Pratiques expérimentales et modélisation: quelques questions didactiques posées par la simulation informatique

MICHEL HENRY*

Résumé
Nous abordons la question du statut didactique de la simulation informatique d´un échantillonnage aléatoire dans une population statistique. Nous nous interrogeons sur l´efficacité didactique des dialectiques nées de la présence massive de l´ordinateur dans les situations proposées aux élèves.
Mots-clé: enseignement de la statistique; simulação informatique; résolution expérimentale des problèmes.

Resumo
Abordamos a questão do estatuto didático da simulação informática de uma amostra aleatória em uma poplulação estatística. Questionamos a eficácia didática das dialeticas que derivam da introdução massissa do computador nas situações propostas aos alunos.
Palavras-chave: *ensino da estaística; simulação informática; resolução experimental de problemas.*

Abstract
We approach the question of the didactic status of the computational simulation of a random sample in a statistical population. We question the didactic efficacy of the dialectics derived from the mass introduction of computers in situations proposed to students.
Key-words: *teaching of statistics; computational simulation; experimental problem-solving.*

Intentions

Dans ce article, on aborde la question du statut didactique de la simulation informatique d'un échantillonnage aléatoire dans une population statistique. L'option expérimentaliste très marquée de la présentation des nouveaux programmes de statistique et probabilités des lycées pose de nouvelles questions pour l'organisation de cet enseignement. On peut notamment s'interroger sur l'efficacité didactique des dialectiques

* IREM de Franche-Comté. E-mail: michel.henry@univ-fcomte.fr

nées de la présence massive de l'ordinateur dans les situations proposées aux élèves:

– simulation informatique de données statistiques, approche expérimentale de leur gestion et compréhension de concepts de base: prélèvements aléatoires dans une population, échantillonnage, caractères, distributions de fréquences, fluctuations d'échantillonnage,
– apports de l'ordinateur comme générateur puissant de nombres au hasard (suites de chiffres pseudo-aléatoires) et introduction à la modélisation probabiliste d'expériences aléatoires,
– résolution expérimentale de problèmes et savoirs théoriques impliqués.

Remarques sur la simulation informatique

Une définition:

> 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. (Encyclopédia Universalis)

L'ordinateur est-il un véritable générateur de hasard ? En principe, pas pour le moment, tant que des phénomènes physiques, comme l'exploitation du bruit électronique par exemple, ne sont pas introduits à cette fin. Actuellement, les nombres pseudo-aléatoires que l'ordinateur (ou la calculatrice) fournit sont déterminés, dès lors qu'il a commencé leur calcul sur la base d'une initialisation (randomisation) qui, en principe, détermine à chaque fois des suites différentes. Mais la complexité de leur détermination, supposant un calcul lourd, rend impossible leur prévision par tout autre moyen humain. "*On a alors le phénomène fortuit*" selon l'expression de Poincaré. **Tout se passe donc comme si** les chiffres fournis par la fonction Random de l'ordinateur étaient issus d'un tirage aléatoire de boules numérotées de 0 à 9 dans une urne. La condition est que cette génération pseudo-aléatoire vérifie différents tests d'uniformité.

En tout cas, les moyens de contrôle à notre disposition ne permettent pas d'invalider cette **hypothèse,** et nous pouvons déclarer que l'ordinateur **simule** les tirages "au hasard" successifs de ces boules de l'urne. Le problème d'obtenir un tel comportement de l'ordinateur est un problème de spécialiste: on sait que la suite "aléatoire" de ces chiffres générés est en fait périodique sur une très longue période, mais pour notre simulation, nous limitant à un nombre raisonnable de données, ce problème ne se pose pas.

Mais, par exemple, en "simulant" un sondage sur Excel, ayant précisé notre modèle probabiliste sous-jacent (loi de Bernoulli $B(1, p)$ pour la variable parente X_0 où p est la proportion dans la population P à estimer), ayant introduit la valeur p dans l'ordinateur, avons-nous réellement simulé quelque chose ? Ou avons-nous seulement vérifié que le résultat de la simulation (conforme à ce qui était donc attendu), 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. Dans notre exemple, il permet une approche expérimentale des situations de sondages, et si la proportion p est cachée aux élèves, nous avons un outil de résolution de problèmes jouant le même rôle que les calculettes graphiques quand elles tracent des courbes représentatives de fonctions données, ce qui suppose une bonne dose de confiance (parfois mal placée car trahie dans certaines conditions particulières) dans le bon fonctionnement de cet outil.

Mais notre question est plus fondamentalement didactique. L'ordinateur nous permet d'abord de travailler rapidement sur de vastes séries statistiques, ce qui donne aux résumés statistiques (moyennes, écarts types etc.) toute leur importance. Il nous permet ensuite une présentation animée du fonctionnement et de l'interaction des notions de fréquence et de probabilité abordées auparavant, soit théoriquement, soit au travers d'expériences pratiques en nombre trop limité pour pouvoir déboucher valablement sur une bonne compréhension de la loi des grands nombres.

Même si, dans son fonctionnement réel, l'ordinateur ne fait pas intervenir de 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 faire comprendre **en actes** ces notions de fréquence, de fluctuations d'échantillonnages et de probabilité.

Mais on ne peut se satisfaire de la seule exploitation de la puissance et la rapidité de l'ordinateur pour générer des nombres aléatoires, permettant de présenter aux élèves une grande richesse de nouvelles expériences aléatoires, car cela n'aurait qu'un intérêt limité. Son intérêt didactique, **en tant qu'outil de simulation**, tient plus essentiellement en ce qu'il nous oblige à analyser la situation aléatoire en jeu, à émettre des hypothèses de modèle, notamment sur la loi de probabilité idoine pour représenter l'intervention du hasard dans l'expérience réelle (par exemple sur le choix de la valeur de la probabilité de Bernoulli à implanter), et à traduire ces hypothèses en instructions informatiques pour que l'ordinateur nous permette de résoudre des problèmes éventuellement inaccessibles par le calcul a priori. De ce point de vue, cela suppose la compréhension du processus de modélisation et d'interpréter les résultats obtenus, rapportés aux hypothèses de modèle introduites. Simulant ainsi une expérience aléatoire réellement effectuée, au-delà de l'exploration statistique, l'ordinateur devient un outil didactique majeur pour l'apprentissage de la modélisation en probabilités.

Modélisation, statuts de l'activité mathématique et didactique

Lors d'une émission télévisée sur les travaux de Paul-Emile Victor au Groenland, l'ethnologue Claude Levi-Strauss déclarait:

> La difficulté pour un ethnologue, c'est d'être à la fois dedans et dehors, il faut qu'il vive pleinement avec les gens dont il veut étudier le mode de vie, mais il faut aussi qu'il se comporte comme un observateur extérieur à eux, sinon il ne fait plus d'ethnologie.

Il apparaît que l'enseignant de mathématiques dans le second degré se trouve confronté à une difficulté similaire à l'ethnologue, lors de l'enseignement des probabilités. En effet partant d'une observation, puis d'une description de la réalité, par étapes successives, l'enseignant doit aider ses élèves à s'extraire peu à peu de cette réalité pour construire progressivement un modèle mathématique.

Or ces étapes successives posent beaucoup de problèmes à cet enseignant, car il a un pied dans l'observation de la réalité et l'autre dans la construction du modèle. Souvent il ne sait plus très bien où se situe son enseignement, ce qu'il traduit parfois par la déclaration: «Les probabilités et les statistiques au lycée, ce n'est pas vraiment des mathématiques».

Cela pose certaines questions didactiques, notamment:

> – Est-ce que le passage par l'observation de la réalité est nécessaire pour construire le modèle ?
>
> – Ne pourrait-on pas se placer directement dans le modèle au moyen d'axiomes pour l'étudier, puis utiliser ce modèle par la suite dans des descriptions de réalités ?

Or, les mathématiques trouvent une part de leur légitimité comme discipline de service, par le transfert de leurs concepts et méthodes pour résoudre des problèmes externes, posés par le développement des connaissances dans d'autres secteurs de l'activité humaine.

Sur le plan didactique, ce transfert suppose, au cours de l'apprentissage, d'avoir relié ces concepts, qui dans le cadre mathématique reçoivent des définitions précises, aux notions et idées générales qui se dégagent du travail d'exploration de la réalité et de sa modélisation.

Cette nécessité épistémologique converge avec l'objectif didactique de donner du sens aux objets introduits, en les reliant aux expériences et aux notions pré construites des sujets apprenants.

Notion de modèle

Le terme de «modèle» est utilisé dans des sens assez variés, voire contraires. La notion de modèle à laquelle je me réfère, a émergé au cours des années 60 lorsque est apparue la nécessité de mieux préciser la distinction entre le sujet du monde réel que l'on étudie et son idéalisation. On ne retient pour cette étude que certains des aspects caractéristiques de la réalité, qui semblent être pertinents et que l'on simplifie en une abstraction de cette réalité. Mais nous distinguerons une simple description ou une représentation de la réalité d'un modèle qui, dans sa structure, contient une part de connaissances théoriques permettant d'évaluer, interpréter et généraliser cette réalité. Cette remarque nous conduit à la définition suivante, qui distingue le modèle en tant que structure abstraite de la symbolique utilisée pour le décrire:

> [...] un modèle est une interprétation abstraite, simplifiée et idéalisée issue d'une description d'un objet réel, ou d'un système de relations, ou d'un processus évolutif de la réalité.

Ce modèle peut être représenté dans différents systèmes de signes: images, schémas, langages ou symbolismes, s'inscrivant dans différents registres de représentations, plus ou moins isomorphes.

Par exemple, on peut présenter un modèle par une analogie, en y introduisant des objets idéalisés de la réalité. Cela veut dire que dans un vocabulaire courant, les objets du modèle sont doués de propriétés caractéristiques bien définies. Je parle alors de modèles «pseudo-concrets». C'est le cas notamment des modèles d'urnes en probabilités, où l'hypothèse implicite est l'équiprobabilité des boules dans un tirage «au hasard».

Parmi les différents registres de représentations, le langage et le symbolisme mathématique permettent des descriptions puissantes sur lesquelles peuvent opérer des propriétés et des algorithmes généraux. Nous les appellons «modèles mathématiques». Souvent, ils nous sont tellement familiers que nous n'en voyons pas d'autres, et nous avons tendance à confondre ces représentations avec les objets idéaux en jeu, lesquels sont souvent confondus avec la réalité qu'ils modélisent. C'est particulièrement le cas en géométrie.

Un modèle de base en probabilités: l'urne de Bernoulli

Les situations aléatoires les plus élémentaires de la réalité consistent en la réalisation ou non d'un événement fortuit à l'issue d'une expérience aléatoire. Toutes ces situations, d'un point de vue probabiliste, peuvent être représentées par un modèle d'urne de Bernoulli. C'est une urne fictive, parfaite, en ce sens que les boules qu'elle contient sont conçues comme rigoureusement identiques, ne différant que par la couleur, blanche ou noire par exemple. « Tirer une boule de cette urne » est une expérience de pensée qui repose sur l' **hypothèse de modèle** que toutes les boules ont « la même chance » d'être tirées (équiprobabilité postulée). On peut réifier une urne de Bernoulli en présentant un bocal opaque garni de billes de deux couleurs. Si les conditions expérimentales sont soigneusement contrôlées, l'hypothèse de modèle semblera légitime. Dans le vocabulaire de la réalité, l'urne de Bernoulli est un objet idéalisé, muni de propriétés théoriques implicites. Je le qualifie de « modèle pseudo concret ».

Un seul paramètre caractérise une urne de Bernoulli: la proportion p des boules blanches parmi l'ensemble des boules. La définition la plus élémentaire de la probabilité d'un événement, introduite au niveau de ce modèle pseudo concret, (somme des probabilités des issues élémentaires qui constituent l'événement) justifie l'affirmation suivante: « la probabilité de tirer une boule blanche de l'urne de Bernoulli est égale à p ».

La formalisation de ce modèle (son expression dans le registre symbolique des mathématiques) consiste à considérer un ensemble fini Ω abstrait (si on veut, on peut considérer que ses éléments représentent les tirages éventuels des différentes boules de l'urne de Bernoulli), sur lequel on considère la distribution uniforme de la probabilité, notée P et appelée à ce niveau « probabilité sur Ω ». On distingue ensuite une partie A de Ω (représentant les boules blanches) et on traduit l'hypothèse de modèle par la donnée Card $(A) = p.$ Card (Ω). L'axiomatique du modèle probabiliste général conduit alors à $\mathrm{P}(A) = p$. On a ainsi constitué un « modèle probabiliste » de la situation aléatoire donnée.

Processus de modélisation

Dans un processus de modélisation, je distingue donc trois étapes, pour l'analyse didactique. Chacune de ces étapes relève de compétences différentes et donc de contrats didactiques différents.

Pour préciser cela, caricaturons un peu la pratique courante: le professeur présente une situation dans des termes naïfs du langage courant. Cette présentation se réduit souvent à «l'habillage» d'un problème mathématique qu'il désire proposer aux élèves. Ainsi la description d'une réalité complexe disparaît pour laisser la place à la proposition d'un modèle déjà là, exprimé en termes pseudo-concrets. Le professeur attend ensuite des élèves qu'ils traduisent ce modèle en termes mathématiques, il les aide au besoin pour interpréter la question posée en un problème interne que contractuellement les élèves doivent résoudre. Ils doivent enfin produire, si possible, un petit commentaire replaçant leurs résultats dans les termes pseudo-concrets du modèle. Il revient au professeur de donner les prolongements possibles et de porter sur la situation réelle évoquée les appréciations qui peuvent se dégager.

Si l'on veut introduire en mathématiques une véritable démarche expérimentale, il convient au contraire de ne pas négliger la *première étape* de la modélisation au **niveau de la situation concrète**: l'observation d'une situation réelle et sa description en termes courants.

Cette description est déjà une sorte d'abstraction et de simplification de la réalité, dans la mesure où certains choix sont faits, pour ne retenir que ce qui semble pertinent de cette situation vis-à-vis du problème étudié. Cette description est d'ailleurs pilotée par ce que j'appellerai un regard théorique, c'est-à-dire une connaissance de type scientifique s'appuyant sur des modèles généraux préconstruits, pour apprécier justement ce qui se révélera pertinent.

La démarche expérimentale consiste aussi à pouvoir agir sur la réalité, afin d'en étudier les évolutions et les invariants. Il faut donc pouvoir mettre en œuvre une expérimentation programmée par ce que je désigne par **«protocole expérimental»**, c'est-à-dire l'ensemble des instructions à suivre pour réaliser cette expérience et éventuellement la reproduire.

De nouvelles compétences sont alors attendues, qui peuvent être objet de formation: savoir décrire une situation porteuse d'un problème (par exemple, l'évolution des files d'attente devant les caisses d'un supermarché), savoir mettre en œuvre un protocole expérimental, et recueillir les effets obtenus, savoir organiser les données recueillies, savoir lire une statistique (par exemple, pointer les files à intervalles réguliers).

Puis il s'agit de traduire cette description en un système simplifié et structuré: c'est le niveau du **modèle pseudo-concret**. Cela se traduit par l'appel à un modèle général dont les conditions de transfert sont maîtrisées. En didactique, nous appelons cela «contextualisation» d'un savoir ancien.

Dans l'exemple précédent, il faut dégager les hypothèses pertinentes pour décrire les arrivées des clients, notamment le nombre moyen d'arrivées par unité de temps. Cette construction est guidée par un premier niveau de connaissances théoriques du phénomène étudié (processus de Poisson) et par les outils mathématiques disponibles. Elle conduit à poser des hypothèses de modèle (indépendance des arrivées...).

On peut alors passer à la *deuxième étape*: la **mathématisation** ou **formalisation** du modèle. Cela suppose que les élèves soient capables de représenter le modèle dans la symbolique propre aux mathématiques. Puis ils doivent savoir interpréter la question posée en un problème purement mathématique et savoir faire appel aux outils mathématiques adaptés pour résoudre le problème abstrait (fonction exponentielle, intégration, raisonnement par récurrence...).

Enfin, il convient, en *troisième étape*, de pouvoir revenir à la question posée pour traduire dans les termes du modèle pseudo-concret, les résultats mathématiques obtenus, leur donner du sens pour dégager des réponses et **relativiser ces réponses** par rapport aux hypothèses de modèle (formuler la loi de Poisson pour le nombre d'arrivées de clients dans un intervalle de temps donné) ; il faut ensuite interpréter ces réponses pour apprécier leur validité et leur étendue dans la situation concrète (décider de l'ouverture ou de la fermeture d'une caisse pour réguler les files d'attentes). Ces compétences peuvent faire l'objet de formation dans diverses disciplines. Elles prennent un aspect spécifique en mathématiques du fait du caractère particulièrement abstrait des outils que l'on désire mettre en œuvre.

Mise en œuvre d'une activité de simulation

Pour exemplifier les éléments théoriques précédents, le problème est illustré par la mise en œuvre d'une simulation pour déterminer une loi de probabilités comme modèle théorique représentatif d'une population statistique, que l'on peut approcher par l'observation expérimentale d'échantillons. Je propose de déterminer un modèle exponentiel (par l'obtention du meilleur paramètre possible) adéquat pour décrire une population de Poisson. On part de l'objectif du programme de terminal sciênce(TS) d'étudier la désintégration radioactive dans une masse de matière fissile.

L´activité consiste en la génération par simulation des effectifs d'atomes désintégrés par unités de temps (phénomène poissonien), le modèle pour cette simulation ayant été implanté au préalable dans les ordinateurs. Les données extraites des échantillons ainsi générés sont comparées aux probabilités théoriques issues d'une loi de Poisson. Le meilleur paramètre de Poisson ainsi obtenu, via la loi exponentielle correspondante, permet alors le calcul des probabilités des temps d'attente pour évaluer, par exemple, la demi-période de l'élément radioactif, ou mettre en œuvre des techniques de datation.

Prérequis

Cette activité suppose une certaine familiarité avec le vocabulaire de la statistique enseigné au lycée, les bases du calcul des probabilités (variable aléatoire et lois, espérance mathématique, densité, loi exponentielle vue en TS). Un petit document présente la loi de Poisson et sa relation avec la loi exponentielle. Côté technique, il suppose une certaine familiarité avec Excel et sa fonction ALEA.

Bibliographie

BATANERO, C. (2001). *Didactica de la Estadistica,* Departamento de Didactica de la Matematica, Universidad de Granada (consultable en ligne).

BOROVCNIK, M. et PEARD, R. (1996). "Probability". In: BISHOP, A. J. et alii. (eds.). *International Handbook of Mathematics Education.* Dordrecht, Kluwer Academic Publishers.

CHAPUT, B. (2003). *Probabilités au lycée*. Comission Inter-Irem Statistique e Probabilités. Paris, APMEP (n. 143).

DRESS, F. (1997). *Probabilités Statistique*. Paris, Dunod.

HARTHONG, J. (1996). *Probabilités & statistiques*, Paris, Diderot.

HENRY, M. (org.) (1992/1994). *CII Statistique et Probabilités*. Actes de l´Universités d'Été de Statistique Inférentielle, La Rochelle, 1-5 Septembre, Rouen, 29 Août-2 Septembre, Irem de Rouen Èditeur.

______(org.) (2001). *Autour de la modélisation en probabilités*. Besançon, Presses Universitaires Franc-Comtoises.

______(2003). Des lois continues en TS, pourquoi et pour quoi faire? *Repères-IREM* n. 51. Metz, Topiques Éditions.

KAPADIA, R. et BOROVCNIK, M. (1991). *Chance Encounters: Probability in Education*. Dordrecht, Kluwer Academic Publishers.

KAUFFMANN, P. (1994). *Information, Estimation, Tests*. Paris, Dunod.

RENYI, A. (1966). *Calcul des probabilités.* Paris, Dunod (rééd. Jacques Gabay, Paris, 1992).

SAPORTA, G. (1990). *Probabilités, Analyse des données et Statistique*. Paris, Technip.

Recebido em mar./2005; aprovado em abr./2005.