

À propos du programme de statistique en Seconde : remarques sur la simulation informatique

Michel Henry(*)

1. Peut-on simuler le hasard ?

Cette question n'est pas anodine⁽¹⁾. Elle nous renvoie à notre interprétation de ce que nous appelons le hasard. Hasard sauvage à la Cournot (la tuile qui, tombant fortuitement du toit d'une maison près de laquelle vous passez, vous arrive sur la tête) ou hasard bénin, hasard des générateurs de hasard familiers intervenant dans des processus reproductibles⁽²⁾. C'est notamment le cas du hasard des prélèvements « au hasard » dans une population statistique pour en extraire des échantillons aléatoires à partir desquels de grands théorèmes probabilistes permettent de prendre des décisions rationnelles⁽³⁾. Le développement au vingtième siècle de la statistique mathématique a rendu possible des applications puissantes, supposant la production de chiffres « au hasard » en grands nombres. Il n'est pas concevable de les obtenir par des tirages « au hasard » (i.e. équiprobables) de boules numérotées de 0 à 9 dans une urne. Avant l'ère des ordinateurs, on utilisait des tables de « nombres au hasard », par exemple celles de la Rank Corporation, désuètes maintenant, mais encore pratiques en classe pour certaines activités.

On exploite donc les performances de calcul et la rapidité de l'ordinateur ou de la calculette pour obtenir des chiffres au hasard, qui apparaissent avec une probabilité de 1/10 suivant la loi uniforme discrète dite équirépartie. Mais l'ordinateur est-il un véritable générateur de hasard ? En principe, pas encore, tant que des phénomènes physiques aléatoires, comme l'exploitation du bruit électronique par exemple, ne sont pas introduits à cette fin. Actuellement, les nombres pseudo-aléatoires donnés par l'ordinateur ou la calculette sont déterminés dès lors que leur calcul a été initialisé, opération qui produit à chaque fois des suites différentes⁽⁴⁾.

(*) IREM de Franche-Comté.

(1) Signalons à ce sujet la sortie récente de l'ouvrage de Nicolas GAUVRIT : *Vous avez dit hasard ? - entre mathématiques et psychologie* (Belin, 2009).

(2) Voir par exemple : LAHANIER-REUTER, D. (1999). *Conceptions du hasard et enseignement des probabilités et statistiques*, col. Éducation et formation, Presses Universitaires de France.

(3) Sur le hasard en statistique, on pourra lire l'article de Jean-Louis PIEDNOIR : *DU HASARD : Des statistiques aux probabilités, des probabilités à la statistique*, dans ce même dossier.

(4) Sur l'obtention de chiffres au hasard avec une calculette, voir l'article d'Alain LADUREAU : *Utiliser une calculatrice comme générateur de hasard pour résoudre des problèmes de mathématiques*, dans *Statistique au lycée*, vol. 2, p. 81-92, brochure APMEP n° 167, 2007.

Mais la complexité de leur production, supposant un calcul lourd, rend impossible leur prévision directe par tout calculateur humain, même prodige. *On a alors le phénomène fortuit* selon l'expression de Henri Poincaré. Tout se passe donc comme si les chiffres fournis par la fonction ALEA de l'ordinateur ou RANDOM de la calculette étaient effectivement issus de tirages au hasard de boules dans l'urne précédente. La condition est que cette génération pseudo-aléatoire vérifie différents tests d'uniformité assez sophistiqués⁽⁵⁾. En tout cas, les moyens de contrôle à notre disposition ne permettent pas d'invalider cette hypothèse et, faisant relativement confiance au matériel informatique mis à notre disposition, nous pouvons déclarer que l'ordinateur simule la loi uniforme discrète. Le problème d'obtenir un tel comportement de l'ordinateur est une affaire de spécialistes. Les algorithmes utilisés ont été progressivement améliorés. Cependant, la suite pseudo aléatoire des chiffres ainsi produits est en fait périodique sur une très longue période. Pour nos simulations, limitant nos besoins à un nombre raisonnable de données (quelques dizaines de milliers), ce problème ne se pose pas.

Les premières méthodes laissaient à désirer. Alain Ladureau cite l'exemple de l'algorithme proposé par John Von Neumann vers 1940 : on se donne un nombre entier positif d'au plus 6 chiffres, on l'élève au carré et on complète éventuellement le nombre ainsi obtenu en ajoutant à gauche le nombre de zéros nécessaires pour obtenir un nombre de 12 chiffres. On garde les 6 chiffres du milieu du nombre précédent et on recommence, on construit ainsi une suite d'entiers ayant tous 6 chiffres grossièrement équirépartis. Mais si l'on choisit sans précaution le nombre initial, on peut arriver très vite à une suite périodique, de période courte, qui nous donne donc une simulation peu crédible d'un comportement aléatoire ! Par exemple, avec 123 456, on obtient toujours 625 000 au bout de 146 nombres.

2. Peut-on résoudre des problèmes par simulation ?

Voici un exemple simple de simulation exploitable en classe, inspiré par le problème du Grand Duc de Toscane : quand on lance trois dés et que l'on fait la somme des points obtenus, il semble que la somme 10 apparaisse plus souvent que 9, alors qu'elles sont obtenues l'une et l'autre par 6 configurations parmi les 56 observables. Pour le vérifier sur un tableur, on garnit chacune des trois premières colonnes A, B, C par la répétition un millier de fois de la fonction ALEA.ENTRE.BORNES(1;6), la quatrième colonne D fait la somme ligne par ligne de ces trois premières, la cinquième et la sixième comptent les fréquences du 9 et du 10 sur 1000 lancers par les fonction NB.SI(D1:D1000;9) et NB.SI(D1:D1000;10). En répétant ces 1000 lancers, on constate que les sommes 10 apparaissent quasi systématiquement plus souvent que les 9, observant une fréquence moyenne d'environ 12,5% pour le 10 contre 11,5% pour le 9. Galilée avait donné une explication de cette observation en distinguant les configurations observables des « manières de tomber », c'est-à-dire des triplets possibles. Il y en a $6^3 = 216$ équiprobables, 25 d'entre eux donnent une somme de 9 (11,7%) et 27 une somme de 10 (12,5%).

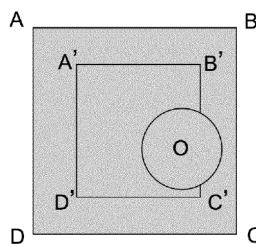
(5) Cf. l'étude de Bernard Parzys : *Quelques questions à propos des tables et des générateurs de hasard*, dans *Statistique au lycée*, vol. 1, p. 181-194, brochure APMEP n° 156, 2005.

Cet exemple de simulation est une bonne approche de la notion de probabilité, moins intuitif que la simulation des lancers d'un seul dé, apportant la réponse à un problème non trivial pour les élèves. Cette simulation donne des fréquences proches des probabilités des sommes 10 et 9, calculables a priori par dénombrement des « triplets favorables » dans un contexte d'équiprobabilité. Elle est pour les élèves une sorte d'introduction intuitive à la loi des grands nombres.

Mais avons-nous réellement simulé quelque chose ? Ou avons-nous seulement vérifié que le résultat de la simulation, par son adéquation aux probabilités calculées a priori, confirme que le fabricant de l'ordinateur et le concepteur du logiciel ont bien rempli leurs cahiers des charges ? Sans doute, en réalité c'est ce que nous avons fait. Cependant, ne négligeons pas l'intérêt de l'outil informatique. La simulation permet en fait de résoudre des problèmes trop difficiles ou impossibles à traiter directement.

Prenons l'exemple du jeu de Franc-Carreau. Ce jeu a été proposé pour la première fois en 1733 par le naturaliste Georges-Louis Leclerc, comte de Buffon. Il consiste à jeter un écu au-dessus d'un carrelage. Les joueurs parient sur sa position : tombera-t-il sur un seul carreau (à franc-carreau), ou rencontrera-il un des joints séparant les carreaux ?

Si on désigne par ABCD le carreau dans lequel le centre O de l'écu est tombé, la condition franc-carreau est géométriquement élémentaire : elle est réalisée si et seulement si le centre du disque tombe à l'intérieur du carré A'B'C'D', homothétique du carré ABCD par rapport à son centre. Buffon considère comme évident que la *probabilité* de « franc-carreau » est égale au rapport des aires des carrés A'B'C'D' et ABCD.



En simulant sur l'écran de l'ordinateur⁽⁶⁾ le jeu du Franc-Carreau, ayant implanté le modèle probabiliste sous-jacent (répartition uniforme de la probabilité sur un carré de côté c représentant le carreau sur lequel le centre de l'écu est tombé), la fréquence des franc-carreaux obtenue sur un millier de lancers, se présente conforme à ce qui était attendu par le calcul a priori dans le cas du jet d'un écu en forme de disque de rayon r : rapport de l'aire du carré de même centre, de côté $c - 2r$, à l'aire c^2 du carreau.

Mais, remarquons qu'en remplaçant l'écu par un triangle, la probabilité de faire franc-carreau n'est pas calculable (facilement) a priori. Cependant, mis en confiance par la bonne performance de l'ordinateur dans le cas du disque, en simulant ce jeu du triangle, nous inférons qu'il se comportera à nouveau raisonnablement et fournira pour le triangle à franc-carreau une fréquence stabilisée à peu près égale à la probabilité qui pourrait éventuellement être calculée a priori⁽⁶⁾.

Nous avons ainsi un outil de résolution de problèmes jouant le même rôle que les caleuses graphiques quand elles tracent des courbes représentatives de fonctions données, ce qui suppose une bonne dose de confiance dans le bon fonctionnement de cet outil.

(6) Cette simulation a été réalisée sur Cabri-Géomètre avec des élèves de troisième par Cileda COUTINHO pour sa thèse soutenue à l'UJF de Grenoble en 2001. La loi uniforme continue sur le carré a été simulée par une loi uniforme discrète par discrétisation fine de la surface.

3. Intérêt didactique de la simulation, le processus de modélisation

Mais notre question est plus fondamentalement didactique. L'ordinateur nous permet d'abord de travailler sur de vastes séries statistiques, ce qui donne sens aux résumés statistiques et montre leur pertinence. Il nous permet ensuite une présentation animée du fonctionnement et de l'interaction des notions de fréquence et de probabilité introduites maintenant en troisième, soit théoriquement dans un contexte d'équiprobabilité, soit au travers d'expériences pratiques en nombres assez grands pour pouvoir déboucher valablement sur une bonne compréhension en acte de la loi des grands nombres.

Même si en effectuant une simulation, l'ordinateur en réalité ne fait pas intervenir la notion de probabilité (car il se limite à exhiber les effets sur les fréquences affichées du principe d'équirépartition des chiffres pseudo-aléatoires qu'il génère, principe qui entre dans ses spécifications), l'usage que nous en faisons dans la classe, comme pseudo-générateur de hasard, permet de mieux appréhender ces notions de fréquence, de fluctuations d'échantillonnages, de variabilité des paramètres empiriques et de probabilité.

Mais on ne peut se satisfaire de la seule exploitation de la puissance et la rapidité de l'ordinateur pour produire des chiffres aléatoires, bien que cela permette de présenter aux élèves une grande richesse de nouvelles expériences aléatoires virtuelles. Cela, me semble-t-il, n'aurait qu'un intérêt limité. Son intérêt didactique, en tant qu'outil de simulation, suppose plus fondamentalement la compréhension du processus de modélisation :

- description et analyse de la situation aléatoire en jeu,
- élaboration d'un protocole expérimental, c'est-à-dire l'ensemble des éléments caractérisant l'expérience du point de vue probabiliste : liste des résultats observables, actions à effectuer pour pouvoir affirmer que l'on répète *la même expérience*, caractérisation des événements auxquels on s'intéresse en fonction des issues qui les réalisent,
- explicitation d'hypothèses de travail pour contrôler la pertinence du modèle en voie d'élaboration,
- interprétation en termes d'hypothèses de modèle, notamment sur la loi de probabilité adéquate pour représenter l'intervention du hasard dans l'expérience réelle, et traduction de ces hypothèses en instructions informatiques pour que l'ordinateur nous permette de résoudre des problèmes éventuellement inaccessibles par le calcul a priori,
- exploitation théorique du modèle pour dégager des propriétés reflétant des phénomènes observables dans la réalité,
- enfin, il convient d'interpréter les résultats obtenus dans la simulation en les rapportant aux hypothèses de modèle introduites.

Simulant ainsi une expérience aléatoire réellement effectuée, au-delà de l'exploration statistique, en ponctuant chaque étape de ce processus, l'ordinateur devient un outil didactique majeur pour l'apprentissage de la modélisation en probabilités.

Les différentes étapes de ce processus de modélisation méritent d'être pointées sur des exemples un peu plus élaborés que les exemples élémentaires traités en classe comme les lancers de dés. Les exemples suivants, auxquels je renvoie les lecteurs intéressés, supposent des exposés assez détaillés qui ne peuvent trouver leur place ici.

Dans mon article *Des lois continues, pourquoi et pour quoi faire ?* paru dans le n° 51 de *Repères IREM* d'avril 2003 (disponible en ligne sur le site de *Repères* : http://www.univirem.fr/spip.php?article=71&id_numero=51&id_article_reperes=35), on trouvera plusieurs exemples de modélisations non triviales, notamment :

- Le jeu du Franc-Carreau, accessible en classe de troisième. Cet exemple suppose un peu de soin dans la rédaction des hypothèses de travail. Pour en faire une simulation, il convient de remplacer le modèle de la loi uniforme continue en un modèle discret où la définition classique de la probabilité justifie le calcul intuitif de Buffon qui repose sur une notion de probabilité géométrique.
- Le problème de l'aiguille de Buffon (quand on lance une aiguille sur un parquet, quelle est la probabilité qu'elle rencontre une des rainures). Cette situation est interprétée dans une représentation cartésienne des deux paramètres aléatoires qui déterminent la position de l'aiguille : la distance de son milieu à la rainure la plus proche et l'angle qu'elle forme avec celle-ci. Le modèle de la loi uniforme continue sur cette représentation permet à Buffon d'obtenir une réponse faisant intervenir l'aire inscrite sous un arc de sinusé. Cet exemple est tout à fait accessible en terminale.
- Le phénomène de la radioactivité naturelle. Il a fait l'objet de TP de physique en terminale S⁽⁷⁾. Sa modélisation fait intervenir la loi de la durée de vie sans vieillissement étudiée en mathématiques. L'article de *Repères* propose un approfondissement de cet exemple qui conduit à l'étude des processus de Poisson. La simulation de ce phénomène fait intervenir une loi binomiale dans un modèle discrétisé. Elle est présentée dans l'article de *Statistique au lycée*⁽⁸⁾, les fichiers sur Excel (Simulation-Poisson.xls) et OpenOffice (Simulation-Poisson.odt) sont disponibles sur le site de la commission inter IREM *Statistique et Probabilité* à l'adresse : <http://sierra.univ-lyon1.fr/irem/CII-Stat/V2.htm>.
- Dans le même volume de *Statistique au Lycée*, page 64, on trouvera un exemple de simulation du problème de régulation des feux dans un carrefour⁽⁹⁾. Cet exemple de modélisation de files d'attente, accessible en BTS ou IUT, simule la somme de variables exponentielles indépendantes. Cette simulation est présentée sur les tableurs Excel et OpenOffice dont on trouvera les fichiers (Feux.xls et Feux.odt) à l'adresse <http://sierra.univ-lyon1.fr/irem/CII-Stat/V2.htm>.

(7) Ce TP est décrit dans l'article de Dominique GAUD : *Quelques éclairages sur la radioactivité* dans *Repères-IREM* n° 65, octobre 2006, p. 17-42.

(8) Cet exemple est traité en détails dans l'article : *Simulation d'un processus de Poisson : exemple de la radioactivité naturelle*, dans *Statistique au lycée*, vol. 2, p. 101-114, brochure APMEP n° 167, 2007.

(9) Cf. l'article de Brigitte CHAPUT et Daniel VAGOST : *Simulations sur calculatrice et tableur*, dans *Statistique au lycée*, vol. 2, p. 49-66, brochure APMEP n° 167, 2007.