

Espacios de trabajo en simulación d'expérience aléatoire au lycée : une étude de cas

Working Spaces in Simulation of Random Experiments in High School: A Case Study

Bernard Parzysz

RESUMEN

Hoy la enseñanza de probabilidades se apoya en gran medida en la informática, en especial para simular diferentes experimentos aleatorios. Este artículo, que se refiere al marco teórico de los espacios de trabajo matemáticos de Kuzniak, estudia un tipo emblemático de tareas con frecuencia propuesto a los alumnos franceses de entre 16 y 17 años. Para realizar estas tareas, el programa de hojas de cálculo se utiliza alternadamente como generador de datos al azar, como calculadora y como herramienta lógica. Por esta razón, se otorgaron muchas indicaciones «técnicas» al alumno sobre el uso del programa. Simultáneamente, él tiene que elaborar un modelo del experimento, después simularlo con ayuda del programa y para lograrlo, él debe navegar a través de varios paradigmas probabilísticos y estadísticos, y por lo tanto, a través de varios espacios de trabajo, pero en cuanto a esto no se le brinda ninguna indicación explícita que lo ayude a identificar el paradigma correcto. Además, el enunciado de la tarea es a menudo ambiguo, lo que vuelve difícil determinar los paradigmas en los cuales se debe trabajar.

PALABRAS CLAVE:

- Probabilidades
- Hoja de cálculo
- Experimento aleatorio
- Simulación
- Modelización
- Espacios de trabajo matemático

ABSTRACT

Today the teaching of probability relies considerably on the use of software, particularly when it comes to simulation of diverse random experiments. This article, which refers to the theoretical framework of Kuzniak's mathematical working spaces, studies an emblematic type of tasks frequently proposed to the French students aged 16-17 years. To accomplish these tasks, spreadsheets are used alternately as random event generator, calculator and logical tool. For this reason, the student was provided with many "technical" indications on the use of the software. Simultaneously, he must elaborate a model of the experiment, then simulate it by means of

KEY WORDS:

- Probability
- Spreadsheet
- Random experiment
- Simulation
- Modeling
- Mathematical working spaces



the software, and to do it, he must explore through various probabilistic and statistic paradigms, and hence through many working spaces, but no explicit indication helping him to identify the right paradigm is offered. Additionally, the task verbalization is often ambiguous, which turns difficult to determine the paradigms on which he is supposed to act.

RESUMO

O ensino actual das probabilidades faz uma grande utilização da informática, especialmente para simular experiências aleatórias variadas. Este artigo, que se refere ao quadro teórico dos espaços matemáticos de trabalho de Kuzniak, estuda um tipo de tarefas emblemáticas propostas frequentemente a alunos franceses do 11º ano de escolaridade. Para executar essas tarefas, a planilha integrando gráficos foi utilizada sucessivamente como gerador de casos aleatórios, como calculadora ou como ferramenta lógica. Por esta razão, foram dadas ao aluno muitas indicações “técnicas” sobre a utilização do software. Ao mesmo tempo, o aluno deve desenvolver um modelo da experiência, fazer uma simulação utilizando o software. Para realizar tais tarefas, o aluno deve navegar através de vários paradigmas probabilísticos e estatísticos, e, portanto, através de vários espaços de trabalho, sem que lhe seja dada qualquer indicação explícita para o ajudar a identificar o paradigma correcto. Além disso, o enunciado da tarefa é muitas vezes ambíguo o que dificulta a determinação dos paradigmas em que é suposto trabalhar.

PALAVRAS CHAVE:

- *Probabilidades*
- *Planilha*
- *Experiência aleatória*
- *Simulação*
- *Modelização*
- *Espaço de trabalho matemático*

RÉSUMÉ

Aujourd’hui, l’enseignement des probabilités fait grand usage de l’informatique, en particulier pour simuler des expériences aléatoires variées. Cet article, qui se réfère au cadre théorique des espaces de travail mathématiques de Kuzniak, étudie un type emblématique de tâches fréquemment proposé aux élèves français de Première. Pour réaliser ces tâches, le tableur-grapheur est utilisé tour à tour comme générateur de hasard, comme calculatrice et comme outil logique. Pour cette raison, on fournit à l’élève beaucoup d’indications « techniques » sur l’usage du logiciel. Simultanément, il doit élaborer un modèle de l’expérience, puis le simuler à l’aide du logiciel, et pour ce faire il doit naviguer à travers plusieurs paradigmes probabilistes et statistiques, et donc à travers plusieurs espaces de travail, mais ici aucun indice explicite ne lui est donné pour l’aider à identifier le bon paradigme. En outre, l’énoncé de la tâche est souvent ambigu, ce qui rend difficile la détermination des paradigmes dans lesquels on est censé travailler.

MOTS CLÉS:

- *Probabilités*
- *Tableur*
- *Expérience aléatoire*
- *Simulation*
- *Modélisation*
- *Espaces de travail mathématique*

1 Introduction

L'enseignement des probabilités a considérablement évolué dans le cursus secondaire français (Parzysz, 1997), passant d'une conception purement « laplacienne » de la notion de probabilité, requérant l'équiprobabilité des événements élémentaires, à une approche beaucoup plus nuancée, et élargie, intégrant une conception de type « fréquentiste », dans laquelle la probabilité d'un événement lié à une épreuve aléatoire apparaît comme « limite » de la fréquence d'apparition de cet événement dans une suite de répétitions de l'épreuve. De même qu'en géométrie, la question de la modélisation apparaît comme un élément important de cette approche, et je me suis plus particulièrement attaché à voir si la notion d'espace de travail géométrique (ETG) de Houdement et Kuzniak (2006) pouvait se transposer, *mutatis mutandis*, dans le domaine des probabilités (Parzysz, 2011), en espace de travail probabiliste (ETP).

Poursuivant cette démarche d'investigation, et partant du fait que, aujourd'hui, les outils technologiques tiennent une aussi grande place dans un domaine que dans l'autre, je me suis intéressé aux usages du tableur-grapheur dans l'enseignement des probabilités au lycée, et plus particulièrement au rôle de la simulation informatique dans la démarche de modélisation. Ainsi que le fait remarquer Trouche (2005) :

Le développement des outils informatiques a eu des effets très importants sur le développement de certaines branches des mathématiques ou sur le développement de nouveaux domaines (...) et a donné un nouveau statut aux aspects expérimentaux de la recherche (...). Ces effets s'exercent aussi sur les mathématiques enseignées. (...) Les outils ont aussi des effets importants sur les modes de travail des élèves. (...) Enfin les outils mis en œuvre dans l'enseignement ont des effets profonds sur la conceptualisation. » (p. 267).

C'est à cet aspect de l'activité mathématique en classe que je vais m'intéresser ici, en prenant comme support une activité de type courant proposée dans un manuel récent de la classe de première scientifique (programme entré en vigueur en septembre 2011).

2 Cadre théorique

Mon cadre théorique de référence comprend bien sûr la notion de paradigme inspirée de Kühn, appliquée à la géométrie par Houdement et Kuzniak (2006) et

que j'ai entrepris de transposer aux probabilités en suivant le parallèle fait par Henry (1999) entre les deux domaines. Dans l'article cité plus haut, je définissais plusieurs paradigmes probabilistes mis en œuvre, de façon plus ou moins explicite, dans l'enseignement français, au collège et au lycée :

- celui de la « réalité », c'est-à-dire de l'expérience concrète effectivement réalisée à l'aide d'objets matériels (dés, pièces de monnaie, roue de loterie, jetons sortis d'une urne, etc. ;
- un paradigme P1, issu d'une première modélisation dans laquelle on associe à l'expérience concrète une liste des issues prises en compte et un protocole expérimental précis (expérience pseudo-concrète), assurant que l'expérience pourra être répétée dans les mêmes conditions, la répétition donnant lieu à des observations permettant d'attribuer une chance d'apparition à chacune des différentes issues ;
- un paradigme P2, obtenu en posant un « regard probabiliste » (Henry, 1999) sur l'expérience, dans lequel on définit l'expérience aléatoire générique, ainsi que la notion de probabilité. On entreprend alors d'étudier les propriétés de cette probabilité, c'est-à-dire qu'on construit une algèbre des événements, illustrée sur des modèles classiques (en particulier des « modèles d'urne ») et les principales lois de probabilité (binomiale, exponentielle, géométrique, normale...). Les outils associés sont : la démonstration de type mathématique, les techniques de calcul usuelles, divers registres de représentation (tableau à double entrée, arbre pondéré, diagramme ensembliste, graphiques statistiques variés...), ainsi que les outils du cadre de la statistique descriptive (SD), transposés à celui des probabilités (lien entre fréquence et probabilité, entre moyenne et espérance mathématique...). Le programme de probabilités de première scientifique¹ incite d'ailleurs à opérer un tel rapprochement : « À l'aide de simulations et d'une approche heuristique de la loi des grands nombres, on fait le lien avec la moyenne et la variance d'une série de données ». Mais il s'agit cependant de bien distinguer les cadres : SD d'un côté et P1/P2 de l'autre, comme nous le verrons plus loin (§ 4.).

Je me référerai également à la notion d'espace de travail mathématique (Kuzniak, 2011), avec ses trois composantes :

- un espace réel constitué d'objets matériels, auquel est associé un processus de visualisation ;
- un ensemble d'artefacts (outils, instruments) permettant d'agir sur ces objets, auquel est associé un processus de construction ;

¹ B.O. spécial n° 9 du 30 septembre 2010.

- un système théorique de référence, auquel est associé un processus discursif (argumentation, preuve).

3 Un exercice emblématique

La simulation fait désormais explicitement partie des programmes de mathématiques, et en particulier dans le domaine des probabilités :

On peut simuler la loi géométrique tronquée avec un algorithme. (...) On peut simuler la loi binomiale avec un algorithme.

(Programme de première scientifique)

Le recours aux représentations graphiques et aux simulations est indispensable. (...) Des activités algorithmiques sont menées (...), notamment pour simuler une marche aléatoire. (...) La simulation de sondages sur tableur permet de sensibiliser aux fourchettes de sondage.

(Programme de terminale scientifique²).

Comme le laisse entendre la référence à un algorithme, le mot « simulation » est ici à prendre dans le sens de « simulation informatique ». Ceci témoigne du souci des concepteurs des programmes de prendre en compte l'évolution technologique, mais restreint le sens initial du terme, qui est la substitution d'une chose (succédané) à une autre, afin qu'elle en tienne lieu lors de l'accomplissement d'une action donnée³. Dans le cas qui nous occupe ici, on peut néanmoins préciser davantage, comme le fait ce manuel de lycée : « *Simuler une expérience aléatoire, c'est remplacer cette expérience par une autre, plus économe, plus rapide, et qui permet d'obtenir des résultats analogues.* » (Belin, 2000). La difficulté consiste alors à pouvoir assurer que les résultats obtenus seront « analogues ». Un autre manuel précise : « *Pour simuler une expérience, on décrit l'expérience aléatoire et on la modélise : on choisit un modèle de tirage aléatoire de nombres qui doit avoir les mêmes propriétés que le phénomène simulé.* » (Borreani, 2004). C'est donc l'existence d'un modèle probabiliste commun à l'expérience aléatoire et à son succédané qui va garantir cette analogie, selon le modèle triadique de la figure 1 :

² B.O. spécial n° 8 du 13 octobre 2011.

³ Faute d'un marteau, on peut par exemple utiliser une pierre pour enfoncer un clou.

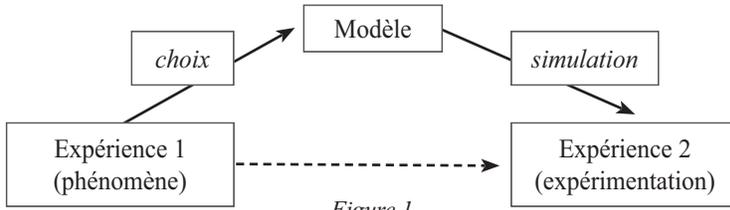


Figure 1

On peut ainsi utiliser un dé pour jouer à pile ou face ; il suffit par exemple de décider qu'un point pair correspondra à pile et un point impair correspondra à face. À l'équiprobabilité supposée d'apparition des six issues du dé correspondra l'équiprobabilité d'apparition d'un nombre pair et d'un nombre impair.

Un type d'exercice couramment proposé par les manuels comporte la simulation d'une situation réelle dans laquelle différentes issues peuvent se produire de façon imprévisible, selon le schéma général suivant :

- 1° simuler la situation à l'aide d'un tableur,
- 2° observer le résultat de la simulation,
- 3° énoncer une conjecture sur la probabilité d'une des issues possibles de l'expérience, ou sur la valeur possible d'un paramètre qui lui est lié.

Nous allons étudier un exemple typique de ce type d'exercice, qui fait l'objet d'une séance de travaux pratiques proposée par le manuel de première scientifique *Transmath Première S* (Barra, Barros, Bénizeau & Morin 2011). Dans ce qui suit, les extraits de l'énoncé figurent en encadré. Il s'agit donc, en fait et pour reprendre la terminologie de Kuzniak, d'étudier un *ETM idoine potentiel* relatif au domaine des Probabilités. Cet exercice présente en outre l'intérêt d'articuler la simulation de l'expérience avec la résolution du problème.

3.1. Première partie : la situation

(Situation réelle évoquée) « On dispose de trois dés tétraédriques parfaits : un bleu, un rouge et un vert⁴. On lance ces trois dés et on s'intéresse au nombre de 4 obtenus. Une issue est un nombre à trois chiffres, par exemple 1 4 3. On considère le jeu suivant : si on obtient trois fois le nombre 4, on gagne 36 € ; si on l'obtient deux fois, on gagne 2 € ; sinon, on perd 1 €. L'objectif de ce TP est d'évaluer le gain moyen que l'on peut espérer sur une série de 2500 lancers. »

⁴ L'énoncé est accompagné d'une photographie couleur montrant les trois dés : on peut (doit) penser que les sommets de chaque dé sont numérotés de 1 à 4 et que le point marqué correspond au sommet qui est au-dessus du plan horizontal sur lequel on lance les dés, mais ce n'est pas totalement évident.

Commentaire. La situation « concrète » est décrite : éléments matériels en jeu (le manuel présente même une photographie des trois dés), règle du jeu. Et une question relative à ce jeu est posée.

L'espace de travail est constitué des objets matériels évoqués (les dés), des actions sur ces objets (2500 lancers), de la règle du jeu et de la grandeur à laquelle on s'intéresse (le gain moyen).

(Modèle semi-concret dans P1) « *On fait l'hypothèse de l'équiprobabilité des issues.* »

Commentaire. Cette hypothèse permet d'attribuer une probabilité à chacune des issues envisagées ; elle introduit en fait un modèle probabiliste de la situation. Ce passage par P1 est nécessaire, car c'est un *modèle* que l'on simule : « *la simulation a pour préalable de choisir un modèle* » (Portail national des professionnels de l'éducation, 2009, p. 7). L'explicitation du modèle est destinée à assurer que c'est bien la *même expérience* que l'on répète, puisque c'est toujours le même protocole qui sera mis en œuvre (Parzys, 2005). Mais dans la pratique le modèle est souvent escamoté : on fait comme si on passait directement de la situation réelle à la simulation. Ce « tour de passe-passe » est facilité par le fait qu'on ne considère que des épreuves aléatoires (lancer d'une pièce, d'un dé, tirage d'une boule dans une urne...) pour lesquelles il existe un modèle canonique, qui est alors considéré comme « transparent ». Mais le risque est alors que l'élève cherche par la suite à simuler directement la situation réelle, sans se rendre compte qu'il lui faut pour cela faire des hypothèses, c'est-à-dire élaborer un modèle, aussi peu élaboré soit-il.

Dans le modèle annoncé ici « une issue est un nombre à trois chiffres » et, comme chacun des dés peut amener 1, 2, 3 ou 4, il y a $4^3 = 64$ issues possibles. En vertu de l'hypothèse faite, chaque issue est donc affectée de la probabilité $1/64$.

L'espace de travail de référence, dans P1, est constitué de l'ensemble des issues et de la probabilité associée à chacune d'elles ; on travaille dans le registre numérique, et plus précisément dans le cadre des rationnels.

3.2. *Deuxième partie : simulation*

La simulation est devenue aujourd'hui un outil usuel de l'enseignement des probabilités. Dans ce domaine particulier, il s'agit de, pour des raisons d'économie et de performance, de faire exécuter par l'ordinateur - ou la calculatrice - une longue séquence d'épreuves censées être analogues à l'expérience aléatoire simulée. Dans la programmation de l'épreuve par la machine intervient le générateur

« aléatoire » de celle-ci, lequel est en fait parfaitement déterministe⁵. Mais « tout se passe comme si » les nombres produits étaient aléatoirement et uniformément répartis dans l'intervalle $[0 ; 1[$.

(Tableur) « *Simulez plusieurs fois cette expérience (...). Comment évolue la moyenne des gains en fonction du nombre de lancers ? Quel gain moyen peut-on espérer au bout de 2500 lancers ?* »

Commentaire. Le manuel est très détaillé sur cette phase, dont voici le début :

- a) Ouvrez une feuille de calcul et complétez les cellules A2 à D2 avec les formules adéquates.
- b) En E2, dénombrez les apparitions du 4 dans la plage de cellules B2 : D2.

	A	B	C	D	E	F	G	H
1	Expérience n°	Dé bleu	Dé rouge	Dé vert	Sorties du 4	Gain		Moyenne des gains
2								
3								

- c) En F2, faire afficher le gain correspondant à l'issue obtenue à l'expérience n° 1. Saisissez l'instruction conditionnelle : $=SI(E2=3;36;SI(E2=2;2;-1))$. Justifiez qu'elle décrit bien la règle du jeu.
- d) Dans la colonne H, on désire afficher la moyenne des gains obtenue en fonction du nombre de lancers. Quelle formule devez-vous saisir ? (...)
-

Une différence saute aux yeux par rapport à la géométrie : la guidance de l'élève est très forte, et il reste cantonné au rôle d'exécutant. Cette hypertrophie des instructions a pour fonction principale d'éviter à l'élève non familier du tableur d'être bloqué (en effet, s'il travaille seul chez lui, qui viendra l'aider ?). Les intitulés des colonnes sont donnés, et la première tâche de l'élève consiste à introduire la fonction « aléa » dans les cellules B2, C2, D2 ; il a ensuite à totaliser le nombre de 4 obtenus dans ces trois cellules (fonction « somme »). Il doit ensuite - plus délicat - utiliser une instruction conditionnelle pour faire calculer le gain correspondant ; le manuel vient alors à son aide en lui fournissant la formule à entrer. Enfin, il faut faire calculer la moyenne des gains, ce que l'élève sait - en

⁵ La machine affiche en effet les termes successifs d'une suite récurrente, donc parfaitement déterministe (voir par exemple [Parzysz, 2005]).

principe - faire depuis le collège. Mais « *faciliter la prise en main des logiciels ne confère pas les compétences requises pour mener à bien des tâches complexes. La transparence que les interfaces sont censées fournir dépend de la proximité de ce qu'elles proposent avec les représentations initiales du sujet.* » (Rabardel, 1995). La demande de justification de la question c) a précisément pour objet de tester que l'élève a compris le sens de l'instruction donnée à la machine, mais elle risque d'être purement rhétorique, dans la mesure où l'élève est livré à lui-même. Cette question des compétences relatives à l'outil technologique est explicitement prise en compte dans un certain nombre de manuels, qui présentent des fiches spécifiques sur la prise en main des calculatrices (algorithmique), tableurs et logiciels de géométrie dynamique.

Notons au passage que cette partie soulève une autre question, indissociable de la simulation à l'aide d'outils technologiques. La première question demande en effet de « compléter les cellules A2 à D2 » : il s'agit donc pour l'élève d'introduire la fonction ALEA dans les cases B2, C2 et D2⁶. Même si cette « boîte noire » ne lui pose pas problème, il faut malgré tout qu'il la considère comme un *véritable* « générateur de hasard », ce qu'elle n'est pas. Le problème, de nature déontologique, se situe plutôt du côté de l'enseignant : lorsqu'on demande à l'élève de faire une conjecture sur la base des résultats affichés, sur quoi porte cette conjecture ? Sur ce qui est effectivement demandé (comme on le ferait en réalisant effectivement l'expérience) ou sur la conformité du générateur par rapport à ce qu'il est censé produire, *i.e.* des valeurs d'une variable aléatoire équirépartie sur $[0 ; 1[$? Le recours au générateur impose donc de « faire comme si » il était réellement aléatoire et, en attendant les ordinateurs quantiques qui seront capables de fournir du *vrai* hasard, de lui faire confiance.

3.3. *Quel modèle ?*

Au-delà du guidage qui reflète la difficulté que les auteurs du manuel attribuent aux diverses tâches dévolues à l'élève, il faut remarquer que le modèle implémenté *n'est pas* celui qui est annoncé. En effet, l'observation des intitulés des colonnes montre que l'équiprobabilité supposée par le tableur est celle des quatre sommets pour chaque dé, et non celle des 64 nombres à trois chiffres possibles. Une procédure de simulation plus « congrue » – selon Duval – au modèle annoncé aurait consisté à faire afficher le nombre à trois chiffres obtenu de façon aléatoire et équiprobable et à lui associer le gain correspondant. Dans le contexte de la simulation, j'ai naguère plaidé pour la recherche d'un modèle aussi « proche » que possible de la situation réelle, au moins dans les débuts de l'apprentissage (Parzysz, 2009) : la notion de schéma d'expérience, première étape vers la notion de modèle probabiliste, peut alors se construire

⁶ Par exemple sous la forme $ENT(1+4*ALEA())$.

sur l'analogie constatée entre les tableaux du grapheur correspondant à des expériences diverses. Dans le cas qui nous occupe, les deux modèles indiqués ci-dessus sont bien sûr, du point de vue de la variable aléatoire G considérée, équivalents. Plus précisément :

- dans le modèle annoncé, les issues sont les 64 nombres à trois chiffres, dont un seul comporte trois chiffres 4, 9 comportent exactement deux chiffres 4, 27 comportent exactement un chiffre 4 et 27 n'en comportent aucun. La loi de probabilité de G est donc donnée par $P(G=36)=1/64, P(G=2)=9/64, P(G=-1)=27/32$.
- dans le modèle entré dans le tableur, on a trois dés à quatre sommets numérotés 1, 2, 3, 4 et, pour chacun : $p(1)=p(2)=p(3)=p(4)=1/4$. Pour chaque lancer des trois dés on note le nombre total de 4 obtenus, de façon à lui attribuer un gain.

Mais le premier modèle aurait évidemment été assurément plus compliqué à implémenter dans le tableur, et c'est pourquoi les auteurs ont - mais implicitement - préféré une autre modélisation. Il n'en reste pas moins que la phrase « Une issue est un nombre à trois chiffres » apparaît, non seulement superflue, mais malheureuse puisqu'elle oriente l'élève vers un modèle différent de celui mis en œuvre dans le tableur. Il en va d'ailleurs de même pour la distinction des trois dés par leur couleur, qui n'apparaît pas *a priori* pertinente ici et dont la véritable fonction (implicite) est d'identifier chacun des trois dés, l'un donnant le chiffre des centaines, un autre le chiffre des dizaines et le troisième le chiffre des unités. En définitive, pour améliorer la congruence entre la situation réelle et la procédure de simulation, il aurait sans doute été préférable de se référer à trois lancers successifs du même dé, ce qui aurait conduit à un tableau tel que ci-dessous :

	A	B	C	D	E	F	G	H
1	Expérience n°	1 ^{er} lancer	2 ^{ème} lancer	3 ^{ème} lancer	Sorties du 4	Gain		Moyenne des gains
2								
3								

On va à présent se situer dans le domaine de la statistique descriptive (SD) : le tableur va fournir une série de données sur laquelle on fera effectuer des calculs (ici, une moyenne).

En définitive, les fonctionnalités du logiciel (instrumentalisation) ici mises en œuvre sont :

- le générateur de hasard
- l'outil de recopie (2500 épreuves)
- le calculateur (somme, moyenne)
- l'outil logique (instruction conditionnelle).

Et on peut également y inclure le grapheur, puisqu'on trouve aussi dans l'énoncé la consigne :

-
- f) *Sélectionnez les colonnes A et H, puis représentez le gain moyen en fonction du nombre n d'expériences. »*
-

Le registre des graphiques cartésiens, lié au processus de visualisation, possède ici, comme c'est le cas avec les logiciels de géométrie, un aspect dynamique, grâce au calcul sur ordre (touche F9) qui fait apparaître les fluctuations d'échantillonnage et la convergence de façon bien plus nette que le seul tableau. C'est un outil important, mais qui a néanmoins des limites - c'est le cas de le dire -, étant donné que la limite de la série des fréquences, dont on postule l'existence, n'est pas nécessairement accessible (en particulier si elle n'est pas décimale).

3.4. *Quel(s) espace(s) de travail ?*

Au départ de cette phase de simulation, et en se plaçant dans le modèle effectivement mis en œuvre par le logiciel, l'espace de travail idoine (virtuel) se situe à l'articulation entre P1 et SD ; il s'agit de « traduire » les hypothèses probabilistes (ici, l'équiprobabilité des issues pour chaque dé) en instructions pour le tableur. Les registres en jeu sont la langue naturelle (par exemple : « la probabilité d'obtenir 4 avec le dé rouge est $\frac{1}{4}$ ») et le langage symbolique propre au tableur.

On travaille ensuite dans SD, avec des notions de statistique descriptive (somme, moyenne) et avec l'aide du tableur, dont l'apprentissage de l'utilisation est sans aucun doute l'un des objectifs du travail entrepris. La tâche de l'élève, relativement complexe, est de construire un tableau dans un but précis : faire afficher le gain moyen sur 2500 parties, dans le but de proposer une conjecture. Il s'agit donc d'abord d'élaborer une stratégie lui permettant d'y parvenir. L'aide fournie dans le manuel, même si elle élimine le côté heuristique de la démarche, permet de décortiquer cette stratégie en trois parties ; elle concerne, non seulement l'aspect « technique » (syntaxe du tableur) mais aussi la mise en œuvre de la démarche (copie d'écran) :

- 1° traitement individuel des trois dés :
 - lancer le dé
 - afficher le point marqué
- 2° mise en commun des trois dés :
 - totaliser le nombre de 4 obtenus
 - calculer le gain du joueur
- 3° répétition de l'expérience
 - itérer 2500 fois le processus
 - calculer le gain moyen du joueur.

Finalement, on voit qu'un tel exercice fait travailler l'élève dans plusieurs cadres différents, et en conséquence son espace de travail évolue de façon notable au fil de la résolution : partant d'une situation réelle, il passe d'abord dans le cadre probabiliste (pour pouvoir faire fonctionner le tableur), puis dans celui de la statistique descriptive (pour faire calculer des paramètres associés à la distribution statistique obtenue), puis repasse dans le cadre probabiliste (via la loi des grands nombres), et revient enfin à la situation réelle pour énoncer le résultat obtenu. Les registres qui interviennent sont : la langue naturelle, le numérique, le symbolique et celui des graphiques cartésiens.

On a donc, pour la partie « simulation » de cet exercice, le schéma de la figure 2.

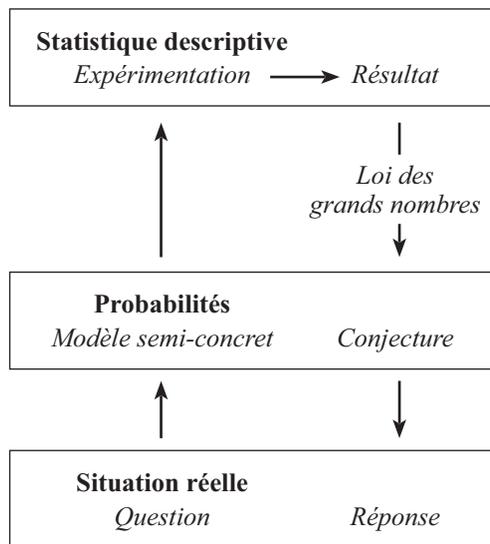


Figure 2

Remarque. Ici, P1 (assorti d'une « formulation naïve de la loi des grands nombres ») est suffisant, car c'est le modèle semi-concret qu'on implémente dans le tableur : il suffit d'identifier les différentes issues possibles et de leur affecter un nombre compris entre 0 et 1 (et de façon telle que la somme de ces nombres soit égale à 1). Mais on aurait aussi pu, à un autre niveau, se situer dans P2 et considérer trois variables de Bernoulli mutuellement indépendantes (une par dé), prenant la valeur 4 avec la probabilité 1/4.

3.5. Troisième partie : résolution

Après la partie « Conjecturer », cet exercice comporte une partie « Démontrer », qui constitue de fait un autre exercice.

(Résolution probabiliste dans P2)

- a) *Quel est le nombre d'issues possibles lors d'une expérience ?*
- b) *Dénombrer les issues favorables à la sortie de trois numéros 4 ; de deux numéros 4.*
- c) *On appelle G la variable aléatoire qui indique le gain algébrique en euros du joueur. Dressez le tableau de la loi de probabilité de G , puis calculez $E(G)$.*
- d) *Quel lien faites-vous entre le résultat de la simulation et l'espérance mathématique de G ?*

Commentaire. Il s'agit cette fois d'un exercice classique de probabilités. On abandonne la simulation pour la résolution d'un problème. La tâche de l'élève consiste d'abord - questions *a* et *b* - à déterminer l'espace probabilisé dans lequel il aura à effectuer les calculs (ici, l'ensemble des 64 issues possibles muni de l'équiprobabilité), puis - questions *c* et *d* - à dresser la table de la loi de probabilité de la variable aléatoire G définie sur cet espace et à calculer son espérance mathématique. Pour finir, on lui demande de faire le lien entre la partie « statistique » (simulation) et la partie « probabilités ». Le seul résultat qu'il peut faire intervenir est une « *formulation naïve de la loi des grands nombres (...)* : lorsque le nombre d'expériences augmente, la fréquence empirique se rapproche de la probabilité. » (Ressources « Probabilités » pour la classe de troisième p. 8). Or, ici on a $E(G) = \sum x_i \cdot p(x_i)$, tandis que la moyenne des gains est $\sum x_i \cdot f(x_i)$ (i variant de 1 à 3 puisque G prend 3 valeurs) ; c'est donc en réalité la convergence de la *distribution* de fréquence vers la *distribution* de probabilité qui permet de conclure à la convergence de la moyenne vers l'espérance, et non la seule considération de la fréquence d'une issue particulière. Il s'agit là aussi d'une conséquence de la loi des grands nombres, mais elle nécessite de considérer la distribution de fréquences *dans son ensemble*.

La démarche peut cette fois être schématisée par la figure 3.

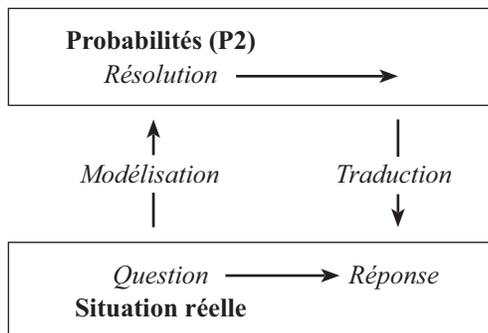


Figure 3

Remarque 1. Ici, P2 est nécessaire car on utilise des définitions et des résultats de P2 et on effectue des calculs. L'espace de travail adéquat est constitué d'éléments de P2, comprenant des notions (variable aléatoire, loi de probabilité, espérance mathématique) et des outils de dénombrement (formule ou arbre).

Remarque 2. Cet exercice constitue de fait une approche de la loi binomiale (et c'est peut-être d'ailleurs l'un des objectifs des auteurs). On peut en effet considérer qu'on a trois variables de Bernoulli mutuellement indépendantes (une par dé) prenant la valeur 4 avec la probabilité $1/4$; la somme X de ces trois variables suit alors la loi binomiale $B(3, 1/4)$ et la loi de probabilité de G est donnée par $P(G=36)=P(X=3)=\binom{3}{3}3^0/4^3=1/64$, $p(G=2)=\binom{3}{2}3^1/4^3=9/64$, et donc $P(G=-1)=27/32$. D'où $E(G)=0$ (les valeurs numériques des gains ont été choisies de façon que le jeu soit équitable, ce qui est peut-être un autre objectif, plus lointain, des auteurs). Mais bien sûr, muni de ces connaissances, on n'a plus besoin de simuler le jeu.

À l'issue de la séance, l'élève a été confronté à deux démarches successives : l'une de type expérimental (simulation) et l'autre de type théorique (calcul probabiliste). La question se pose donc du lien réel entre cette seconde partie (résolution dans P2) et la première (simulation dans SD via P1), d'autant plus que celle-ci n'est pas nécessaire puisque le calcul fournit à lui seul la réponse exacte.

La jonction ne s'opère que dans la dernière question (*Quel lien faites-vous entre le résultat de la simulation et l'espérance mathématique de G ?*). Mais, ici, le recours à la simulation était-il utile - sinon nécessaire - pour conjecturer le résultat ? C'est d'autant moins évident que le but de la résolution n'est pas de démontrer que l'espérance de G est nulle, mais de calculer cette espérance (et de constater qu'elle est nulle). L'idée des auteurs était peut-être en fait d'ordre « méta » : montrer aux élèves que le recours au tableur peut être utile pour conjecturer la réponse à un problème de probabilités ou, comme ici, pour contrôler la pertinence d'un modèle probabiliste⁷.

4 Statistique et probabilités

Les deux facettes de l'activité étudiée apparaissent complémentaires, comme le montrent, d'une part le lien explicitement établi dans la dernière question, et d'autre part l'opposition « Conjecturer / Démontrer » dans les sous-titres.

⁷ Par exemple en prenant pour issues les triplets de points marqués, comme le suggère la coloration des dés.

L'analogie des formules donnant, d'une part la moyenne d'une distribution statistique, et d'autre part l'espérance d'une variable aléatoire, semble, on l'a vu, un bon moyen de passer de l'une à l'autre, mais à condition de ne pas les confondre et de savoir à tout moment dans quel cadre on se situe. Les probabilités s'intéressent en effet à un modèle *théorique*, tandis que la statistique descriptive étudie une série de données issue d'un phénomène *observé*. Et, dans le cas d'une simulation sur tableur, ce phénomène consiste en la *réalisation* de l'itération d'un modèle probabiliste. Par exemple, dans le cas de l'exercice étudié, la moyenne des gains observée sur les 2500 épreuves peut être par exemple de 0,0544, tandis que l'espérance mathématique de gain de la variable est exactement 0.

Or, dans le cas qui nous occupe, la distinction entre le cadre de la statistique (SD) et celui des probabilités (P1) n'est pas rendue explicite, ne serait-ce que par la terminologie utilisée : le « gain moyen » dont il est question au départ n'est pas le même que le « gain moyen » qu'on représente graphiquement : le premier ressortit à P2 (puisque - dit le texte - on ne peut que l'« évaluer ») tandis que le second se situe dans SD. Qui plus est, une phrase telle que « Quel gain moyen peut-on espérer au bout de 2500 lancers » est difficilement compréhensible car, dès lors qu'on aura fait effectuer ces lancers par le tableur, on connaîtra la valeur du gain moyen correspondant aux 2500 lancers réalisés. Et, s'il s'agit de l'espérance de gain - suggérée par le verbe « espérer » - indiquer le nombre de lancers est superflu. Il y a en fait ici un « écrasement » du cadre probabiliste sur le cadre statistique, favorisé par la démarche envisagée par les auteurs : on veut apparemment, sans faire une étude probabiliste dans P2, se faire une idée de (« évaluer ») la valeur du gain moyen sur un grand nombre d'épreuves (ici, 2500), et pour cela on effectue une simulation. Si on observe (avec l'aide de la représentation graphique), qu'en augmentant le nombre d'épreuves la moyenne des gains est tantôt positive, tantôt négative, mais que sa valeur absolue devient de plus en plus faible (« Comment évolue la moyenne des gains en fonction du nombre de lancers ? »), on sera amené à conjecturer que cette moyenne fluctue autour de 0 (convergence de la suite des fréquences lorsque le nombre d'épreuves augmente indéfiniment). Cette fois encore, la démarche « naturelle » aurait consisté, non pas à *augmenter* le nombre de lancers, mais à *répéter* des séries de 2500 lancers, puisque la question initiale comporte explicitement cette référence⁸. Mais, de toute façon, on se situe uniquement dans le cadre statistique.

Venons-en maintenant à la résolution du problème de probabilités demandée dans la dernière partie de l'exercice, qui a pour but l'obtention de

⁸ On remarquera qu'ici non plus la congruence sémantique n'est pas prise en compte.

l'espérance mathématique de gain. La loi des grands nombres va encore intervenir ici, mais de façon différente : la limite (espérance de gain nulle) étant donnée, on peut prévoir que, à l'issue d'un grand nombre d'épreuves (en l'occurrence 2500), la moyenne des gains aura sans doute une faible valeur absolue. On se situe donc, au départ, dans le cadre probabiliste, pour en tirer une conclusion d'ordre statistique.

Nous avons en définitive pu voir que, au cours d'un tel exercice l'élève se trouve, selon le moment considéré, face à trois expériences différentes : l'expérience réelle (lancer de dés matériels), l'expérience modélisée (en vue de la simulation) et l'expérience simulée (réalisée par le tableur), associés à des cadres de référence différents et donc à des espaces de travail différents. Elles sont en interaction les unes avec les autres, et il importe de savoir à tout moment de laquelle il est question. On peut rapprocher ceci de ce qui se passe lors de la résolution d'un problème de géométrie élémentaire, qui fait intervenir de façon dialectique les paradigmes G1 et G2 : dans un problème posé dans G2, on commence par réaliser un dessin ($G2 \rightarrow G1$) qui servira de support pour aider à faire des conjectures et/ou imaginer des stratégies de résolution, avec si nécessaire d'autres dessins (travail dans G1) ; puis, toujours avec l'aide du dessin on imagine une démonstration, complète ou partielle ($G1 \rightarrow G2$) que l'on met ensuite en forme (travail dans G2), et on réitère éventuellement ce processus cyclique après chaque avancée dans la résolution du problème.

5 Conclusion

Dans l'exercice étudié ici interviennent trois cadres différents : celui de la « réalité » (la situation évoquée), celui de la statistique descriptive SD (associée au tableur) et celui des probabilités, présent sous deux paradigmes : P1 (modèle semi-concret) et P2, modèle plus élaboré dont l'étude fait l'objet du travail au niveau du lycée. Dans ce type d'activité, l'espace de travail personnel de l'élève est donc constamment amené à évoluer, tout d'abord dans une démarche de modélisation, puis dans une démarche de simulation du modèle élaboré, ensuite dans un retour au modèle pour conjecturer. Et, au cours de cette démarche, apparaissent tour à tour trois « expériences » distinctes. Il s'agit donc d'une démarche particulièrement complexe, même si la conscience de cette complexité peut parfois échapper à l'enseignant. Comme le faisait naguère remarquer Bruillard (1997) :

Il faut remarquer une dissymétrie importante entre les élèves et les enseignants. Ces derniers, experts dans leur domaine, décodent directement, sans même y penser, les résultats qui leur sont présentés. Ce n'est pas le cas pour les élèves, en cours d'apprentissage, qui ne disposent pas toujours des connaissances requises.

D'autre part, dans ce type de tâches les fonctions du tableur sont multiples, car il intervient tour à tour comme générateur de hasard, comme calculateur et comme outil logique. L'ordinateur apparaît donc ici, non pas comme une simple commodité permettant d'obtenir les résultats une longue série d'épreuves répétées plus rapidement qu'en les réalisant réellement (c'est-à-dire comme un *outil*), mais en tant que partie prenante dans la démarche de modélisation (c'est-à-dire comme un *instrument*) :

Adopter une perspective instrumentale, c'est se centrer sur les activités dans lesquelles l'apport des instruments est essentiel : des moyens, certes attachés aux tâches, non en tant que compléments ou substituts, mais comme parties intégrantes.

(Bruillard, 1997).

Ce processus d'instrumentation nécessite du temps, celui de la genèse instrumentale par laquelle, selon Rabardel (1995) « *les représentations relatives aux systèmes techniques dans leur fonction instrumentale se construisent solidairement et en étroite articulation avec les représentations relatives au réel sur et dans lequel l'instrument permet d'agir.* » (p. 167). Mais il faut aussi que l'élève soit à même de disposer des éléments conceptuels permettant d'entreprendre la démarche de modélisation ; nous avons vu que, dans le cas présent, un modèle assez peu sophistiqué (le modèle pseudo-concret P1) est sans doute suffisant pour mettre en œuvre la simulation, mais que la congruence sémantique entre l'expérience concrète et le modèle implémenté dans le tableur n'est pas prise en compte. Plus généralement, la considération explicite, dans l'espace de travail idoine mis en place par l'enseignant, des divers paradigmes du domaine en jeu, ainsi que des diverses fonctions que le tableur est amené à y remplir, devrait pouvoir permettre de gérer de façon moins empirique les rapports de l'élève avec le tableur dans son espace de travail personnel.

Enfin, la comparaison de ce processus avec la démarche mise en œuvre dans la résolution d'un problème de géométrie avec l'aide d'un logiciel de géométrie dynamique permettrait sans doute de dégager un certain nombre de points communs, mais peut-être aussi des différences, ne serait-ce que le fait que, contrairement aux logiciels de géométrie, les tableurs-grapheurs n'ont pas été conçus spécifiquement pour l'enseignement. Mais ceci est une autre histoire...

References

- Barra, R.; Barros, J.-M.; Bénizeau, P. & Morin, J. (2011). *Transmath Première S*. Paris, France : Nathan.
- Belin. (2000). *Math seconde avec thèmes d'étude* [manuel scolaire]. Paris, France : Belin.
- Borreani, Jacqueline (collectif). (2004). *Maths 2^e* [manuel scolaire]. Turin, Italie : Magnard.
- Bruillard, E. (1997). L'ordinateur à l'école : de l'outil à l'instrument in Pochon, L. & Blanchet, A. (Eds.). *L'ordinateur à l'école : de l'introduction à l'intégration* (pp. 99-118). Neuchâtel : IRDP.
- Portail national des professionnels de l'éducation. (2009). *Ressources pour la classe de seconde : probabilités et statistiques* [document en ligne d'utilisation libre pour les enseignants]. Rérépéré à : <http://www.youscribe.com/catalogue/tous/education/cours/ressources-pour-la-classe-de-seconde-probabilites-et-statistiques-375226>
- Henry, M. (1999). L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie. *Repères-IREM*, 36, 15-34.
- Houdement, C. & Kuzniak, A. (2006). Paradigmes géométriques et enseignement de la géométrie. *Annales de Didactique et de Sciences Cognitives*, 11, 175-193.
- Kuzniak, A. (2011). L'espace de travail mathématique et ses genèses. *Annales de Didactique et de Sciences Cognitives*, 16, 9-24.
- Parzysz, B. (1997). L'enseignement de la statistique et des probabilités dans l'enseignement secondaire, d'hier à aujourd'hui, in Enseigner les probabilités au lycée. IREM de Reims (Eds.) *Commission inter-IREM Statistique et Probabilités*, 3, 17-38.
- Parzysz, B. (2005). Quelques questions à propos des tables et des générateurs aléatoires in Chapat, B. & Henry, M. (Eds.). *Statistique au lycée*, 1, 181-200.
- Parzysz, B. (2009). Des expériences aléatoires au modèle, via la simulation. *Repères-IREM*, 74, 91-103.
- Parzysz, B. (2011). Quelques questions didactiques de la statistique et des probabilités. *Annales de Didactique et de Sciences Cognitives*, 16, 127-147.
- Rabardel, P. (1995). *Les hommes et les technologies. Une approche cognitive des instruments contemporains*. Paris, France : Armand Colin.
- Trouche, L. (2005). Des artefacts aux instruments, une approche pour guider et intégrer les usages des outils de calcul dans l'enseignement des mathématiques. *Actes de l'Université d'été de Saint-Flour « Le calcul sous toutes ses formes »*, 265-289. Rérépéré à : <http://irem.univ-rouen.fr/sites/default/files/u17/Trouche%20St%20Flour.doc>

Autor

Bernard Parzysz

Laboratoire de Didactique André Revuz, Université Paris-Diderot, Université d'Orléans, France