémoire » de la suite considérée, c'est-à-dire qu'un élément quelconque de cette suite (nombre, chiffre, couple de chiffres, etc.) ne dépend pas de ceux qui le précèdent [9].

3. Simulation et modèle

Nous avons vu que le choix d'une « bonne » simulation présuppose l'association d'un modèle probabiliste à l'expérience étudiée, modèle qui servira ensuite à déterminer la simulation. Prenons par exemple le cas élémentaire de la simulation d'une série de lancers successifs d'une pièce de monnaie pour étudier la distribution des fréquences d'apparition de ses deux faces, et faisons les hypothèses suivantes :

(H1) Un résultat sera « pile » ou « face » selon le côté de la pièce visible à l'issue du lancer.

7 Il en va de même pour les tables de hasard, puisqu'elles sont produites par ordinateur.

8 On se place ici dans le cas où l'affichage de la machine fournit un nombre de la forme 0,xxx… avec n décimales.

9 Dans ce cas, la suite est dite « bien enchevêtrée », ce qui signifie que les écarts entre deux occurrences successives d'un même élément suivent une loi géométrique.

(H2) Si la pièce tombe sur la tranche, le lancer sera annulé.

(H3) La pièce est « équilibrée », c'est-à-dire que « pile » et « face » ont autant de chances d'être obtenus.

(H4) On lance la pièce suffisamment fort pour que le résultat ne soit pas prévisible.

Nous avons ainsi constitué ce que Michel Henry nomme un « *modèle pseudo-concret* », c'est-à-dire « *une sorte d'abstraction et de simplification de la réalité perçue dans sa complexité, dans la mesure où certains choix sont faits pour ne retenir que ce qui semble pertinent de cette situation vis-à-vis du problème étudié* » [Henry 1999] [10]. Dans le cas présent, nous ne nous intéressons pas au type de pièce (monnaie, valeur faciale…), nous la supposons équilibrée (H3), nous négligeons le cas où elle tomberait sur la tranche (H2) et nous définissons un protocole d'expérience (H4).

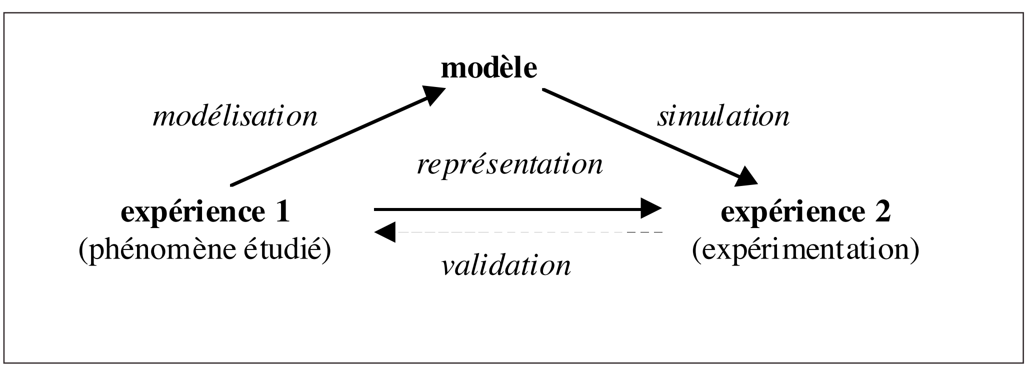
Les hypothèses ci-dessus vont alors être traduites en termes de productions du générateur aléatoire, par l'application consistant à associer à chaque nombre x fourni par la machine l'une des deux issues possibles pour la pièce :

(H1) et (H2) — A $x < 0{,}5$ on associe « face », et à $x \geq 0{,}5$ on associe « pile » (par exemple),

(H3) est *ipso facto* vérifiée puisqu'on suppose l'équirépartition des nombres (« acte de foi »),

(H4) correspond au caractère « aléatoire » du générateur (*id°*).

La correspondance entre expérience et simulation correspond donc au schéma ternaire ci-dessous. Mais le problème est qu'il faut attendre la classe de Première pour aborder véritablement la notion de modèle, et la question de savoir si une simulation est acceptable demeure entière à ce stade. Pour cette raison, le document d'accompagnement de Seconde préconise de court-circuiter le modèle dans un premier temps : « *Pour les élèves, une première étape importante est d'établir un lien entre l'expérience et une simulation de cette expérience.* » (p. 21), et, pour pouvoir comparer expérience et simulation, d'avoir recours au dispositif suivant : « *la classe pourra être partagée en deux : dans un groupe, les élèves feront effectivement des expériences, dans l'autre, les élèves simuleront l'expérience.* » (*ibid.*). Ceci doit effectivement permettre d'établir si une simulation fournit bien les mêmes résultats que



10 Sur ce point voir aussi [Henry 2001].

l'expérience, mais, si l'on veut aller plus loin en direction de la notion de modèle, il est nécessaire de l'accompagner par une réflexion se proposant d'expliciter certaines des hypothèses qui sous-tendent expérience et simulation, ainsi que la façon dont elles se correspondent (hypothèses « de modèle »), comme nous l'avons fait ci-dessus.

4. Procédures de simulation

La mise en œuvre d'une simulation sur calculatrice ou tableur nécessite l'établissement d'une procédure qui sera implémentée sur la machine, et il arrive fréquemment que plusieurs procédures différentes puissent être envisagées. Considérons l'exemple suivant, extrait d'un manuel de Seconde [11] :

> *Énoncé*
>
> A l'aide d'une table de nombres au hasard, simuler l'expérience suivante : on forme au hasard une famille de deux enfants, puis on s'intéresse à la composition de la famille (deux garçons, deux filles, une fille et un garçon) (…)
>
> *Solution*
>
> Pour simuler cette expérience, on peut convenir que tout [chiffre] pair de la table permet d'obtenir un garçon et que tout [chiffre] impair permet d'obtenir une fille. On partage la suite de nombres de la table en tranches de deux chiffres : chaque tranche fournit une famille de deux enfants. (…)

On peut relever dans la solution deux hypothèses « de modèle » implicites qui sous-tendent la simulation effectuée grâce à la table au hasard, hypothèses qui résultent de notre « acte de foi » :

1° Du fait que, suite à l'équirépartition supposée, on a autant de chances d'obtenir un chiffre pair qu'un chiffre impair, il résulte que, pour chaque naissance, on a autant de chances d'avoir un garçon qu'une fille.

2° Du fait qu'on a admis qu'un chiffre ne dépend pas de ceux qui le précèdent, il résulte que, pour chaque famille, le genre du second enfant n'est pas lié à celui du premier enfant.

On peut éventuellement s'interroger sur le bien-fondé de ces hypothèses, mais il n'en reste pas moins que ce sont celles qui sont faites sur la naissance dans les familles de deux enfants, et que leur identification constitue un premier pas en direction du modèle.

Revenons maintenant au jeu de Croix ou Pile. On peut *a priori* envisager plusieurs procédures de simulation pour cette expérience, qui toutes supposent la pièce équilibrée. Par exemple :

Procédure 1. Produire un aléa [12] à deux valeurs équiréparties C et P :

— si on obtient C, on a gagné ;

— sinon, produire un autre aléa du type précédent :

  - si on obtient C, on a gagné ;
  - sinon, on a perdu.

La réalisation de cette procédure sur un tableur fournit un tableau du type donné ci-contre (*Tableau 1*) ; on choisit ici d'associer C à un nombre inférieur à 0,5 et P à un nombre supérieur ou égal à 0,5).

[N.B. Il convient bien sûr de contrôler la pertinence de la simulation, c'est-à-dire la com-

11 Collection Indice, Bordas 2004 (p. 165).

12 J'entends par « produire un aléa » le fait de demander à la machine de fournir un nombre au hasard, c'est-à-dire appuyer sur la touche random (calculatrice) ou faire agir la fonction aléa (tableur).

Partie	Aléa 1	Résultat	Aléa 2	Partie
1	0,858	relancer	0,691	perdu
2	0,212			gagné
3	0,035			gagné
4	0,963	relancer	0,289	gagné
5	0,631	relancer	0,807	perdu

Tableau 1

Partie	Aléa 1	Aléa 2	Partie
1	0,858	0,691	perdu
2	0,212	0,978	gagné
3	0,035	0,724	gagné
4	0,963	0,289	gagné
5	0,631	0,807	perdu

Tableau 2

parer avec l'expérience réelle et s'assurer qu'elle n'est pas en contradiction avec elle.]

Procédure 2. Produire deux aléas à deux valeurs équiréparties C et P :

— si les deux produisent P, on a perdu ;

— sinon, on a gagné.

On obtient cette fois un tableau du type ci-dessus (*Tableau 2*), dans lequel on a opéré le même choix que dans la première procédure.

Comme on le voit, cette seconde procédure est différente de la première. Elle est certes plus simple à mettre en œuvre (on n'a pas deux tests « SI » imbriqués), mais elle fait intervenir un second lancer qui n'a pas toujours réellement lieu. On peut alors se poser la question de savoir si cette « entorse » aux conditions de l'expérience ne risque pas de poser problème à certains élèves. C'est ce que nous allons tenter de voir maintenant sur un exemple célèbre.

5. Un autre exemple historique

La question posée ci-dessus n'est pas de pure forme, car elle s'est posée historiquement, avec, dans le rôle du héros involontaire, Gilles Personne de Roberval.

La notion de probabilité, on le sait, est censée avoir pris naissance à l'occasion d'une correspondance de 1654 entre Blaise Pascal et Pierre Fermat. Au cours de ces échanges, les deux savants s'intéressent à plusieurs jeux de hasard, et notamment au fameux « problème des partis », qui avait déjà fait couler beaucoup d'encre à l'époque. Voici de quoi il s'agit : deux joueurs jouent l'un contre l'autre à un jeu de hasard équitable. Le gagnant sera celui qui le premier aura remporté un nombre fixé de parties. Or, le jeu doit être interrompu alors que qu'il manque deux parties à l'un et trois à l'autre. Comment répartir les mises ?

Ce qui va nous occuper plus particulièrement ici est une lettre de Pascal, datée du 24 août. Celui-ci commence par énoncer le problème et expose la solution de Fermat, dont la première étape est de déterminer en combien de parties au maximum le jeu sera nécessairement terminé :

> « *Voici comment vous procédez, quand il y a deux joueurs. Si deux joueurs jouant en plusieurs parties, se trouvent en cet état qu'il manque deux parties au premier et trois au second, pour trouver le parti, il faut (dites-vous) voir en combien de parties le jeu sera décidé absolument.* »

Puis, Pascal explique pourquoi — toujours selon Fermat — il suffira de quatre parties :

> « *Donc pour voir combien quatre parties se combinent entre deux joueurs, il faut ima-*

giner qu'ils jouent avec un dé à deux faces (puisqu'ils ne sont que deux joueurs), comme à croix et pile, et qu'ils jettent quatre de ces dés (parce qu'ils jouent en quatre parties) ; et maintenant il faut voir combien ces dés peuvent avoir d'assiettes différentes. »

Pascal écrit alors les $2^4 = 16$ possibilités (« *assiettes* »), qui montrent que cinq d'entre elles sont favorables à l'un des joueurs et onze à l'autre. D'où la répartition des mises. Mais c'est la suite qui est digne d'intérêt, eu égard à mon propos :

« *Je communiquai votre méthode à nos messieurs ; sur quoi M. de Roberval me fit cette objection.*

Que c'est à tort que l'on prend l'art de faire le parti, sur la supposition qu'on joue en quatre parties ; vu que quand il manque deux parties à l'un et trois à l'autre, il n'est pas de nécessité que l'on joue quatre parties, pouvant arriver qu'on n'en jouera que deux ou trois, ou, à la vérité, peut-être quatre ; et ainsi qu'il ne voyoit pas pourquoi on prétendait de faire le parti juste sur une condition feinte, qu'on jouera quatre parties ; vu que la condition naturelle du jeu est qu'on ne jouera plus dès que l'un des joueurs aura gagné ; et qu'au moins si cela n'étoit faux, cela n'étoit pas démontré. De sorte qu'il avait quelque soupçon que nous avions fait un paralogisme. »

La raison pour laquelle Roberval récuse la méthode de Fermat est que celle-ci met en jeu des « lancers virtuels ». Il considère que le jeu original, telle qu'il se déroule en réalité, et sa simulation par Fermat à l'aide du lancer simultané de quatre pièces, constituent des expériences différentes, qui ne peuvent se réduire l'une à l'autre. Cette objection s'applique également à la procédure 2 envisagée plus haut pour simuler le jeu de Croix ou Pile. En effet, cette procédure fait intervenir un second aléa *dans tous les cas*, même lorsque c'est inutile parce que le joueur A a déjà gagné suite au premier lancer. Ce qui a gêné Roberval risque également de gêner certains élèves, s'ils ne sont pas convaincus que la seconde procédure se ramène à la première. Il convient alors de « négocier » le passage de la première procédure à la seconde, et pour cela la comparaison des tableaux 1 et 2 peut se révéler utile [13]. On peut en effet constater que dans le second tableau la colonne « Résultat » a disparu : elle est devenue inutile puisqu'on rejoue systématiquement. De plus, dans la colonne « Aléa 2 » du premier tableau, certaines cases sont vides : celles correspondant aux cas où l'on a déjà gagné au premier lancer. Les résultats du tableau 2 figurant dans ces cases sont donc inutiles, mais elles ne changent rien pour le reste. C'est d'ailleurs cet argument que rétorque Pascal à Roberval :

« *Je lui dis donc ainsi :*
N'est-il pas clair, que les mêmes joueurs n'étant pas astreints à jouer les quatre parties, mais voulant quitter le jeu dès que l'un aura atteint son nombre, peuvent, sans dommage, ni avantage, s'astreindre à jouer les quatre parties entières, et que cette convention ne change en aucune manière leur condition ? Car si le premier gagne les deux premières parties de quatre, et qu'ainsi il ait gagné, refusera-t-il de jouer encore deux parties, vu que s'il les gagne, il n'a pas mieux gagné ; et s'il les perd, il n'a pas moins gagné. (…)
Certainement il est aisé de considérer qu'il est absolument égal et indifférent à l'un et à l'autre de jouer en la condition naturelle

13 C'est bien sûr à dessein que j'ai utilisé les mêmes données numériques dans les deux tableaux.

à leur jeu, qui est de finir dès qu'on aura son compte, ou de jouer les quatre parties entières ; donc, puisque ces deux conditions sont égales et indifférentes, le parti doit être tout pareil en l'un et en l'autre. Or il est juste quand ils sont obligés de jouer quatre parties comme je l'ai montré ; donc il est juste aussi en l'autre cas.»

En appliquant ceci à nos deux procédures, on peut dire, à l'instar de Pascal, que la procédure 1 correspond à la « *condition naturelle* » du jeu, c'est-à-dire au jeu tel qu'il se pratique effectivement, tandis que la procédure 2 introduit un artifice (les « lancers virtuels ») destiné à faciliter la mise en place de la simulation. Mais cette modification peut rendre nécessaire de montrer, comme on l'a vu avec les tableaux, que son introduction ne modifie en rien les résultats.

6. Expériences, simulations et procédures

Revenons maintenant à l'actualité de l'enseignement secondaire. Le document d'accompagnement de Seconde préconise l'établissement d'expériences de référence pour les élèves : « *Pour les élèves, lancer des pièces, des dés ou obtenir une liste de nombres au hasard avec une calculatrice sont des expériences aléatoires de référence.* » (p. 21). Parmi celles-ci on peut sans doute inclure les expériences ci-contre. Pour simuler chacune de ces quatre expériences, nous allons nous efforcer de déterminer une procédure aussi « proche » que possible de celle de l'expérience elle-même, et nous la mettrons en œuvre sur le tableur.

(E1) *on lance 4 fois successivement une pièce de monnaie et on note le nombre de piles obtenu.*

(E2) *on lance simultanément 4 pièces de monnaie et on note le nombre de piles obtenu.*

(E3) *dans un sac contenant 10 billes (5 blanches et 5 noires) on effectue 4 tirages successifs d'une bille (avec remise), et on note le nombre de billes blanches obtenu.*

(E4) *on laisse tomber une bille dans une planche de Galton à 5 godets et on note le numéro du godet dans lequel elle tombe.*

Chacune de ces expériences est recommencée un certain nombre de fois. Quelle sera la répartition des nombres notés ?

Expérience (E1)

Hypothèses. La pièce est équilibrée. Un lancer n'est pas affecté par les précédents et est sans influence sur les suivants.

Procédure

a) Produire 4 fois successivement un aléa à deux valeurs équiréparties P (pile) et F (face) (on décide par exemple qu'un nombre inférieur à 0,5 correspondra à « face » et qu'un nombre supérieur ou égal à 0,5 correspondra à « pile »).

b) Totaliser le nombre de P obtenus.

Tableau

Aléa	Résultat	Nombre de piles
0,652	P	
0,837	P	
0,091	F	
0,763	P	3

Tableau 3

Expérience (E2)

Hypothèses. Les pièces sont équilibrées. Aucune n'influe sur une autre.

Aléa 1	*Résultat 1*	*Aléa 2*	*Résultat 2*	*Aléa 3*	*Résultat 3*	*Aléa 4*	*Résultat 4*	*Nombre de piles*
0,652	P	0,837	P	0,091	F	0,763	P	3

Tableau 4

Procédure

a) Produire 4 aléas à deux valeurs équiréparties P et F.

b) Totaliser le nombre de P obtenus.

(Voir Tableau 4)

Expérience (E3)

Hypothèses. A chaque tirage toutes les boules ont la même chance d'être tirées, donc on a autant de chances de tirer une boule blanche qu'une noire. Un tirage n'est pas affecté par les précédents et est sans influence sur les suivants.

Procédure

a) Produire 4 fois successivement un aléa à 2 valeurs équiréparties B (blanche) et N (noire).

b) Totaliser le nombre de B obtenus.

Tableau

Aléa	Résultat	Nombre de blanches
0,652	B	
0,837	B	
0,091	N	
0,763	B	3

Tableau 5

Expérience (E4)

La planche qui porte son nom fut imaginée par Francis Galton [14] pour illustrer la distribution binomiale. Ce dispositif consiste en une planche verticale munie de clous placés en quinconce et de godets (*fig. 1*). Lorsqu'on laisse tomber une bille au-dessus du clou supérieur, elle rebondit sur les clous et à chaque fois part, soit vers la droite, soit vers la gauche ; elle aboutit enfin dans un des godets.

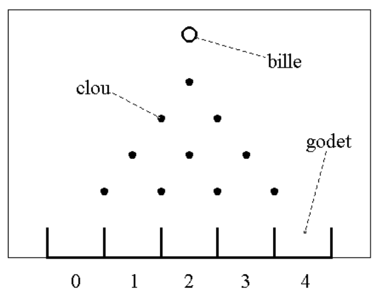


Figure 1

En fait, la planche de Galton n'est qu'une réalisation imparfaite du modèle binomial [Parzysz 2005].

Hypothèses. Sur chaque clou, une bille a autant de chances de partir vers la droite que vers la gauche. Son trajet ne dépend pas de ceux des billes lancées avant elle et est sans influence sur le trajet des suivantes.

Procédure

a) Produire 4 fois successivement un aléa à 2 valeurs équiréparties D (droite) et G (gauche) ;

b) Totaliser le nombre de D obtenus (ce qui donne le numéro du godet où aboutit la bille).

14 Francis Galton (1822-1911), cousin de Charles Darwin, fut un touche-à-tout de génie (il a entre autres inventé le sac de couchage). Il s'est intéressé à la géographie, à la psychologie, à la biologie et à la statistique. Karl Pearson fut son disciple.

Tableau :

Aléa	*Résultat*	*Numéro du godet*
0,652	D	
0,837	D	
0,091	G	
0,763	D	3

Tableau 6

La comparaison des procédures fait apparaître que celles de (E1), (E3) et (E4) sont identiques aux notations près ; cependant, identifier les procédures de (E1) et de (E2) nécessite d'admettre que produire quatre aléas équivaut à produire quatre fois successivement un même aléa, ce qui est vrai en vertu des propriétés attribuées au générateur mais peut ne pas paraître évident à tous 15. La comparaison des quatre tableaux fait plus clairement apparaître leur analogie : en effet, en raison de l'analogie des procédures, les tableaux 3, 5 et 6 sont identiques aux notations près, mais le tableau 4 apparaît simplement comme la version « horizontale » du tableau « vertical » 3. La conjugaison du registre langagier des procédures et de celui des tableaux apparaît ainsi susceptible d'aider les élèves à se persuader, au-delà des différences de surface, de l'analogie profonde entre les 4 expériences, à travers leurs simulations, qui correspondent à un même *schéma d'expérience*.

Plus précisément, la comparaison des tableaux permet d'établir une correspondance terme à terme, et dans le même ordre, entre leurs éléments (les cases et leur contenu), ce qui caractérise la congruence sémantique [16] définie par Duval [Duval 1995]. L'analogie que l'on peut constater *de visu* tient bien sûr à l'identité du modèle probabiliste sous-jacent, et l'accès au modèle, via le schéma d'expérience, est précisément une finalité que vise à terme l'enseignement à ce niveau. En fait, il s'agit de quatre traitements par des schémas et notations différentes d'un même processus aléatoire. Les tableaux montrent que, quelle que soit l'expérience E1, …, E4 considérée et quelle que soit sa représentation, le nombre affiché est en fin de compte le même. Finalement, la démarche décrite ci-dessus peut être représentée par le schéma suivant (*fig. 2*) :

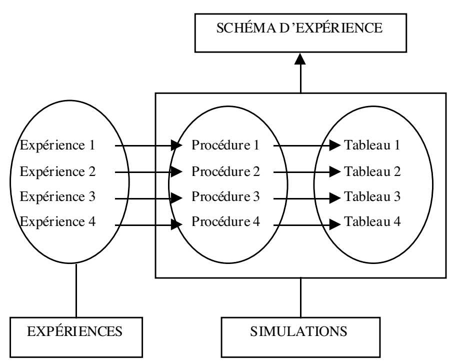


Figure 2

7. Conclusion

Pour les élèves, le passage de l'expérience aléatoire à sa simulation ne va pas de soi ; en particulier, des modèles préconçus peuvent

15 Lors d'un atelier où je faisais cette réflexion, un participant me fit remarquer que, lorsque des enfants jouent au jeu de l'oie, qui nécessite en principe le lancer simultané de deux dés, ils ne sont pas gênés lorsqu'ils n'en ont qu'un et se contentent alors de lancer deux fois l'unique dé. Cet argument ne me semble pas convaincant pour pouvoir affirmer que les deux expériences (lancer simultané de deux dés, deux lancers successifs d'un dé) leur apparaissent comme identiques. On peut seulement dire qu'ils considèrent que la seconde expérience, comme la première, préserve l'équité entre les joueurs. Il serait sans doute intéressant d'étudier la réaction de deux enfants à la proposition de jouer l'un contre l'autre, l'un lançant simultanément deux dés et l'autre lançant deux fois le même dé.

16 R. Duval a défini trois critères pour la congruence sémantique : possibilité d'une correspondance sémantique entre les éléments signifiants, univocité sémantique terminale (la correspondance se fait terme à terme), appréhension des unités dans le même ordre.

venir s'interposer et fausser la démarche [17]. C'est pourquoi je pense qu'il faut, dans les premiers temps, consacrer suffisamment temps à sa mise en place, en mettant en évidence les hypothèses qui le sous-tendent et qui justifient la procédure de simulation, et en s'efforçant de faire « coller » celle-ci, autant que faire se peut, à celle de l'expérience qu'on étudie, sans vouloir imposer aux élèves des procédures, certes plus simples ou plus efficaces, mais qui ne feront pas sens pour eux.

L'accès à la notion de modèle, qui est une finalité visée à terme dans le cycle terminal, peut être préparé par l'étude et la simulation d'expériences aléatoires diverses correspondant au même modèle probabiliste [18]. La comparaison des procédures, des tableaux et des hypothèses sous-jacentes doivent permettre aux élèves de se convaincre de l'analogie que présentent ces expériences sous leurs apparences diverses, et de déboucher sur la notion de schéma d'expérience, constituant en quelque sorte une classe d'équivalence d'expériences aléatoires, comme par exemple « lancer de n dés / n lancers successifs d'un dé » ou « tirage simultané de n boules d'une urne / n tirages successifs sans remise d'une boule d'une urne ». Les tableaux réalisés sur ordinateur pour mettre en œuvre les simulations, dont la structure permet facilement une comparaison visuelle, apparaissent bien adaptés à la mise en évidence des schémas d'expérience.

Bien entendu, je n'ai fait ici qu'exprimer quelques idées générales sur une démarche qui mériterait d'être étudiée plus en détail et dont la mise en œuvre dans les classes pourrait se heurter à un certain nombre de problèmes (lourdeur du programme, disponibilité de la salle informatique, recherche d'expériences adéquates, etc.). Mais je considère que le passage de l'expérience aléatoire à sa simulation est une question importante de l'enseignement de la statistique, qui ne peut être considérée comme « transparente » sous peine de compromettre l'accès de certains élèves à la « pensée statistique ».

Bibliographie

Dodge, Y. (1993) *Statistique. Dictionnaire encyclopédique*. Ed . Dunod, Paris.

Duval, R. (1995) Sémiosis et pensée humaine. Registres sémiotiques et apprentissages intellectuels. Ed. Peter Lang, Berne.

Henry, M. (1999) L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie. *Repères-IREM* 36 pp. 15-34.

Henry, M. (2001) Modélisation d'une situation aléatoire. Notion de modèle et modélisation dans l'enseignement in *Autour de la modélisation en probabilités* (dir. M. Henry) Ed. Presses Universitaires de Franche-Comté pp. 149-159.

Parzysz, B. (2007) Expérience aléatoire et simulation : le jeu de croix ou pile. Relecture actuelle d'une expérimentation déjà un peu ancienne. *Repères-IREM* 66, pp. 27-44.

17 C'est ainsi, par exemple, que certains élèves de Seconde, pour simuler le lancer de deux pièces équilibrées *via* le générateur aléatoire, attribuent *a priori* la même probabilité aux trois modalités PP, PF et FF.

18 Mais cela, seul le professeur le sait.

Parzysz, B. (2005) Quelques questions à propos des tables et des générateurs aléatoires, in B. Chaput & M. Henry (éd.) in *Statistique au lycée 1* (*Les outils de la description statistique*) pp. 181-199. Ed. APMEP / ADIREM.

Parzysz, B. (2005) Du modèle à sa réalisation. La planche de Galton réalise-t-elle vraiment une distribution binomiale ? in B. Chaput & M. Henry (éd.) *Statistique au lycée 1* (*Les outils de la description statistique*) pp. 201-210. Ed. APMEP / ADIREM.

Pascal, B. (1819) Œuvres. Tome 4e. Nouvelle édition. Chez Lefèvre, libraire à Paris.