Simulations d'expériences aléatoires en classe :

un enjeu didactique pour comprendre la notion de modèle probabiliste, un outil de résolution de problèmes(*)

Michel Henry(**)

Dans cet article, j'aborde la question des enjeux didactiques de la simulation informatique en classe comme partie prenante essentielle de l'enseignement de la statistique. La pratique de premières simulations simples permet aux élèves de consolider leur appréhension de la nature fréquentiste de la notion de probabilité, qu'ils ont naïvement éprouvée dès le plus jeune âge avec des jeux de hasard. Elle prépare aussi la compréhension des conditions de prises de décisions dans le domaine de l'incertain, au vu de données statistiques issues d'un échantillonnage aléatoire.

Dans cette optique, je propose un ensemble de cinq exemples de simulations inspirées de problèmes historiques, dans une progression conçue pour les élèves de la seconde à la terminale. Ces exemples vont de l'estimation de probabilités calculables *a priori* par dénombrements ou par l'introduction d'hypothèses de modèle ad hoc, à l'estimation fréquentiste d'une probabilité impossible à calculer d'avance.

I – Place de la simulation dans l'enseignement de la statistique dans le second degré

En collège, dès la classe de Sixième, les moyens informatiques, calculettes, ordinateurs et tableurs, sont utilisés pour exercer les élèves aux représentations et traitements de donnée statistiques (tableaux, diagrammes, histogrammes, effectifs, séries classées, calculs de fréquences, de moyennes, médianes et quartiles).

Le programme de seconde (BO n° 30 du 23 juillet 2009) propose de poursuivre cet enseignement dans le même esprit :

Statistique descriptive, analyse des données : « *Utiliser un logiciel (par exemple, un tableur) ou une calculatrice pour étudier une série statistique* ».

Dans le cadre de l'échantillonnage : « *Faire réfléchir les élèves à la conception et la*

(*) Cet article est la transcription d'un exposé que j'ai présenté au séminaire de didactique de Besançon en janvier 2011. Il complète et prolonge les quelques pages publiées dans le dossier « Les Probabilités » du Bulletin Vert n° 484 de septembre-octobre 2009. Il doit beaucoup à la contribution que nous avons proposée, Bernard Parzysz et moi-même, au groupe de travail sur la pensée statistique lors du colloque CERME 7 de février 2011.
On en trouvera une version en ligne avec des liens sur les fichiers Excel dans le n° 26 de septembre 2011 de MathémaTICE à l'adresse http://revue.sesamath.net/spip.php?article353.

(**) CII Statistique et probabilités. michel.henry@univ-fcomte.fr

mise en œuvre d'une simulation ; sensibiliser les élèves à la fluctuation d'échantillonnage, aux notions d'intervalle de fluctuation et d'intervalle de confiance et à l'utilisation qui peut en être faite ».

Échantillonnage : *« Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice ».*

Dans le cadre des probabilités : *« Étudier et modéliser des expériences relevant de l'équiprobabilité. Proposer un modèle probabiliste à partir de l'observation de fréquences dans des situations simples ».*

En première S, on introduit les notions de variables aléatoires discrètes et de lois de probabilités. Le programme indique notamment (JO du 28 août 2010) : *« 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 ».*

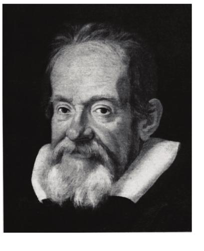
Cette orientation de travailler sur des séries statistiques obtenues par simulation n'est pas propre aux programmes français. En témoignent ces recommandations de quatre organisations américaines (AMS, American Mathematical Society en 2001, ASA, American Statistical Association en 2005, MAA, Mathematical Association of America en 1991, NCTM, National Council of Teachers of Mathematics en 1991) [Papaieronymou] :
« Les professeurs de mathématiques doivent être capables de planifier et conduire des expériences et des simulations, en distinguant les probabilités expérimentales et théoriques, de déterminer des probabilités expérimentales, d'utiliser des probabilités expérimentales et théoriques pour formuler et résoudre des problèmes de probabilités, et utiliser des simulations pour estimer les solutions de problèmes de hasard.
Les professeurs de mathématiques de l'enseignement secondaire devraient être capable de construire un modèle utilisant une probabilité théorique qui peut être comparée à des résultats expérimentaux, ce qui est essentiel pour l'étude du concept de fréquence ».

II – Un premier exemple simple de simulation : le problème du Grand Duc de Toscane

Voici le problème que le Grand Duc de Toscane posa à Galilée vers 1620 :

Comment parier sur la somme des points obtenus avec 3 dés ?

Portrait de Galilée



On peut préciser ainsi la question qui intriguait les joueurs de l'époque :

« Bien que le 9 et le 12 se décomposent en autant de façon que le 10 et le 11, si bien qu'ils devraient être considérés comme ayant la même chance, on voit néanmoins que la longue observation a fait que les joueurs estiment plus avantageux le 10 et le 11 plutôt que le 9 et le 12 ».

Question pour des élèves de seconde : est-il possible que « la longue observation » ait permis aux joueurs invétérés de remarquer les différences de fréquences entre les sommes 9 et 10 par exemple ?

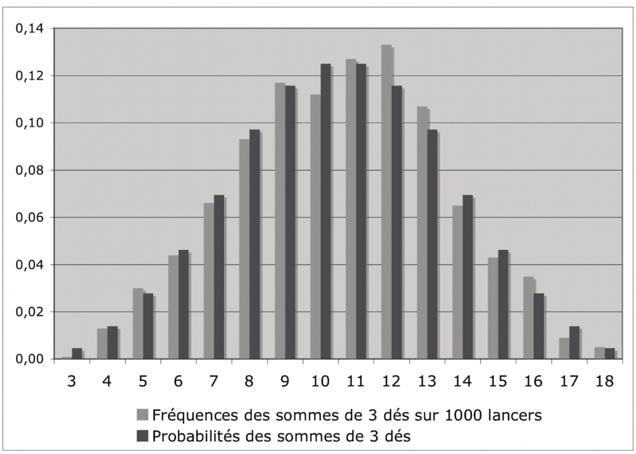
On peut répondre à cette question en simulant le lancer de trois dés un très grand nombre de fois. L'hypothèse de modèle est que les faces des trois dés sont parfaitement équiprobables. La simulation proposée ici sur le tableur Excel permet de répéter instantanément des suites de 1000 lancers :

Problème du Grand Duc de Toscane : sommes de 3 dés					
Dé 1 cellules B2:B1001	**Dé 2** cellules C2:C1001	**Dé 3** cellules D2:D1001	**Sommes des faces des trois dés** cellules E2:E1001	**Fréquence des 9 sur 1000 lancers** cellule F2	**Fréquence des 10 sur 1000 lancers** cellule G2
1+ENT(6*ALEA())	1+ENT(6*ALEA())	1+ENT(6*ALEA())	B2+C2+D2	=(NB.SI(E2:E1001;9)) /1000	=(NB.SI(E3:E1003;10)) /1000

Les histogrammes des fréquences obtenues, tels que celui qui suit, ne permettent pas de bien voir une différence significative entre celles des sommes 9 et 10.

Combien de fois les joueurs italiens avaient-ils pu jouer ? On peut penser qu'ils connaissaient le calcul de combinatoire (déjà connu au XIII^{e} siècle) que Galilée a présenté au Grand Duc dans sa réponse, montrant que parmi les 16 sommes possibles réalisées par les 216 triplets possibles avec trois dés, le 9 est obtenu avec 6 combinaisons réalisées par 25 triplets (probabilité : 0,116) alors que le 10, également décomposable en 6 combinaisons, l'est par 27 triplets (probabilité : 0,125).

Programmant ainsi les calculs des fréquences des sommes de 3 à 18, voici un exemple (réellement obtenu) des fréquences observées sur une simulation de 1000 lancers, comparées aux valeurs calculées des probabilités correspondantes : la prépondérance des sommes 10 et 11 sur les sommes 9 et 12 n'apparaît pas évidente…



III - Détermination d'une probabilité : trois contextes

1- Équiprobabilité postulée des événements élémentaires

La probabilité d'un événement aléatoire est le rapport du nombre des cas favorables qui réalisent cet événement à celui de tous les cas possibles (Premier principe de Laplace).

« Mais cela suppose les divers cas également possibles. S'ils ne le sont pas, on déterminera d'abord leurs possibilités respectives… Alors la probabilité sera la somme des possibilités de chaque cas favorable » (Deuxième principe [Laplace]).

Quelle définition pour la « possibilité » ?

On peut distinguer trois positions épistémologiques [Batanero, Henry & Parzysz, 2005] :

– Option objectiviste : les symétries du système générateur du hasard considéré engendrent l'équiprobabilité sur les issues possibles. La probabilité d'un événement est objectivement déterminée par le premier principe.

– Option subjectiviste : dans l'ignorance absolue des conditions de réalisation des issues de l'expérience, *c'est-à-dire telles que nous soyons également indécis sur leur existence* (Laplace), le plus raisonnable est de postuler l'équiprobabilité (principe de raison insuffisante).

– Option de la modélisation : les conditions de l'expérience permettent de proposer un modèle d'équiprobabilité dont la pertinence devra être contrôlée. On peut associer à une expérience le modèle de l'Urne de Bernoulli, contenant des boules équiprobables de deux couleurs dans une proportion donnée [Henry, 2001].

2- Estimation fréquentiste basée sur la loi des grands nombres

Une même expérience aléatoire est répétée un nombre n de fois suffisamment grand. On enregistre la fréquence F_n des issues réalisant un événement donné de probabilité notée p.

Alors (théorème de Bernoulli), il y a une probabilité aussi voisine de 1 que l'on veut, que l'écart entre la fréquence F_n des issues réalisant l'événement et sa probabilité p soit plus petit que tout ε donné (convergence en probabilité de F_n vers p).

Cette fréquence observée F_n peut donc être prise comme « mesure » à ε près pour estimer la probabilité p de l'événement, avec un risque inférieur à α de se tromper, par l'encadrement de confiance :

$$\mathrm{P}\left(F_n - \varepsilon < p < F_n + \varepsilon\right) > 1 - \alpha \quad \left(1 - \alpha \text{ est le niveau de confiance}\right).$$

Alfred Renyi en tire une « définition fréquentiste » de la probabilité [Renyi, 1966] :

« *La probabilité d'un événement est le nombre autour duquel oscille la fréquence de l'événement considéré…* »

Cette « définition fréquentiste » confond deux domaines qu'il faut pourtant bien séparer :
– le domaine de la réalité où l'on observe les fréquences F_n de réalisations d'un événement au cours de *n* répétitions d'une même expérience aléatoire,
– le domaine théorique (mathématique) où les objets sont définis abstraitement.

3- Méthode Bayésienne

La valeur de la probabilité d'un événement relève d'une appréciation subjective propre à chacun pouvant être corrigée par les résultats expérimentaux en application de la formule de Bayes. L'introduction du point de vue bayésien dans l'enseignement secondaire est actuellement objet de controverses. Nous n'y entrerons pas ici. Pour approfondir ces questions, on pourra consulter Wikipedia :
http://fr.wikipedia.org/wiki/Théorème_de_Bayes.

Face à ces diverses interprétations de la notion de probabilité, il nous faut clarifier les enjeux de cet enseignement dans le second degré. Il faut dépasser le « langage des chances » ainsi que le débat « philosophique » entre objectivistes et subjectivistes, tout en présentant conjointement la notion de probabilité sous ses deux visages, classique et fréquentiste.

Le point de vue de la modélisation réalise cet enjeu, donne des clés didactiques et contribue à la formation de la démarche scientifique : observation de la réalité - description - hypothèses - modèle abstrait - développement théorique - résolution de problèmes - interprétation dans le contexte réel - validation expérimentale.

La probabilité y est axiomatiquement définie comme un objet théorique, quantifiant idéalement la possibilité d'un événement calculée *a priori* ou estimée expérimentalement.

Une initiation au processus de modélisation fait donc partie des enjeux de l'enseignement secondaire de la statistique et des probabilités. La mise en œuvre de simulations concourt à cet objectif [Girard & Henry, 2005].

IV - Contextes didactiques et exemples de simulations

1- Notion de simulation

Les programmes font donc largement appel à la simulation informatique.

Le document d'accompagnement des programmes de première précisait [GEPS, 2001] :

« Modéliser consiste à associer un modèle à des données expérimentales, alors que simuler consiste à produire des données à partir d'un modèle prédéfini. Pour simuler une expérience, on associe d'abord un modèle à l'expérience en cours, puis on simule la loi du modèle ».

On trouve cette définition de la simulation dans l'Encyclopaedia Universalis :

« La simulation est l'expérimentation sur un modèle. C'est une procédure de

recherche scientifique qui consiste à réaliser une reproduction artificielle (modèle) du phénomène que l'on désire étudier, à observer le comportement de cette reproduction lorsque l'on fait varier expérimentalement les actions que l'on peut exercer sur celle-ci, et à en induire ce qui se passerait dans la réalité sous l'influence d'actions analogues ».

Il convient donc de faire d'abord le choix d'un modèle pour l'implanter dans les instructions de calcul d'un ordinateur.

L'approche fréquentiste suppose de reproduire une même expérience aléatoire dans les mêmes conditions. Il y a dans cette affirmation beaucoup d'implicites : peut-on remplacer un dé par un autre ? Ou par un autre générateur aléatoire comme une calculette (1+INT(6*ran#) sur Casio ou 1+FLOOR(6*rand()) sur TI) ou un ordinateur avec un tableur (sur Excel : 1+ENT(6*ALEA()) ou ALEA.ENTRE.BORNES(1;6)) ?

Beaucoup d'élèves ne reconnaissent pas la similarité entre des expériences d'apparences différentes, mais qui se réfèrent implicitement au même modèle probabiliste.

Dans les conditions de la simulation, celle-ci peut-elle vraiment remplacer l'expérience réelle et donner des réponses de nature probabiliste ?

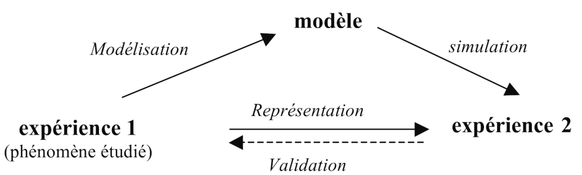
Bernard Parzysz a relevé que la notion de simulation est diversement présentée par les manuels dans lesquels on peut trouver quatre sortes d'interprétations [Parzysz, 2007] :
– *La simulation est un substitut de l'expérience ou une simple représentation à l'écran.*
– *Il doit y avoir une analogie entre l'expérience et sa simulation.*
– *La simulation économise du temps par rapport à l'expérience.*
– *La simulation est un modèle de l'expérience.*

Il faut comprendre le statut de la simulation : à partir d'un protocole expérimental, on dégage des hypothèses de modèle et on programme une simulation de ce modèle.

Les données expérimentales seront confrontées aux résultats de cette simulation pour adopter ou rejeter ce modèle. C'est une opération difficile mais cohérente avec le point de vue de la modélisation. Elle suppose en effet une introduction minimale à la théorie des tests d'hypothèses, notamment à la notion de risque acceptable quand on prend une décision à partir de données statistiques aléatoires.

2- La modélisation

Selon Parzysz [2009], on a le schéma suivant :



Parzysz note qu'il y a un réel problème pour les élèves qui ne savent pas ce qu'est un modèle probabiliste. Ce schéma triangulaire est alors compris comme linéaire, l'expérience 2 étant conçue comme une simple représentation de l'expérience 1.

Pour valider l'usage d'une simulation, Parzysz propose de partager la classe en deux groupes : dans un groupe les élèves procèdent à la répétition de l'expérience réelle (15 élèves × 20 expériences par exemple), dans l'autre groupe les élèves font tourner la simulation sur un ordinateur. La comparaison entre les résultats des deux groupes conduit à réfléchir sur les hypothèses de travail issues du protocole expérimental et leur correspondance avec les hypothèses de modèle implantées dans la simulation.

3- Simulations pour expliciter le processus de modélisation

Quelques premiers exemples simples de simulations conduisent les élèves à une meilleure compréhension de ce qu'est un modèle probabiliste et à s'intéresser au processus de modélisation tel que je l'avais décrit dans le bulletin n° 484 de l'APMEP [Henry, 2009] :
– décrire et analyser une expérience aléatoire,
– expliciter un protocole expérimental, i. e. l'ensemble des éléments qui définissent l'expérience d'un point de vue probabiliste, permettant d'affirmer que l'on peut répéter la même expérience dans les mêmes conditions,
– expliciter des hypothèses de travail en vue de contrôler la pertinence du modèle en construction,
– interpréter les caractéristiques de l'expérience réelle en termes d'hypothèses de modèle (notamment les probabilités représentant le caractère aléatoire de l'expérience),
– transposer ces hypothèses en instructions informatiques,
– exploiter ce modèle théorique pour en tirer des propriétés relatives au phénomène qui peuvent être observées dans la réalité,
– interpréter enfin les résultats de la simulation en les relativisant par les hypothèses de modèle.

Ainsi, une simulation informatique suppose d'implanter un modèle dont la pertinence reste à contrôler. La simulation est appelée à jouer un grand rôle dans le développement de la pensée statistique chez les élèves. L'exemple suivant présente deux modèles possibles pour le même problème.

4- Deuxième exemple : le problème « croix ou pile » de D'Alembert

D'Alembert fut associé à Diderot pour éditer l'Encyclopédie du XVIII^{e} siècle. D'Alembert y signe cet article contestataire devenu célèbre [D'Alembert, (1754)] :

Croix ou Pile (analyse des hasards)

« Ce jeu, qui est très connu, et qui n'a pas besoin de définition, nous fournira les réflexions suivantes. On demande combien il y a à parier qu'on amènera croix en jouant deux coups consécutifs. La réponse qu'on trouvera dans tous les auteurs, et suivant les principes ordinaires, est celle-ci. Il y a quatre combinaisons :

Premier coup	*Second coup*
croix	*croix*
pile	*croix*
croix	*pile*
pile	*pile*

De ces quatre combinaisons, une seule fait perdre et trois font gagner ; il y a donc 3 contre 1 à parier en faveur du joueur qui jette la pièce. [...]
Cependant cela est-il bien exact ? [...] ne faut-il pas réduire à une les deux combinaisons qui donnent croix au premier coup ? Car, dès qu'une fois croix est venu, le jeu est fini, et le second coup est compté pour rien. Ainsi, il n'y a proprement que trois combinaisons possibles :

croix,	*premier coup*
pile, croix,	*premier et second coup*
pile, pile,	*premier et second coup.*

Donc il n'y a que 2 contre 1 à parier. »

Bernard Parzysz a expérimenté ce problème avec des élèves de première ES [Parzysz, 2007], il écrit :

« Ce problème met les élèves dans l'embarras : comment départager les réponses ? (Notons qu'il ne peut y avoir d'argument purement mathématique, car l'équiprobabilité, à ce niveau, est subjective). Les élèves font spontanément des essais avec des pièces et sont vite convaincus que le joueur a de bonnes chances de gagner.

Mais quel est le bon pari : 3 contre 1 ou 2 contre 1 ?

Le nombre de parties jouées dans la classe est insuffisant pour trancher entre 2/3 (0,66) et 3/4 (0,75), il en faudrait plus de 1000. C'est un problème constant lors d'une estimation d'une probabilité par une évaluation fréquentiste.

Une simulation informatique va-t-elle permettre de résoudre le problème ?

Le protocole expérimental est bien posé par d'Alembert. Ajoutons que la pièce est bien équilibrée et correctement lancée (hypothèse de travail), ce qui conduit au modèle de Bernoulli pour un lancer avec p = 1/2 (hypothèse de modèle).

Comme souvent, il y a plusieurs possibilités (plusieurs modèles possibles). Les deux algorithmes qui suivent interprètent-ils rigoureusement le jeu de d'Alembert ?

Algorithme 1

Produire une variable aléatoire (alea 1) à deux issues équiprobables :
(C : aléa 1 < 0,5) et (P : aléa 1 ≥ 0,5).

- Si C arrive, gagné.
- Si P arrive, produire une même v. a. (alea 2) :
  + si C arrive, gagné ;
  + si P arrive, perdu.

Exemple de parties :

Aléa 1	Résultat	Aléa 2	?
.732	rejouer	.813	perdu
.307			gagné
.042			gagné
.967	rejouer	.274	gagné
.766	rejouer	.819	perdu

Algorithme 2

Produire deux mêmes variables aléatoires (alea 1, aléa 2) à deux issues (C) et (P) équiprobables.

- Si on obtient (P, P), perdu.
- Sinon, avec (C, C), (C, P) ou (P, C), gagné.

Exemple de parties :

Aléa 1	Aléa 2	?
.732	.813	perdu
.307	.785	gagné
.042	.624	gagné
.967	.274	gagné
.766	.819	perdu

Pour une simulation, le second algorithme est plus facile à programmer, mais il introduit un deuxième jet systématique qui n'est pas nécessaire en réalité.

Certains élèves n'acceptent pas cette deuxième procédure, de même que Roberval contestait la solution de Fermat au problème des partis, lorsqu'il prenait en compte des parties « feintes ».

L'observation des dernières colonnes des deux tableaux, remplis à la demande, montre que l'on peut supprimer la colonne « résultats » du premier sans rien changer à la colonne « ? » des deux tableaux.

On peut donc remplacer l'algorithme 1 par l'algorithme 2 sans changer le jeu du point de vue probabiliste ».

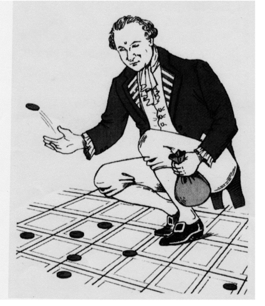
Croix ou pile de d'Alembert						
aléa 1 cellules A3:A1002	aléa 2 cellules B3:B1002	Gagné = 0, perdu = 1 cellules C3:C1002	Fréquence F_n des gains sur n = 1000 lancers cellule D3	Borne inf de l'intervalle de confiance à 95% cellule E3	Borne sup de l'intervalle de confiance à 95% cellule F3	Test : 0,66 = 0, 0,75 = 1
=ALEA()	=ALEA()	=SI(A1>0,5;0;1) *SI(B1>0,5;0;1)	=1–(SOMME(C3:C1002))/1000)	=D3–1/RACINE(1000)	=D3+1/RACINE(1000)	=SI(0,667<E3;1;0)

L'intervalle de confiance au niveau de 95% pour la probabilité de gagner est

$$\left]\mathrm{F}_n - \frac{1}{\sqrt{n}} ; \mathrm{F}_n + \frac{1}{\sqrt{n}}\right[.$$

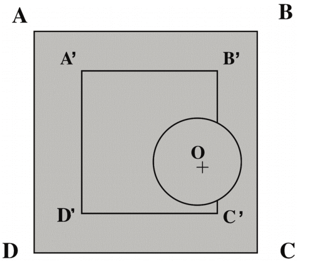
Cela signifie que dans 95% des échantillons simulés, la probabilité de d'Alembert sera comprise dans cet intervalle. Le test consiste donc à rejeter la valeur 0,66 comme étant dans moins de 5% des cas inférieure à la borne inf de l'intervalle de confiance.

5- Troisième exemple : le jeu du « franc-carreau » de Buffon



Avec le jeu du franc-carreau, Buffon introduit en 1733 une nouvelle notion de probabilité, appelée « probabilité géométrique » (loi uniforme sur un domaine plan) [Buffon]. Cette situation peut être présentée à différents niveaux. Les élèves de seconde peuvent en élaborer une simulation telle que celle qui suit.

Une pièce de monnaie est lancée sur un carrelage constitué de carreaux carrés. Un joueur parie sur la position de la pièce ; si elle s'arrête sur un seul carreau (franc-carreau) : gagné ; si elle rencontre un ou plusieurs joints : perdu.
Quelle est la probabilité de faire « franc-carreau » ?



Pour simuler cette situation, il nous faut faire des hypothèses de modèle. On considère le carreau représenté par le carré ABCD où le centre O de la pièce est tombé,
– on a « franc-carreau » si les distances du centre O aux côtés du carré sont supérieures au rayon de la pièce,
– la probabilité est de loi uniforme sur le carré ABCD. A′B′C′D′ est le carré qui délimite la zone dans laquelle peut se trouver le centre O de la pièce pour que l'on ait « franc-carreau ».

Ce problème est résolu théoriquement en comparant les aires des carrés A′B′C′D′ et ABCD. On obtient la probabilité géométrique : $A'B'^2/AB^2$.

Dans une simulation, avec AB = 1 et le rayon de la pièce égal à ¼, on a A′B′ = 1/2 et, dans ce modèle, la probabilité de « franc-carreau » vaut 1/4.

Avec la simulation de ce modèle, on obtient pour plusieurs séries de 10 000 lancers, des fréquences de « franc-carreau » égales à 0.2564 ; 0.2515 ; 0.2445 ; 0.2497 ; …

Le jeu du Franc-carreau de Buffon. La cellule A2 est réservée pour choisir le rayon de l'écu			
Abscisse de O cellules C2:C10001	Ordonnée de O cellules D2:D10001	O dans A'B'C'D' ? cellules E2 :E10001	Fréquence de Franc-carreau sur 10 000 lancers cellule B2
=ALEA()	=ALEA()	=SI(ET(A2<C2;C2<1-A2;A2<D2;D2<1-A2);1;0)	=(SOMME(E2:E10001))/10000

Ayant fait un calcul *a priori* de la probabilité (0,25 dans l'exemple) de faire « franc-carreau », les élèves peuvent évaluer la qualité de cette simulation, qui, malgré les fluctuations d'échantillonnage, donne une estimation (ponctuelle) assez bonne de cette probabilité. Ils peuvent alors éprouver une certaine confiance dans les performances de ce type de simulations, à condition que les tailles des échantillons soient assez grandes, confiance qui sera ensuite quantifiée par une estimation par intervalle.

V - La simulation comme outil de résolution de problème

1- Intérêt didactique des simulations

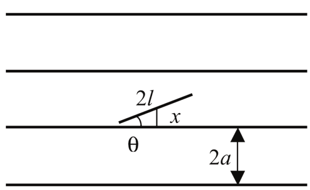
Les simulations permettent de travailler sur de grandes séries statistiques, donnant du sens aux résumés statistiques (paramètres de position et de dispersion, diagrammes et histogrammes) et montrant leur pertinence.
– Elles mettent en œuvre un processus de modélisation.
– Elles permettent une présentation dynamique de l'interaction entre les notions de fréquence et de probabilité.
– Elles permettent de répéter en classe une expérience aléatoire un assez grand nombre de fois pour induire une bonne compréhension de la loi des grands nombres.
– Un intérêt supplémentaire non négligeable est qu'elles permettent de résoudre des problèmes trop difficiles ou impossibles à traiter directement « à la main ».

Examinons deux exemples.
Le premier est « l'aiguille de Buffon », montrant les limites de la simulation.
Le deuxième est inspiré du jeu du « franc-carreau », où l'écu est remplacé par une baguette.

2- L'aiguille de Buffon : un résultat surprenant dans un modèle ad hoc

À la suite du problème du franc-carreau, Buffon donne une solution géométrique du problème bien connu dit de l'aiguille (une baguette dans le texte original). Ce problème peut être étudié en terminale.

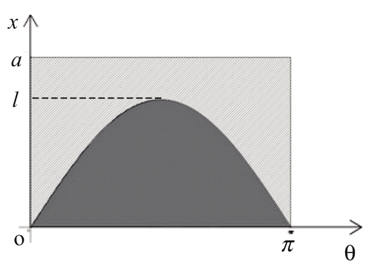
Sur un parquet formé de planches de largeur $2a$, séparées par des rainures droites, parallèles et équidistantes, on jette une aiguille de longueur $2l$, avec $l < a$. Quelle est la probabilité que l'aiguille coupe l'une des rainures ?



Hypothèses de travail :
Soit x la distance du milieu de l'aiguille à la rainure la plus proche. x prend une valeur aléatoire quelconque dans $[0, a]$. Soit θ l'angle des droites formées par cette rainure et l'aiguille. θ prend une valeur aléatoire quelconque dans $[0, \pi]$. L'aiguille, jetée au hasard, coupe une rainure si $x < l \sin \theta$ (événement A).

Buffon propose une représentation cartésienne de cette condition :

Dans cette représentation du pavé $\Omega = [0, \pi] \times [0, a]$, rectangle hachuré, cet événement A est représenté par la partie grisée délimitée par la courbe d'équation $x = l \sin \theta$.



Buffon fait implicitement l'hypothèse de modèle que la probabilité est uniformément répartie sur le pavé Ω. On a donc :

$$P(A) = \frac{\text{aire grisée}}{\text{aire du rectangle}} = \frac{1}{\pi a}\int_0^{\pi} l \sin\theta \, d\theta = \frac{2l}{\pi a}.$$

Remarques sur cette solution de Buffon :

Ainsi, le nombre π intervient dans l’expression d’une probabilité, non rationnelle, ce qui peut être surprenant si l’on s’en tient à la définition classique comme rapport de nombres de « cas ». Buffon a donc introduit une notion nouvelle de probabilité géométrique.

Dans ce problème, le modèle uniforme choisi ne s’applique pas à la situation géométrique d’origine, mais à sa représentation cartésienne. Aurions-nous le même résultat avec une représentation en coordonnées polaires ?

Quelle légitimité le modèle uniforme sur le rectangle cartésien a-t-il en l’occurrence ? Outre les lois uniformes pour x et θ, il fait intervenir une hypothèse d’indépendance entre x et θ. Ces hypothèses de modèle sont-elles pertinentes ?

L’expérimentation numérique par simulation informatique ne permettra pas d’y répondre, car ayant implanté ces hypothèses dans le programme, il ne sera pas surprenant d’observer que la fréquence stabilisée (sur 10 000 jets) de l’événement tend à confirmer le résultat théorique.

On pourrait imaginer donner une valeur approchée pour π à partir de la fréquence observée, comme avec la méthode de Monte-Carlo. C’est très peu performant comme on peut le voir avec la simulation.

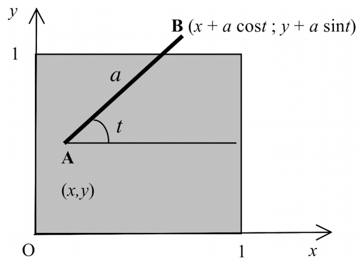
L’aiguille de Buffon						
l/a A3	x cellules B3:B10002	θ cellules C3:C10002	Coupe une raie = 1 cellules D3:D100002	Fréquence F_n des coupures sur 10 000 lancers cellule E3	Valeur de Buffon de la probabilité $2l/\pi a$ cellule F3	Valeur estimée de π par cette méthode cellule G3
0,8	ALEA()	PI()*ALEA()	SI((B3<A$3*SIN(C3));1;0)	(SOMME(D3:D10002))/10000	2*A3/PI()	2*A3/E3
0,8	0,042452161	2,968076081	1	0,5123	0,5093	3,1262

3- Un problème plus difficile : la baguette à franc-carreau

On suppose maintenant qu’une baguette – et non un disque – est lancée sur un carrelage constitué de carreaux carrés.

Dans ce cas, la probabilité de faire « franc-carreau » ne peut pas être calculée aussi facilement que dans le problème de Buffon. Le calcul reviendrait à une comparaison de volumes sous une hypothèse de répartition uniforme dans un pavé de R^3 que seuls des mathématiciens aguerris savent faire.

Modélisation : On représente la baguette par un segment [AB] de longueur a, et on considère le carreau où l’extrémité A est tombée. Ce carreau est représenté par le carré unité dans un repère cartésien orthonormé.



Soient x et y les coordonnées aléatoires de A ($0 \leq x < 1$, $0 \leq y \leq 1$) et t l'angle aléatoire entre l'axe des x et le vecteur AB ($0 \leq t \leq 2\pi$).

La baguette est à « franc-carreau » si les conditions suivantes sont vérifiées :
$0 \leq x + a \cos t < 1$ et $0 \leq y + a \sin t < 1$

Le volume en x, y, t inclus dans le produit cartésien $[0,1] \times [0,1] \times [0,2\pi]$ défini par ces conditions, n'est pas facile à calculer.

Mais une simulation de cette expérience est très accessible pour des élèves de terminale, avec l'hypothèse de modèle de la répartition uniforme de (x, y, t) dans le pavé $[0,1] \times [0,1] \times [0,2\pi]$.

Voici une réalisation possible de cette simulation.

Dans cet exemple, a = 0,44 ; sur 10 000 lancers, on a obtenu une fréquence de « franc-carreau » de 0,499, ce qui donne un intervalle de confiance à 95% : [0,489 ; 0,509].

Baguette à franc-carreau						
Longueur de la baguette AB Cellule A2	abscisse x de A dans le carré unité cellules B3:B10003	ordonné y de A dans le carré unité cellules C3:C10003	angle t de la baguette AB avec l'axe des x cellules D3:D10003	abscisse de l'extrémité B cellules E3:E10003	ordonnée de l'extrémité B cellules F3:F10003	Franc-carreau = 1 Cellules G3:G10003
0,44	ALEA()	ALEA()	2*3,14*ALEA()	B3+A$2*COS(D3)	C3+A$2*SIN(D3)	SI((E3>0)*ET(E3<1)*ET(F3>0)*ET(F3<1);1;0)
	0,306	0,477	5,786	0,692	0,267	1

Pour terminer, je vous propose de résoudre le problème suivant.

L'unité de longueur étant le côté d'un carreau, à l'aide de simulations, estimer à 0,01 près la longueur que doit avoir la baguette pour que la probabilité qu'elle tombe à franc-carreau soit égale à un p donné (par exemple 0,1 ; … ; 0,9) ?

V – Une petite conclusion

Le lecteur qui aura eu la patience de me lire jusqu'ici aura pu noter l'évolution du statut de la probabilité au fil de ces exemples :

– une donnée théorique calculable *a priori* par dénombrement des différents triplets possibles dans le problème du Grand Duc et vérifiable approximativement par l'instrument de mesure qu'est la simulation (modulo l'hypothèse que les dés sont parfaits),

– un moyen de trancher une controverse malicieusement introduite par d'Alembert dans la Grande Encyclopédie, avec la seule hypothèse que la pièce est bien équilibrée, la prise de décision relevant d'une première démarche de test d'hypothèse,

– une valeur interprétant géométriquement les « chances » de faire franc-carreau à ce jeu, dès lors que l'on considère que le centre de l'écu peut tomber en n'importe quel endroit d'un carreau, ceci de manière uniforme. Pour réaliser une simulation, il devient nécessaire d'expliciter ce modèle probabiliste admis par Buffon, pour constater que, dans ce modèle, cette simulation redonne approximativement cette valeur facilement calculée,

– une estimation ponctuelle de la fameuse probabilité de Buffon dans le problème de l'aiguille, accessible dès la classe de première, moyennant l'hypothèse d'uniformité (et d'indépendance) des deux variables en jeu, alors que le calcul théorique passe par celui d'une intégrale du niveau terminale S, à condition d'accepter le modèle de répartition uniforme de la probabilité dans une représentation cartésienne de l'événement considéré,

– une réponse à un problème (d'école) inaccessible par le calcul (quelle longueur pour la baguette ?), alors que sa résolution par simulations est possible en terminale S sans grandes difficultés.

Dans cette progression, il apparaît de plus en plus nécessaire d'expliciter les modèles sous- jacents aux simulations effectuées. Ces exemples peuvent être aussi considérés comme un apprentissage « en acte » du processus de modélisation.

Références

Batanero, C., Henry, M., & Parzysz B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school : Challenges for teaching and learning* (p. 20-42). New York: Springer.

Buffon, G. L. Leclerc (1733). Solution de problèmes sur le jeu du franc-carreau, Mémoire présenté à l'Académie Royale des Sciences. In *Essai d'arithmétique morale, Histoire Naturelle, générale et particulière*, Supplément, tome quatrième, Paris, Imprimerie Royale, 1777, 46-148.

D'Alembert, J. Le Rond (1754). *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, Diderot et d'Alembert eds, tome IV, article Croix ou Pile : http://diderot.alembert.free.fr/C.html

Groupe d'Experts sur les Programmes Scolaires, Direction de l'Enseignement Scolaire (2001). *Accompagnement des programmes de mathématiques, classe de Première des séries générales*. Paris: Centre National de Documentation Pédagogique (CNDP). En ligne : www.sceren.fr/produits/detailsimp.asp?ID=84491.

Girard, J. C. & Henry, M. (2005). Modélisation et simulation en classe, quel statut didactique ? *Statistique au lycée*, vol. 1 (Chaput, B. & Henry, M., eds), 147-159. Paris : APMEP, brochure nº 156.

Henry, M. (2001). Notion de modèle et modélisation dans l'enseignement, *Autour de la modélisation en probabilités*, Besançon, Presses Universitaires Franc-Comtoises, 149-159.

Henry, M. (2009). À propos du programme de statistique en Seconde : remarques sur la simulation informatique, *Bulletin de l'APMEP* n° 484, septembre-octobre 2009, 598-602.

Kahane, J. P. (2002) (Ed.), *L'enseignement des sciences mathématiques : Rapport au Ministre de l'Éducation Nationale* de la Commission de Réflexion sur l'Enseignement des Mathématiques. Paris : Odile Jacob.

Laplace, Pierre-Simon (1825). *Essai philosophique sur les probabilités*. Bourgois, Paris 1986. Dans *Théorie analytique des probabilités* en 2 vol. Jacques Gabay, Paris 1995.

Papaieronymou, I. (2009). Recommended Knowledge of Probability for Secondary Mathematics Teachers. *Proceedings of CERME 6* (Working Group 3 on “ stochastic thinking ”). On line : http://www.inrp.fr/editions/editions-electroniques/cerme6/.

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, 7-44.

Parzysz, B. (2009). Des expériences au modèle, via la simulation. *Repères-IREM* 74, 91-103.

Renyi, A. (1966). *Calcul des probabilités*, Dunod. Réédition : Jacques Gabay, Paris 1992.