Recibido el 23 de septiembre de 2021, aceptado el 7 de marzo de 2022

03

EL PAPEL DE LA SIMULACIÓN INFORMÁTICA EN LA ARTICULACIÓN DE LOS CAMPOS PROBABILÍSTICO Y ESTADÍSTICO

THE ROLE OF COMPUTER-AIDED SIMULATION IN ARTICULATING THE DOMAINS OF PROBABILY AND STATISTICS

LE RÔLE DE LA SIMULATION INFORMATIQUE DANS L'ARTICULATION DES DOMAINES PROBABILISTE ET STATISTIQUE

Assia Nechache[1]
Bernard Parzysz[2]

[1, 2]CY Cergy Paris Université - LDAR (Francia)
[1]Correo: assia.nechache@cyu.fr
[1]ORCID https://orcid.org/0000-0001-9650-1261

[2]Correo: parzysz.bernard@wanadoo.fr
[2]ORCID https://orcid.org/0000-0002-5340-4418

PädiUAQ

PädiUAQ | Vol. 5 Núm. 10 | julio-diciembre | Facultad de Ingeniería | Universidad Autónoma de Querétaro

RESUMEN

Este artículo se centra en el uso y el papel de la simulación informática de experimentos aleatorios en la enseñanza secundaria, tomando el caso de la Francia como ejemplo. De hecho, uno de los aspectos esenciales de la enseñanza de la probabilidad en los centros de secundaria es el lugar que se otorga a la modelización. Vincular un experimento aleatorio real con la teoría de la probabilidad lleva a cuestionar la modelización en términos de proceso. Además, el uso de herramientas tecnológicas (ordenador, calculadora), cada vez más presentes en la enseñanza de las matemáticas, lleva a plantear la cuestión de la modelización de la realidad mediante la simulación de experimentos aleatorios. La transición de los experimentos aleatorios a su simulación es una cuestión fundamental en la enseñanza de la estadística, ya que permite a los estudiantes acceder al "pensamiento estadístico".
Los retos de la programación de simulaciones en la enseñanza de la probabilidad están de primera importancia. La aplicación de simulación informática recurre a dos campos diferentes: probabilidad y tambien estadística, que intervienen a su vez en el proceso de resolución de problemas. Es probable que esta ambigüedad, inherente al enfoque, cree confusión en los alumnos entre lo que es observacional y lo que es teórico. Además, el uso de tecnología necesita el uso de la programación algorítmica, cuyo estatuto arriesga en consecuencia de ser tambien une fuente de ambigüedad. Pero una vez explicada con claridad puede utilizarse para introducir conceptos probabilísticos, basándose en su analogía con conceptos estadísticos ya conocidos.

Palabras clave: Simulación; Informática; Probabilidad; Estadística; Espacio de trabajo matemático.

ABSTRACT

This article focuses on the use and role of computer simulation of random experiments in secondary education, taking the case of France as an example. Indeed, one of the essential aspects of the teaching of probability in secondary schools is the place given to modelling. Linking a real random experiment with probability theory leads to questioning the modelling in terms of process. Moreover, the use of technological tools (computer, calculator), which are increasingly present in mathematics teaching, leads to the question of modelling reality via the simulation of random experiments. The transition from random experiments to their simulation is a fundamental issue in the teaching of statistics, as it allows students to access "statistical thinking". Thus the stakes of programming simulations in the teaching of probability are most important. The iImplementaztion of a computrer-aided simulation requires making use of two domains: probability, of course, but also statistics, which intervene in turn in the problem-solving process. This co-existence of two domains conceals an ambiguity, inherent in the approach, is likely to create confusion in students between what is observational and what is theoretical. Moreover, the use of technology requires also the use of algorithmics-programming, the status of which is at risk to be a source of confusion as well.

Keywords: Simulation; Computer science; Probability; Statistics; Mathematical Work Space.

RÉSUMÉ

Le propos de cet article porte sur l'usage et le rôle de la simulation informatique d'expériences aléatoires dans l'enseignement secondaire, en prenant comme exemple le cas de la France. En effet, l'un des aspects essentiels de l'enseignement des probabilités dans le secondaire est la place donnée à la modélisation. La mise en lien d'une expérience aléatoire réelle avec la théorie des probabilités conduit à interroger la modélisation en termes de processus. En outre, le recours aux outils technologiques (ordinateur, calculatrice), de plus en plus présents dans l'enseignement des mathématiques, conduit à aborder la question de la modélisation du réel *via* la simulation d'expériences aléatoires. Or le passage de l'expérience aléatoire à sa simulation est une question fondamentale de l'enseignement de la statistique, car il est une voie d'accès des élèves à la «pensée statistique». Les enjeux de la programmation de simulations dans l'enseignement des probabilités sont donc de première importance. La mise en oeuvre de la simulation informatique fait appel aux deux domaines: probabilités, bien sûr, mais aussi statistique, qui interviennent tour à tour dans la démarche de résolution de problème. Cette coexistence des deux domaines recèle une ambiguïté fondamentale. Celle-ci est inhérente à la démarche, est susceptible de créer chez les élèves une confusion entre ce qui relève de l'observation et ce qui relève de la théorie. En outre, le recours à la technologie requiert également, l'usage de l'algorithmique-programmation, dont le statut risque aussi en conséquence d'être source de confusion.

Mots-clés: Simulation; Informatique; Probabilités; Statistique; Espace de Travail Mathématique.

INTRODUCTION

Cet article s'intéresse à la simulation informatique dans l'enseignement des probabilités de façon générale, en s'appuyant en particulier sur le cas de la France. Le type de tâches lié à la simulation met en jeu la co-occurrence et l'articulation de trois domaines distincts: les probabilités, la statistique et l'informatique[1]. Cependant, si la zone d'intervention de la programmation y est clairement visible en raison de sa nature même, il s'avère que l'articulation entre les deux autres domaines mathématiques reste le plus souvent totalement implicite. Ce sont plus précisément les enjeux didactiques de la simulation informatique de situations aléatoires par les élèves que l'on se propose d'étudier ici, en en mettant en évidence à la fois l'intérêt et les difficultés.

Du point de vue des programmes scolaires en France, avec le développement du calcul informatique, il y a une forte insistance sur les traitements des situations aléatoires faisant appel aux modèles numériques avec l'usage des simulations informatiques d'expériences aléatoires; en particulier lorsque les modèles probabilistes théoriques (lois de probabilités) ne sont pas encore disponibles pour les élèves. L'utilisation d'un ordinateur ou d'une calculatrice permet d'expérimenter une expérience aléatoire par simulation, en la répétant un grand nombre de fois, de façon beaucoup plus aisée et rapide que par une expérimentation réelle. La simulation est donc un outil intéressant pour établir des conjectures, résoudre des problèmes trop difficiles ou impossibles à traiter directement dans un modèle probabiliste de type analytique. La justification des résultats (de l'expérience simulée) obtenus est alors fondée sur la loi des grands nombres qui permet d'affirmer qu'en réalisant une expérience aléatoire un grand nombre de fois, de manière indépendante, les fréquences observées — qui relèvent donc du domaine de la statistique — se rapprochent des probabilités à évaluer. Par conséquent, deux domaines mathématiques, la statistique et les probabilités, sont en constantes et étroites interactions. Ces interactions sont traduites par des changement de domaine (Montoya et Viver, 2014) et par conséquent des changements d'espace de travail mathématique entre celui des probabilités et de la statistique. Mais ces interactions ne sont pas clairement explicitées dans l'enseignement. Cette situation implique alors une ambiguïté fondamentale causée par la coexistence de deux domaines différents. Elle est susceptible de créer chez les élèves une confusion entre ce qui relève de l'observation et ce qui relève de la théorie. Cependant, nous pensons qu'une fois clairement explicitées elles peuvent être utilisées pour l'introduction de certains concepts probabilistes, en s'appuyant justement sur leur analogie avec des concepts statistiques déjà connus.

1 Sous ce vocable nous incluons à la fois l'algorithmique – ensemble de connaissances permettant la transformation d'une tâche mathématique en une suite ordonnée d'opérations élémentaires, en vue de sa réalisation par une machine (ordinateur, calculatrice) – et la programmation, permettant de coder cette suite d'opérations en une liste d'instructions rédigée dans un langage compréhensible par une machine donnée.

Cet article, qui se réfère au cadre théorique des Espaces de Travail Mathématique (Kuzniak et al., 2016) et à la notion de paradigmes probabilistes P1 et P2 (Nechache & Parzysz, 2019) étudie le rôle de la simulation informatique des expériences aléatoires dans l'enseignement secondaire, et plus précisément son rôle dans l'articulation des deux domaines mathématiques que constituent la statistique et les probabilités et dans la connexion des deux ETM associés respectivement à ces deux domaines. Pour ce faire, nous commençons par préciser la nature épistémologique de la simulation et les outils informatiques utilisés dans l'enseignement pour explorer des expériences aléatoires. Cela permet de préciser la manière dont les deux domaines, probabilités et statistique, sont articulés. Ensuite, nous expliquons les enjeux didactiques de la simulation dans l'enseignement. Enfin, à partir de l'étude de quelques manuels scolaires nous expliciterons le travail mathématique de simulation attendu des élèves dans cet espace idoine potentiel[2].

NATURE ÉPISTÉMOLOGIQUE DE LA SIMULATION D'EXPÉRIENCE ALÉATOIRE

La simulation est une manière de représenter une situation de façon analogique en utilisant des outils mathématiques. Une définition plus précise de la simulation serait alors:

La méthode statistique permettant la reconstitution fictive de l'évolution d'un phénomène. C'est une expérimentation qui suppose la construction d'un modèle théorique présentant une similitude de propriétés ou de relations avec le phénomène faisant l'objet de l'étude. (Dodge, 1993)

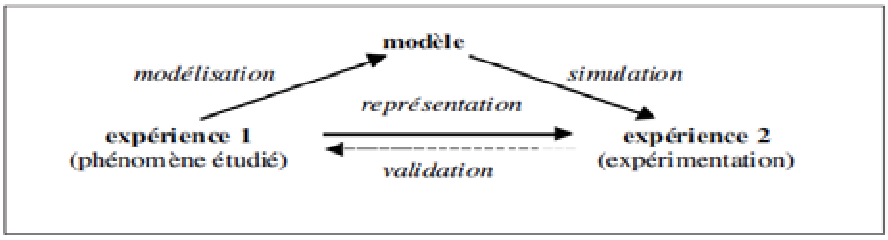

Figure 1. Schéma ternaire de la simulation (Parzysz, 2009).

Il est donc nécessaire d'élaborer, préalablement à la simulation, un modèle théorique de l'expérience étudiée. C'est en effet ce modèle qui va être simulé, et non pas l'expérience elle-même. Il en résulte qu'on se trouve ici en présence d'un schéma ternaire, et non binaire (Figure 1).

Dans ce schéma, l'expérience 1 est celle qu'il s'agit de reproduire sous une autre forme, celle de l'expérience 2. Pour ce faire on commence par définir un modèle mathématique de l'expérience 1, puis on recherche une autre expérience répondant au même modèle. Dans le cas qui nous occupe ici, les expériences en jeu sont des expériences aléatoires et le modèle de référence est pris dans la théorie des probabilités.

Par conséquent, avant de simuler une expérience aléatoire, il est nécessaire de concevoir – ou de choisir parmi plusieurs disponibles – un modèle théorique de cette expérience. Comme nous l'avons indiqué plus haut, la simulation consiste à remplacer l'expérience aléatoire initiale par une autre expérience qu'on considère comme « équivalente », ce qualificatif signifiant ici que la seconde expérience relève du même modèle probabiliste. C'est donc la pertinence du modèle probabiliste déterminant la simulation qui assure l'adéquation entre les deux expériences. En d'autres termes et *par définition*, c'est le fait que l'expérience aléatoire réelle (expérience 1 de la figure 1) et l'expérience aléatoire simulée (expérience 2 de la figure 1) ressortissent au même modèle, qui assure leur équivalence.

Pour prendre un exemple très simple, supposons qu'on veuille jouer à pile ou face. Pour être équitable, le jeu suppose que chacun des deux côtés de la pièce a une chance sur deux d'être obtenu. Pour s'en assurer, on peut procéder à une série nombreuse de lancers de la pièce, selon un même protocole expérimental, et si le nombre d'apparitions de chaque côté est sensiblement le même on dira que la pièce est bien équilibrée et on acceptera le modèle. Cette phase de test du modèle, toute qualitative qu'elle soit, est importante avec les débutants en probabilités, car non seulement elle valide le modèle, mais elle met en œuvre la loi des grands nombres en faisant apparaître la convergence de la fréquence observée vers la probabilité supposée, constituant ainsi une première confrontation entre réalité et modèle, c'est-à-dire entre le domaine de la statistique et celui des probabilités. On aborde de ce fait, même si c'est de façon tout à fait marginale, la statistique intérentielle.

2 ETM idoine potentiel correspond à ce que les auteurs ou un enseignant prévoit a priori, qui montre un travail mathématique attendu a priori

Supposons maintenant qu'on veuille jouer à pile ou face mais qu'on soit dépourvu de pièce de monnaie. Si on dispose d'un dé ordinaire à six faces, on peut l'utiliser pour *simuler* le jeu, en décidant par exemple qu'obtenir un point pair correspondra à pile et un point impair à face. C'est en effet le même modèle probabiliste, celui de l'équiprobabilité de deux éventualités, qui régit les deux expériences. Le modèle associé au dé – supposé non pipé – est celui de l'équiprobabilité d'obtenir chacune des six faces. Ce modèle implique celui dans lequel on a la même probabilité d'obtenir un nombre pair ou un nombre impair, et finalement le même modèle correspond aux deux expériences aléatoires.

Si on dispose d'une calculatrice, on pourra utiliser la touche « random » pour simuler le jeu, en décidant par exemple qu'obtenir un nombre inférieur à 0,5 correspondra à pile et que sinon ce sera face. Dans ce cas, on pourra même aller plus vite et plus loin en programmant un nombre élevé de lancers de la pièce, et ainsi obtenir immédiatement ce qu'on n'aurait obtenu qu'au bout d'un temps beaucoup plus long à l'aide d'une pièce ou d'un dé. La simulation informatique apparaît ainsi comme un *substitut* de l'expérience aléatoire étudiée.

Comme nous allons le voir, la simulation informatique *d'expériences aléatoires* peut jouer un rôle important dans l'acquisition de la notion de modèle probabiliste. En effet, la comparaison des procédures de simulation associées à diverses expériences aléatoires, ainsi que celle des tableaux qu'elles produisent dans un tableur, peuvent faire apparaître des similitudes conduisant à considérer comme légitime la substitution d'une expérience à une autre, et à dégager l'idée d'un schéma d'expérience commun, sur lequel pourra ensuite s'élaborer la notion de modèle (Parzysz, 2011).

OUTILS INFORMATIQUES POUR EXPLORER DES EXPÉRIENCES ALÉATOIRES

En France, dès le début de l'enseignement secondaire (dès l'âge de 11 ans), les élèves sont amenés à utiliser des outils informatiques tels que la calculatrice et l'ordinateur. Ces outils, et plus particulièrement le tableur-grapheur, sont mobilisés dans un premier temps dans l'Espace de Travail Mathématique de la Statistique (ETMS) visant à traiter des données statistiques résultant d'enquêtes réelles ou fictives: classement des données, regroupement en classes, calcul d'effectifs et de fréquences, calcul de paramètres : moyenne, médiane, quartiles, écart-type, etc.), représentations graphiques (representamen) de types variés (diagramme en bâtons, diagramme circulaire, etc.). Outre les éléments de logique élémentaire nécessaires pour la programmation, cet espace de travail se limite à la statistique descriptive: les données sont des faits certains, observés, et il en résulte que les paramètres associés ne sont sujets à aucune fluctuation.

Par la suite, ordinateur ou calculatrice (artefacts technologiques) sont utilisés pour produire un échantillon résultant de la simulation (dite informatique) de répétitions d'une même expérience aléatoire. On conçoit aisément l'intérêt premier, pour l'enseignant, du recours à cet outil, qui permet d'obtenir rapidement à l'aide d'artefacts technologiques, et à peu de frais un très grand nombre de données qu'on pourra ensuite utiliser pour étudier des situations aléatoires. Cependant, ces données ne sont pas obtenues de façon "naturelle", c'est-à-dire comme résultats d'une expérimentation physique; elles sont générées artificiellement par un artefact technologique. Il importe donc que l'enseignement des probabilités prenne en compte les rapports entre l'expérience aléatoire réelle et l'expérience simulée, afin que le recours à la technologie apparaisse légitime, et non comme un tour de passe-passe. Cette situation nécessite alors de concevoir un Espace de Travail Mathématique probabiliste (ETMP) idoine explicitant l'intérêt de l'usage d'un artefact technologique pour faciliter le passage entre les deux expériences aléatoires.

Commençons par détailler la démarche de programmation intervenant dans le processus de simulation. Il s'agit d'utiliser le générateur pseudo-aléatoire de l'artefact technologique à l'intérieur d'un programme ayant pour objet de fournir une suite de résultats dont on pense qu'elle aurait pu être obtenue avec l'expérience réelle. Le programme doit donc comprendre une liste d'issues, une instruction indiquant la probabilité associée à chacune d'elles, ainsi qu'une instruction d'itération de la procédure. Cette opération nécessite donc la donnée d'un *modèle probabiliste* de l'expérience simulée, et l'artefact technologique ne fera qu'appliquer ce modèle. L'établissement de celui-ci constitue par conséquent un élément crucial de la démarche, et cette première partie du processus se situe tout entière dans L'ETMP.

Comme avec l'expérience réelle, la réponse donnée par l'artefact technologique est une suite d'issues, autrement dit un échantillon, qu'il s'agira ensuite d'étudier en vue de la résolution du problème posé. On est donc à présent passé dans l'ETMS. Mais, dans la pratique, le passage entre l'ETMP et l'ETMS est-il si apparent? De fait, en observant une feuille de données statistiques, réalisée par exemple sous tableur, on ne peut pas savoir si ces données ont été introduites manuellement, à la suite d'une expérimentation réelle (cf. Figure 2 à gauche), ou si elles ont été générées par l'artefact technologique lui-même (cf. Figure 2 à droite). En effet, les deux procédures, quoique distinctes, sont visualisées de façon identique, puisqu'elles utilisent les mêmes represen-tamens (le tableau, les données affichées). La seule différence entre ces deux procédures réside dans la façon dont les données ont été obtenues. En outre, le prestige accordé à l'artefact technologique par les élèves tend à renforcer l'aspect "absolu" du résultat.

Pour contrecarrer cette tendance on peut par exemple:

- soit provoquer un recalcul de la feuille.
- soit aller rechercher la source d'une donnée dans sa cellule.

Cette différence, si elle n'est pas clairement explicitée aux élèves, peut représenter un obstacle pour l'interprétation des données. En effet, le nombre "5" (cf. Figure 2 à gauche), obtenu par une expérimentation réelle d'une expérience aléatoire, n'a pas la même signification que le nombre "5" (Figure 2 à droite) obtenu par la simulation de cette même expérience. Un moyen de rendre explicite le fait qu'on se trouve maintenant dans l'ETMS et non plus dans l'ETMP consiste à faire produire plusieurs échantillons, et non un seul. En outre, le travail au sein d'une classe permet de faire comparer les échantillons obtenus par les élèves ou les groupes d'élèves. Ainsi, le fait qu'ils soient différents "relativisera" le résultat obtenu. Cette différence portant sur la signification des données ne se limite donc pas au seul domaine de la statistique, mais necessite *de facto* celui des probabilités. Par ailleurs, la recherche de stratégies de simulation pour résoudre des problèmes de probabilités (et de statistique) constitue une véritable alternative à la résolution "classique" et peut offrir une ouverture pour des élèves peu à l'aise avec la formalisation mathématique. On voit ainsi que le recours à la simulation informatique permet une présentation dynamique du fonctionnement et de l'interaction des deux espaces de travail ETMS et ETMP.

Le recours à la simulation implique pour les élèves des difficultés que l'on doit prendre en compte. L'une d'entre elles – voir ci-dessus – porte sur le fait de distinguer clairement ce qui est observé (qui est sujet à variations) et ce qui appartient à la théorie (qui est fixe). Avec une "vraie" simulation (*i.e.* manuelle), la distinction est plus évidente car on travaille avec du concret (lancer de pièces tirage de boules…). De plus, par la suite le traitement des données issues de la simulation mettra en jeu des paramètres statistiques qui sont formellement très voisins de paramètres probabilistes (fréquence ? probabilité, moyenne ? espérance, variance ? variance, etc.).

Un lien raisonné entre les deux notions de *fréquence* et de *probabilité* s'établira ainsi selon deux modalités possibles: l'une en constatant "matériellement" le phénomène de stabilisation de fréquences; l'autre en utilisant un tableur pour simuler l'expérience aléatoire. Ces modalités relèvent toutes deux d'une approche fréquentiste, et elles visent à établir

EXPÉRIENCE RÉELLE

N°	Valeur
1	5
2	1
3	1
4	6
5	4
6	6
7	5
8	3
9	4
10	2

5
Valeur numérique observée

SIMULATION

N°	Valeur
1	5
2	1
3	1
4	6
5	4
6	6
7	5
8	3
9	4
10	2

=ALEA.ENTRE.BORNES(1;6)
Probabilité fournie par le générateur

Figure 2. Série de 10 lancers d'un dé ordinaire non pipé.

un lien entre les deux paradigmes SD et P1 (Nechache & Parzysz, 2019) au moyen de la gestion du saut conceptuel entre l'observé et le théorique. Mais elles sont cependant fort différentes, car dans l'étude de la répétition de l'expérience réelle la probabilité est le point d'aboutissement de la démarche, tandis que dans la simulation informatique elle se situe au contraire au départ, sous la forme du modèle mis en oeuvre. Pour cette raison, il nous semble préférable que la simulation informatique n'intervienne que dans un second temps, lorsque la notion de *modèle* d'expérience aléatoire a préalablement été dégagée au moyen d'expérimentations *réelles* (ibid.). La simulation informatique pourra ensuite permettre, par comparaison avec les résultats de l'expérience originale, de tester la validité du modèle introduit dans l'artefact technologique.

Par ailleurs, lorsqu'on réalise une simulation informatique, le modèle probabiliste implémenté peut fort bien ne pas être perçu comme tel, particulièrement dans les débuts de l'enseignement des probabilités, lorsqu'on travaille en ne disposant que du paradigme P1 dans lequel l'expérience est définie par un protocole expérimental, une liste d'issues possibles et une valeur de probabilité affectée à chacune d'entre elles (Henry, 1999, Parzysz, 2009). Par exemple, pour le lancer d'un dé ordinaire à six faces, non pipé, le point marqué s'obtiendra à l'aide d'une fonction pseudo-aléatoire donnant la même chance d'apparition à toute valeur entière comprise entre 1 et 6. Cette fonction est censée correspondre à une distribution de probabilité équiprobable, mais formellement sa programmation est quasi identique à celle d'une fonction mathématique déterministe, à ceci près que la valeur prise ne dépend pas d'une autre valeur du tableau mais du générateur pseudo-aléatoire (Figure. 3)

Même si, dans son fonctionnement réel, l'ordinateur se limite à exhiber les effets sur les fréquences affichées du principe d'équirépartition des nombres pseudo-aléatoires qu'il génère, l'usage qui en est fait dans la classe en tant que (pseudo-)générateur de hasard permet de faire comprendre en acte les notions de fréquence, de fluctuation d'échantillonnage et de probabilité (Henry, 2011), à condition toutefois de prendre certaines précautions.

Ce qui précède montre qu'on ne saurait se satisfaire de la seule exploitation de la puissance et de la rapidité des moyens informatiques pour exhiber des nombres aléatoires et permettre de présenter aux élèves une grande variété d'expériences aléatoires, ce qui n'aurait qu'un intérêt limité. En tant qu'outils de simulation, l'intérêt didactique des outils informatiques – comme nous l'avons vu sur l'exemple étudié plus haut – tient plus fondamentalement à ce qu'ils contraignent à analyser la situation aléatoire en jeu, à émettre sur celle-ci des hypothèses de modèle, notamment sur la loi de probabilité adéquate pour représenter l'intervention du hasard dans l'expérience réelle (par exemple, pour la simulation d'un sondage sociologique à deux éventualités, sur le choix de la valeur de la probabilité *p* de la variable de Bernoulli à introduire dans le programme), et à traduire ces hypothèses en instructions informatiques pour que l'ordinateur permette de résoudre des problèmes difficilement accessibles par le calcul au niveau où se situent les élèves (Henry, 2011).

Comme on vient de le voir, la mise en oeuvre de la simulation informatique fait appel à deux domaines mathématiques: probabilités, bien sûr, mais aussi celui de la statistique, qui interviennent tour à tour dans la démarche de résolution de problème. Cette

FONCTION DÉTERMINISTE

x	f(x)
1	2
2	3
3	5
4	5
5	6
6	5
7	5
8	3
9	2
10	1

=1+ENT(5*ABS(SIN(A6*PI()/10)))

FONCTION ALÉATOIRE

N°	Valeur
1	2
2	3
3	5
4	5
5	6
6	5
7	5
8	3
9	2
10	1

=ALEA.ENTRE.BORNES(1;6)

Figure 3. Programmation d'une fonction déterministe: $x \rightarrow f(x) = \overline{1 + \left\| 5.\left|\sin(\frac{\pi x}{10})\right| \right\|}$ et d'une fonction aléatoire équiprobable sur [1;6].

coexistence des deux domaines recèle une ambiguïté fondamentale. Celle-ci est susceptible de créer chez les élèves une confusion entre ce qui relève de l'observation et ce qui relève de la théorie, mais une fois clairement explicitée elle peut être utilisée pour l'introduction de concepts probabilistes, en s'appuyant justement sur leur analogie avec des concepts statistiques déjà connus.

RÔLE DE LA SIMULATION D'EXPÉRIENCE ALÉATOIRE DANS L'ENSEIGNEMENT

Comme nous l'avons dit plus haut, la simulation d'expérience aléatoire fait appel à la notion de modèle probabiliste, et par conséquent à un processus de modélisation permettant d'y parvenir. Ceci suppose la compréhension de ce processus et aussi la disponibilité de connaissances théoriques relatives aux deux domaines mathématiques cités précédemment. Un problème se pose ici, car l'artefact technologique ne fait que produire un échantillon d'issues correspondant au modèle qu'on y a introduit au départ, et on ne peut donc tester que la qualité du générateur. Illustrons-le avec un exemple simple, observé dans des classes de lycée. Soit à simuler le lancer de deux dés pour étudier la distribution de la somme des points obtenus. Si l'élève pense que les 11 issues ont la même chance d'apparaître, c'est ce modèle qu'il entrera dans son programme informatique …et bien sûr la simulation produira un résultat compatible avec l'équiprobabilité des issues. Le choix du modèle est donc une étape essentielle de la démarche de simulation, qu'on ne saurait sous-estimer.

On pourrait alors penser que la simultation informatique ne présente que peu d'intérêt, en ce sens qu'elle se contente de produire les effets du modèle qu'on lui a fourni[3]. Cependant, selon Henry, elle presente un intérêt didactique majeur pour la compréhension de la notion de modèle probabiliste, et constitue également un outil de résolution de problèmes. Pour lui, en effet, la pratique de la simulation d'expérience aléatoire "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"* (Henry, 2011, p. 536). La pratique de la simulation informatique en classe rend possible un travail sur de grandes séries statistiques, donnant véritablement du sens aux résumés statistiques (paramètres de position et de dispersion, diagrammes et histogrammes) et montrant ainsi leur pertinence. Ce qui n'interdit pas de travailler *ensuite* sur de petits échantillons dans le but d'étudier et d'établir les propriétés des paramètres. Elle permet aussi de répéter une expérience aléatoire un nombre de fois suffisamment élevé pour induire une bonne compréhension de la loi des grands nombres et mettre en œuvre un processus de modélisation (plus ou moins explicite). Comme nous l'avons souligné dans la section précédente, cette pratique aide à une présentation dynamique de l'interaction entre les notions de fréquence et de probabilité. Un intérêt supplémentaire non négligeable est qu'elle offre l'occasion aux élèves de résoudre des problèmes *a priori* trop difficiles pour eux, ou impossibles à traiter directement "à la main". En outre, le recours à la technologie est l'occasion d'initier ou de poursuivre la familiarisation des élèves avec l'algorithmique et la programmation.

3 Elle ne fait que tester la qualité du générateur implanté.

LE TRAVAIL MATHÉMATIQUE ATTENDU EN LIEN AVEC LA SIMULATION INFORMATIQUE

Pour rendre compte du travail mathématique attendu lorsque les élèves sont confrontés à des tâches faisant intervenir la simulation informatique dans l'enseignement secondaire, on étudie dans cette partie les activités[4] proposées par cinq manuels français de la classe de seconde (grade 10), traitant le programme de 2019 actuellement en vigueur. Sous la rubrique "*Échantillonnage*", ce programme de cette classe stipule que "*l'objectif est de faire percevoir, sous une forme expérimentale, la loi des grands nombres, la fluctuation d'échantillonnage et le principe de l'estimation d'une probabilité par une fréquence observée sur un échantillon*" (Programme de la classe de seconde, 2019, p.14). Il s'agit ici, plus précisément, de simuler une expérience aléatoire en s'aidant de l'ordinateur (avec le logiciel Python ou le tableur).

1- Un modèle prototypique d'exercice

Une première constatation est l'existence d'un schéma emblématique d'exercice sur la simulation d'expérience aléatoire, qui est réalisé plus ou moins complètement selon les manuels et le type d'activité considéré (TD, exercices). Ce schéma est organisé selon cinq parties:

4 Par ce terme nous entendons ici les travaux dirigés, exercices et problèmes.

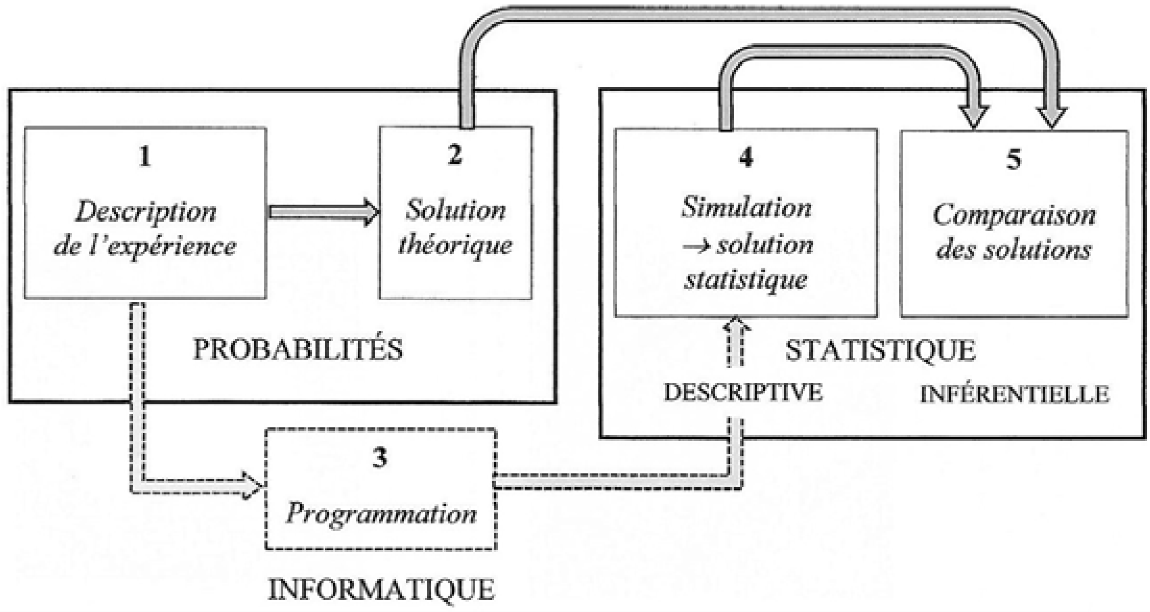

Figure 4. Changements de domaines dans le schéma prototypique d'exercice.

1. Description sommaire de l'expérience aléatoire (dans le paradigme P1) et question sur un événement associé.
2. Résolution théorique (dans le paradigme P2)
3. Programmation (avec le logiciel Python ou Excel)
4. Simulation de *n* épreuves et fréquence de l'événement considéré.
5. Comparaison des résultats du 2° et du 4°.

(N.B. Le qualificatif « sommaire » appliqué à la description de l'expérience aléatoire se justifie notamment par le fait que les conditions de l'expérience ne sont que rarement précisées.)

Les deux premières parties[5] correspondent à un exercice-type de probabilités. La troisième constitue souvent une aide à la prise en main du logiciel, car la mise en forme est détaillée et s'accompagne de questions intermédiaires destinées à guider et à contrôler le travail de l'élève. Une fois ce travail effectué, le programme est mis en œuvre pour réaliser un certain nombre de simulations de l'expérience aléatoire et observer la fréquence obtenue pour l'événement considéré. Enfin, une dernière question demande de comparer ce résultat avec la probabilité théorique calculée au 2°.

Du point de vue des domaines mathématiques abordés, les deux premières parties se situent clairement dans celui des probabilités, la troisième dans celui de l'algorithmique-programmation et la quatrième dans celui de la statistique descriptive. La dernière partie, quant à elle, esquisse un pas en direction de la statistique inférentielle (estimation). Elle doit normalement être l'occasion d'une prise de conscience et d'une reflexión sur la simulation, en montrant:

qu'une simulation produit une distribution voisine de la distribution théorique, quoique différente ;

que des simulations répétées fournissent des résultats différents, quoique voisins.

Ainsi, la mise en œuvre de ce schéma d'exercice suppose des changements de domaines mathématiques, partant d'abord de celui des probabilités pour se diriger vers la statistique inférentielle, en transitant au passage par la statistique descriptive (Figure 4).

En réalité, l'intérêt des auteurs se focalise au premier chef sur l'activité de programmation, car la maîtrise de ce type de programme permettra ensuite à l'élève de simuler d'autres expériences aléatoires. Ceci est clair dans certains intitulés d'activités: "*Comprendre un programme*" (Hyperbole, p. 316), "*Comprendre une fonction en langage Python simulant une expérience à deux issues*" (Transmath, p. 334). Qui plus est, la confrontation du résultat de la simulation avec le modèle théorique n'est même pas toujours demandée.

Ainsi, l'objectif principal de ces activités semble être d'apprendre aux élèves à programmer une expérience aléatoire (avec le tableur qu'ils connaissent déjà, mais aussi et surtout avec le logiciel Python qui est nouveau). On pourrait

5 La deuxième partie est parfois repoussée après la simulation.

presque dire que la simulation n'est là que pour s'assurer que le programme fonctionne convenablement. Par rapport à un programme classique, la seule différence est le fait que la fonction qu'on introduit fait intervenir le générateur pseudo-aléatoire du logiciel et ne se réfère pas à une autre colonne du tableau; il n'y a donc pas de raison justifiant une étude particulière de la construction d'un tel programme. C'est en fait au niveau de son utilisation, c'est-à-dire de la mise en œuvre de *n* itérations, que se situe la spécificité, et c'est plutôt sur ce point que devrait se porter l'attention, de façon à engager une réflexion. Mais on a souvent affaire à un *glissement* (au sens de Carranza) dans lequel le domaine des probabilités n'a comme fonction effective que d'introduire la situation conduisant à la rédaction du programme, sa mise en œuvre servant plutôt de moyen de contrôle de la correction de sa syntaxe que de résolution du problème posé.

2- Une variante

Dans un certain nombre de cas, la première étape du schéma précédent est précédée par une donnée statistique servant à introduire l'expérience aléatoire. Exemple (Sésamath, p. 324): " *D'après un rapport de l'ONU, 16,6% de la population mondiale vit en en Afrique*". On considère ensuite "le *tirage au sort d'un individu dans la population mondiale selon qu'il vit ou non en Afrique*" (ibid.)

Il s'opère ici un passage subreptice du domaine de la statistique descriptive (fréquence d'un caractère) à celui des probabilités (tirage au hasard). Qui plus est, on peut même se demander si certains auteurs ne jouent pas, de façon plus ou moins inconsciente, sur

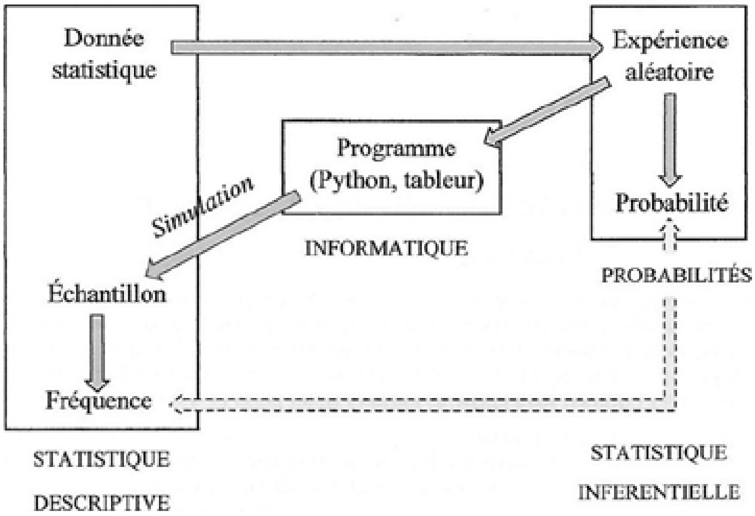

Figure 5. Schéma général des activités de simulation informatique.

l'ambiguïté du vocabulaire et de la notation. Ils notent en effet *p* la *proportion* des individus présentant un certain caractère dans la population considérée, proportion qui est numériquement égale à la *probabilité p* d'obtenir un individu possédant cette propriété dans un tirage au hasard (supposé équiprobable); c'est en particulier le cas pour l'exercice cité ci-dessus. Ainsi, la même lettre *p* désigne un nombre réel qui correspond à deux objets appartenant à deux domaines distincts: une donnée statistique dans l'un (une proportion, c'est-à-dire une fréquence), et une probabilité dans l'autre. Le programme de Seconde lui-même contribue à cette ambiguïté, en utilisant les trois termes dans une même phrase: "*Si p est la probabilité d'une issue et f sa fréquence observée dans un échantillon* [de taille *n*], *calculer la proportion des cas où l'écart entre p et f est inférieur ou égal à* $\frac{1}{\sqrt{n}}$" (Programme de la classe de seconde, 2019, p. 14).

Le passage par le domaine de l'informatique est bien visible, à la fois sous la forme du programme lui-même et par son utilisation pour la simulation. En revanche, le passage initial du domaine de la statistique à celui des probabilités, ainsi que le passage en sens inverse lors de la simulation, peuvent faire croire qu'on reste en permanence dans le domaine statistique. En effet, le détour par le domaine des probabilités n'apparaît que sous la forme de l'utilisation d'une fonction aléatoire dans le programme. Au final, pour cette variante du schéma l'ensemble de la démarche peut être schématisé sous la forme du diagramme suivant, étant entendu que certaines parties peuvent éventuellement en être omises (à l'exception de la simulation, bien sûr).

3- La congruence sémantique (Duval, 1995)

Le passage de l'expérience aléatoire au programme informatique nécessite le plus souvent une adaptation de celle-ci, qui constitue de fait un changement de modèle probabiliste. Par exemple (Hyperbole, p. 316):

> " *L'alphabet grec compte 24 lettres dont 5 voyelles. On choisit au hasard une lettre de cet alphabet et on note s'il s'agit d'une voyelle ou d'une consonne.*
>
> *a) Quel est le rôle de la fonction ci-contre écrite en langage Python?*

```
1. from random import *
2.
3.  def Alea():
4. n=randint(1,24)
5. if n<=5:
6.    A=1
7. else :
8.    A=0
9. return A (…)"
```

On s'aperçoit que les lettres de l'alphabet grec sont remplacées par les entiers de 1 à 24, les voyelles par les nombres de 1 à 5 et les consonnes par les nombres de 6 à 24. De façon générale, l'idée générale est de renommer les issues de l'expérience aléatoire de façon que leur distribution de probabilité soit identique à celle de l'expérience aléatoire étudiée. L'important est donc ici de construire une distribution rédigée sous une forme accessible à la machine – c'est-à-dire qu'elle sera capable de traiter –, qui soit *sémantiquement congruente* à la distribution originale. Ce type de transformation, rendu nécessaire pour pouvoir être compris par la machine, n'est pas forcément transparent pour l'élève, et mériterait lui aussi d'être interrogé. C'est bien sûr encore la notion de modèle probabiliste qui est sous-jacente à cette opération, et il est souhaitable que les élèves y aient été préparés plus tôt (voir plus haut).

4- La confrontation

Comme nous l'avons dit, parallèlement à la solution statistique fournie par l'ordinateur à la question posée, certains des exercices demandent une résolution probabiliste, soit au début, soit à la fin. C'est loin d'être toujours le cas, ce qui conforte l'idée que l'objectif de tels exercices n'est pas tant d'exploiter la situation probabiliste en confrontant la réponse statistique avec la réponse probabiliste, que de faire programmer une simulation de l'expérience aléatoire pour obtenir une estimation de la réponse théorique.

On peut cependant noter qu'il est parfois demandé, conformément au programme:

- de comparer le résultat statistique à la solution théorique : "*Comparer le résultat avec le résultat théorique*" (Transmath, p. 318), " *Comparer aux probabilités calculées dans le modèle* " (Métamaths, p. 297) ;
- d'estimer une probabilité à partir d'une simulation : « *quel modèle d'expérience aléatoire semble convenir ?* » (Hyperbole, ex. 53 p. 317)
- de produire plusieurs échantillons, afin d'observer la fluctuation d'échantillonnage : "*Si on exécute de nouveau la fonction N echantillons (N,n,p) on obtiendra d'autres échantillons et donc un résultat différent* " (Sésamath, p, 336), " *Appuyer sur la touche F9 (ou ctrl+maj+F9) pour réaliser plusieurs simulations et interpréter les résultats obtenus* " (Barbazo, p. 335) ;
- d'observer " *la conséquence de l'augmentation de la taille des échantillons* " (Barbazo, p. 344).

Certains auteurs indiquent parfois des mises en garde, comme: "*Il faut être conscient que le résultat d'une simulation dépend de l'échantillon. Par conséquent, si on change d'échantillon, on peut avoir un résultat bien différent*" (Transmath, p. 337) ou "*Il en faut pas se contenter de faire une unique simulation pour déterminer une probabilité*" (ibid.). Mais le fait que les deux résultats se situent dans des domaines différents n'est que rarement exposé clairement. Cependant, des procédés linguistiques tels que l'emploi du qualificatif "statistique" pour qualifier les échantillons pourraient sans doute contribuer à réduire la confusion.

5- Des formulations qui interrogent

Pour clore ce paragraphe, signalons que certaines formulations figurant dans les manuels posent problème, comme par exemple "*simuler un nombre aléatoire entre 0 et 1*" (Transmath, p. 331) ou "*on considère l'expérience aléatoire qui consiste à lire le nombre renvoyé par cette fonction*" (Hyperbole, p. 316), ou encore "*Vérifier que l'on obtient des nombres qui semblent pris au hasard entre 1 et 6*" (Métamaths, p. 296). En effet, pour ce qui est de la première phrase le nombre n'est pas simulé, puisqu'il apparaît à l'écran; en revanche, ce qui l'est, c'est le caractère aléatoire de ce nombre. Dans le même ordre d'idées, pour la deuxième formulation ce n'est pas la *lecture* du nombre qui est aléatoire, mais sa *génération*. La troisième phrase, quant à elle, émet un doute sur le caractère aléatoire de la génération des nombres fournis par l'ordinateur. Dans les trois cas, une interrogation apparaît en filigrane, due au fait que – contrairement aux élèves – les auteurs sont conscients que ce qui est présenté comme un générateur aléatoire de nombres réels est en réalité un générateur déterministe de nombres décimaux calibrés. Fournir, même très sommairement, des éclaircissements sur le mode de génération des nombres par les fonctions *alea* et *random*, puis permettre aux élèves de vérifier que "tout se passe comme si" on avait affaire à de l'aléatoire véritable aurait sans doute été plus efficace. Notons

PädiUAQ | Vol. 5 Núm. 10 | julio-dicembre | Facultad de Ingeniería | Universidad Autónoma de Querétaro

au passage que ces formulations particulières pourraient témoigner d'un certain malaise des auteurs vis-à-vis de la "cohabitation", dans un même problème, des deux domaines, probabiliste et statistique.

On peut finalement retenir de cette incursion dans les manuels, d'une part que les tâches de simulation informatique y sont assez stéréotypées, et d'autre part que leur objectif premier se réduit souvent à l'apprentissage de la programmation d'une expérience aléatoire. En particulier, les changements de domaines entre les probabilités et la statistique – qui interviennent dans les deux sens – ne sont pas spécialement travaillés, contribuant ainsi à installer une confusion préjudiciable à une bonne compréhension de la nature de l'aléatoire. En outre, les éléments théoriques justifiant l'intérêt du recours à la simulation pour l'étude d'expériences aléatoires ne sont pas souvent explicités. On peut résumer la situation en disant que le glissement opéré vers la programmation s'accompagne en contrepartie d'une perte de sens sur la simulation, et au-delà sur la théorie des probabilités elle-même.

CONCLUSION

Il est incontestable que la simulation informatique d'expériences aléatoires est aujourd'hui un élément incontournable de l'enseignement-apprentissage des probabilités, son principal intérêt étant de donner très rapidement accès à un grand nombre de répétitions d'une même expérience aléatoire (ou tout au moins supposées telles), présentant ainsi une aproche fréquentiste, dynamique, de la probabilité. Mais nous avons pu constater que la richesse du sujet était insuffsamment exploitée dans les manuels. On pourrait en effet constater, bien plus systématiquement que ne le font les manuels, la proximité des distributions statistiques obtenues par simulation avec la distribution probabiliste théorique. On pourrait observer son évolution en fonction de la taille des échantillons et, lorsque la distribution théorique n'est pas disponible, utiliser cette convergence pour modéliser l'expérience en définissant une distribution empirique sur cette base. Ces points ne sont pas totalement absents des manuels consultés mais ils sont, à notre avis, largement sous-exploités.

Par rapport aux activités classiques de résolution de problèmes liés aux expériences aléatoires, une caractéristique de la simulation est de ne pas se cantonner au seul domaine des probabilités mais de faire également intervenir, dans l'exploitation de la situation, non seulement l'informatique mais aussi la statistique. Comme nous l'avons vu, cette "cohabitation" des domaines probabiliste et statistique intervient tout au début de la démarche de programmation, lorsqu'il s'agit d'introduire un modèle probabiliste *a priori* dans l'artefact technologique; elle peut aussi intervenir lors de la recherche d'un modèle *a posteriori* issu de la simulation. Or, dans les manuels, si la seconde occurrence est explicite, la première est totalement transparente. Comme en outre un certain nombre de concepts de la statistique descriptive présentent de fortes similitudes avec des concepts probabilistes, il en résulte un risque réel de confusion entre les deux domaines, c'est-à-dire entre les éléments observés et les éléments théoriques. Une telle confusion peut conduire à des erreurs d'interprétation, et il importe qu'elle soit travaillée de façon spécifique. Notons toutefois qu'une fois les domaines bien distingués, on pourra au contraire utiliser leurs similitudes pour transposer certaines notions de l'un à l'autre, comme par exemple les paramètres de position et de dispersion, ou les représentations graphiques.

Dans le cadre de cet article nous avons étudié le travail mathématique attendu dans des ETM idoines potentiel conçus par des auteurs de manuels. Nous avons pu constater que ces ETM ne permettent pas de mettre en évidence les différents domaines mathématiques en jeu et les distinguer de façon claire. Nous avons vu aussi que le souci de transparence sur les domaines de référence était loin d'être mis en application dans les manuels. Qu'en est il alors dans les classes? Au regard de ces considérations, il apparaît nécessaire de prendre connaissance du travail mathématique réellement produit dans les ETM idoines conçus et mis en oeuvre par des enseignants.

Enfin nous restreignant au seul domaine des probabilités, nous avons également vu que la programmation d'une expérience aléatoire nécessitait souvent une adaptation de celle-ci pour pouvoir être "comprise" par l'artefact technologique, adaptation qui, elle non plus, n'est pas toujours transparente. Ce remplacement d'une expérience aléatoire par une autre qui lui est equivalente, qui repose sur la notion de modèle probabiliste relevant du paradigme P2, nécessite lui aussi un apprentissage, au centre duquel se trouve la notion de congruence sémantique.

En perspective, notre objectif sera donc de concevoir et de mettre en oeuvre des tâches faisant intervenir la simulation informatique d'expériences aléatoires. Ces tâches doivent être conçues de manière à articuler de façon claire et explicite les deux domaines mathématiques des probabilités et de la statistique avec celui de l'informatique, sans perdre de vue l'objectif principal qui est de donner du sens aux concepts probabilistes et statistiques dans l'apprentissage de ces deux domaines. La mise en oeuvre de ces tâches dans les classes au sein des ETM idoines devrait permettre aux élèves d'accéder à la pensé statistique et probabiliste.

RÉFÉRENCES

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

Duval, R. (1995). *Sémiosis et pensée humaine. Registres sémiotiques et apprentissages intellectuels*. Bern: Peter Lang.

Henry, M. (1999). L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie. *Repères*-IREM, 36, 15-34. http://www.univ-irem.fr/exemple/reperes/articles/36_article_245.pdf

Henry, M. (2011) Simulations d'expériences aléatoires en classe. Un enjeu didactique pour comprendre la notion de modèle probabiliste, un outil de résolution de problèmes. *Bulletin de l'*APMEP, 496, 536-550.

Kuzniak, A.; Tanguay, D.; & Elia, I. (2016). Mathematical working spaces in schooling: An introduction. ZDM-*Mathematics Education*, 48(6). 721-737.

Montoya-Delgadillo, E.; Vivier, L. (2014). Les changement de domaine dans le cadre des espaces de travail mathématique. *Annales de Didactique et de Sciences cognitives*, 19, 73-101.

Nechache, A.; & Parzysz, B. (2019) Le jeu des paradigmes dans l'ETM probabiliste. *Actes du 6e Symposium sur les Espaces de Travail Mathématique*, 179-191. Valparaiso: Pontificia Universidad Católica.

Parzysz, B. (2009). De l'expérience aléatoire au modèle, via la simulation. *Repères*-IREM, 74, 91-103.

Parzysz, B. (2011). Quelques questions didactiques de la statistique et des probabilités. Strasbourg, *Annales de Didactique et de Sciences cognitives*, 16, 127-147. http://mathinfo.unistra.fr/fileadmin/upload/IREM/Publications/Annales_didactique/vol_16/adsc16-2011_006.pdf

Programme de la classe de seconde, 2019. Bulletin officiel spécial n°1 du 22 janvier 2019.

https://www.education.gouv.fr/bo/19/Special1/MENE1901631A.htm

Manuels scolaires de la classe de seconde (programme 2019) consultés

Barbazo 2de. Baheux, M. *et al*. Ed. Hachette Education.

Hyperbole 2de. Bachimont, M. *et al*. Ed. Nathan.

Métamaths 2de. Alory, S. *et al*. Ed. Belin Eduction.

Sésamath 2de. Bau, D. *et al*.. Ed. Magnard.

Transmath 2de. Briset, A. *et al*. Ed. Nathan.