LA MODÉLISATION EN PROBABILITÉS : QUELS ENJEUX ET QUELLE PLACE DANS L'ENSEIGNEMENT SECONDAIRE EN FRANCE ?

Charlotte DEROUET[1]
LISEC UR 2310, Université de Strasbourg, UL, UHA

Résumé. Cet article questionne les enjeux et la place de la modélisation en probabilités, dans l'enseignement secondaire (collège et lycée) en France. En appui sur des travaux de recherche en didactique des mathématiques sur la modélisation notamment, nous présentons des éléments théoriques dont un cycle de modélisation dans lequel on distingue trois mondes : la réalité, le monde pseudo-concret et le monde mathématique (ici probabiliste). Sur l'exemple simple du lancer d'un dé, nous illustrons les différents concepts introduits. Nous menons ensuite une étude du point de vue institutionnel avec une analyse des programmes, des documents Ressources et des sujets de DNB et baccalauréat de l'année 2023 pour identifier les aspects de modélisation travaillés dans l'enseignement secondaire en probabilités. Nous montrons que, suivant les niveaux scolaires, certaines étapes du processus de modélisation sont privilégiées et d'autres peu travaillées, mais que l'ensemble du processus n'est jamais étudié dans les exemples analysés. Nous concluons en apportant quelques pistes pour l'enseignement.

Mots-clés. Probabilités, modélisation mathématique, enseignement secondaire.

Abstract. This article examines challenges and the place of probability modeling in secondary education (grade 7 to grade 12) in France. Based on research in didactics of mathematics on modelling in particular, we present theoretical elements including a modelling cycle in which three worlds are distinguished: reality, the pseudo-concrete world and the mathematical (probabilistic in this case) world. Using the basic example of throwing a dice, we illustrate the various concepts introduced. We then carry out a study from an institutional point of view, analysing the curricula, resource documents and exams of the DNB and baccalauréat in 2023 to identify the modelling aspects covered in secondary education in probability. We show that, depending on the school level, certain stages of the modelling process are given priority and others little attention, but that the whole process is never studied in the examples analysed. We conclude with some suggestions for teaching.

Keywords. Probability, mathematical modelling, secondary education.

Introduction

Les programmes de mathématiques, de l'école élémentaire au lycée, mettent en avant six compétences mathématiques : *Chercher*, *Modéliser*, *Représenter*, *Raisonner*, *Calculer*, Communiquer. Le préambule du document Ressource Éduscol du cycle 4 sur la compétence *Modéliser* débute ainsi :

> *La compétence « Modéliser », si on la prend dans son acception la plus large, renvoie pour le mathématicien au fait d'utiliser un ensemble de concepts, de méthodes, de théories mathématiques qui vont permettre de décrire, comprendre et prévoir l'évolution de phénomènes externes aux mathématiques* (MENESR, 2016b, p. 1).

Nous partageons cette vision de la modélisation. Nous y reviendrons plus loin. Dans les classes, s'il y a bien un domaine mathématique où des « phénomènes externes aux mathématiques » (pour reprendre l'expression de l'extrait ci-dessus), ou plus généralement le réel, sont présents, ce sont les probabilités. Si ce n'est tout à fait le réel, ce sont bien des contextes ancrés dans le réel. Nous allons donc, dans cet article, nous pencher sur la question de la modélisation en

[1] charlotte.derouet@unistra.fr

probabilités. Les probabilités fournissent des outils pour modéliser des systèmes réels afin de comprendre et faire des prédictions sur le réel à l'aide de ces modèles (par exemple Henry, 2001a ; Chaput *et al.*, 2011).

Dans le document Ressource cité précédemment, toute une section est consacrée aux probabilités, dont nous avons extrait ce passage :

> *« Aborder les questions relatives au hasard à l'aide de problèmes simples » est un des objectifs du programme de cycle 4. Pour autant, la simplicité des problèmes ne doit pas cacher la complexité du modèle proposé. Il convient de ne pas précipiter le passage au modèle probabiliste et aux calculs à l'intérieur de ce modèle, et il convient surtout de ne pas prendre les calculs menés à l'intérieur du modèle pour des preuves ou des démonstrations de phénomènes réels. On ne prouve pas que « le dé est truqué », ou qu'« il vaut mieux miser sur l'obtention d'un 7 comme somme de deux dés que sur l'obtention d'un 2 », mais on effectue des calculs à l'intérieur du modèle mathématique qui aboutissent à une conclusion mathématique rigoureuse. Les conclusions que l'on peut en tirer quant aux dés réels ou aux paris réels ne relèvent pas des mathématiques et doivent être considérées avec recul et prudence* (MENESR, 2016b, p. 3).

Cet extrait nous semble particulièrement riche et intéressant mais mérite, selon nous, d'être plus explicite et exemplifié pour une bonne compréhension de ce qu'il recouvre. Nous reviendrons en partie 2 sur les différents points mentionnés, après avoir présenté nos appuis théoriques.

Notre objectif, dans cet article, est de mettre en évidence la place et les enjeux de la modélisation en probabilités dans l'enseignement secondaire, sous un angle institutionnel. Cet article s'inscrit dans la continuité de la réflexion menée par la Commission inter-IREM Statistique et Probabilités à la fin des années 1990 (Henry, 2001a).

Dans un premier temps (partie 1), nous présenterons les éléments théoriques de didactique des mathématiques que nous mobilisons pour aborder la question de la modélisation en probabilités, en les illustrant sur l'exemple simple du lancer d'un dé. À la lumière de ces éléments théoriques, nous reviendrons sur l'extrait ci-dessus, présenterons notre méthodologie (partie 2) puis nous analyserons les programmes du cycle 4[2] (grades 7 à 9) à la terminale (grade 12), les documents Ressources ainsi que les sujets du diplôme national du brevet (DNB) et du baccalauréat 2023, du point de vue de la modélisation mathématique (partie 3). Nous pourrons enfin conclure sur la place et les enjeux de la modélisation en probabilités dans l'enseignement secondaire en France et dégager quelques pistes pour l'enseignement.

1. Modélisation en probabilités : quelques éléments théoriques de didactique des mathématiques

Nous allons illustrer les différents éléments théoriques présentés à l'aide d'un exemple simple, très rencontré dans les classes, celui du lancer d'un dé. Nous faisons ce choix pour être au plus proche de l'extrait du document Ressource qui insiste sur les « problèmes simples ».

1.1. Un exemple simple : le lancer d'un dé

La situation est la suivante : « Je lance mon dé préféré (un dé rouge légèrement abîmé à force d'être lancé !) à un instant t. Quel va être le résultat du dé ? » Le lancer de mon dé préféré, qui paraît être une situation simple amène en fait à un problème complexe qui dépend de manière très sensible des conditions initiales (aspect du dé ; rigidité de la surface sur laquelle le dé

[2] L'enseignement des probabilités débute au cycle 4.

atterrit ; hauteur, vitesse, direction du lancer ; moiteur des mains du lanceur…). La sensibilité du résultat à toutes les conditions initiales empêche de prévoir le résultat, quels que soient les instruments d'observation et la puissance des calculs (Henry, 2001c). De ce constat, soit on abandonne toute idée de pouvoir « dire quelque chose » sur l'issue de cette situation, soit on aborde cette question selon une autre approche, non plus déterministe mais aléatoire. La question serait alors : « Quels sont les résultats de mon dé préféré que l'on a le plus ou moins de chance d'obtenir ? ».

Konold *et al.* (2011), repris notamment par Thibault et Martin (2018), ont introduit le concept de *vraie probabilité*. Selon notre interprétation, dans le cas de mon dé préféré, la vraie probabilité d'obtenir 5 serait une *« valeur unique et spécifique […] mais qui demeure abstraite puisqu'elle ne peut être connue avec exactitude »* (Thibault & Martin, 2018, p. 14) que l'objet réel considéré, ici mon dé préféré à l'instant *t*, porterait en lui et qui représenterait la « chance » d'obtenir 5 en le lançant. Cette vraie probabilité[3] serait donc une valeur inconnue qui appartient au monde réel, intrinsèque à l'objet réel étudié et susceptible de changer dans le temps, notamment à cause des effets d'usure. Elle n'est pas une valeur définie par une approche probabiliste particulière (*cf.* partie 1.4), au contraire elle est la valeur abstraite vers laquelle les résultantes des différentes approches probabilistes, prises individuellement ou de façon combinée, devraient tendre à leur manière (Thibault & Martin, 2018). Nous y reviendrons.

Quoi qu'il en soit, une reconsidération de la question que l'on se pose, ainsi que des simplifications de la situation étudiée vont être nécessaires pour pouvoir espérer traiter le problème mathématiquement, car les vraies probabilités associées à mon dé préféré (si tant est qu'elles existent) ne peuvent être connues.

1.2. Modèle, modélisation

D'après Girard (2003), un modèle est une simplification de la réalité. Pour construire un modèle, on utilise certains paramètres de la situation, tout en négligeant certains aspects de l'expérience réelle (comme les frottements en physique par exemple). Dans la même lignée, Henry (2001b) définit un modèle comme *« une interprétation abstraite, simplifiée et idéalisée d'un objet du monde réel, ou d'un système de relations, ou d'un processus évolutif issus d'une description de la réalité »* (p. 151). Nous nous appuyons sur cette définition pour la suite.

Pour construire un modèle, des simplifications et des idéalisations vont donc être nécessaires. Dans le cas d'une approche aléatoire d'une situation, la situation réelle (expérience concrète) va être simplifiée en une expérience aléatoire (nous y reviendrons). Il est important d'avoir en tête qu'un modèle n'est pas la réalité. Israël (1996) précise que *« non seulement un seul modèle peut décrire différentes situations réelles, mais le même fragment de réalité peut être représenté à l'aide de modèles différents »* (p. 11). Suivant les idéalisations et les simplifications faites, nous pouvons associer à la situation des modèles différents (pour la situation de mon dé préféré aussi). En appui sur les travaux d'Israël, Yvain-Prébiski (2018) définit la modélisation comme une *« démarche de construction d'un modèle en langage mathématique permettant de mettre en relation les éléments choisis d'un fragment de réalité en lien avec la question à étudier »* (p. 65). La modélisation est donc un processus. Dans la définition de Yvain-Prébiski, l'accent est mis sur la construction du modèle, mais le processus de modélisation peut être pris de façon plus globale. De nombreux travaux en didactique des mathématiques cherchent à décrire le processus de

[3] La notion de *vraie probabilité* nous semble intéressante pour éclairer la suite de notre propos dans l'article, cependant elle est à prendre avec du recul car supposer qu'il existe une vraie probabilité est un vrai débat (philosophique) en soi.

modélisation et les représentations proposées sont essentiellement sous forme de cycle. Nous prenons comme point de départ le cycle de modélisation de Blum et Leiß (2007), présenté en figure 1.

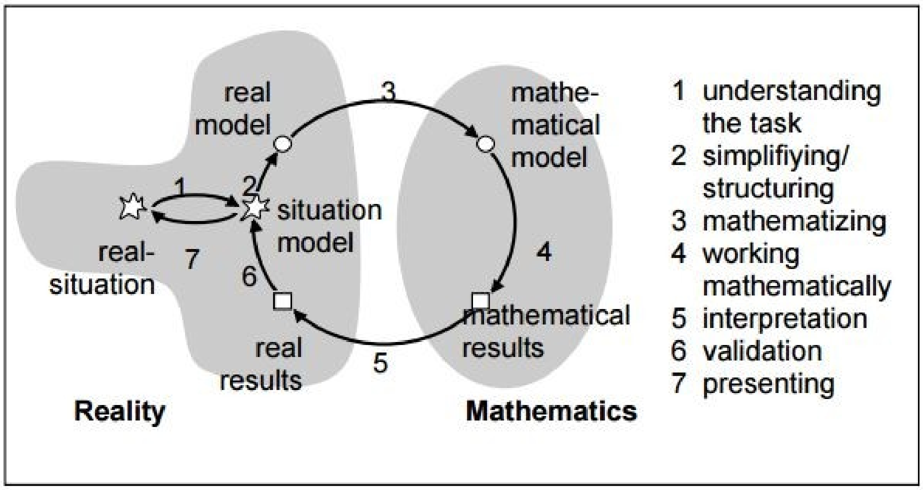


Figure 1 *: Cycle de modélisation proposé par Blum et Leiß (2007).*

Ce cycle de modélisation est très répandu dans les travaux en didactique des mathématiques portant sur l'enseignement et l'apprentissage de la modélisation mathématique. Partant d'une situation réelle (A) pour laquelle on s'intéresse à une certaine problématique, le solveur comprend la situation (étape 1 de la figure 1), en n'en retenant que certains aspects (de façon consciente ou non), et se construit un modèle de situation (B). Ce modèle de situation est ensuite simplifié et structuré (étape 2) pour obtenir un *real model* (Kaiser, 1995), traduit par modèle réel (C). En plus de la simplification et la structuration, il nous semble important d'ajouter l'idéalisation, aspect important dans la définition de modèle (Henry, 2001b) sur laquelle nous prenons appui dans cet article. Ce modèle idéalisé et simplifié associé à un problème « réel » est ensuite traduit mathématiquement (étape 3) pour arriver au modèle mathématique et au problème mathématique (D). Une fois dans le monde des mathématiques, à partir de ce modèle mathématique, un traitement mathématique (étape 4) est enfin possible et permet d'aboutir à des résultats mathématiques (E), qui seront ensuite interprétés (étape 5) en « résultats réels » (F). Ces résultats seront validés ou non (étape 6) par rapport au modèle de situation : soit les résultats sont acceptables pour décrire le problème réel, soit des modifications sont à apporter dans les simplifications et un nouveau modèle « réel » peut être reproposé… Puis, si les résultats obtenus semblent cohérents, ils seront présentés comme généralisations ou prédictions sur la situation réelle (étape 7). Ces généralisations ou prédictions sur la situation réelle (étape 7), dans des situations complexes spécifiques d'une discipline autre que mathématique (par exemple, la météorologie, la vulcanologie, la chimie…), ne relèvent plus tout à fait du mathématicien car elles nécessitent des connaissances spécialisées de la situation étudiée pour décider de son acceptabilité pour décrire le phénomène étudié (Henry, 2001d).

Le modèle de Blum et Leiß (2007) nous semble tout à fait pertinent pour analyser le processus de modélisation. Cependant, l'utilisation des termes que nous avons traduits par « modèle réel » et « résultats réels » semblent sous-entendre que nous parlons de la réalité, aspect renforcé par le fait que le cycle de modélisation laisse apparaître deux mondes, la réalité et le monde mathématique. Ces points nous paraissent porteurs d'ambiguïtés. Pour cette raison, nous avons adapté légèrement ce cycle de modélisation (Derouet, 2022) afin de lever ces ambiguïtés en nous appuyant sur les travaux de Coulange (1998) et de Henry (1999). Coulange (1998) propose, dans la démarche de modélisation, de séparer le monde de la réalité (extra-mathématique) du monde mathématique par un troisième monde, le monde « pseudo-concret » dans lequel :

> *On appelle modèle pseudo-concret un modèle intermédiaire (en langage naturel ou éventuellement sous forme d'un schéma) entre la situation réelle et le modèle mathématique à construire. C'est en quelque sorte un premier niveau d'abstraction de la « réalité » invoquée, qui n'est d'ailleurs pas fixe la plupart du temps : […] un modèle pseudo-concret peut être plus ou moins proche de la situation réelle considérée ou du modèle mathématique à construire* (Coulange, 1998, p. 36).

Elle emprunte cette idée de « pseudo-concret » à Henry (1999), qui définit un modèle pseudo-concret comme *« une sorte d'abstraction et de simplification de la réalité perçue dans sa complexité, 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 »* (p. 28).

De cette description de la démarche de modélisation, nous retenons le monde pseudo-concret et le modèle pseudo-concret, qui nous paraissent plus appropriés pour lever les ambiguïtés que nous avons identifiées dans le cycle de Blum et Leiß (2007). Il ne s'agit pas d'un changement profond vis-à-vis de ce cycle mais plus d'une adaptation permettant une explicitation. Cela nous amène à considérer le cycle de modélisation en figure 2.

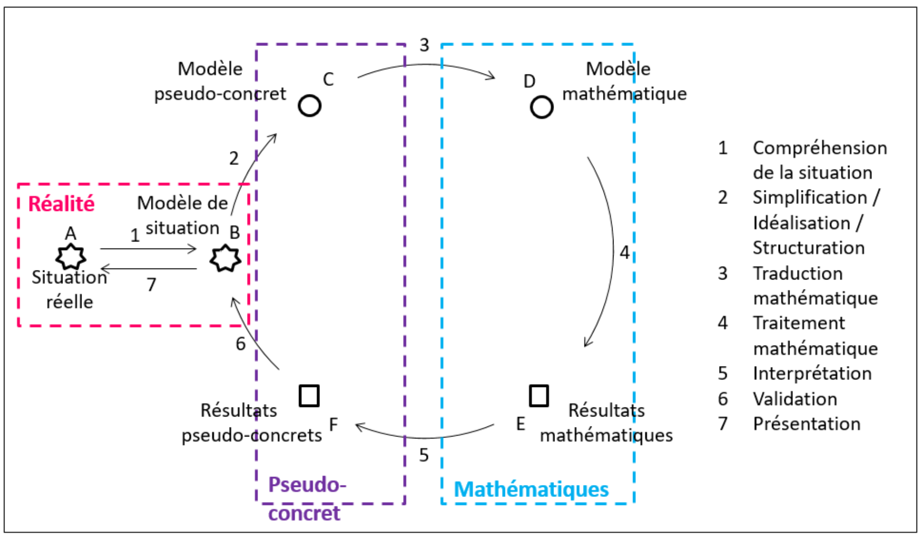


***Figure 2** : Cycle de modélisation retenu (Derouet, 2022).*

La modélisation est l'ensemble du processus. Ce processus n'est pas nécessairement linéaire, comme pourrait le laisser penser la figure, mais au contraire dynamique : il peut sauter des étapes, faire des allers-retours (Yvain-Prébiski, 2018). En revanche, ce cycle peut permettre d'analyser *a priori* une situation et peut aussi permettre *a posteriori* de débusquer des blocages ou des types de démarche d'élèves.

Dans la situation du lancer de mon dé préféré, le modèle de situation va déjà écarter plusieurs éléments de la situation réelle de départ. Tout d'abord, la couleur du dé peut être écartée, le moment du lancer aussi, ensuite le fait que ce soit moi ou quelqu'un d'autre qui lance le dé peut ne pas être un élément pris en considération… On pourrait donc garder pour le modèle de situation (étape B de la figure 2) seulement que quelqu'un lance un dé que l'on considère « classique », c'est-à-dire à 6 faces et que l'on pourrait trouver dans la boîte d'un jeu de société. La problématique initiale qui est de déterminer le résultat du dé va, elle aussi, changer et devenir

« quels sont les résultats du dé que l'on a le plus ou moins de chance d'obtenir ? » (et non plus ceux de mon dé préféré). Dans ce cas, la situation a été dépersonnalisée (il ne s'agit plus d'un dé en particulier, ni d'un lanceur en particulier…) du fait notamment que nous n'avons pas accès à toutes ces informations, et la question a été adaptée aussi du fait que la situation va être considérée comme relevant du « hasard ».

Un modèle pseudo-concret de la situation du lancer de mon dé préféré (étape C de la figure 2) serait donc un lancer de dé simplifié, idéalisé… Il s'agit alors de définir ce qu'est un lancer de dé simplifié/idéalisé. Plusieurs choix sont en fait possibles, et aboutiront à des modèles éventuellement différents.

1.3. Expérience aléatoire, hypothèses de travail et hypothèses de modèle

Expérience aléatoire

Dans le cadre d'une interprétation objective de la probabilité (Hacking & Dufour, 2004), la modélisation en probabilités repose avant tout sur une expérience aléatoire. Selon Henry (2001c), une expérience aléatoire est :

- une expérience reproductible,
- dont le résultat n'est pas connu d'avance et peut varier si l'on répète cette expérience « dans les mêmes conditions »,
- dont on peut dresser la liste des issues possibles.

Cette question de la reproductibilité est cruciale. L'expérience réelle ne peut en aucun cas être reproductible, elle est nécessairement unique, ne serait-ce que du point de vue temporel (Henry, 2001c). Il convient alors de préciser ce que l'on prend en compte pour vérifier cet aspect « reproductible » : *« De la complexité naturelle, l'expérimentateur ne retiendra pour la description de son expérience aléatoire, que ce qui lui semble pertinent pour indiquer comment la réaliser »* (Henry, 2001c, p. 163). Il est clairement fait mention ici d'une simplification du réel : on évince certains aspects pour n'en retenir que d'autres. Henry (1997) met bien en évidence la distinction entre réalité et modèle en lien avec l'expérience aléatoire : *« Pour approfondir [la notion d'expérience aléatoire] on est immanquablement amené à séparer la description de situations réelles, des modèles simplifiés qui permettent de les mathématiser. »* (p. 55)

Henry (2001c) associe à la notion d'expérience aléatoire la notion de protocole expérimental, c'est-à-dire *« le texte des instructions à respecter pour pouvoir affirmer que l'on a bien réalisé l'expérience aléatoire, objet de l'étude »* (p. 164). Cela permet de définir l'aspect reproductible de l'expérience. Il précise :

> *une expérience aléatoire est donc la mise en œuvre dans des conditions expérimentales déterminées par un protocole, d'un processus évolutif pour un système matériel dont le comportement est sensible par rapport aux conditions initiales, de telle sorte que l'on ne peut prévoir son état en fin de parcours* (*ibid.*, pp. 164-165).

Il distingue alors l'expérience *« concrète »*, qui relève de la réalité, de l'expérience aléatoire *« abstraite »*, qui sera donnée dans les mêmes termes naïfs de la réalité, mais qui cette fois revêt un sens plus précis en désignant des objets abstraits. Ce passage du modèle de situation au modèle pseudo-concret (étape 2 de la figure 2), par le biais de choix à opérer sur le réel, est guidé par un regard théorique, qui suppose des connaissances mathématiques préalables (Henry, 2001c). Ces choix opérés pour déboucher sur un modèle pseudo-concret correspondent à ce qu'Henry (2003) nomme « hypothèses de travail ».

Hypothèses de travail, hypothèses de modèle

Pour comprendre la démarche de modélisation qui permet d'analyser des situations réelles à l'aide de connaissances probabilistes théoriques, Henry (2003) distingue deux types d'hypothèses qui relèvent des choix à faire pour passer de la réalité au modèle mathématique : les *hypothèses de travail* et les *hypothèses de modèle*. Les hypothèses de travail, à relier au modèle pseudo-concret (étape C de la figure 2), permettent de définir l'expérience aléatoire (pseudo-concrète) en précisant notamment les résultats possibles et le protocole expérimental associé (Parzysz, 2009) dans un langage courant. Les hypothèses de modèle, à relier au modèle probabiliste théorique (modèle mathématique, étape D de la figure 2) permettent, quant à elles, de définir l'expérience aléatoire mathématique dans le langage probabiliste en termes d'issues, d'univers, de loi de probabilité, de variable aléatoire, de probabilité…[4] Certaines hypothèses sont simplificatrices, d'autres peuvent être des données de la situation.

Les hypothèses de travail et de modèle introduites dans le contexte des probabilités peuvent tout à fait être généralisées à d'autres domaines mathématiques. En probabilités, on peut noter que les hypothèses peuvent aussi être dépendantes de l'approche des probabilités considérées.

1.4. Deux approches pour construire le modèle probabiliste

Une fois l'expérience aléatoire décrite, il faut déterminer les probabilités élémentaires, c'est-à-dire les probabilités associées aux issues constituant l'univers, dont la somme est égale à 1. Plusieurs possibilités existent. Dans cet article, nous aborderons seulement deux d'entre elles qui relèvent de deux approches de la probabilité rencontrées au collège-lycée : l'approche laplacienne (ou classique) et l'approche fréquentiste[5]. Ces deux approches relèvent d'une interprétation objective des probabilités (Hacking & Dufour, 2004). Les probabilités, résultantes des différentes approches probabilistes, devraient tendre d'une certaine manière vers la vraie probabilité associée à chaque face de mon dé préféré (*cf.* partie 1.1).

Approche laplacienne

Le premier cas de figure pour déterminer les probabilités, cité par Henry (2001c), est le suivant :

> *Le contexte de l'expérience aléatoire, les symétries des objets matériels utilisés, permettent de ramener tous les résultats à un ensemble de n issues[6] qui sont considérées comme également possibles. On fait alors l'hypothèse d'équiprobabilité. On dit aussi que l'on admet la distribution uniforme des probabilités sur l'univers de ces n issues.*
>
> *Cela revient à donner aux probabilités de chaque issue la même valeur* $\frac{1}{n}$ (p. 168).

La question de la symétrie des objets matériels en jeu relève des hypothèses de travail (modèle pseudo-concret) et celles-ci sont ensuite associées à une hypothèse de modèle d'équiprobabilité (modèle mathématique). Cela relève de l'approche laplacienne (ou classique, ou théorique) des

[4] Henry (2003) illustre sur plusieurs exemples la description des hypothèses de travail et de modèle, notamment sur l'exemple simple du lancer de dé (pp. 6-7).

[5] Les travaux en didactique des mathématiques distinguent trois approches probabilistes : l'approche laplacienne, l'approche fréquentiste et l'approche subjective. Les deux premières approches relèvent d'une interprétation objective de la probabilité. La troisième approche, subjective, s'appuie sur des croyances personnelles de l'individu qui étudie la situation. Elle ne sera pas abordée ici car elle n'est pas (ou très peu) rencontrée dans l'enseignement secondaire en France. Se reporter à Thibault et Martin (2018) pour plus de détails sur ces approches.

[6] Nous avons corrigé une coquille présente dans le texte original.

probabilités, approche objective *a priori*. La probabilité d'un événement associé à diverses issues est, dans ce cas, définie comme le rapport du nombre de celles qui le produisent à leur nombre total (Parzysz, 2011). Le modèle est construit *a priori*, sans nécessité de recourir à l'expérience concrète. Elle repose sur une hypothèse simplificatrice de travail de symétrie des objets considérés.

Revenons à la situation du lancer de mon dé préféré. De l'expérience concrète du lancer de mon dé qui s'est transformée en expérience concrète du lancer d'un dé (modèle de situation, *cf.* partie 1.1), nous allons simplifier et idéaliser la situation en décrivant l'expérience aléatoire (protocole expérimental et issues) pour obtenir un modèle pseudo-concret, qui sera ensuite traduit en langage mathématique pour aboutir à un modèle mathématique (*cf.* tableau 1).

<table>
<tr><th>Modèle pseudo-concret
Hypothèses de travail (HT)</th><th>Modèle mathématique
Hypothèses de modèle</th></tr>
<tr><td>(HT1) Le dé a 6 faces (1, 2, 3, 4, 5 et 6).
(HT2) On lance le dé suffisamment fort pour que le résultat ne soit pas prévisible.
(HT3) Un résultat du dé sera 1, 2, 3, 4, 5 ou 6 suivant la face visible à l'issue du lancer.
(HT4) Si le dé n'atterrit pas à plat, le lancer est annulé.
(HT5) Le dé est correctement construit dans une matière suffisamment homogène pour bien accepter les symétries en jeu, plaçant les six faces dans des conditions équivalentes[7] : le dé est alors considéré comme un cube parfait.</td><td>On considère l'expérience aléatoire du lancer de dé, pour lequel on s'intéresse au caractère « résultat du dé ».
L'univers considéré est : $\Omega=\{1;...;6\}$ (6 éléments)[8].
Les 6 issues possibles de Ω sont supposées toutes équiprobables. La probabilité est distribuée sur Ω suivant la loi représentée par le tableau :
<table>
<tr><td>w_i</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>p_i</td><td>$\frac{1}{6}$</td><td>$\frac{1}{6}$</td><td>$\frac{1}{6}$</td><td>$\frac{1}{6}$</td><td>$\frac{1}{6}$</td><td>$\frac{1}{6}$</td></tr>
</table></td></tr>
</table>

Tableau 1 *: Premier modèle du lancer d'un dé.*

L'équiprobabilité des issues de l'expérience aléatoire est une hypothèse de modèle qui résulte de l'hypothèse de travail HT5, et plus particulièrement du fait que le dé est considéré comme un « cube parfait ». On voit bien ici apparaître une idéalisation forte. Le dé, qui au départ dans la situation réelle était mon dé préféré légèrement abîmé dans la réalité, puis devenu « un » dé, est considéré dans le monde pseudo-concret comme un cube parfait. D'ailleurs, au-delà du fait que le dé soit abimé ou non, les constellations d'un dé sont différentes sur chaque face et donc la quantité de matière sur chaque face est différente dans la réalité (et ceci a certainement une influence sur le résultat du dé, même minime). L'expérience aléatoire dans le monde pseudo-concret devient **le** lancer de dé régulier (ou équilibré), expérience aléatoire avec un dé, qui est maintenant un objet abstrait et idéalisé. La probabilité, par exemple, d'obtenir 3, dans ce modèle, est $\frac{1}{6}$. Cette probabilité, appelée parfois probabilité théorique, dans le sens où elle est obtenue avec une approche laplacienne[9], n'est pas à confondre avec la *vraie probabilité* (Thibault & Martin, 2018).

[7] Nous prenons appui sur les hypothèses de travail proposées par Henry (2003, p. 7).

[8] Au lieu de considérer $\Omega=\{1;...;6\}$, on pourrait considérer les six constellations du dé. En effet, ici nous avons déjà fait un « raccourci » en assimilant les constellations à l'écriture chiffrée associée.

[9] Qui peut parfois aussi être appelée approche théorique.

Approche fréquentiste

Une autre approche pour construire le modèle probabiliste est envisageable, il s'agit de l'approche fréquentiste : *« par l'approche fréquentielle*[10]*, c'est à travers l'augmentation de la taille de l'échantillon dans le cadre d'une expérimentation qu'on observe une fréquence relative pour approcher la vraie probabilité »* (Thibault & Martin, 2018, p. 14). Dans cette approche, la probabilité se rapporte à *« une probabilité objective, résultant de nombreuses observations de [l'] événement »* (Parzysz, 2011). Dans ce cas, à partir des fréquences observées, il existe en fait plusieurs possibilités pour choisir le modèle :

> • *Soit on décide de prendre comme probabilité de chaque issue la valeur numérique de la fréquence obtenue (comme proposé ci-dessus par Henry, 2001c) ; mais alors, si on effectue une nouvelle fois « un grand nombre de fois » l'expérience aléatoire, on obtiendra une autre valeur ;*
>
> • *Soit on décide de prendre une autre distribution « proche » de celle trouvée empiriquement, par exemple en prenant comme probabilité de chaque issue des rationnels simples ;*
>
> • *Soit, si les réalisations peuvent laisser penser à un modèle d'équiprobabilité, on teste l'hypothèse d'équiprobabilité (test statistique) et, si on ne peut la récuser, on l'accepte* a posteriori (Derouet, 2022, p. 98).

Le modèle est construit à partir d'« un grand nombre » de réalisations de l'expérience concrète, donc d'un échantillon de taille aussi grand que choisi par le solveur.

L'approche fréquentiste est souvent reliée au théorème mathématique appelé la loi forte des grands nombres. Cependant le lien n'est pas direct. Il nous semble important de clarifier cela. La loi forte des grands nombres est un théorème mathématique, qui de façon vulgarisée peut être énoncé comme ceci :

> *Si A est un événement de probabilité p, issue possible d'une expérience aléatoire E, et si l'on reproduit dans les mêmes conditions un grand nombre de fois cette expérience E, alors on peut considérer que la suite observée des fréquences F_n des réalisations de A au cours des n premières expériences tend vers p quand n tend vers l'infini, car il n'y a aucune chance d'observer le contraire* (Henry, 2009, p. 81).

Ce théorème n'a de validité que dans le monde mathématique, il s'applique donc aux modèles mathématiques. Or l'approche fréquentiste prend, elle, appui sur le monde réel, en expérimentant sur les objets réels un grand nombre de fois et en observant les fréquences d'apparition d'un événement A. Cependant, est-ce que, dans la réalité, la fréquence d'apparition d'un événement A va converger vers une valeur limite ? Cela repose sur un postulat d'existence (Ventsel, 1971, cité dans Henry, 2009) : on suppose que dans la réalité, comme en mathématiques, la fréquence d'un événement observée (dans la réalité) converge vers une valeur limite à l'infini, chose que l'on ne peut pas vérifier expérimentalement car le concept d'infini est théorique et non réalisable en pratique (Ross, 2007, cité dans Thibault & Martin, 2018). De façon empirique, on retrouve tout de même cette stabilisation des fréquences dans le cas d'expériences réelles « reproductibles », ce qui rend légitime cette approche. On pourrait parler de loi des grands nombres empirique, sorte de loi de la nature. Dans un premier temps, cette loi des grands nombres empirique justifie l'approche fréquentiste car il y a bien un phénomène naturel de stabilisation des fréquences. De plus, comme les probabilités associées aux issues de l'expérience aléatoire (monde des mathématiques) que l'on va choisir pour modéliser la situation réelle étudiée vérifieront, elles, la loi forte des grands nombres (théorème mathématique), si l'on souhaite que notre modèle décrive « au mieux » la situation réelle, il paraît raisonnable de choisir comme probabilité de l'événement A un nombre égal ou proche de la fréquence d'apparition de l'événement A sur un

[10] Terme utilisé au Québec, plutôt que « fréquentiste ».

grand nombre d'expériences concrètes observées. Donc, dans un second temps, la loi forte des grands nombres (théorème mathématique) justifie aussi l'approche fréquentiste, dans la mesure où la répétition de l'expérience dans le monde réel (approche fréquentiste) se substitue à la répétition de l'expérience aléatoire mathématique du modèle mathématique en construction (loi forte des grands nombres). Cependant, il faut bien avoir conscience qu'il y a la loi des grands nombres empirique d'un côté et la loi forte des grands nombres (théorème mathématique) de l'autre, et que chacune ne s'applique pas aux mêmes objets.

L'approche fréquentiste est mobilisée notamment pour modéliser des situations comme le lancer de punaise, le lancer d'osselet, le lancer de lego, situations pour lesquelles une hypothèse de symétrie ne semble pas raisonnable (et donc pour lesquelles une approche laplacienne n'aurait pas de sens). Cependant, bien que dans le cas du lancer de dé, une hypothèse simplificatrice de symétrie du dé est assez naturelle, l'approche fréquentiste est aussi possible. Dans ce cas,

> *on peut [...] déterminer la probabilité de chaque issue ou événement en effectuant « un grand nombre de fois » l'expérience aléatoire, et en retenant pour valeur numérique de cette probabilité la valeur des fréquences observées de cette issue ou événement* (Henry, 2001c, pp. 168-169),

ou une valeur proche.

Dans le cas du lancer de mon dé préféré, nous pouvons arriver aux hypothèses présentées dans le tableau 2.

<table>
<tr><th>Modèle pseudo-concret
Hypothèses de travail</th><th>Modèle mathématique
Hypothèses de modèle</th></tr>
<tr><td>(HT1) Le dé a 6 faces (1, 2, 3, 4, 5 et 6).
(HT2) On lance le dé suffisamment fort pour que le résultat ne soit pas prévisible.
(HT3) Un résultat du dé sera 1, 2, 3, 4, 5 ou 6 suivant la face visible à l'issue du lancer.
(HT4) Si le dé n'atterrit pas à plat, le lancer est annulé.
(HT5) Le dé lancé 1 000 fois, dans les conditions ci-dessus, a donné les fréquences observées suivantes :
$f_1=0{,}162$; $f_2=0{,}172$; $f_3=0{,}164$; $f_4=0{,}164$; $f_5=0{,}162$; $f_6=0{,}176$.</td><td>On considère l'expérience aléatoire du lancer de dé, pour lequel on s'intéresse au caractère « résultat du dé ».
L'univers considéré est : $\Omega=\{1;...;6\}$ (6 éléments)[11].
Les 6 issues possibles de Ω sont supposées toutes équiprobables. La probabilité est distribuée sur Ω suivant la loi représentée par le tableau :
<table><tr><td>w_i</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>p_i</td><td>0,162</td><td>0,172</td><td>0,164</td><td>0,164</td><td>0,162</td><td>0,176</td></tr></table></td></tr>
</table>

Tableau 2 *: Deuxième modèle du lancer d'un dé.*

Le modèle mathématique résultant d'une approche fréquentiste est *a priori* tout aussi valable, du point de vue de la construction d'un modèle, qu'un modèle issu d'une approche laplacienne. Les deux approches permettent à leur manière d'approcher la vraie probabilité (et non la probabilité théorique). Les deux modèles (*cf.* tableau 1 et tableau 2) reposent simplement sur des hypothèses différentes.

Pour mener l'expérience réelle (en appui sur les hypothèses arrêtées dans le modèle pseudo-concret), il conviendrait aussi (en classe notamment) de se mettre d'accord dans le protocole expérimental sur l'importance ou non de la couleur du dé, de la taille du dé, si on s'intéresse à mon dé en particulier ou à un dé en général, si on peut lancer chacun un dé différent, sur le type de surface sur laquelle on fait l'expérience (sur le cahier, sur la table directement, sur le sol, dans

[11] Au lieu de considérer $\Omega=\{1;...;6\}$, on pourrait considérer les six constellations du dé. En effet, ici nous avons déjà fait un « raccourci » en assimilant les constellations à l'écriture chiffrée associée.

une boîte)… On voit bien ici que les conditions de l'expérience aléatoire vont alors déjà nous éloigner de la vraie probabilité, dans le sens où ce n'est plus mon dé préféré ou un dé classique que l'on lance mais peut-être plusieurs dés différents (de taille, de couleur… différentes) et pas tous lancés au même moment.

Avec les hypothèses de travail présentées dans le tableau 2, et notamment les fréquences observées, on pourrait envisager de faire un test statistique pour tester l'hypothèse d'équiprobabilité[12] (troisième point de la citation (Derouet, 2022, p. 98) précédente). Le test du khi-deux (dans ce cas) nous amènerait à ne pas rejeter l'hypothèse d'équiprobabilité, et donc à considérer à partir du même modèle pseudo-concret un modèle mathématique d'équiprobabilité comme dans le premier modèle (tableau 1, à droite) et non comme dans le deuxième (tableau 2). Dans ce cas, il s'agirait de valider le modèle d'équiprobabilité (modèle 1) en appui sur les données réelles observées (étape 6 de la figure 2).

On pourrait aussi ajouter, dans la construction du modèle, une prise en compte de l'influence des creux sur les faces du dé (liées aux constellations) comme hypothèse de travail et considérer par exemple que les probabilités doivent être croissantes par rapport au résultat du dé.

Bien que plusieurs modèles puissent raisonnablement apparaître, nous pouvons remarquer qu'ils sont relativement « proches » (le test du khi-deux nous l'a montré). L'hypothèse simplificatrice de symétrie semble donc tout à fait raisonnable dans le cas du lancer de dé et sera plus commode pour les différents traitements mathématiques qui peuvent en découler. Cela explique pourquoi une fois introduit dans les classes, une situation de lancer d'un dé, renvoie (dans les exercices par exemple) de façon implicite au modèle du lancer d'un dé du tableau 1.

1.5. La simulation informatique dans la modélisation en probabilités

Dogme (1993) définit la simulation comme :

> *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* (p. 476).

La simulation informatique nécessite donc au préalable le choix d'un modèle mathématique. Parzysz (2009) explique qu' *« une « bonne » simulation présuppose l'association d'un modèle probabiliste à l'expérience étudiée, modèle qui servira ensuite à déterminer la simulation »* (p. 94). Par exemple, une fois que l'on a choisi d'associer la loi uniforme sur $\{1 ; \dots ; 6\}$ au lancer de dé par une approche laplacienne (modèle du tableau 1), on peut entrer dans le tableur la commande « =ALEA.ENTRE.BORNES(1;6) » (on peut parler d'un modèle informatique) pour simuler informatiquement des réalisations de lancers de dé ou encore lancer un dé numérique (simulateur de dé), qui n'est rien d'autre qu'une interface qui renvoie des réalisations d'une loi uniforme sur {1 ; … ; 6}. La simulation ne permet pas de construire un modèle mathématique associé à une situation réelle, elle est simplement l'incarnation d'un modèle mathématique préalablement choisi. Si la simulation est à l'initiative de l'élève, ce modèle préalable a été construit par le biais d'une approche laplacienne (*cf.* tableau 1). La simulation permet alors par exemple de visualiser la loi des grands nombres, la fluctuation d'échantillonnage (propriétés du monde mathématique)… Mais elle ne permet pas de conclure que la loi uniforme modélise bien la réalité, seulement de confirmer que les fréquences relatives des réalisations simulées suivant une loi uniforme tendent bien vers 1/6 (loi des grands nombres). Comme le précisent Thibault et Martin (2018),

[12] Pour un complément sur les tests d'hypothèses, se reporter par exemple à Bonneval (2002).

cet outil technologique [un simulateur informatique] nous permet d'observer la fréquence relative se stabiliser non pas autour de la vraie probabilité, mais plutôt vers la probabilité calculée dans l'approche théorique[13] *en fonction de ce qui a été programmé* (p. 16).

Dans le cas de l'étude de variables aléatoires « complexes » (pour lequel la loi associée n'est pas connue *a priori* des élèves par exemple), la simulation informatique peut permettre d'estimer des probabilités d'événements. Pour la somme de deux dés par exemple, on peut choisir le modèle d'équiprobabilité pour chacune des deux épreuves (lancer du premier dé et lancer du deuxième dé, selon le modèle du tableau 1, par une approche laplacienne) composant l'expérience aléatoire du lancer de deux dés, puis simuler des réalisations du lancer de deux dés sur le tableur et calculer les sommes associées (*cf.* figure 3). En calculant la fréquence d'apparition des différentes sommes possibles, on peut alors estimer les probabilités relatives à la somme de deux dés, d'après la loi des grands nombres, si notre échantillon est de taille suffisamment grande. Il ne s'agit pas d'une approche fréquentiste pour construire un modèle (le modèle est déjà présent au préalable : loi uniforme pour le lancer de chaque dé, et donc la variable aléatoire donnant la somme est la somme de deux lois uniformes), mais en revanche les fréquences obtenues permettent d'estimer des probabilités d'événements tel que « la somme des deux dés est égale 4 », en appui sur la loi des grands nombres.

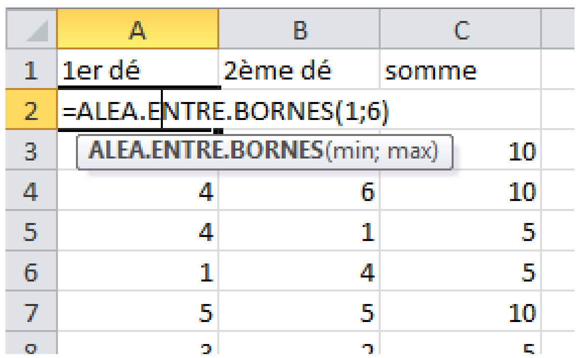


Figure 3 *: Simulation du lancer de deux dés et observation de la somme.*

Bien que reposant toutes les deux sur des liens faits entre fréquences et probabilités, il nous semble important de distinguer l'approche fréquentiste, qui a pour but de construire un modèle pseudo-concret puis mathématique en appui sur des expériences réelles (d'où un aller-retour entre le monde de la réalité et le monde pseudo-concret et même un chevauchement des deux mondes dans la figure 4), et la simulation informatique, qui intervient dans le monde mathématique et permet d'estimer des probabilités d'événements dans un ou des modèles initiaux choisis au préalable, en appui sur des fréquences de réalisations de variables aléatoires simulées (*cf.* figure 4).

[13] Approche laplacienne.

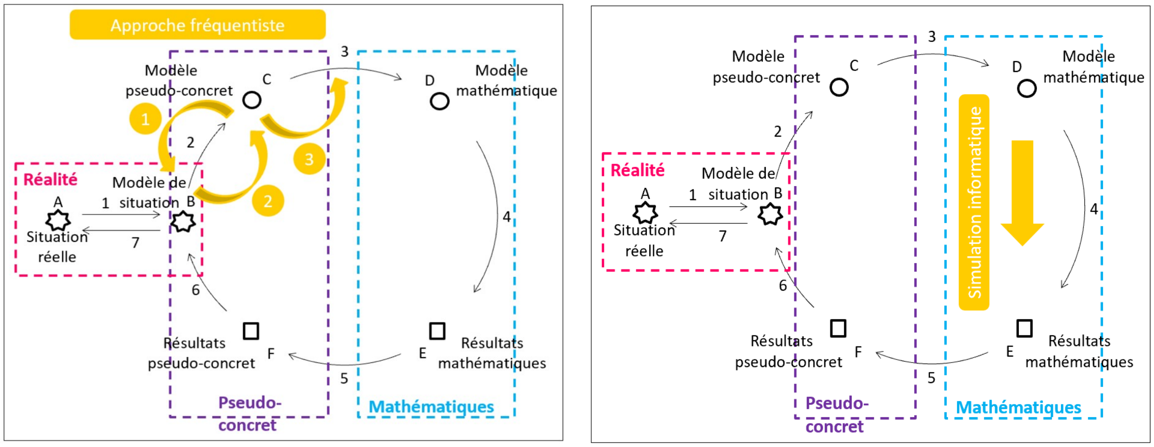


Figure 4 *: Comparaison de la place de l'approche fréquentiste et de la simulation informatique dans le cycle de modélisation (d'après Derouet, 2022).*

Après cette distinction, nous souhaitons aborder les étapes d'interprétation et de validation du processus de modélisation (étapes 5 et 6 de la figure 2).

1.6. Interprétation et validation

Si l'on revient à la situation du lancer de mon dé préféré, chaque étape du processus de modélisation est accompagnée d'un problème et donc de questions, puis de résultats (*cf.* figure 5).

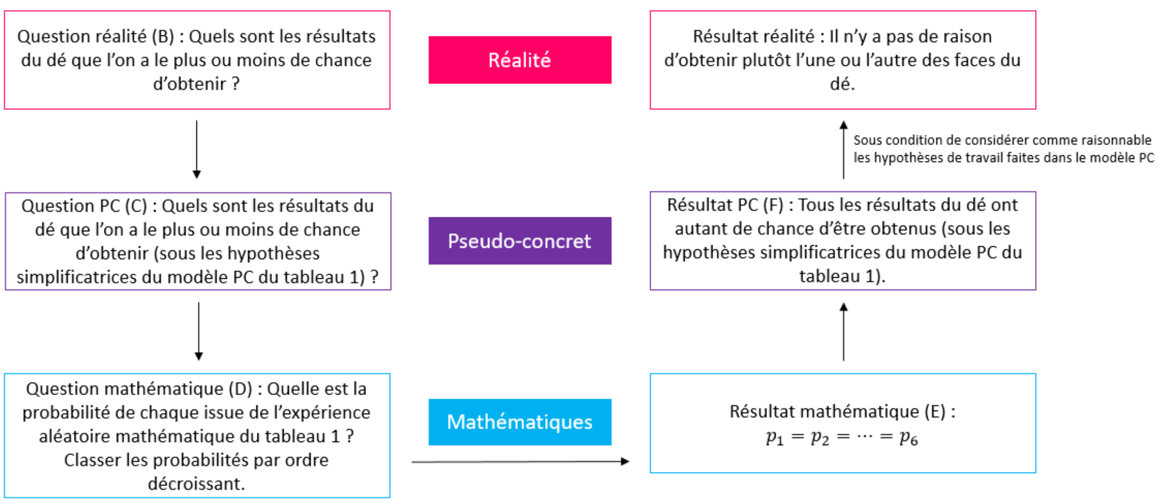


Figure 5 *: Questions et résultats dans les différents mondes dans le cadre du modèle du tableau 1.*

On peut se demander ce que signifie « avoir autant de chance ». Dans le cadre d'une interprétation objective de la probabilité, qui repose sur une interprétation de type fréquentiste (Hacking & Dufour, 2004), cela signifie que si l'on reproduisait une infinité de fois l'expérience, les fréquences d'apparition des résultats seraient les mêmes. La question de la validation dans la réalité (étape 6 de la figure 2) vient ensuite avec une confrontation qui peut être faite entre les données réelles et les données obtenues à partir du modèle, notamment pour évaluer si l'écart entre les fréquences observées (réalité) et les probabilités calculées à partir du modèle ou les fréquences simulées est suffisamment faible pour être acceptée dans le but de donner une réponse à notre problématique de départ. Dans l'enseignement secondaire, cette comparaison peut être faite par exemple visuellement en appui sur des graphiques, puis avec des intervalles de confiance/de fluctuation.

2. Vers une analyse institutionnelle

2.1. Interprétation de l'extrait du document Ressource *Modéliser*

À la lumière des éléments théoriques exposés en première partie, nous souhaitons revenir sur l'extrait du document Ressource *Modéliser* (MENESR, 2016b), présenté dans l'introduction, en proposant notre interprétation de différents passages. Nous écrivons en italique l'extrait du document Ressource (MENESR, 2016b), suivi de notre interprétation en lien avec le cycle de modélisation et le codage de la figure 2.

« La simplicité des problèmes ne doit pas cacher la complexité du modèle proposé » (p. 3).

Les problèmes étudiés peuvent concerner des situations *a priori* simples : lancer d'un dé, d'une pièce, tirage au sort… Par exemple, le lancer de dé peut sembler être une situation très simple pourtant c'est la complexité de ce phénomène (forte dépendance aux conditions initiales) qui nous empêche de prévoir le résultat. L'étude du lancer de dé se fera donc à travers la construction d'un modèle pseudo-concret (C) puis mathématique (D).

« Il convient de ne pas précipiter le passage au modèle probabiliste et aux calculs à l'intérieur de ce modèle […] » (p. 3).

Il est important de bien distinguer la situation réelle (A/B) et le modèle pseudo-concret (C) ou mathématique (D) qui permet de l'étudier, et pour cela de bien prendre le temps de détailler le modèle pseudo-concret (C) en précisant les hypothèses de travail qui le décrivent et ne pas se précipiter dans le traitement mathématique (4).

« […] et il convient surtout de ne pas prendre les calculs menés à l'intérieur du modèle pour des preuves ou des démonstrations de phénomènes réels » (p. 3).

Les résultats (E) obtenus dans le monde mathématique à partir du modèle mathématique (D) ne sont pas les résultats réels, répondant au problème réel de départ. Les résultats mathématiques dépendent en effet des simplifications et idéalisations faites dans le modèle pseudo-concret et de ce fait ne répondent qu'au problème pseudo-concret considéré (C).

« On ne prouve pas que « le dé est truqué », ou qu'« il vaut mieux miser sur l'obtention d'un 7 comme somme de deux dés que sur l'obtention d'un 2 », mais on effectue des calculs à l'intérieur du modèle mathématique qui aboutissent à une conclusion mathématique rigoureuse » (p. 3).

Les résultats (E et F) ne disent « rien » sur la réalité (A/B), ils sont valides dans le monde mathématique, sous les hypothèses de travail que l'on a faites dans le monde pseudo-concret (qui n'est pas la réalité).

« Les conclusions que l'on peut en tirer quant aux dés réels ou aux paris réels ne relèvent pas des mathématiques et doivent être considérées avec recul et prudence » (p. 3).

Le retour à la réalité (6) dépend de considérations qui dépassent les mathématiques et qui finalement ne concernent plus seulement le mathématicien (Henry, 2001d, p. 157) car des connaissances dans le monde de la situation de départ peuvent être nécessaires. Les résultats disent des choses sur le réel, plus ou moins fiables… À nous ensuite de décider la part de confiance que l'on met dans les résultats obtenus dans ce modèle, confiance que l'on peut parfois quantifier dans le modèle.

« Pour des élèves à qui on demande un effort d'abstraction et de manipulation de concepts, il est important de connaître et d'éprouver régulièrement l'efficacité des outils qu'ils apprennent à maîtriser, au travers en particulier de cette troisième phase de confrontation des résultats du modèle avec la réalité empirique » (p. 2).

La phase de confrontation des résultats du modèle avec la réalité empirique peut être travaillée en probabilités en appréciant l'écart entre les données empiriques (réelles) et les résultats obtenus avec le modèle pour évaluer la fiabilité du modèle choisi. Pour cela, on peut regarder si les réalisations simulées (avec le modèle mathématique choisi) sont bien en adéquation avec les données récoltées par exemple.

L'interprétation de ces extraits met bien en évidence l'importance du processus modélisation en probabilités avec les différentes étapes constitutives de la démarche, ainsi que la distinction entre réalité et modèle. Cela semble constituer de véritables enjeux de compréhension même des probabilités. Les éléments présents dans ce document Ressource transparaissent-ils dans les programmes et les autres documents accompagnant les programmes ? dans les sujets d'examens nationaux ?

2.2. Méthodologie

Afin de questionner la place de la modélisation en probabilités dans l'enseignement secondaire, nous allons mener une analyse institutionnelle à travers l'étude des programmes, des documents Ressources et des sujets de d'examens nationaux. Pour le collège, nous analyserons :

- le programme actuel de mathématiques (MENJS, 2020), ainsi que les repères de progressivité et les attendus de fin d'année du cycle 4 (MENJ, 2019a/b/c/d),
- les documents Ressources du collège disponibles sur Éduscol (MENESR, 2016a ; MENJ/DGESCO, 2016 ; MENJS, 2022),
- les exercices portant sur les probabilités des neufs sujets du diplôme national du brevet (DNB) de la session 2023[14] (métropole et centres étrangers),

et, pour le lycée :

- les programmes actuels de seconde et de spécialité mathématiques de première et de terminale du lycée général (MENJ, 2019e/f/g)[15],
- les exercices portant sur les probabilités des vingt-et-un sujets de baccalauréat 2023 de la spécialité mathématiques[16].

Le fait d'analyser l'ensemble des sujets d'examens nationaux (DNB et baccalauréat) d'une même année (sujets de métropole et des centres étrangers) permet d'avoir un point de vue global sur ce qui peut être attendu des élèves à la fin du collège et du lycée en spécialité mathématiques. L'analyse sera faite du point de vue de la modélisation probabiliste, à partir des éléments théoriques présentés en partie 1. Les critères étudiés pour notre analyse sont les suivants :

[14] Sujets disponibles sur le site de l'APMEP : https://www.apmep.fr/Brevet-2023

[15] Aucun document Ressources n'existe actuellement concernant les probabilités pour ces niveaux scolaires.

[16] Sujets disponibles sur le site de l'APMEP : https://www.apmep.fr/Annales-Terminale-Generale

- explicitation ou non de certaines étapes du processus de modélisation, distinction réalité/modèle ;
- nature des situations abordées (réelle, pseudo-concrète, mathématique) ;
- modèles de référence étudiés ;
- présence/absence des hypothèses de travail ;
- nature des problèmes posés (R, PC, M) ;
- nature de l'approche des probabilités pour construire le modèle ;
- présence/absence et rôle de la simulation informatique.

Nous présentons ci-dessous les conclusions de cette analyse institutionnelle, tout d'abord du point de vue curriculaire (programmes et documents accompagnant les programmes) puis du point de vue des examens nationaux (DNB et baccalauréat) pour le collège et ensuite pour le lycée, que nous illustrons ponctuellement par des exemples. Nous ne présentons que les tendances générales mises en lumière par les analyses ; notre description ne prétend donc pas à l'exhaustivité.

3. Place de la modélisation en probabilités dans l'enseignement secondaire

3.1. Au collège : l'illusion du réel

Bilan de l'analyse curriculaire du cycle 4

Le processus de modélisation, plus particulièrement certaines étapes constitutives du processus de modélisation, sont présentes dans le document Ressource *Modéliser*, point de départ de cet article, et dans le préambule du programme du cycle 4 (MENJS, 2020). Dans ce dernier, une partie est consacrée aux six compétences mathématiques et pour la compétence *Modéliser*, il est écrit :

- *Reconnaître un modèle mathématique (proportionnalité, équiprobabilité) et raisonner dans le cadre de ce modèle pour résoudre un problème* (étapes 3 et 4 du cycle de modélisation en figure 2) ;
- *Traduire en langage mathématique une situation réelle* (étapes 1, 2 et 3 de la figure 2) ;
- *Comprendre et utiliser une simulation numérique ou géométrique* (étape 4 de la figure 2) ;
- *Valider ou invalider un modèle, comparer une situation à un modèle connu* (par exemple un modèle aléatoire) (étape 6 de la figure 2) (pp. 129-130).

En revanche, dans la partie spécifique aux probabilités du programme et dans les différents documents accompagnant le programme, on peut remarquer l'absence de mention de ces étapes et d'éléments concernant la distinction réel/modèle.

Pour préciser les types de situations et de problèmes abordés, on trouve dans le programme : *« aborder les questions relatives au hasard à partir de problèmes simples »*, « *calculer des probabilités dans des cas simples (par exemple évaluation des chances de gain dans un jeu) »* (MENJS, 2020, p. 133) ; dans les attendus de fin d'année de cinquième : *« [calculer] des probabilités dans des situations simples d'équiprobabilité »*[17] (MENJ, 2019c) ; ou encore dans le document Ressource intitulé *Comprendre et utiliser des notions élémentaires de probabilités*[18] (MENESR, 2016a) : *« situations familières aux élèves »* et *« situations de la vie courante (jeux,*

[17] On peut noter que le terme « équiprobabilité » n'est pas présent dans le programme.

achats, structures familiales, informations apportées par les médias, etc.) » (p. 1). Pour les différentes situations en question, elles semblent parfois relever de la réalité (vie quotidienne), parfois du monde pseudo-concret et parfois du monde mathématique (situations d'équiprobabilité). Les exemples trouvés dans les différents documents Ressources relèvent finalement essentiellement de situations et de problèmes du monde pseudo-concret (et non réels) autour du jeu, où l'on travaille avec des objets idéaux. Il s'agit essentiellement de situations de lancer de dé, de tirage dans un sac ou dans une urne, de roue de loterie (modèles pseudo-concrets de référence). Les hypothèses de travail sous-jacentes aux différents modèles pseudo-concrets étudiés sont rarement entièrement explicitées. Il peut même parfois manquer l'expression « au hasard ». Cela pourrait permettre un travail de l'étape 2 du cycle de modélisation, mais ce n'est jamais ce qui est attendu.

Sans que soit citée explicitement la (ou les) approche(s) privilégiée(s) pour construire les modèles mathématiques, dans la grande majorité des problèmes, les expériences aléatoires étudiées relèvent du modèle d'équiprobabilité et donc une hypothèse (pas toujours explicite) de symétrie a été faite. Il s'agit donc majoritairement d'une approche laplacienne. On trouve cependant explicitement une approche fréquentiste dans les annexes *Un exemple de tâche avec prise d'initiative : Jeu équitable ?* et *Le jeu du franc-carreau* (MENESR, 2016a). En revanche, cette approche est toujours en première étape et semble plus être là pour permettre à l'élève de se familiariser avec le jeu, plutôt que pour lui permettre effectivement de construire un modèle probabiliste. En effet, ensuite il y a un basculement vers un modèle équiprobable (approche laplacienne). Dans les différents documents, l'approche fréquentiste ne semble donc pas avoir la même valeur que l'approche laplacienne pour construire un modèle.

Dans les attendus de fin d'année de troisième (MENJ, 2019a), il est écrit que *« [l'élève] fait le lien entre stabilisation des fréquences et probabilités »* (le mot « stabilisation » a été ajouté par rapport au programme). Les repères annuels de progression du cycle 4 (MENJ, 2019d) précisent que, pour la classe de troisième, *« [l]e constat de la stabilisation des fréquences s'appuie sur la simulation d'expériences aléatoires à une épreuve à l'aide d'un tableur ou d'un logiciel de programmation »*, ce qui se rapproche d'une illustration de la fluctuation d'échantillonnage, de l'intervalle de fluctuation et de la loi des grands nombres par la simulation informatique (dans le cadre d'un modèle mathématique choisi au préalable) plus que de l'approche fréquentiste.

Un autre exemple de réussite liée à cette compétence est cité :

> *On donne les fréquences d'apparition de chaque face d'un dé pour 10 000 lancers.*
> *L'élève interprète les résultats en les comparant aux probabilités théoriques* (MENJ, 2019a).

La nature des lancers du dé est ambiguë : s'agit-il de lancers réels ou simulés ? Les « probabilités théoriques » font référence aux probabilités issues d'une approche théorique (modèle du tableau 1), c'est-à-dire laplacienne (*cf.* partie 1.4), à ne pas confondre avec ce que l'on a appelé la vraie probabilité. Dans le premier cas, s'il s'agit d'expériences réelles de lancers de dé, il s'agirait donc de questionner la validité du modèle 1 au vu des fréquences réelles observées (qui semblent se stabiliser). Dans le second cas, s'il s'agit de lancers simulés informatiquement, il est attendu d'interpréter les résultats en appui sur la loi des grands nombres.

Illustration de manques concernant la description du modèle pseudo-concret

Dans la fiche consacrée à des *Exemples de questions flash* (MENESR, 2016a), un des exercices (*cf.* figure 6) est consacré à « des roues » (d'après le titre).

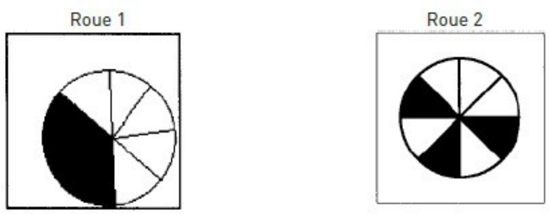
[18] Titre reprenant l'intitulé du thème concernant les probabilités du programme de cycle 4.

Exercice 3 : avec des roues

1. Quelle couleur a le plus de chance d'être obtenue quand on fait tourner la roue ci-dessous ?

2. Avec la roue 1, la couleur noire a :
   a. autant de chance ;
   b. plus de chance ;
   c. moins de chance ;

d'être obtenue qu'avec la roue 2.



3. On place une fève au hasard dans une galette des rois. On partage cette galette en huit parts égales.

Quelle probabilité a-t-on de tomber sur la part ayant la fève ?

***Figure 6** : Extrait 1 de la fiche* Exemples de questions flash *(MENESR, 2016a).*

Pour les deux premières questions, il s'agit d'étudier une expérience aléatoire « tourner une roue de loterie et identifier la partie sur laquelle la roue s'arrête », même si l'image donnée pourrait peut-être plus faire penser à une cible qu'à une roue de loterie. Des roues différentes interviennent (avec des secteurs colorés différents). Cependant dans tous les cas, la roue (considérée implicitement comme un disque) est partagée en huit parties de même aire et de même forme. Il est donc attendu de considérer que l'on a autant de chance de tomber sur une partie ou une autre de la roue (hypothèse de symétrie des différentes parties), et donc de construire le modèle par une approche laplacienne : l'expérience aléatoire suit une loi uniforme sur chacune des 8 parties. Cela permet ensuite de considérer les événements en jeu dans les questions et répondre dans le cadre de ce modèle. Les situations étudiées sont dans le monde pseudo-concret, cependant tous les éléments détaillés ci-dessus (expérience aléatoire et hypothèses de travail sous-jacentes[19]) sont passés sous silence.

La question 3 change de contexte, en s'intéressant au tirage d'une part de galette. Ici, on sort du contexte du jeu pour se rapprocher d'une situation de la « vie quotidienne », cependant, la situation présentée peut être ramenée au modèle pseudo-concret étudié précédemment. On peut repérer que, selon l'énoncé, l'expérience aléatoire en jeu est la position de la fève dans la galette (« On place une fève au hasard dans une galette des rois »), en revanche par rapport à la question posée l'expérience aléatoire qui nous intéresse véritablement sans être écrite c'est le tirage au hasard d'une part de galette. Pour aborder ce problème, des hypothèses (simplificatrices) de travail sont à faire : on suppose que la fève n'est pas à la frontière entre deux parts (ou l'on

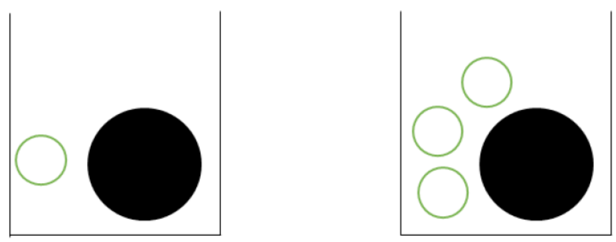
[19] Ici, elles ne sont pas toutes explicitées : il faut notamment supposer que la roue est tournée suffisamment fort…

considère que la fève est un point…), que l'on choisit indifféremment une part plutôt qu'une autre (dans la réalité, les parts ne sont jamais exactement identiques et cela influence sûrement le choix que l'on fait d'une part…). Cet exemple simple pourrait être l'occasion de réfléchir avec les élèves à des questions relatives à la modélisation et de passer du temps à expliciter des hypothèses de travail et de modèle.

C'est seulement dans l'exercice suivant de la même annexe qu'est proposée en remarque une discussion sur les hypothèses simplificatrices, avec notamment une réflexion autour de l'expression « indiscernable au toucher » (figure 7). Mais cela semble anecdotique et peu étayé pour l'enseignant.

Exercice 4 : avec des urnes

1. A-t-on plus de chance de tirer une boule blanche dans :
   a. une urne A qui contient 3 boules toutes blanches ?
   ou
   b. une urne B qui contient 500 boules blanches et une boule rouge ?
2. A-t-on plus de chance de tirer une boule blanche ou une boule noire dans chacune des urnes ci-dessous ?



Remarque : il n'y a pas de « bonne réponse » attendue. L'intérêt est de mettre en évidence l'expression « indiscernable au toucher ». On peut ensuite reprendre le même exercice avec des boules de même taille, indiscernables au toucher.

***Figure** 7 : Extrait 2 de la fiche* Exemples de questions flash *(MENESR, 2016a).*

Ces exemples montrent des situations du monde pseudo-concret, mais pour lesquelles l'expérience aléatoire et les hypothèses de travail ne sont pas explicitées.

Analyse globale des sujets du DNB 2023

Sur les neuf sujets du DNB 2023[20] (métropole et centres étrangers), tous proposent *a minima* une question qui concerne les probabilités, et *a maxima* un exercice complet pour certains sujets. Tous les exercices sont contextualisés, faisant intervenir des modèles pseudo-concrets de référence : il est notamment question six fois de tirage de boules (ou billes, ou jetons) dans une urne (ou un sac), deux fois de tourner une roue de loterie, une fois d'un tirage de carte. Contrairement à ce que l'on a pu trouver dans les documents Ressources, il n'y a pas une seule fois un lancer de dé ou de pièce. Pour tous les exercices, comme repéré dans les documents Ressources, les hypothèses de travail décrivant l'expérience aléatoire ne sont pas toutes clairement explicitées. Les modèles mathématiques, et notamment le modèle d'équiprobabilité, n'est jamais mentionné tel quel dans les sujets et n'est jamais demandé en question (est-ce attendu dans les réponses des élèves aux autres questions ?). Les questions sont très majoritairement (trois quarts des questions) dans le monde mathématique (avec la présence du mot « probabilité ») : « Quelle est la probabilité de… » ou « Montrer que la probabilité de…

[20] Disponibles sur le site de l'APMEP : https://www.apmep.fr/Brevet-2023

est… ». Les élèves doivent être capable de dénombrer les issues de l'expérience aléatoire et les issues favorables à l'événement considéré (sans même mentionner explicitement que l'on se situe dans un modèle d'équiprobabilité) pour appliquer la formule. L'approche présente dans le DNB est donc l'approche laplacienne. Un seul sujet propose une simulation informatique, il s'agit d'une simulation dans le cas de lois uniformes pour aborder la notion de fluctuation d'échantillonnage.

Il n'est jamais question de revenir à l'interprétation des résultats mathématiques dans le monde pseudo-concret. Les exercices ne sont jamais autour d'une problématique ; le monde de la réalité est absent : ce qui peut sembler compréhensible dans le cadre d'un examen.

Analyse d'un exercice en particulier

Il s'agit ici de présenter plus en détail l'analyse d'un exercice du corpus. Cet exercice est un des deux exercices portant entièrement sur les probabilités. Dans l'exercice (DNB Polynésie, juin 2023, en annexe 1), il s'agit de l'étude de deux jeux (*cf.* figure 8) correspondant chacun à une expérience aléatoire à une épreuve décrite dans l'énoncé : la première « piocher au hasard une boule dans un sac et noter la lettre inscrite sur la boule » (expérience aléatoire 1), la seconde « tourner la roue et noter le nombre inscrit sur le secteur pointé par la flèche » (expérience aléatoire 2). Les deux expériences sont étudiées indépendamment dans la partie A de l'exercice, puis sont étudiées de façon combinée comme une expérience aléatoire à deux épreuves dans la partie B. Les deux épreuves sont indépendantes (monde mathématique) mais non identiques ; les épreuves ne sont pas de même nature (piocher une boule et tourner une roue), les issues différentes (lettre et nombre) et pas en même nombre. On peut noter que l'indépendance des issues est totalement implicite dans l'énoncé. Les hypothèses relatives aux deux expériences aléatoires sont relativement explicites « boules indiscernables au toucher » et « secteurs angulaires identiques ». L'énoncé est dans le monde pseudo-concret. Il s'agit de traiter un problème sur des objets abstraits (et non une vraie roue ou un vrai sac). Les questions sont toutes dans le monde mathématique car elles mettent en jeu la notion de probabilité. Le passage dans le monde mathématique se fait au travers de l'utilisation (explicite ou implicite) de la formule de Laplace pour déterminer les probabilités sous la condition d'équiprobabilité des événements élémentaires : cependant, pour le jeu 1, il faut noter que l'équiprobabilité n'est pas associée à l'expérience aléatoire 1 citée précédemment pour laquelle il s'agit de « noter la lettre » (en effet il y a plusieurs boules avec la même lettre) mais l'expérience aléatoire où il s'agit d'identifier la boule tirée (boule 1, boule 2, …, boule 5). Il est donc fort probable que l'équiprobabilité soit passée sous silence par les élèves, mais utilisée en acte. Le retour au monde pseudo-concret n'est pas présent dans cet exemple.

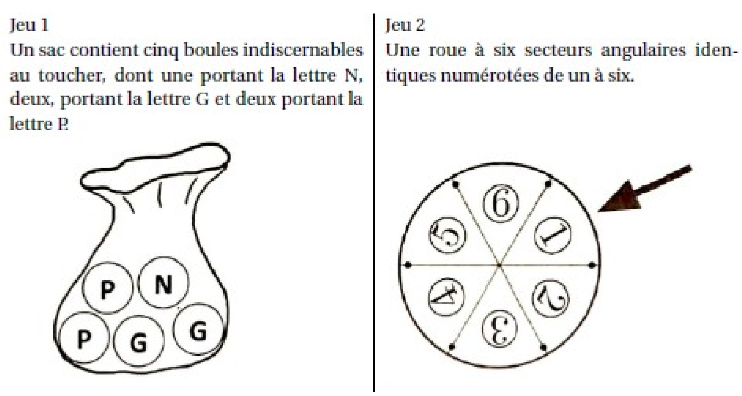


Figure 8 *: Extrait du Brevet des collèges Polynésie 23 juin 2023 (exercice 3).*

3.2. Au lycée : travail sur les modèles mathématiques

Analyse curriculaire

Le thème des probabilités est abordé dans tous les programmes de mathématiques du lycée (excepté dans l'option mathématiques expertes). Nous nous focalisons ici sur la seconde et la spécialité mathématiques en classe de première et de terminale du lycée général.

Contrairement au programme de cycle 4, les programmes de mathématiques de seconde et de première (spécialité mathématiques) insistent bien sur la distinction entre réalité et modèle, avec des éléments explicites dans l'introduction du thème *Statistiques et probabilités* :

> *En classe de seconde, on formalise la notion de loi (ou distribution) de probabilité dans le cas fini en s'appuyant sur le langage des ensembles et on précise les premiers éléments de calcul des probabilités.* ***On insiste sur le fait qu'une loi de probabilité*** *(par exemple une équiprobabilité)* ***est une hypothèse du modèle choisi et ne se démontre pas***[21]*. Le* ***choix du modèle peut résulter d'hypothèses implicites d'équiprobabilité*** *(par exemple, lancers de pièces ou dés équilibrés, tirage au hasard dans une population)* ***qu'il est recommandable d'expliciter*** *;* ***il peut aussi résulter d'une application d'une version vulgarisée de la loi des grands nombres, où un modèle est construit à partir de fréquences observées pour un phénomène réel*** *(par exemple : lancer de punaise, sexe d'un enfant à la naissance). Dans tous les cas,* ***on distingue nettement le modèle probabiliste abstrait et la situation réelle*** (programme de seconde, MENJ, 2019f, p. 13).
>
> *Comme en seconde,* ***on distingue nettement modèle et réalité****. Ainsi,* ***une hypothèse d'indépendance fait partie d'un modèle : elle peut être un point de départ théorique ou être la conséquence d'autres hypothèses théoriques. Lorsque le modèle est appliqué à une situation réelle*** *(par exemple, lancer de deux dés physiques)****, l'indépendance fait partie de la modélisation et résulte de l'analyse de la situation physique*** (programme de spécialité mathématiques de première, MENJ, 2019e, p. 13).

Ces extraits insistent bien sur le fait que l'équiprobabilité ou l'indépendance d'expériences aléatoires sont des hypothèses, associées au modèle choisi. Il est préconisé dans un des extraits de rendre explicite les « hypothèses implicites d'équiprobabilité ». On voit donc ici des éléments relatifs au modèle pseudo-concret (étape C de la figure 2), avec l'explicitation des hypothèses de travail, pour le distinguer de la situation réelle (étape A/B de la figure 2), ainsi qu'au modèle mathématique (étape D de la figure 2), avec les hypothèses de modèle. Ceci met bien en évidence des éléments constitutifs de la démarche de modélisation. En revanche, il n'est pas question de l'interprétation (étape 5 de la figure 2), ni de la validation (étape 6 de la figure 2).

Dans le programme de seconde, on voit aussi apparaître clairement (même si ce n'est pas en ces termes) la construction des modèles par l'approche laplacienne et par l'approche fréquentiste :

> • *Utiliser des modèles théoriques de référence (dé, pièce équilibrée, tirage au sort avec équiprobabilité dans une population) en comprenant que les probabilités sont définies* a priori.
>
> • *Construire un modèle à partir de fréquences observées, en distinguant nettement modèle et réalité* (MENJ, 2019f, p. 14).

Pour le premier point, ce qui est appelé ici « modèles théoriques de référence » renvoie au concept de modèles pseudo-concrets de référence et est associé à la construction de modèles par une approche laplacienne (hypothèse de symétrie). Pour le deuxième point, il est mentionné la construction d'un modèle « à partir de fréquences observées », il s'agit ici de l'approche fréquentiste ayant pour objectif de construire un modèle en appui sur des observations (dans la réalité). Il est bien mentionné de ne pas confondre réalité et modèle. Le contenu de seconde est

[21] Mis en gras par nos soins.

relativement proche de celui du cycle 4 mais beaucoup plus explicite sur la place de la modélisation en probabilités et est en cohérence avec le document Ressource *Modéliser* du cycle 4 (MENESR, 2016b).

En terminale, le programme de spécialité mathématiques insiste sur le fait que la partie *Probabilités* consiste à *« [diversifier] et [approfondir] les modèles probabilistes rencontrés »* (MENJ, 2019g, p. 16). Il s'agit de travailler le traitement mathématique (étape 4 de la figure 2) au sein du monde mathématique.

La simulation apparaît dans les trois programmes et est associée à un travail sur la loi des grands nombres. En seconde, dans la partie *Échantillonnage*, il est écrit :

> *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* (MENJ, 2019f, p. 14).

Le travail sur la simulation informatique se poursuit en première et en terminale, et la loi des grands nombres est formalisée en terminale.

Il n'existe pas, actuellement, de document Ressource sur les probabilités en complément des programmes de seconde, première et terminale spécialité mathématiques. Nous ne pouvons donc pas aller plus loin dans l'analyse pour la seconde et la première. Les sujets du baccalauréat nous informent sur la terminale.

Analyse globale des sujets de baccalauréat 2023 de spécialité mathématiques

Comme nous l'avons vu dans l'analyse du programme, le travail semble axé sur la partie mathématique de la modélisation en terminale spécialité mathématiques.

Sur les vingt-et-un sujets de baccalauréat 2023 de la spécialité mathématiques, tous proposent un exercice portant sur les probabilités (*a minima* une question). L'ensemble des exercices est contextualisé avec un modèle pseudo-concret donné avec des hypothèses de travail (mais souvent incomplètes). Il ne s'agit pas ici de modèles pseudo-concret de référence comme au DNB (lancer de dé, lancer de pièce…) mais de situations se rapprochant de la vie quotidienne ou professionnelle, cependant régulièrement associée au modèle pseudo-concret de référence de tirage avec remise. Les choix d'hypothèses de modèle (étape D de la figure 2) dans le cas de la loi binomiale ne sont pas toujours précisés dans les sujets, parfois la loi binomiale est admise directement sans faire de lien avec le modèle pseudo-concret.

Sur plus de 100 questions (113 questions précisément) portant sur les probabilités, les trois quarts d'entre elles concernent le calcul d'une probabilité ou la détermination du paramètre n dans une loi binomiale pour qu'une certaine probabilité donnée soit supérieure ou inférieure à une valeur fixée. Il s'agit donc très majoritairement de questions de traitement mathématique (étape 4 de la figure 2) dans un modèle probabiliste donné.

Pour les questions restantes,

- 10 % des questions portent sur la construction complète ou partielle d'un arbre pondéré (étape D de la figure 2) ;
- 9 % portent sur l'identification de la loi de probabilité en jeu et/ou de ses paramètres, elles relèvent du passage entre le modèle pseudo-concret et le modèle mathématique (étape 3 de la figure 2) ;
- 6 % portent sur de l'interprétation de résultats mathématiques (étape 5 de la figure 2) : il

s'agit notamment de l'interprétation de l'espérance mathématique (4 fois) et d'une probabilité (2 fois).

L'analyse des sujets de baccalauréat montrent bien une prédominance d'un travail dans le monde mathématique en fin de lycée, avec essentiellement du calcul de probabilités, dans le cas d'une expérience aléatoire (mathématique) à deux ou trois épreuves non nécessairement identiques ou dans le cas d'une variable aléatoire suivant une loi binomiale. Le modèle mathématique est déjà donné (mis ou non en relation avec un modèle pseudo-concret) ou à reconnaître par exemple dans le cas de la loi binomiale avec un modèle pseudo-concret standardisé (avec un tirage avec remise par exemple). Bien que le travail de l'élève soit dans le monde mathématique, les exercices sont toujours contextualisés dans le monde pseudo-concret avec des hypothèses de travail pour décrire des modèles pseudo-concrets (non forcément toutes présentes) ou avec parfois même des éléments de modèles probabilistes (avec directement la présence de probabilités). En revanche, les choix faits ne sont jamais questionnés ou remis en question (validation du modèle, étape 6). Les modèles reposent toujours sur une hypothèse d'équiprobabilité (pas forcément directement sur l'expérience aléatoire explicitée mais sur une autre, préalable, par exemple « on prélève au hasard… »). Il s'agit donc d'une approche laplacienne. Nous avons pu voir que très peu de questions concernent une interprétation des résultats mathématiques en résultats pseudo-concrets. Et comme au collège, la question de la validité d'un modèle ou rien qu'une réflexion sur les choix faits pour le modèle n'est jamais présente au baccalauréat. Bien que le programme de seconde et de première spécialité laissent penser qu'une telle réflexion pourrait avoir lieu.

La simulation est absente des sujets de baccalauréat.

Analyse d'un exercice en particulier

L'exemple choisi (issu du baccalauréat Amérique du Nord, mars 2023, en annexe 2) est présenté à titre illustratif. En appui sur les données d'une enquête statistique sur la production de déchets d'une entreprise, il s'agit de construire un modèle probabiliste pour étudier l'expérience aléatoire qui consiste à *« prélever au hasard un déchet et à donner sa nature (minéral et non dangereux, non minéral et non dangereux, dangereux, recyclable, non recyclable) »*. Pour associer le modèle probabiliste attendu dans l'exercice, des hypothèses simplificatrices de travail assez fortes sont nécessaires (et ici non explicitées) :

- HT1. Au moment du prélèvement, tous les déchets peuvent être prélevés indifféremment, donc notamment on considère que le lieu et le moment du prélèvement n'ont pas d'influence sur le type de déchets prélevés.
- HT2. On considère qu'à tout moment et en tout lieu de l'entreprise, les déchets sont répartis suivant les fréquences données dans l'énoncé.
- HT3. On considère que les fréquences observées dans l'échantillon de l'enquête sont les fréquences observables dans la population des déchets, à tout moment et en tout lieu de l'entreprise.

Ces hypothèses simplificatrices pourraient être questionnées du point de vue de la réalité. Le modèle est construit en appui sur une approche laplacienne : on suppose en effet l'équiprobabilité de tirer l'un ou l'autre des déchets (HT1). Les paramètres sont ensuite choisis par l'énoncé en appui sur l'enquête statistique (très certainement inventée).

Sur la partie B concernant la loi binomiale, il est écrit *« Afin de contrôler la qualité de la collecte dans l'entreprise, on prélève un échantillon de 20 déchets pris au hasard dans la production. On* *suppose que le stock est suffisamment important pour assimiler le prélèvement de cet échantillon à un tirage avec remise »*. Ici, on voit bien apparaître une hypothèse simplificatrice de travail, à savoir de faire le choix de considérer que le stock est suffisamment important (ce qui n'est pas nécessairement le cas dans la réalité, et d'ailleurs que signifie « suffisamment important » ?) afin de se ramener à un modèle pseudo-concret de tirage aléatoire avec remise (que l'on peut qualifier de référence en terminale), que l'on peut ensuite associer à la loi binomiale. Il s'agit d'un tirage au hasard (hypothèse d'équiprobabilité dans le choix) avec remise de 20 boules dans une urne (ou déchets dans l'entreprise), la composition de l'urne n'étant pas explicitée mais la proportion de boules notées « recyclable » devant être 0,6514. La question de l'indépendance des tirages n'est pas explicitée. Ce modèle pseudo-concret de tirage avec remise très souvent utilisé en terminale pour modéliser des situations « réelles », la réflexion autour des hypothèses de travail que cela implique et la question de la validité du modèle ne sont pas abordées.

Cet exemple particulier permet d'illustrer des manques d'explicitation et parfois de rigueur que l'on peut retrouver dans d'autres sujets.

3.3. Bilan de l'analyse et questionnement

À travers cette analyse, on observe qu'une réflexion sur la modélisation semble être plus présente dans les programmes du lycée que du collège.

Au collège, le travail de modélisation est essentiellement centré sur le passage du modèle pseudo-concret avec une hypothèse de symétrie (pas forcément très explicite) à un résultat mathématique avec le calcul de probabilités (passage de C à E, *cf.* figure 9 à gauche), sans nécessairement expliciter le modèle mathématique d'équiprobabilité. L'approche laplacienne des probabilités est majoritaire. Les programmes de seconde et de première insistent bien sur les hypothèses simplificatrices qui sous-tendent un modèle et laissent penser qu'un travail sur la description du modèle pseudo-concret est attendu. Le travail en terminale spécialité mathématiques tend ensuite à se centrer sur le monde mathématique et le travail de modélisation est essentiellement axé sur le traitement mathématique (passage de D à E, *cf.* figure 9 à droite).

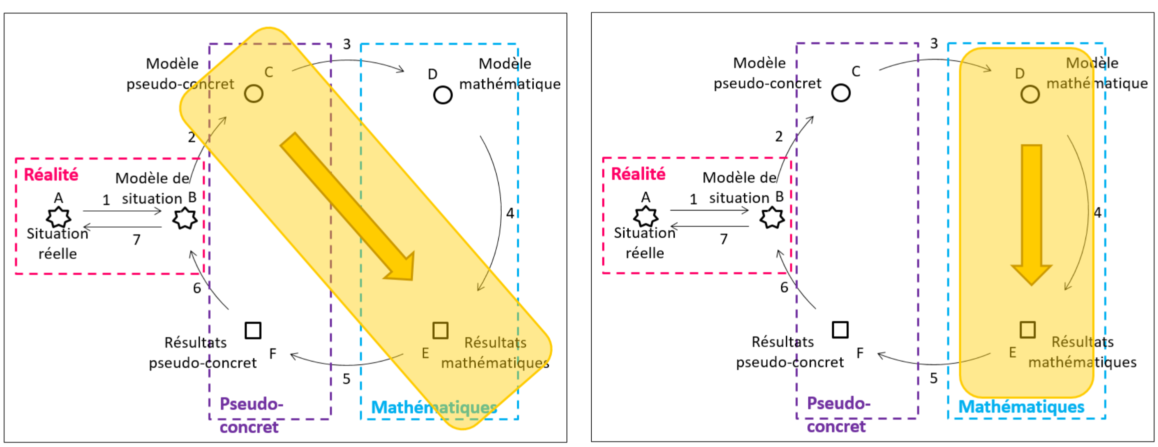


Figure 9 *: Étapes du cycle de modélisation travaillées au DNB (à gauche) et au baccalauréat (à droite).*

Les modèles pseudo-concrets étudiés, que l'on pourrait qualifier de modèles pseudo-concrets de référence (lancer d'un dé, lancer d'une pièce, tirage dans une urne… puis tirage sans remise en terminale), ont des hypothèses de travail très peu souvent explicitées (et jamais entièrement) et ce, même dans les documents Ressources, proposant des activités à mettre en œuvre en classe.

De plus, on peut remarquer la forte présence du mot « au hasard » pour déduire des modèles pseudo-concrets un modèle d'équiprobabilité. On peut se demander si cette systématique équivalence dans l'intitulé du modèle pseudo-concret (par exemple tirage au hasard dans une urne) entre « au hasard » et équiprobabilité ne peut pas renforcer chez les élèves des conceptions erronées, à savoir le biais d'équiprobabilité (Lecoutre, 1985). De plus, le manque d'explicitation des hypothèses de travail peut renforcer une confusion entre ce qui relève de la réalité et ce qui relève du modèle.

Les hypothèses n'étant jamais entièrement explicitées, la question de la validité de ces modèles l'est encore moins. Aucune réflexion de cet ordre n'est présente dans les documents analysés, bien que certains exemples de tâches aient au départ une problématique réelle dans les documents Ressources du cycle 4. Il semble que les éléments présents dans le document Ressource *Modéliser* (MENESR, 2016b) du cycle 4, comme la distinction entre réalité et modèle et le processus complet de modélisation, ne sont pas incarnés dans les autres documents analysés.

De plus, au-delà de l'implication de la réalité, entre monde pseudo-concret et monde mathématique, peu d'allers-retours existent, le passage du modèle pseudo-concret au modèle mathématique est souvent pris en charge dans les énoncés (ou implicite par exemple au DNB) et il n'y a pratiquement jamais de retour dans le monde pseudo-concret avec une interprétation des résultats mathématiques.

Conclusion

La modélisation fait donc partie intégrante de l'enseignement des probabilités, tout du moins dans l'enseignement secondaire. Les extraits du document Ressource *Modéliser* étudiés peuvent laisser penser que l'institution a conscience des enjeux de la modélisation en probabilités et souhaite mettre en pratique, dans les classes, la démarche de modélisation. Cependant, à travers l'analyse des programmes du collège et du lycée, des différents documents Ressources ainsi que des sujets de DNB et baccalauréat, nous avons montré globalement la place peu visible du processus complet de modélisation, et notamment l'on retrouve beaucoup d'ambiguïtés et de confusions dans les situations étudiées avec une distinction entre réalité et modèle souvent peu lisible, l'absence de description explicite des modèles pseudo-concrets étudiés… Girard (2001) avait déjà repéré ces confusions dans les classes au début des années 2000, il semblerait qu'elles restent toujours présentes. Il conviendrait de poursuivre cette étude en étudiant les manuels scolaires et les pratiques effectives en classe pour vérifier.

Cet article aura, nous l'espérons, contribué à la clarification de certains points, utiles aux enseignants et aux formateurs, concernant la modélisation en probabilités. Le cycle de modélisation utilisé ici peut aussi servir de grille d'analyse des situations données en classe, en formation des enseignants.

Nous concluons en proposant quelques pistes pour l'enseignement des probabilités, et plus particulièrement de la modélisation en probabilités :

- proposer des situations réelles avec une vraie problématique, une question claire, pour laquelle une modélisation probabiliste pourra apporter une (voire des) réponse(s) sous un certain angle (par exemple, comme dans le document Ressource de cycle 4, avec la recherche d'une stratégie gagnante) et reformuler cette question au fur et à mesure du processus ;

- explicitér en classe le protocole expérimental associé à l'expérience aléatoire étudiée, en précisant les hypothèses de travail et de modèle, notamment aux débuts de l'enseignement des probabilités (Parzysz, 2009), et faire prendre conscience que d'autres choix auraient pu être faits et auraient pu amener à d'autres modèles ;
- travailler des problèmes de modélisation « complexes » (notamment plus complexes que le lancer d'un dé qui est peut-être assez intuitif pour les élèves) pour lesquels l'explicitation des hypothèses deviendra indispensable ;
- comparer des modèles, via la simulation informatique, pour voir s'ils sont « relativement proches ». Par exemple, cela peut permettre de visualiser qu'il est raisonnable de considérer un modèle pseudo-concret de tirage avec remise plutôt que sans remise ;
- Questionner la pertinence d'un modèle choisi, notamment en confrontant les simulations informatiques de ce modèle (ou la loi de probabilité considérée) avec les données réelles.

Le premier point insiste sur la problématique, la question réelle initiale, qui va guider le travail de modélisation du début à la fin du processus. La question initiale n'est pas « quelle est la probabilité ? » (problème mathématique), mais bien un vrai problème réel, par exemple, « quelle est la stratégie gagnante ? ».

Pour revenir au deuxième point, l'idée n'est pas d'expliciter les hypothèses de travail des situations classiques à chaque fois. Par exemple, pour le lancer de dé : une fois que la classe s'est mise d'accord, il suffirait de se dire « maintenant quand on parlera du lancer de dé, on se placera dans ce modèle (avec telles hypothèses sous-jacentes) ». Le modèle du lancer de dé deviendra alors pour tous le modèle présenté dans le tableau 1, sans ambiguïté. Il pourra alors avoir le statut de modèle (pseudo-concret) de référence comme indiqué dans le programme de seconde. L'important est d'avoir conscience que d'autres choix auraient pu être fait. Il convient aussi de bien préciser ce qui est entendu par « au hasard » suivant les situations travaillées, et ne pas l'associer exclusivement à l'équiprobabilité pour ne pas renforcer le biais d'équiprobabilité. Ces pratiques permettraient de lever des ambiguïtés et des confusions entre réalité et modèle, pouvant être présentes dans les classes (Girard, 2001) afin que les élèves aient bien conscience que les situations étudiées en mathématiques ne relèvent plus de la réalité. Des travaux internationaux en didactique des mathématiques sur la modélisation (en général) ont aussi mis en évidence que le processus de modélisation n'était pas suffisamment explicité dans l'enseignement (Blum & Borromeo Ferri, 2009), ce qui pouvait conduire à des difficultés pour les élèves.

Enfin, la modélisation fait intervenir d'autres domaines que les mathématiques en lien avec la situation réelle, on peut alors se demander si ces connaissances utiles à la modélisation doivent être enseignées dans la classe de mathématiques. À l'interface entre plusieurs sciences, l'enseignement scientifique en première et terminale est un cadre privilégié *a priori* pour aborder la modélisation et notamment la modélisation probabiliste. Il nous semble qu'au-delà de cet enseignement spécifique, certaines situations réelles ne demandent pas de connaissances trop poussées et peuvent permettre un travail de modélisation (même s'il pourrait paraître naïf pour des experts) avec une réflexion complète même sur la validité, en restant à l'échelle des connaissances des élèves (par exemple le problème du volcan Aso abordable en mathématiques complémentaires dans le cadre des nouveaux programmes, présenté dans Derouet et Alory (2018)).

L'idée de notre propos n'est pas de travailler exclusivement la globalité du processus de modélisation en classe dans l'enseignement et notamment la construction de modèle : bien entendu un travail dans un modèle donné est indispensable et légitime. Il s'agit seulement de ne pas se limiter à cela.

Henry (1999) affirme que *« d'un point de vue didactique, l'enjeu est de faire de la modélisation en probabilités un objectif d'enseignement et d'apprentissage fondé sur la pratique des élèves »* (pp. 33-34). La présence des probabilités dans les programmes n'a fait qu'augmenter et avancer dans la scolarité depuis 1999, il nous semble que — plus que jamais — un enseignement des probabilités centré sur la modélisation pourrait être riche pour les élèves, dès le collège, avec une réelle réflexion en classe sur les différents choix et simplifications possibles (dans le monde du pseudo-concret) pour bien faire prendre conscience aux élèves de la distinction entre réalité et mathématiques (ici probabilités). Au-delà des connaissances en probabilités, ce travail sur la modélisation contribuerait, selon nous, à la formation des citoyens de demain.

Références bibliographiques

Bonneval (2002). Test d'équirépartition : qui a dit khi-deux ? *Bulletin de l'APMEP, 441*, 512-521.
https://www.apmep.fr/Test-d-equirepartition-qui-a-dit

Blum, W. & Borromeo Ferri, R. (2009). Mathematical Modelling: Can It Be Taught And Learnt? *Journal of Mathematical Modelling and Application, 1*(1), 45-58.

Blum, W. & Leiß, D. (2007). How do students and teachers deal with modelling problems? The example "Sugarloaf" and the DISUM Project. Dans C. Haines, P.L. Galbraith, W. Blum & S. Khan (dir.), *Mathematical modelling (ICTMA12) - Education, engineering and economics* (pp. 222-231). Horwood.
https://doi.org/10.1533/9780857099419.5.221

Chaput, B., Girard, J. C. & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. Dans C. Batanero, G. Burrill & C. Reading (dir.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 85-95). Springer.

Coulange, L. (1998). Les problèmes « concrets » à « mettre en équation » dans l'enseignement. *Petit x, 47*, 33-58.

Derouet, C. (2022). Caractérisation de démarches de modélisation probabiliste. *Annales de didactique et de sciences cognitives, 27*, 89-131.
https://journals.openedition.org/adsc/1349

Derouet, C. & Alory, S. (2018). Une séquence d'enseignement articulant les lois de probabilité à densité et le calcul intégral en terminale S. *Repères-IREM, 113*, 45-80.

Dogme, Y. (1993). *Statistique. Dictionnaire encyclopédique*. Dunod.

Girard, J.-C. (2001). Un exemple de confusion modèle-réalité. Dans M. Henry (dir.), *Autour de la modélisation en probabilités* (pp. 145-148). Presses Universitaires de Franche-Comté.

Girard, J.-C. (2003). Modélisation et simulation. *Probabilités au lycée*. APMEP.

Hacking, I. & Dufour, M. (2004). *L'ouverture au probable. Eléments de logique inductive*. Armand Colin.

Henry, M. (1997). Qu'est-ce qu'une expérience aléatoire ? Modélisation en probabilités. Introduction. Dans Commission Inter-IREM Statistique et Probabilités (dir.), *Enseigner les probabilités au lycée. Ouvertures statistiques, enjeux épistémologiques, questions didactiques et idées d'activités* (pp. 55-56). IREM de Reims.

Henry, M. (1999). L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie. *Repères-IREM, 36*, 15-34.

Henry, M. (dir.). (2001a). *Autour de la modélisation en probabilités*. Presses Universitaires de Franche-Comté.

Henry, M. (2001b). Modélisation d'une situation aléatoire. Notion de modèle et modélisation dans l'enseignement. Dans M. Henry (dir.), *Autour de la modélisation en probabilités* (pp. 149-159). Presses Universitaires de Franche-Comté.

Henry, M. (2001c). Notion d'expérience aléatoire. Vocabulaire et modèle probabiliste. Dans M. Henry (Dir.), *Autour de la modélisation en probabilités* (pp. 161-171). Presses Universitaires de Franche-Comté.

Henry, M. (2001d). Notion de modèles et modélisation dans l'enseignement. Dans M. Henry (dir.), *Autour de la modélisation en probabilités* (pp. 149-159). Presses Universitaires de Franche-Comté.

Henry, M. (2003). Des lois de probabilité continues en terminale S, pourquoi et pour quoi faire ? *Repères-IREM, 51*, 5-25.

Kaiser, G. (1995). Realitätsbezüge im Mathematikunterricht - Ein Überblick über die aktuelle und historische Diskussion. Dans G. Graumann, T. Jahnke, G. Kaiser & J. Meyer (dir.), *Materialien für einen realitätsbezogenen Mathematikunterricht* (pp. 66-84). Franzbecker.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N.J. & Kazak, S. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning, 13*(1), 68-86.

Israël, G. (1996). *La mathématisation du réel : essai sur la modélisation mathématique*. Éditions du Seuil.

Lecoutre, M.-P. (1985). Effet d'informations de nature combinatoire et de nature fréquentielle sur les jugements probabilistes. *Recherche en Didactique des Mathématiques, 6*, 193-213.

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. *Annales de Didactique et de Sciences Cognitives, 16*, 127-147.

Ross, S. M. (2007). *Initiation aux Probabilités*. Presses polytechniques et universitaires romandes.

Thibault, M. & Martin, V. (2018). Confusion autour du concept de probabilité. *For the Learning of Mathematics, 38*(1), 12-16.

Ventsel, H. (1973). *Théorie des probabilités*. Mir.

Yvain-Prébiski, S. (2018). *Étude de la transposition à la classe de pratiques de chercheurs en modélisation mathématique dans les sciences du vivant. Analyse des conditions de la dévolution de la mathématisation horizontale aux élèves* [Thèse de doctorat, Université de Montpellier, Montpellier].
https://tel.archives-ouvertes.fr/tel-01956661v1/document

Programmes et documents Ressources

MENESR (2016a). *Comprendre et utiliser des notions élémentaires de probabilités*. Éduscol.
https://eduscol.education.fr/document/17275/download

MENESR (2016b). *Modéliser, cycle 4*. Éduscol.
https://eduscol.education.fr/document/17218/download

MENJ/DGESCO (2016). *Ressources pour le collège. Probabilités*. Éduscol.

MENJ (2019a). Cycle 4. *Mathématiques. Attendus de fin d'année. 3e*. Éduscol.
https://eduscol.education.fr/document/14068/download

MENJ (2019b). *Cycle 4. Mathématiques. Attendus de fin d'année. 4e*. Éduscol.
https://eduscol.education.fr/document/14056/download

MENJ (2019c). *Cycle 4. Mathématiques. Attendus de fin d'année. 5e*. Éduscol.
https://eduscol.education.fr/document/14044/download

MENJ (2019d). *Cycle 4. Mathématiques. Repères annuels de progression*. Éduscol.
https://eduscol.education.fr/document/14080/download?attachment

MENJ (2019e). Programme de mathématiques de première générale. *BO spécial n° 1 du 22 janvier 2019*.
https://eduscol.education.fr/document/24565/download

MENJ (2019f). Programme de mathématiques de seconde générale et technologique. *BO spécial n° 1 du 22 janvier 2019*.
https://eduscol.education.fr/document/24553/download

MENJ (2019g). Programme de spécialité de mathématiques de terminale générale. *BO spécial n° 8 du 25 juillet 2019*.
https://eduscol.education.fr/document/24568/download

MENJS (2020). Programme d'enseignement du cycle des approfondissements (cycle 4). BOEN n° 31 du 30 juillet 2020.
https://eduscol.education.fr/document/621/download

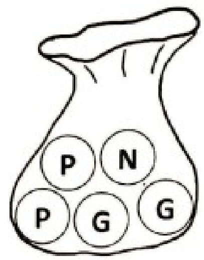
MENJS (2022). *Les guides fondamentaux pour enseigner. La résolution de problèmes mathématiques au collège*. Éduscol.
https://eduscol.education.fr/document/13132/download?attachment

Annexe 1
Exercice de probabilités du DNB Polynésie (juin 2023)

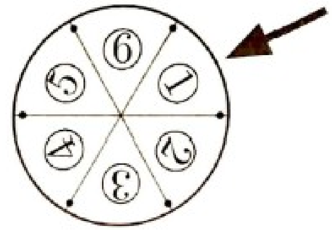
Dans cette exercice, on étudie la probabilité de gain des deux jeux ci-dessous.

Partie A

Jeu 1
Un sac contient cinq boules indiscernables au toucher, dont une portant la lettre N, deux, portant la lettre G et deux portant la lettre P.



Jeu 2
Une roue à six secteurs angulaires identiques numérotées de un à six.



1. On considère le jeu 1.
   On pioche une boule au hasard dans ce sac et on note la lettre inscrite sur la boule choisie.
   On considère qu'on a gagné si on pioche la lettre G.
   Montrer que la probabilité de gagner avec ce jeu et de $\frac{2}{5}$.
2. On considère le jeu 2.
   On fait tourner la roue et on note le nombre d'inscrits sur le secteur pointé par la flèche.
   On considère qu'on a gagné si on s'arrête sur un nombre premier.
   Quelle est la probabilité de gagner à ce jeu?
3. **a.** Quel est le jeu qui présente la plus faible probabilité de gagner?
   b. Proposer une liste de boules à rajouter pour que la probabilité de gagner avec le jeu 1 soit de $\frac{1}{4}$.

Partie B

Dans cette partie, toute trace de recherche sera valorisée.

On choisit finalement de combiner ces deux jeux.
Dans un premier temps, le joueur doit tirer une boule dans le sac du jeu 1.
On doit ensuite faire tourner la roue du jeu 2.
Le joueur gagne un lot s'il a tiré une boule portant la lettre G et si la roue s'arrête sur un secteur angulaire dont le numéro est un nombre premier.
Quelle est la probabilité de gagner à cette combinaison des deux jeux?

Annexe 2

Exercice de probabilités du baccalauréat spécialité mathématiques Amérique du Nord (sujet 2, mars 2023)

Dans un souci d'améliorer sa politique en matière de développement durable, une entreprise a réalisé une enquête statistique sur sa production de déchets.
Dans cette enquête, les déchets sont classés en trois catégories :

— 69 % des déchets sont minéraux et non dangereux;

— 28 % des déchets sont non minéraux et non dangereux;

— les déchets restants sont des déchets dangereux.

Cette enquête statistique nous apprend également que :

— 73 % des déchets minéraux et non dangereux sont recyclables;

— 49 % des déchets non minéraux et non dangereux sont recyclables;

— 35 % des déchets dangereux sont recyclables.

Les parties A *et* B *sont indépendantes et peuvent être traitées séparément.*

Partie A

Dans cette entreprise, on prélève au hasard un déchet. On considère les évènements suivants :

— M : « Le déchet prélevé est minéral et non dangereux »;

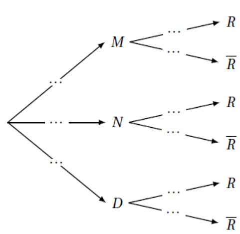
— N : « Le déchet prélevé est non minéral et non dangereux »;

— D : « Le déchet prélevé est dangereux »;

— R : « Le déchet prélevé est recyclable ».

On note $\overline{R}$ l'évènement contraire de l'évènement R.

1. Recopier et compléter l'arbre pondéré ci-dessous représentant la situation de l'énoncé.



2. Justifier que la probabilité que le déchet soit dangereux et recyclable est égale à 0,010 5.

3. Déterminer la probabilité $P\left(M \cap \overline{R}\right)$ et interpréter la réponse obtenue dans le contexte de l'exercice.

4. Démontrer que la probabilité de l'évènement R est $P(R) = 0{,}651\,4$.

5. On suppose que le déchet prélevé est recyclable. Déterminer la probabilité que ce déchet soit non minéral et non dangereux. ***On donnera la valeur arrondie au dix-millième.***

Partie B

On rappelle que la probabilité qu'un déchet prélevé au hasard soit recyclable est égale à 0,651 4.

1. Afin de contrôler la qualité de la collecte dans l'entreprise, on prélève un échantillon de 20 déchets pris au hasard dans la production. On suppose que le stock est suffisamment important pour assimiler le prélèvement de cet échantillon à un tirage avec remise.
On désigne par X la variable aléatoire égale au nombre de déchets recyclables dans cet échantillon.

a. On admet que la variable aléatoire X suit une loi binomiale. Préciser ses paramètres.

b. Donner la probabilité que l'échantillon contienne exactement 14 déchets recyclables. ***On donnera la valeur arrondie au dix-millième.***

2. Dans cette question, on prélève désormais n déchets, où n désigne un entier naturel strictement positif.

a. Donner l'expression en fonction de n de la probabilité p_n qu'aucun déchet de cet échantillon ne soit recyclable.

b. Déterminer la valeur de l'entier naturel n à partir de laquelle la probabilité qu'au moins un déchet du prélèvement soit recyclable est supérieure ou égale à 0,999 9.