CHARLOTTE DEROUET

CARACTERISATION DE DEMARCHES DE MODELISATION PROBABILISTE

Abstract. Characterisation of probabilistic modelling approaches. The modelling process is an essential part of any probabilistic approach. In this article, we propose three different categories of modelling approaches involving a probabilistic model that can be encountered in French secondary education: one based on the Laplacian approach to probability, another on the frequentist approach and the last one based on informal statistical inference. Based on the work of modelling of the German current, the objective of this article is to characterise the three categories of approaches and to identify what place and role statistics plays in these categories. The theoretical reflection is illustrated by an *a priori* analysis of examples of modelling problems. The characterisation is based on the different stages of the chosen modelling cycle and the associated working and model assumptions.

Keywords. Modelling, probability, statistics, Laplacian and frequentist approaches, informal statistical inference.

Résumé. La modélisation est un élément essentiel de toute démarche probabiliste. Dans cet article, nous proposons trois catégories différentes de démarches de modélisation faisant intervenir un modèle probabiliste pouvant être rencontrées dans l'enseignement secondaire français : l'une prenant appui sur l'approche laplacienne des probabilités, une autre sur l'approche fréquentiste et enfin la dernière relevant de l'inférence statistique informelle. S'appuyant sur les travaux autour de la modélisation du courant allemand, l'objectif est de caractériser les trois catégories de démarches et d'identifier quelle place et quel rôle joue la statistique dans ces catégories. La réflexion théorique est illustrée par une analyse *a priori* d'exemples de problèmes de modélisation. La caractérisation se fait en relation avec les différentes étapes du cycle de modélisation choisi et les hypothèses de travail et de modèle associées.

Mots-clés. Modélisation, probabilités, statistique, approches laplacienne et fréquentiste, inférence statistique informelle.

Les travaux de recherche en didactique des mathématiques sur l'activité de modélisation dans l'enseignement des mathématiques sont nombreux, notamment dans le cadre international. Le groupe d'étude pour la modélisation et les applications mathématiques ICTMA[1], affilié à la commission internationale de l'enseignement

[1] International Community of Teachers of Mathematical Modelling and Applications.

ANNALES de DIDACTIQUE et de SCIENCES COGNITIVES, volume 27, p. 89 – 131.

mathématique (ICMI[2]), travaille depuis 1983 sur des questions en lien avec l'intégration de la modélisation et d'applications mathématiques dans les classes, de l'école primaire à l'université (par exemple Blum et al., 2007 ; Kaiser et al., 2011 ; Leung et al., 2021). En France, nous pouvons citer plusieurs travaux de recherche sur la modélisation mathématique en tant qu'objet d'apprentissage : la thèse de Rodriguez (2007) porte sur la démarche de modélisation en physique et en mathématiques et la récente thèse de Yvain-Prébiski (2018) sur le travail de mathématisation horizontale (Treffers, 1986) nécessaire pour envisager un traitement mathématique d'une situation ancrée dans le réel et de sa possible transposition dans les classes de collège et de lycée, dans le contexte des sciences du vivant. Des travaux ont aussi été menés autour des pratiques de modélisation et de la démarche de modélisation à l'école primaire (Cabassut et Wagner, 2011 ; Wozniak, 2012 ; Adjiage et Rauscher, 2013).

Dans cet article, nous nous focalisons sur le domaine des probabilités dans une perspective de modélisation probabiliste. Nous partageons le point de vue de Henry (1999) selon lequel « d'un point de vue didactique […] en probabilités, peut-être plus que dans d'autres domaines, les situations de modélisation sont essentielles » (Henry, 1999, p. 25). Les probabilités fournissent des outils pour modéliser des systèmes réels afin de comprendre et faire des prédictions sur ces modèles (par exemple, Henry, 2001a ; Chaput et al., 2011 ; Greer et Mukhopadhyay, 2005 ; Konold et Kazak, 2008). Des travaux internationaux ont montré l'importance de la modélisation probabiliste pour amener les étudiants vers une meilleure acculturation de la pensée probabiliste (Pratt, 2011). Pfannkuch et al. (2016) ont étudié les pratiques de sept professionnels utilisant des modèles probabilistes dans différents domaines (file d'attente/réseaux, écologie, probabilités, commerce, hydroélectricité, agriculture, gestion d'opérations) et ont dégagé des éléments pour caractériser la modélisation probabiliste et la pensée probabiliste. Ils concluent en disant que, pour développer au mieux la pensée probabiliste, il est important de trouver un équilibre entre l'utilisation de modèles théoriques, la construction de modèles empiriques et l'exploration des comportements des modèles (p. 34). Le potentiel de problèmes de modélisation pour construire de nouvelles notions probabilistes comme la fonction de densité a été illustré dans nos travaux (Derouet, 2019).

L'objectif de cet article est de caractériser différentes catégories de démarches de modélisation faisant intervenir des modèles probabilistes pouvant être rencontrées dans une classe de mathématiques de collège ou de lycée en France.

Dans une première partie, nous présenterons nos points d'appui théoriques concernant la modélisation mathématique, que nous préciserons ensuite de manière spécifique à la modélisation probabiliste. À l'issue de cette partie, nous formulerons

[2] International Commission on Mathematical Instruction.

nos questions de recherche. Les trois parties suivantes illustreront notre réflexion théorique sur la modélisation probabiliste. Puis nous dégagerons des éléments liés à la place de la statistique dans les différentes démarches. Enfin, nous conclurons cet article par une discussion et des perspectives.

1. Modélisation et probabilités

Dans cette partie, nous présentons les éléments théoriques sur lesquels s'appuie notre réflexion. Après avoir défini la modélisation mathématique et retenu un cycle de modélisation, nous précisons des éléments spécifiques à la modélisation probabiliste.

1.1. Modélisation mathématique

Dans cet article, nous définissons la modélisation mathématique comme « une démarche de construction d'un modèle en langage mathématique[3] permettant de mettre en relation les éléments choisis d'un fragment de réalité en lien avec la question à étudier » (Yvain-Prébiski, 2018). Par modèle, nous entendons « 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 » (Henry, 2001b, p. 151). Cette définition distingue le modèle en tant que structure abstraite de la symbolique utilisée pour le décrire (Henry, 2001b). Comme dans les travaux autour de la modélisation du courant allemand (par exemple Kaiser (1995)), nous considérons les problèmes de modélisation comme des problèmes complexes et ouverts issus du monde réel.

1.2. Cycle de modélisation retenu

Comme l'a montré Yvain-Prébiski (2018) dans son travail de thèse, différents cycles décrivant le processus de modélisation se retrouvent dans la littérature, détaillant plus ou moins les différentes étapes constitutives du processus suivant les objectifs d'apprentissage visés dans les recherches. Dans cet article, nous prenons comme point de départ le cycle de modélisation (figure 1) proposé par Blum et Leiss (2007), qui complète le cycle de modélisation de Kaiser (1995). Ce cycle est, selon Hankeln et Hersant (2020), « reconnu comme le plus pertinent pour l'analyse cognitive. Il tient compte de travaux sur la modélisation en mathématiques appliquées (Pollak, 1979 ; Burghes, 1986), en linguistique (Kintsch et Greeno, 1985) et en psychologie cognitive (Staub et Reusser, 1995) » (p. 41).

[3] À comprendre : « exprimé dans un langage des mathématiques ».

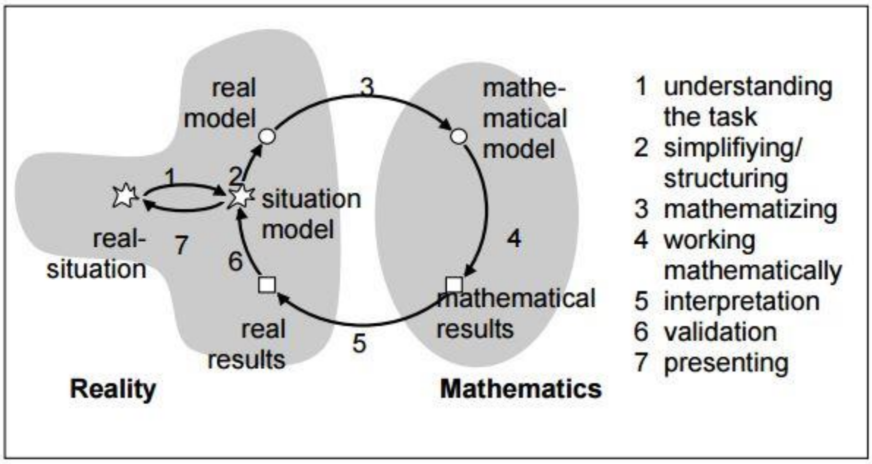

Figure 1. Le cycle de modélisation proposé par Blum et Leiss (2007)

Le point de départ du processus est une situation du monde réel, que le modélisateur (souvent l'élève dans notre contexte) lit et « comprend », qui va amener à la construction d'une représentation mentale de la situation réelle (Borromeo Ferri, 2006), appelé *situation model* que l'on pourrait traduire par modèle de situation ou encore situation générique (qui s'apparente à une situation modèle). La situation de départ, bien qu'issue du monde réel, comporte déjà des simplifications, car nous n'avons pas accès à l'intégralité des informations relatives à cette situation. Le réel est trop complexe et ne peut être appréhendé dans toute sa complexité. Un individu n'a finalement accès qu'à une partie du réel, sa réalité. Ensuite au niveau du modèle de situation (ou situation générique), la situation a été simplifiée pour ne garder que ce qui est important pour le modélisateur, au regard de la question qu'il se pose, d'où le caractère générique de la situation décrite à ce stade. Les simplifications sont plus ou moins automatiques, mais sont le fait de choix conscients ou non du modélisateur. La situation est ensuite idéalisée (2), simplifiée ou structurée pour obtenir un *real model* (Kaiser, 1995) traduit par modèle réel. Le modèle réel est mathématisé (3), c'est-à-dire traduit en langage mathématique pour obtenir un modèle mathématique de la situation de départ. Un traitement mathématique (4) entraîne des résultats mathématiques, interprétés en « résultats réels » (5) qui seront validés ou non par rapport au modèle de situation (6). Puis, si les résultats obtenus semblent cohérents, ils seront présentés comme prédictions sur la situation réelle (7).

En s'appuyant sur les travaux de Chevallard (1989) et Henry (1999), Coulange (1998)[4] propose, dans la démarche de modélisation, de séparer le domaine réel (extramathématique) du domaine mathématique par un troisième domaine, le domaine « pseudo-concret » :

> 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 domaine pseudo-concret et le modèle pseudo-concret qui nous paraissent plus appropriés que le terme *real model* dans les cycles de modélisation proposés par Kaiser (1995) ou Blum et Leiss (2007). Cela nous amène à considérer le cycle de modélisation en figure 2.

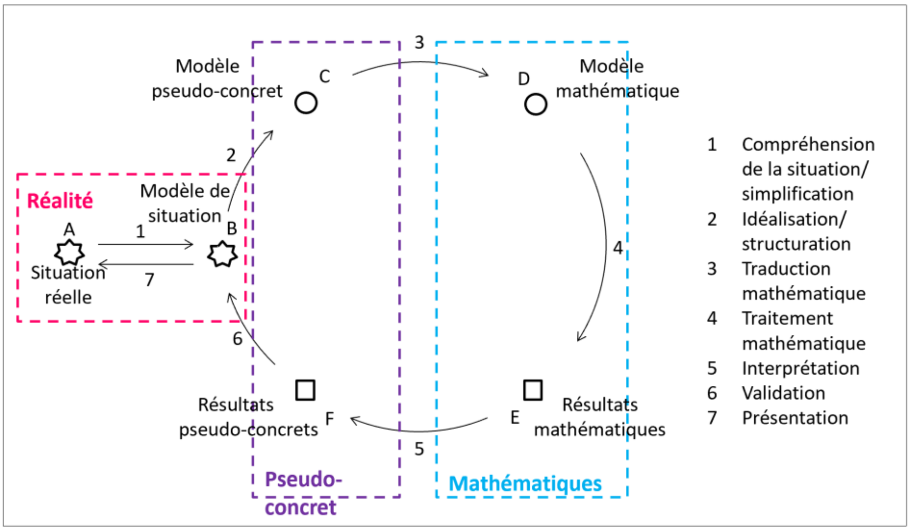

Figure 2. Cycle de modélisation retenu

[4] Dans son article, Coulange (1998) s'appuie sur les propos tenus par Henry lors d'une présentation au séminaire Didatech du 5 février 1997.

Pour préciser la distinction entre modèle pseudo-concret et modèle mathématique, nous reprenons Henry (2001b ; 2010) :

> [Le] modèle peut être représenté dans différents systèmes de signes : images, schémas, langages ou symbolismes, s'inscrivant dans différents registres de représentations, plus ou moins isomorphes.
>
> Par exemple, on peut représenter un modèle par une analogie, en y introduisant des objets idéalisés de la réalité. Cela veut dire que dans un vocabulaire courant, les objets du modèle sont doués de propriétés caractéristiques bien définies. Nous parlerons alors de modèles « pseudo-concrets ». C'est le cas notamment des modèles d'urnes en probabilités, où l'hypothèse implicite est l'équiprobabilité des boules dans un tirage « au hasard ».
>
> Parmi les différents registres de représentations, le langage et le symbolisme mathématiques permettent des descriptions puissantes sur lesquelles peuvent opérer des propriétés et des algorithmes généraux. Nous les appellerons « modèles mathématiques ». Souvent, ils nous sont tellement familiers que nous n'en voyons pas d'autres, et nous avons tendance à confondre ces représentations avec les objets idéaux en jeu, lesquels sont souvent confondus avec la réalité qu'ils modélisent. (Henry, 2001b, p. 152)

Il précise et illustre :

> Un modèle peut être présenté dans un vocabulaire courant renvoyant à des objets réels, mais qui dans le modèle sont doués de propriétés caractéristiques idéales et abstraites que j'appelle modèle pseudo-concret (Henry, 2001[b]). Ainsi une urne de Bernoulli est une abstraction d'une urne réelle, dans laquelle les boules sont de deux couleurs, mais pour le reste parfaitement identiques. Elles sont supposées rigoureusement équiprobables (hypothèse de modèle) dans un tirage « au hasard ». (Henry, 2010, p. 42-43)

Dans la suite, nous nous focaliserons sur une partie de la démarche de modélisation allant de la situation réelle, et de la question qui se pose, au modèle mathématique avec la reformulation mathématique de la question, puis au traitement mathématique de la question, sans nous intéresser au retour à la réalité[5]. Ces étapes de la démarche de modélisation renvoient aussi aux questions relatives à la mathématisation horizontale et à la mathématisation verticale que Yvain-Prébiski (2018) définit, en prenant appui sur Treffers (1986) et Freudenthal (1991), comme suit :

---

[5] Le retour à la réalité *via* la validation c'est-à-dire « l'évaluation du degré d'approximation des résultats théoriques obtenus avec les valeurs expérimentales correspondantes et la décision que le modèle est ou n'est pas bien adapté à la situation étudiée » (Henry, 2001b, p. 157) est une étape délicate qui nécessite des connaissances spécialisées des phénomènes étudiées (Henry, 2001b) d'où le fait qu'elle soit peu (voire pas du tout) rencontrée en classe.

> *La mathématisation horizontale* relève du choix d'un fragment de réalité, de l'identification et du choix de certains aspects de ce fragment de réalité susceptibles de relever d'un traitement mathématique puis de leur mise en relation en vue de construire un modèle mathématique. (p. 79)
>
> *La mathématisation verticale* relève du travail mathématique à l'intérieur « *du monde des symboles* », c'est-à-dire du traitement mathématique d'un problème mathématique. (p. 80)

Cependant, dans cet article, nous n'aborderons pas le processus de modélisation selon cette approche.

1.3. Modélisation probabiliste

Notre étude s'intéresse à la modélisation, mais plus particulièrement au cas où interviennent des modèles probabilistes. Selon Biehler (1994) :

> A major point is that the ontological debate of whether something 'is' deterministic or not may not be useful, rather, a situation can be described with deterministic and with probabilistic models and one has to decide what will be more adequate for a certain purpose. (p. 4)

La question est alors de savoir quand il est plus pertinent de choisir un modèle probabiliste plutôt que déterministe suivant l'objectif. Quand la situation réelle de départ s'appuie sur une expérience réelle dont :

- plusieurs résultats de cette expérience sont possibles,
- on ne peut ni prévoir *a priori*, ni calculer le résultat, notamment car celui-ci est très sensible aux conditions initiales,

on va alors préférer associer un degré de certitude que l'on peut avoir dans la réalisation des différents résultats possibles (Girard, 2001 ; Henry, 2003). Dans ce cas, on adopte une modélisation probabiliste (Pratt, 2011, p. 892) mettant en jeu tout d'abord une expérience aléatoire mathématique c'est-à-dire « un modèle d'expérience aléatoire » (Girard, 2001, p. 141).

1.3.1 Expérience aléatoire

Dans le cas de la modélisation probabiliste, le passage de la réalité aux domaines pseudo-concret puis mathématique repose immanquablement sur l'identification de l'expérience aléatoire. Pour Henry (2001c), la notion d'expérience aléatoire désigne « un processus réel de nature expérimentale, où le hasard intervient, avec des issues possibles bien identifiées » (p. 163). Il poursuit :

> On ne sépare donc pas la description d'une expérience aléatoire de la donnée de ces issues possibles que l'on va considérer. Ainsi, dans la nature, il n'y a pas d'expériences aléatoires, il y a des situations complexes ou des systèmes évolutifs (par

> exemple le jet d'une pièce de monnaie, mais aussi la situation météorologique, etc.) qui dépendent de manière sensible des conditions initiales, de telle sorte qu'on ne peut pas déterminer leurs évolutions, quels que soient les instruments d'observation et la puissance des outils de calcul. (Henry, 2001c, p. 163)

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). Il précise ensuite qu'« 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 » (p. 164-165). Il distingue alors l'expérience « concrète », qui relève de la réalité, et 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 (Henry, 2001c, p. 166). Une expérience aléatoire peut être remplacée par une autre qui la représente génériquement, faisant intervenir des objets pseudo-concrets usuels comme la pièce de monnaie, le dé, les cartes à jouer… Ce passage au modèle pseudo-concret par le biais de choix à opérer sur le réel est guidé par un regard théorique, car il suppose des connaissances préalables (Henry, 2001c, p. 166). Ces choix opérés pour déboucher sur un modèle pseudo-concret correspondent à ce qu'Henry (2003) appelle hypothèses de travail.

1.3.2 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, permettent de définir l'expérience aléatoire 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), 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…[6]

1.3.3 Détermination des probabilités élémentaires

Dans le modèle mathématique, 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. Selon Henry (2001c), « Il reste à préciser les conditions dans lesquelles les probabilités élémentaires sont déterminées. Elles relèvent d'hypothèses de modèle. Dans la pratique, on trouve trois catégories d'hypothèses » (p. 168). Le premier cas cité est le suivant :

> a- Le contexte de l'expérience aléatoire, les symétries des objets matériels utilisés, permettent de ramener tous les résultats à un ensemble n d'issues[7] qui sont considérées comme également possibles. On fait alors l'**hypothèse d'équiprobabilité**[8]. […] (ibid.)

Ce premier cas relève de l'approche laplacienne (ou classique) des probabilités, basée sur le premier principe de Laplace, qui postule une égalité de probabilité des événements élémentaires et qui consiste à définir la probabilité d'un événement associé à diverses issues comme le rapport du nombre de celles qui le produisent à leur nombre total[9] (Parzysz, 2011). Le modèle est créé *a priori*, sans nécessité de recourir à l'expérience réelle.

Le deuxième cas présenté par Henry (2001c) est le suivant :

> b- La complexité de l'expérience aléatoire ne permet pas de se ramener à un système d'issues équiprobables […]. On peut alors 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. (p. 168-169)

Ce deuxième cas relève de l'approche fréquentiste des probabilités. Cette dernière 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 :

---

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

[7] Selon nous, lire : « de n issues ».

[8] Les mots en gras l'étaient dans le texte original.

[9] Cette probabilité se retrouve en faisant la somme des probabilités élémentaires.

– 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*. Les tests statistiques ne sont cependant pas dans les programmes de mathématiques de l'enseignement secondaire français.

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

Selon Henry (2001c), les fréquences observées « sont alors conçues comme mesures approximatives de cette probabilité, comme une mesure physique, avec le degré de précision souhaité suivant le nombre d'expériences que l'on consent d'effectuer » (p. 169). Cependant, selon nous, concevoir la fréquence observée comme une approximation de la valeur de la probabilité n'est pas de même nature que la mesure physique d'une grandeur spécifique. En effet, un objet physique a une masse (théorique) unique, on peut mesurer sa masse (grandeur mesurée), ce qui va donner une mesure physique avec un degré de précision. Un autre mesurage pourra donner une autre mesure. En revanche, la question de savoir si une punaise porte « en elle » la probabilité de tomber la pointe vers le haut, comme elle a une unique masse, est moins consensuelle. La probabilité appartient au modèle mathématique et non à la réalité.

Enfin, le troisième cas précisé par Henry (2001c) est le suivant :

> c- On reconnaît dans le processus expérimental une situation relevant d'un modèle standard. Dans ce modèle, les probabilités des issues sont réparties selon une **loi** connue qui dépend d'un ou plusieurs **paramètres**[10]. On **estime** alors ces paramètres à partir d'un **échantillon** de résultats obtenus en répétant l'expérience (estimation inférentielle), ce qui détermine entièrement la loi que l'on accepte alors comme **hypothèse de modèle**. Les probabilités élémentaires sont alors calculées à partir des données ou des propriétés mathématiques connues du modèle. (p. 169)

[10] Les paramètres sont quant à eux inconnus.

Dans ce troisième cas, l'approche est différente. En effet, il s'agit de créer un modèle[11] à partir d'un échantillon donné, de taille fixe, qui peut être plus ou moins grand, mais non contrôlable par le solveur. Le modèle peut être connu du solveur avec les paramètres inconnus à estimer (Henry, 2001c), mais il peut aussi être envisageable que le modèle soit entièrement à construire par le solveur en appui sur les propriétés mathématiques qu'il doit vérifier.

Dans ce cas, il s'agit d'une méthode de statistique inférentielle. Les méthodes d'inférence statistique permettent de tirer des conclusions sur des populations ou des processus à partir d'un échantillon de données (Zieffler et al., 2008). Zieffler et al. (2008) rappellent deux définitions de la statistique inférentielle :

> David Moore (2004) describes statistical inference as moving beyond the data in hand to draw conclusions about some wider universe, taking into account that variation is everywhere and the conclusions are therefore uncertain. Garfield and Ben-Zvi (2008) define statistical inference further by differentiating two important themes in statistical inference, parameter estimation and hypothesis testing, and two kinds of inference questions, generalization (from samples) and comparison and determination of cause (from randomized comparative experiments). In general terms, the first theme is concerned with generalizing from a small sample to a larger population, whereas the second involves determining whether a pattern in the data can be attributed to a real effect. (p. 41)

Dans notre cas, il s'agit du premier thème concernant la généralisation d'un petit échantillon à une (plus large) population.

Les concepts de statistique inférentielle, notamment d'estimation statistique, ne sont pas des attendus des programmes de l'enseignement secondaire français, cependant des méthodes informelles sont possibles. Au niveau international, des chercheurs travaillant sur l'enseignement de la statistique (*statistics education* en anglais) ont introduit et décrit l'inférence statistique informelle (*informal statistical inference*) (par exemple, Pfannkuch, 2005, 2006 ; Ben-Zvi, 2006 ; Zieffler et al., 2008). Zieffler et al. (2008) résume le cadre du raisonnement inférentiel informel (IIR, en anglais) en trois composantes :

> 1. Making judgments, claims, or predictions about populations based on samples, but not using formal statistical procedures and methods (e.g., p-value, t tests);
>
> 2. Drawing on, utilizing, and integrating prior knowledge (e.g., formal knowledge about foundational concepts, such as distribution or average; informal knowledge about inference such as recognition that a sample may be surprising given a particular claim; use of statistical language), to the extent that this knowledge is available; and

[11] Soit à partir d'un modèle connu dont les paramètres sont inconnus comme présenté dans Henry, 2001, mais il serait aussi envisageable de construire le modèle.

> 3. Articulating evidence-based arguments for judgments, claims, or predictions about populations based on samples.
>
> Note that this definition refers to IIR as a process for making inferences that does not utilize the formal methods of statistical inference described earlier and that may or may not include use of formal statistical concepts or language. (p. 45)

Les trois cas de figure décrits par Henry (2001c) sont à relier à trois approches différentes. Cela nous amène à considérer les trois catégories de démarches de modélisation probabiliste associées, toutes envisageables dans l'enseignement secondaire français[12]. Chacun des cas fait intervenir les probabilités et la statistique, mais avec des approches et/ou des rôles différents.

1.3.4 Relation entre probabilité et fréquence : la loi des grands nombres de Bernoulli

La loi des grands nombres, qui permet de justifier théoriquement la relation entre probabilité et fréquence, est depuis 2019 au programme de spécialité Mathématiques de terminale du lycée français, mais reste non abordable sous ses formes relatives à la théorie de la mesure pour des élèves du lycée. Le théorème de Bernoulli est quant à lui une version qui pourrait être accessible à des élèves, même avant la classe de terminale.

D'une urne de Bernoulli contenant t boules dont r blanches (fertiles) et s noires (stériles), on tire n boules avec remises et on compte le nombre de boules blanches obtenues (schéma binomial). D'après Pichard (2001), Bernoulli, dans son ouvrage *Ars Conjectandi*, formule ainsi son théorème :

> On peut concevoir des expériences en un nombre tel qu'il soit plus vraisemblable d'autant de fois que l'on veut […] que le nombre des observations fertiles soit au nombre de toutes les observations dans un rapport ni plus grand que $(r + 1)/t$, ni plus petit que $(r – 1)/t$. (p. 40)

En écrivant $r/t = p$ et $1/t = \varepsilon$,

> on peut exprimer ce théorème de la façon suivante avec des notations modernes : Soit A une issue possible d'une épreuve, de probabilité p. On répète n fois cette épreuve et on note F_n la fréquence de réalisations de A dans ces n épreuves.
>
> Alors pour toute valeur $1 – \alpha$ entre 0 et 1 et tout $\varepsilon > 0$, on peut trouver n_0 assez grand tel que si $n > n_0$, on a : $P(|F_n – p| < \varepsilon) > 1 – \alpha$ ce qui signifie que la suite de variables aléatoires F_n tend « en probabilité » vers la valeur p de la probabilité. (Pichard, 2001, p. 39)

---

[12] Dans le cadre des nouveaux programmes de mathématiques du lycée français de 2019 (enseignement scientifique et option mathématiques complémentaires, de terminale), l'approche bayésienne serait aussi à considérer, mais nous ne le faisons pas dans cet article.

On peut aussi le traduire ainsi :

Si l'on répète une même expérience aléatoire un nombre n de fois assez grand, alors pour tout $\alpha > 0$ et tout $\varepsilon > 0$ aussi petits que l'on veut, la probabilité que la fréquence observée F_n d'un événement de probabilité p soit comprise entre $p - \varepsilon$ et $p + \varepsilon$ est supérieure à $1 - \alpha$: $P(p - \varepsilon < F_n < p + \varepsilon) > 1 - \alpha$.

$1 - \alpha$ est alors le niveau de confiance de l'estimation de p par F_n. En fin de compte, la suite des fréquences tend probablement vers la probabilité.

La loi des grands nombres aussi peut jouer un rôle dans la démarche de modélisation probabiliste, que nous allons interroger en lien avec la place de la statistique.

2. Objectifs et questions de recherche

L'objectif de cet article est de caractériser les trois catégories de démarches de modélisation probabiliste présentées dans la section 1.3.3, à savoir les démarches :

a) prenant appui sur l'approche laplacienne des probabilités,

b) prenant appui sur l'approche fréquentiste,

c) relevant de l'inférence statistique informelle.

Ces trois démarches peuvent être rencontrées dans l'enseignement secondaire français et seront étudiées dans ce contexte[13].

Nous pouvons formuler les deux questions de recherche suivantes :

- QR1 : Dans le cadre de notre cycle de modélisation, comment caractériser chacune de ces trois catégories de démarches de modélisation probabiliste ?
- QR2 : Quelle place tient la statistique et quel rôle joue-t-elle dans les trois catégories de démarches de modélisation probabiliste considérées ?

La réflexion théorique sera illustrée par une analyse *a priori* d'exemples de problèmes de modélisation « complexes », dans le sens où le modèle mathématique à étudier n'est pas connu à l'avance des élèves. Il ne s'agit donc pas, par exemple, de se limiter à une situation réelle de lancer d'un dé. La caractérisation se fera en s'appuyant sur les différentes étapes du cycle de modélisation et, en particulier, sur les trois domaines (réalité, domaine pseudo-concret, domaine des mathématiques) et les hypothèses de travail et de modèle associées.

[13] Nous rappelons que l'approche bayésienne pourrait être considérée dans le cadre des nouveaux programmes de lycée de 2019.

3. Modélisation probabiliste et approche laplacienne

L'exemple choisi pour illustrer cette première catégorie de démarches de modélisation probabiliste est le problème de la rencontre qui a déjà été étudié dans le cadre d'une ingénierie didactique (Derouet et Alory, 2018 ; Derouet, 2019), sous un autre angle que celui de la modélisation probabiliste. Nous en proposons ici une description légèrement différente pour mettre en évidence toutes les étapes du cycle de modélisation, en partant autant que possible d'une situation réelle.

3.1. Description de la situation réelle (A)

> Deux amis, Karine et Olivier, se donnent rendez-vous au café de la gare de Strasbourg demain entre 7h et 8h.

De cette situation réelle, plusieurs questions peuvent émerger, notamment : Qui arrivera le premier ? Combien de temps va attendre le premier arrivé ?...

Nous nous intéresserons ici à la question :

> Combien de temps va attendre le premier arrivé ?

Étant donné que l'on considère une situation réelle, on pourrait chercher à avoir des informations complémentaires concernant cette situation : Quand et d'où partent Karine et Olivier avant de venir au café ? Par quels moyens de transport vont-ils se rendre au café ? Ont-ils un train à prendre après ? Sont-ils connus pour être ponctuels ou souvent en retard ?...

3.2. Modèle de situation

Nous considérons que nous n'avons accès à aucune autre information sur cette situation. De cette situation réelle, dont on ne connait finalement qu'un fragment de réalité, on peut en tirer un modèle de situation, pour laquelle on se pose une question. Par exemple :

> Karine et Olivier (personnages fictifs) se donnent rendez-vous au café de la gare entre 7h et 8h. Combien de temps va attendre le premier arrivé ?

Cette description de la réalité est déjà une simplification de la situation réelle, ne prenant en compte que certains éléments constitutifs de la situation réelle.

L'expérience « concrète » repose sur les instants d'arrivées de Karine et d'Olivier (Karine et Olivier étant des personnages dont les prénoms importent peu).

3.3. Modèle pseudo-concret

Dans ce modèle de situation retenu, le temps d'attente du premier arrivé ne peut pas être prévu même si l'on répète l'expérience réelle une deuxième fois « dans les mêmes conditions » ; c'est cela qui rend cette expérience aléatoire (Girard, 2001, p. 141), d'où le fait de proposer une modélisation probabiliste de la situation. Bien que nous ne soyons pas encore dans le domaine mathématique, cette perspective d'un modèle probabiliste va guider la construction du modèle pseudo-concret. Henry (2001c) parle du « regard théorique », qui oriente les choix pour décrire l'expérience aléatoire que l'on va considérer, dans une perspective de traitement mathématique de la situation.

En effet, lorsque l'on définit un modèle pseudo-concret, on fait des choix pour simplifier et idéaliser le problème. Cela sous-entend que d'autres choix pourraient être faits et amener à d'autres modèles pour la même situation réelle de départ. Plusieurs options sont possibles (associés à des hypothèses de travail différentes), mais certains des choix doivent être guidés par un regard théorique (Henry, 2001c) sur la situation, lié aux connaissances mathématiques disponibles pour l'élève. Par exemple, ici, il va être pertinent de considérer que les personnages puissent arriver à n'importe quel instant indifféremment entre 7h et 8h, non pas parce que cela colle bien à la réalité, mais plutôt, car cela pourra amener à un modèle mathématique disponible pour les élèves, à savoir la loi uniforme. Il serait possible d'envisager de prendre en compte, par exemple, le trait de caractère ponctuel ou retardataire des personnages (en reconsidérant peut-être le modèle de situation), mais il serait ensuite moins aisé de faire concorder un modèle probabiliste disponible pour les élèves, au niveau de l'enseignement secondaire.

Dans le tableau 1, nous présentons des hypothèses de travail (Henry, 2003) et un protocole d'expérience (Parzysz, 2009) permettant de décrire l'expérience aléatoire mathématique choisie, dans le but de garantir que « la description des conditions de l'expérience détermine celle-ci de façon précise et suffisante pour en garantir l'unicité » (Girard, 2001, p. 142).

On considère que les instants d'arrivée des deux personnages ne peuvent être prédits, car ils sont soumis à trop de conditions non connues. D'où le fait que l'on étudiera cette situation à l'aide d'un modèle probabiliste.

Les deux premières hypothèses, H1 et H2, définissent le protocole expérimental et la troisième les résultats possibles. Nous avons ainsi défini l'expérience aléatoire pseudo-concrète. Nous pourrions envisager un retour dans la « réalité » via le modèle de situation, pour produire et reproduire autant de fois que l'on veut cette expérience dans le « monde réel ». Cependant ce monde réel serait fortement contrôlé par le domaine pseudo-concret. Nous pouvons considérer qu'il y a ici un chevauchement entre la réalité et le domaine pseudo-concret.

Tableau 1. Hypothèses de travail du problème de la rencontre

Modèle pseudo-concret
Hypothèses de travail
H1 : On considère que Karine ou Olivier arrive lorsqu'il ou elle franchit la porte du café.
H2 : Si Karine ou Olivier arrive avant 7h ou après 8h, l'expérience est annulée.
H3 : L'instant d'arrivée de Karine ou Olivier est donné de la forme 7h et *n* minutes ou 8h.
H4 : Karine et Olivier peuvent arriver à n'importe quel instant, indifféremment entre 7h (inclus) et 8h (inclus). L'instant d'arrivée de l'un n'influence pas l'instant d'arrivée de l'autre.
H5 : Le temps d'attente du premier arrivé commence quand le premier arrive au café. Karine et Olivier se rencontrent (donc le premier n'attend plus) lorsque le second franchit la porte du café.

Pour revenir sur les choix possibles, pour l'hypothèse H1, nous aurions pu faire un autre choix en considérant que Karine ou Olivier arrive lorsqu'il ou elle s'assoit à une table. Pour l'hypothèse H2, nous pourrions considérer que les personnages puissent arriver en avance ou en retard et élargir le créneau possible des instants d'arrivées. Dans le protocole, nous pourrions considérer en H3 que l'instant d'arrivée est de la forme 7h *n* minutes et *m* secondes.

Comme mentionné plus tôt, le choix fait dans l'hypothèse H4, « à n'importe quel instant, indifféremment entre 7h et 8h », est guidé par le « regard théorique », qui mènera dans le domaine mathématique à un modèle de répartition uniforme de la probabilité. De même, il a été choisi de s'intéresser aux instants d'arrivées de Karine et Olivier et non aux instants d'arrivées du premier et du second. En effet, dans le second cas, l'instant d'arrivée du second arrivé dépend de l'instant d'arrivé du premier, ce qui rendra le traitement mathématique plus complexe et même impossible au niveau de l'enseignement secondaire ; ce qui n'est pas le cas lorsque l'on considère les instants d'arrivées de Karine et Olivier, qui eux sont indépendants l'un de l'autre.

3.4. Modèle mathématique, traitement mathématique

La traduction du modèle pseudo-concret en modèle mathématique s'appuie notamment sur la connaissance de modèles de référence (loi uniforme, loi de Bernoulli, loi exponentielle…), qui ne sont pas forcément connus des élèves au moment considéré. La traduction mathématique peut donc se faire à travers différents registres de représentation sémiotique (langage naturel, registre symbolique des probabilités, registre des tableaux…), mais aussi suivant différents modèles. Par exemple, ici, nous pouvons associer au modèle pseudo-concret un modèle discret ou un modèle continu. Ces différents possibles dépendront notamment des connaissances mathématiques des élèves.

3.4.1 Modèle discret

Nous allons dans cette section considérer un modèle où le temps est discret. À partir des hypothèses de travail choisies dans le tableau 1, nous arrivons aux hypothèses de modèle suivantes (tableau 2).

Une fois ces hypothèses posées, il faut reformuler la question « combien de temps va attendre le premier arrivé ? » pour qu'elle puisse être abordée par un traitement mathématique dans le modèle choisi. Cette question peut être :

> Quelle est la probabilité que le premier arrivé attende entre a et b minutes ? (avec a et b deux nombres entiers compris entre 0 et 60)

Ou encore, dans un autre registre de représentation sémiotique (Duval, 1993), elle peut être reformulée :

> Quelle est la probabilité $P(a \leq T \leq b)$ pour tous a et $b \in \{0 ;\ldots ;60\}$ tels que $a \leq b$?

Cette question pourra alors être traitée mathématiquement avec des outils mathématiques mobilisant différents registres de représentation sémiotique, mais toujours dans le même modèle. Par exemple, nous pouvons envisager une résolution dans le registre des tableaux à double entrée (figure 3) ou encore dans le registre symbolique des probabilités. Le registre des arbres de probabilités n'est en revanche pas très efficace dans ce modèle (l'arbre sera vite illisible et donc inexploitable). Un autre traitement mathématique (au niveau de l'enseignement secondaire en France) serait de travailler à partir de simulation informatique. Nous détaillerons le travail mathématique dans ce cadre dans le cas continu.

Tableau 2. Hypothèses de modèle du problème de la rencontre (cas discret)

<table>
<tr><th>Modèle pseudo-concret</th><th>Modèle mathématique</th></tr>
<tr><td>Hypothèses de travail</td><td>Hypothèses de modèle</td></tr>
<tr><td>H1
H2
H3</td><td>On considère l’expérience aléatoire composée de deux épreuves décrivant l’arrivée de Karine et l’arrivée d’Olivier (épreuves identiques et indépendantes). Dans cette expérience aléatoire, on s’intéresse aux caractères : « instant d’arrivée de Karine » et « instant d’arrivée d’Olivier »[14].
On considère le temps discret (à la minute près). Cette hypothèse relève du modèle mathématique.
L’univers considéré est : Ω = {7h00 ; 7h01 ; 7h02 ; … ; 7h59 ; 8h00}×{7h00 ; 7h01 ; 7h02 ; … ; 7h59 ; 8h00} (61^2 éléments)
[On note Ω_K et Ω_O les ensembles des issues possibles des épreuves respectivement « instant d’arrivée de Karine » et « instant d’arrivée d’Olivier ». $\Omega_K = \Omega_O$ = {7h00 ; 7h01 ; 7h02 ; … ; 7h59 ; 8h00}. On a donc $\Omega = \Omega_K \times \Omega_O$]</td></tr>
<tr><td>H4</td><td>Les 61 issues possibles de $\Omega_{i\in\{K;O\}}$ sont supposées toutes équiprobables. La probabilité est distribuée sur $\Omega_{i\in\{K;O\}}$ suivant la loi représentée par le tableau :
<table><tr><td>$w_{i,j}$</td><td>7h00</td><td>7h01</td><td>7h02</td><td>…</td><td>7h59</td><td>8h00</td></tr><tr><td>$p_{i,j}$</td><td>1/61</td><td>1/61</td><td>1/61</td><td>…</td><td>1/61</td><td>1/61</td></tr></table>Dans le modèle probabiliste, on peut numériser ces issues en introduisant une variable aléatoire $X_{i\in\{K;O\}}$, définie sur $\Omega_{i\in\{K;O\}}$, à valeurs dans {0 ; 1 ; … ; 60} et dont la loi est donnée par les probabilités élémentaires p = 1/61. X_i suit une loi uniforme (discrète) sur {0 ; 1 ; … ; 60}.</td></tr>
<tr><td>H5</td><td>On considère la variable aléatoire $T = | X_K - X_O |$, correspondant au temps d’attente du premier arrivé :
T : $\Omega \rightarrow \{0 ; \dots ; 60\}\times\{0 ; \dots ; 60\} \rightarrow \{0 ; \dots ;60\}$
$(w_{K,j} ; w_{O,j}) \rightarrow (x_{K,j} ; x_{O,j}) \rightarrow |x_{K,j} - x_{O,j}|$</td></tr>
</table>

[14] Comme précisé dans la section précédente, nous choisissons de considérer l’expérience aléatoire rattachée aux instants d’arrivées de Karine et Olivier, qui sont des épreuves indépendantes, plutôt que de considérer les instants d’arrivées du premier arrivé puis du second, qui sont, elles, des épreuves dépendantes. En effet, sinon, le traitement mathématique ne serait pas envisageable, même au niveau de la classe de terminale.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
2	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
3	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
4	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56
5	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
6	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
7	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
8	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
9	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
10	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
11	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
12	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
13	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
14	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
15	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
16	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
17	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
18	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
19	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
20	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
21	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
22	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
23	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
24	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
25	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
26	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
27	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
28	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
29	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
30	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
31	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
32	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
33	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
34	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
35	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
36	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
37	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
38	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
39	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
40	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
41	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
42	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
43	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
44	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
45	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
46	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
47	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
48	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12
49	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11
50	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10
51	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9
52	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8
53	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7
54	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6
55	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5
56	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4
57	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3
58	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1	2
59	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	1
60	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

Figure 3. Tableau à double entrée représentant le temps d'attente du premier arrivé

Pour déterminer les probabilités à l'aide de ce tableau à double entrée, étant donné que toutes les issues possibles de l'expérience aléatoire sont équiprobables, nous pouvons utiliser la formule de Laplace pour déterminer $P(T = k)$, pour tout $k \in \{0 ; \ldots ; 60\}$. On a alors :

$$P(T = k) = \textit{nombre de cas favorables / nombre de cas possibles.}$$

Il s'agit alors ensuite d'un problème de dénombrement pour déterminer pour chaque k le nombre de cas favorables. Pour cela, on peut s'appuyer sur des considérations de symétries. On obtient : $P(T = 0) = 61/61^2 = 1/61$ et

$$\forall\, k \in \{1 ; \ldots ; 60\},\ P(T = k) = (2(61 - k)) / 61^2.$$

Nous rappelons qu'une arrivée après 8h est exclue. Cela nous permet ensuite de déterminer la probabilité $P(a \leq T \leq b)$ pour tous a et $b \in \{0 ; \dots ; 60\}$.

3.4.2. Modèle continu

Nous allons dans cette section choisir un modèle où le temps est continu. À partir des hypothèses de travail considérées dans le tableau 1, nous arrivons aux hypothèses de modèle suivantes (tableau 3).

Tableau 3. Hypothèses de modèle du problème de la rencontre (cas continu)

<table>
<tr><th>Modèle pseudo-concret</th><th>Modèle mathématique</th></tr>
<tr><td>Hypothèses de travail</td><td>Hypothèses de modèle</td></tr>
<tr><td>H1
H2
H3</td><td>On considère l'expérience aléatoire composée de deux épreuves décrivant l'arrivée de Karine et l'arrivée d'Olivier (épreuves identiques et indépendantes). Dans cette expérience aléatoire, on s'intéresse aux caractères : « instant d'arrivée de Karine » et « instant d'arrivée d'Olivier ».
On considère le temps continu. Cette hypothèse relève du modèle mathématique.
L'univers considéré est : $\Omega = [7h ; 8h] \times [7h ; 8h]$
On note Ω_K et Ω_O les ensembles des issues possibles des épreuves respectivement « instant d'arrivée de Karine » et « instant d'arrivée d'Olivier ». $\Omega_K = \Omega_O = [7h ; 8h]$</td></tr>
<tr><td>H4
H4' : Karine et Olivier peuvent arriver à tout instant entre 7h n minutes et 7h $n + 1$ minutes avec « autant de chance », etc.</td><td>Dans le modèle probabiliste, on peut numériser ces issues en introduisant une variable aléatoire $X_{i \in \{K;O\}}$, définit sur Ω_i, à valeurs dans [0 ; 1] et dont X_i suit une loi uniforme (continue) sur [0 ; 1].</td></tr>
<tr><td>H5</td><td>On considère la variable aléatoire $T = | X_K - X_O |$, correspondant au temps d'attente du premier arrivé :
$$T : \Omega \to [0 ; 1] \times [0 ; 1] \to [0 ; 1]$$
$$(w_{K,j} ; w_{O,j}) \to (x_{K,j} ; x_{O,j}) \to | x_{K,j} - x_{O,j} |$$</td></tr>
</table>

Dans ce modèle mathématique, la question peut être posée de cette façon :

> Quelle est la probabilité que le premier arrivé attende entre *a* et *b* heure ? (avec *a* et *b* deux réels compris entre 0 et 1)

Ou encore, dans un autre registre de représentation sémiotique (Duval, 1993), elle peut être reformulée ainsi :

> Quelle est la probabilité $P(a \leq T \leq b)$ pour tous a et $b \in [0 ; 1]$ tels que $a \leq b$?

Théoriquement, la loi de T découle de la loi du couple de variables aléatoires uniformes indépendantes X_K et X_O, que l'on obtient par double intégration. Cependant, ces connaissances mathématiques ne sont pas disponibles chez les élèves du lycée. Plusieurs traitements mathématiques mobilisant des cadres différents (Douady, 1986), cadre probabiliste ou cadre géométrique (figure 4), ou registres différents, comme le registre symbolique probabiliste ou le registre du langage tableur, sont envisageables au niveau lycée. Sur la même idée que le tableau à double entrée dans le cas discret (figure 3), une résolution géométrique avec un travail sur les aires est envisageable (figure 4).

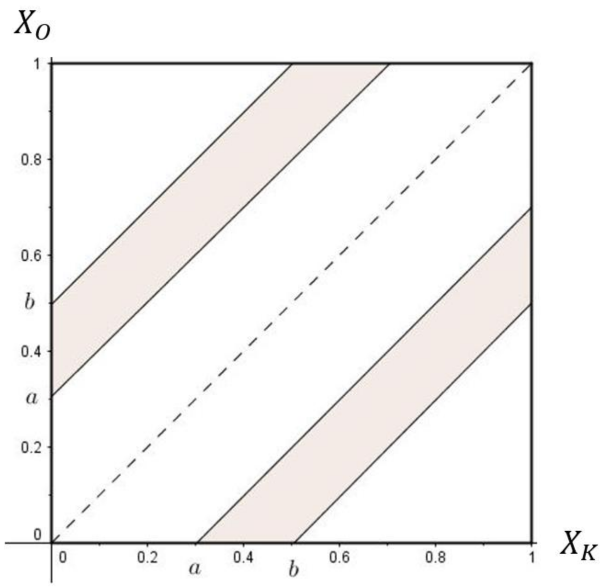

Figure 4. Résolution géométrique (Alory et Derouet, 2018)

Une résolution prenant appui sur la simulation informatique est possible, donc avec un passage par le cadre statistique. À l'aide de la commande =ABS(ALEA() − ALEA()) sur le tableur, on peut simuler des réalisations de la variable aléatoire *T*.

À partir de cette simulation, nous pouvons représenter les données statistiques obtenues sous forme d'un histogramme (figure 5).

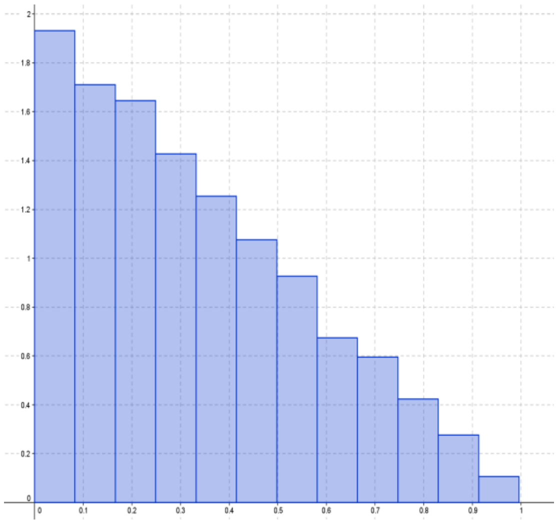

Figure 5. Histogramme d'un échantillon de taille 10 000 avec des classes d'amplitude 5 minutes (Alory et Derouet, 2018)

En appui sur la loi faible des grands nombres[15], le haut de l'histogramme représentant un grand nombre de réalisations de la variable aléatoire T se rapproche de la courbe représentative de la fonction de densité de probabilité f de la variable aléatoire T. En utilisant notamment le fait que l'aire sous la courbe d'une fonction de densité est égale à 1 et en s'appuyant sur l'histogramme, un travail mathématique permet d'arriver au fait que f est la fonction affine définie sur $[0 ; 1]$ par $f(x) = 2 - 2x$ (Derouet et Alory, 2018 ; Derouet, 2019). Connaître f permet alors de calculer $P(a \leq T \leq b) = \int_a^b f(x)dx$, pour tous a et $b \in [0 ; 1]$.

La simulation d'un grand nombre de réalisations des instants d'arrivées des deux personnages permet d'obtenir des fréquences permettant d'approcher les probabilités qui nous intéressent grâce à la loi des grands nombres (plus précisément dans sa forme particulière et simplifiée du théorème de Bernoulli). Ici, grâce à d'autres considérations qualitatives, il est possible de déterminer complètement la loi de T.

[15] Se reporter au théorème de Bernoulli à la section 1.3.4.

4. Modélisation probabiliste et approche fréquentiste

Les expériences classiques pour illustrer l'approche fréquentiste sont le lancer de punaise, le lancer d'osselets, le lancer de capsules de bouteilles… Pour complexifier un peu la situation de départ, nous allons nous intéresser au lancer de deux punaises.

4.1. Description de la situation réelle (A)

Nous partons de la situation réelle suivante, que nous appellerons le problème des trois sœurs :

> Alice, Elya et Juliette sont trois sœurs. Elles doivent mettre la table. Elles décident de jouer à un jeu de hasard pour décider laquelle d'entre elles mettra la table. Alice propose :
>
> « Je vais lancer ces deux punaises de bureau :
>
> Si les deux punaises ont la pointe vers le haut, je mets la table ;
>
> Si les deux punaises ont la pointe qui touche le sol, Elya met la table ;
>
> Si les deux punaises s'arrêtent dans des positions différentes (une la pointe vers le haut, l'autre la pointe touchant le sol), Juliette met la table.
>
> Êtes-vous d'accord ? »
>
> Juliette dit qu'elle n'est pas d'accord, car elle a beaucoup plus de risque de mettre la table que les autres et qu'Alice a peu de chance de la mettre.
>
> Qu'en pensez-vous ?

4.2. Modèle de situation

Nous n'avons pas connaissance ici des caractères de chacune des trois sœurs ni d'autres informations plus personnelles que nous pourrions prendre en compte. On peut donc considérer que les personnes en question et le contexte (dresser la table) n'ont pas d'importance dans la situation. Seul le jeu est à questionner. Le problème pourrait finalement être :

> Lorsqu'on lance deux punaises, y a-t-il beaucoup plus de chance que les deux punaises s'arrêtent dans des positions différentes, plutôt que les deux s'arrêtent dans la même position ?

4.3. Modèle pseudo-concret, modèle mathématique, traitement mathématique

4.3.1 Modèle 1

Nous présentons un premier modèle possible à partir d'hypothèses détaillées (tableau 4).

Tableau 4. Hypothèses de travail et hypothèses de modèle du problème des trois sœurs

<table>
<tr><th>Modèle pseudo-concret</th><th>Modèle mathématique</th></tr>
<tr><td>Hypothèses de travail</td><td>Hypothèses de modèle</td></tr>
<tr><td>H1 : On prend deux punaises de la même fabrication, sans défauts apparents. On considère les punaises indiscernables.</td><td rowspan="4">On considère l'expérience aléatoire du lancer simultané de deux punaises. Dans cette expérience aléatoire, on s'intéresse au caractère « positions d'arrivée des deux punaises ».

L'univers considéré est :

$\Omega = \{HH ; SS ; SH\}$ (H correspond à une punaise qui s'arrête la pointe vers le haut ; S la pointe touchant le sol). Remarque : SH correspond aux punaises qui ont chacune une position différente (SH est équivalent à HS).</td></tr>
<tr><td>H2 : On lance par terre simultanément les deux punaises, suffisamment fort. Dans ces conditions, on ne peut pas anticiper leurs positions d'arrêt.</td></tr>
<tr><td>H3 : Une fois arrêtées, on note les positions des deux punaises.</td></tr>
<tr><td>H4 : Une punaise peut s'arrêter soit la pointe vers le haut, soit avec la pointe touchant le sol (ce sont les deux seules positions possibles). Les positions d'arrivées des deux punaises sont soit : les deux ont la pointe vers le haut, les deux ont la pointe qui touche le sol ou les deux punaises ont une position différente. On ne distingue pas quelle punaise est dans quelle position en vertu de H1.</td></tr>
</table>

La question est alors :

> Quelles sont les probabilités élémentaires associées à cette expérience aléatoire ?

En comparant ces probabilités, nous pourrons avoir des éléments pour analyser le jeu.

Étant donné qu'aucune symétrie dans les résultats ne peut être prise en compte *a priori* pour les objets « punaises », il n'est pas possible de considérer une approche laplacienne. Une fois le protocole expérimental établi (H1, H2, H3), il est possible de revenir dans la « réalité » pour simuler matériellement un grand nombre de fois le lancer de deux punaises comme décrit et de prendre note des résultats des différentes réalisations.

Après 5000 lancers de deux punaises, avec deux punaises du monde réel, dans le cadre d'une expérience du monde pseudo-concret, nous avons obtenu les fréquences d'apparition suivantes : $f(HH) \approx 0{,}3684$; $f(SS) \approx 0{,}1506$; $f(SH) \approx 0{,}4810$.

On peut alors décider de poser les probabilités élémentaires suivantes (les fréquences arrondies à 10^{-2} près) :

w_i	HH	SS	SH
p_i	0,37	0,15	0,48

On a alors : la probabilité $P(SH)$ est un peu plus de 3 fois plus grande que la probabilité $P(SS)$, tandis que $P(HH) \approx 2{,}5\ P(SS)$.

4.3.2 Modèle 2

Nous pourrions envisager un autre protocole expérimental et donc un autre modèle (tableau 5).

Avec ces considérations, nous pouvons considérer un « sous-protocole expérimental », celui associé à l'épreuve « lancer d'une punaise » :

H1' : On prend une punaise d'un certain modèle.

H2' : On lance par terre la punaise suffisamment fort.

H3' : Une fois arrêtée, on note la position de la punaise.

Cela nous permet de simuler matériellement un grand nombre de fois le lancer de punaise.

Après 5000 lancers de la même punaise, nous obtenons une fréquence d'apparition de la pointe de 0,3812. Bien entendu 5000 autres lancers donneraient une fréquence différente, mais espérée proche. Nous sommes ici entre « réalité » et pseudo-concret. Cette fréquence est issue d'expérimentations sur un objet matériel de la « réalité », qui ont été guidées par des considérations et des choix dans le monde pseudo-concret.

Tableau 5. Deuxième proposition d'hypothèses de travail et hypothèses de modèle du problème des trois sœurs

<table>
<tr><th>Modèle pseudo-concret</th><th>Modèle mathématique</th></tr>
<tr><td>Hypothèses de travail</td><td>Hypothèses de modèle</td></tr>
<tr><td>H1 : On prend deux punaises de la même fabrication, sans défauts apparents, mais de couleur différente (par exemple une verte et une jaune). On considère les punaises identiques mis à part la couleur.</td><td rowspan="5">On considère l'expérience aléatoire du lancer simultané de deux punaises (de couleur différente), qui se décompose en deux épreuves simultanées, mais indépendantes : le lancer de punaise verte et le lancer de punaise jaune. Dans cette expérience aléatoire, on s'intéresse au caractère « position d'arrivée des deux punaises », composé de la « position d'arrivée de la punaise verte » et de la « position d'arrivée de la punaise jaune ». En considérant que la position d'arrivée d'une punaise est indépendante de la couleur de celle-ci, on peut dire que l'expérience aléatoire du lancer simultané de deux punaises est composée de deux épreuves identiques et indépendantes de lancer d'une punaise.

L'univers considéré de l'expérience aléatoire est :

$\Omega = \{(H,H) ; (H,S) ; (S,H) ; (S,S)\}$
Remarque : la première coordonnée correspond au résultat de la punaise verte et la deuxième au résultat de la punaise jaune.

L'univers de l'épreuve « lancer d'une punaise » est :

$\Omega' = \{H ; S\}$

Il s'agit d'une épreuve de Bernoulli de paramètre inconnu qui reste à déterminer.</td></tr>
<tr><td>H2 : On lance par terre simultanément les deux punaises, suffisamment fort. Dans ces conditions, on ne peut pas anticiper leurs positions d'arrêt.</td></tr>
<tr><td>H3 : Une fois arrêtées, on note la position de la punaise verte puis celle de la punaise jaune.</td></tr>
<tr><td>H4 : Une punaise peut s'arrêter soit la pointe vers le haut, soit avec la pointe touchant le sol (ce sont les deux seules positions possibles). Les positions d'arrivée des deux punaises sont soit : les deux ont la pointe vers le haut, les deux ont la pointe qui touche le sol ou les deux punaises ont une position différente.</td></tr>
<tr><td>H5 : Le résultat d'une punaise n'influence pas celui de l'autre.</td></tr>
</table>

Nous allons utiliser cette fréquence pour construire le modèle mathématique. Comme présenté dans la section 1.3.3, plusieurs choix sont envisageables. Nous décidons ici de poser $P(S) = 0{,}4$ (la fréquence trouvée arrondie au dixième), notamment pour simplifier les calculs. On pourrait aussi considérer $P(S) = 0{,}3812$ ou encore faire de nouveaux lancers et prendre la nouvelle fréquence obtenue…

Nous avons donc une expérience aléatoire « lancer simultané de deux punaises » constitué de deux épreuves de Bernoulli identiques et indépendantes de paramètre $p = 0{,}4$ (nous considérons l'événement « s'arrêter la pointe touchant le sol » comme l'événement « succès »).

Soit X la variable aléatoire donnant le nombre de punaises qui s'arrêtent la pointe touchant le sol dans l'expérience aléatoire « lancer simultané de deux punaises » :

$$X : \Omega \rightarrow \{0 ; 1 ; 2\}$$

X suit une loi binomiale de paramètres $n = 2$ et $p = 0{,}4$.

Au sein de ce modèle, la question que l'on se pose peut alors se reformuler dans le registre de la langue naturelle, comme suit :

> Quelles sont les probabilités des événements suivants :
>
> - Les deux punaises s'arrêtent la pointe vers le haut (événement A) ;
> - Les deux punaises s'arrêtent avec la pointe touchant le sol (événement B) ;
> - L'une des punaises s'arrête la pointe vers le haut et l'autre avec la pointe touchant le sol (événement C) ?

Ou encore, dans un autre registre de représentation sémiotique, le registre symbolique des probabilités, elle peut être reformulée ainsi :

> Calculer $P(X = 0)$, $P(X = 1)$ et $P(X = 2)$.

Cela permettra alors de comparer les différentes probabilités relatives aux différents cas de figure possibles pour le jeu.

Suivant les connaissances des élèves, plusieurs traitements mathématiques sont alors possibles.

a) Utilisation des formules liées à la loi de Bernoulli :

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{2}{k} 0{,}4^k 0{,}6^{2-k}$, pour tout $k \in \{0 ; 1 ; 2\}$.

b) Utilisation d'un arbre de probabilités (figure 6), des probabilités conditionnelles et utilisation de la formule des probabilités totales :

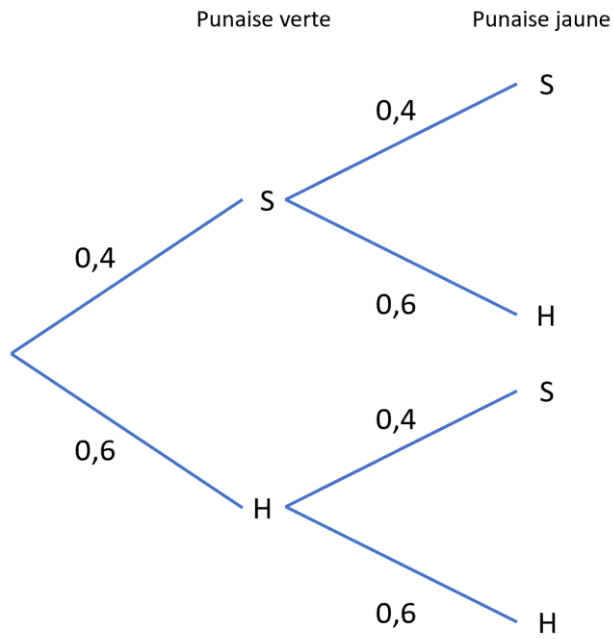

Figure 6. Arbre de probabilité

$P(A) = P((H,H)) = 0{,}6 \times 0{,}6 = 0{,}36$

$P(B) = P((S,S)) = 0{,}4 \times 0{,}4 = 0{,}16$

$P(C) = P((S,H)) + P((H,S)) = 2 \times (0{,}4 \times 0{,}6) = 0{,}48$

On a donc $P(C) = 3 \times P(B)$ et $P(A) = 2{,}25 \times P(B)$.

c) Utilisation de la simulation informatique

Il est aussi envisageable de procéder à l'aide d'une simulation informatique. Avec le tableur, la commande = SI(ALEA() < 0,4 ; P ; T) permet de simuler le lancer d'une punaise (dans le modèle que nous avons choisi avant). Il est alors possible de simuler un grand nombre de fois le lancer de deux punaises et de calculer les fréquences d'apparition des différentes issues.

Ces différents registres (registre symbolique des probabilités, registre des arbres de probabilités, registre du tableur avec un passage par le cadre statistique) permettent de traiter mathématiquement la situation dans le même modèle. Les résultats mathématiques permettent ensuite de retourner dans le domaine pseudo-concret et de conclure dans la réalité.

5. Modélisation probabiliste et inférence statistique informelle

Dans cet exemple, nous allons nous intéresser à une question d'inférence statistique relevant de la généralisation à partir d'un échantillon (Garfield et Ben-Zvi, 2008). L'exemple présenté, que nous appelons le problème du volcan Aso, a déjà été présenté et étudié dans un autre cadre (Alory et Derouet, 2018 ; Derouet, 2019).

5.1. Description de la situation réelle (A)

> Le mont Aso est le volcan le plus vaste du Japon, et aussi un des plus actifs. Quand aura lieu la prochaine éruption de ce volcan ?

5.2. Modèle de situation

Un expert en vulcanologie doit prendre en compte beaucoup de paramètres pour aborder cette question. Ici nous allons limiter l'étude de la situation à partir des simples données des relevés des dates d'éruptions du volcan depuis le XIIIè siècle. Nous pouvons modifier légèrement la question :

> À partir des données des dates des précédentes éruptions, évaluer quand aura lieu la prochaine éruption du volcan Aso.

5.3. Modèle pseudo-concret

Le tableau 6 présente les hypothèses de travail que nous pouvons faire.

Tableau 6. Hypothèses de travail du problème du volcan Aso

Modèle pseudo-concret
Hypothèses de travail
H1 : On ne peut pas prédire avec certitude la date d'éruption de la prochaine éruption volcanique.
H2 : On considère comme date d'éruption du volcan Aso la date de début de l'éruption déclarée par des organismes de vulcanologie reconnus. Nous en retenons ensuite seulement l'année, notée A_i.
H3 : On utilise toutes les données accessibles des éruptions passées.
H4 : On considère que le temps d'attente entre deux éruptions passées est donné par l'intervalle $]A_{i+1} - A_i - 1 ; A_{i+1} - A_i]$.
H5 : Le temps d'attente entre deux éruptions peut être compris entre 0 et une durée aussi grande que l'on veut (si la prochaine éruption n'arrive jamais).

Bien entendu, lors de la récolte des données statistiques, d'autres choix seraient possibles et tout aussi valables (par exemple prendre comme date d'éruption non pas la date de début, mais la date où l'intensité est la plus forte, prendre en compte la date précise et pas seulement l'année…). Nous ne les discutons pas ici. Les données récoltées, depuis le XIIIe siècle, peuvent être représentées sous forme d'un histogramme de fréquences (figure 7).

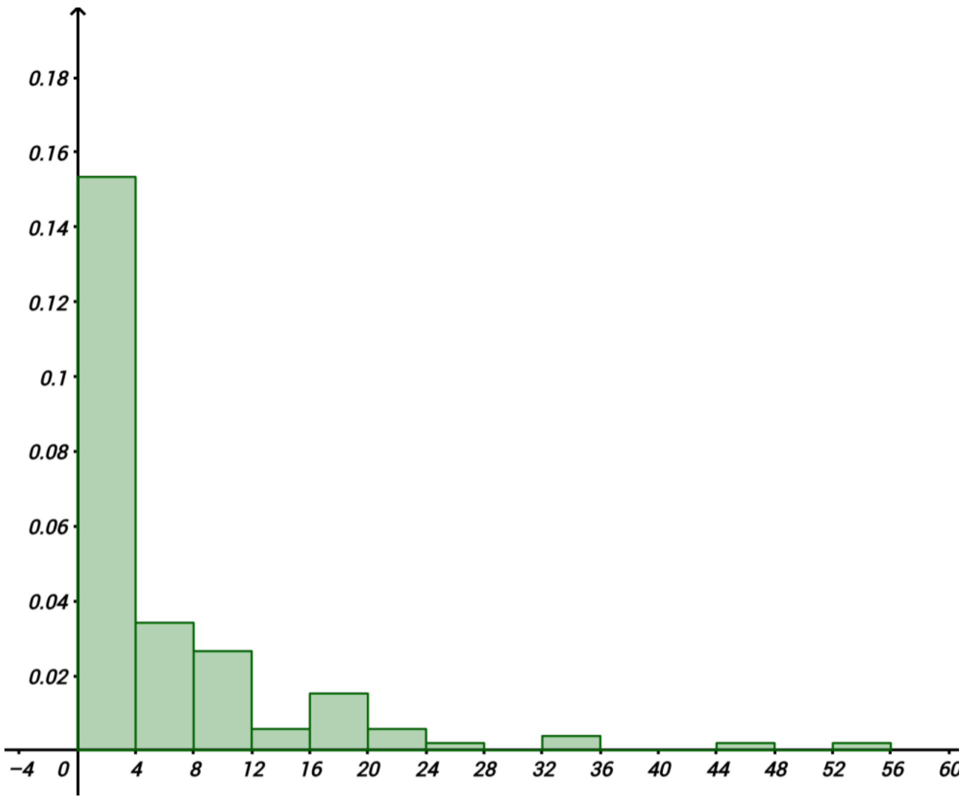

Figure 7. Histogramme des temps d'attente entre deux éruptions consécutives

5.4. Modèle mathématique, traitement mathématique

Comme pour le problème de la rencontre, nous pouvons ici considérer le temps discret ou continu. Nous faisons le choix de le considérer continu dans la suite. À partir des hypothèses de travail considérées (tableau 6), nous pouvons obtenir les hypothèses de modèle suivantes (tableau 7).

Tableau 7. Hypothèses de modèle du problème du volcan Aso

<table>
<tr><th>Modèle pseudo-concret</th><th>Modèle mathématique</th></tr>
<tr><td>Hypothèses de travail</td><td>Hypothèses de modèle</td></tr>
<tr><td>H1</td><td rowspan="3">On considère l’expérience aléatoire de la date d’éruption du volcan Aso.

On considère le temps continu. Cette hypothèse relève du modèle mathématique.

L’univers considéré est : $\Omega = [1200 ; +\infty[$ (nous ne disposons que des données depuis le XIIIe siècle).</td></tr>
<tr><td>H2</td></tr>
<tr><td>H3</td></tr>
<tr><td>H4</td><td rowspan="2">On considère la variable aléatoire X correspondant au temps d’attente entre deux éruptions du volcan :

$$\begin{aligned} X : \quad \Omega^2 &\to [0 ; +\infty[\\ (w_i ; w_{i+1}) &\to w_{i+1} - w_i \end{aligned}$$

La loi de probabilité de X doit suivre une distribution « proche » de la distribution de fréquence des données statistiques récoltées.</td></tr>
<tr><td>H5</td></tr>
</table>

Il s’agit maintenant de choisir une loi de probabilité pour la variable aléatoire à partir, par exemple de l’histogramme construit dans le modèle pseudo-concret (figure 7). Dans les lois à disposition ou abordables par les élèves de lycée, seule la loi exponentielle est possible. En revanche, plusieurs choix sont possibles pour le paramètre (figure 8).

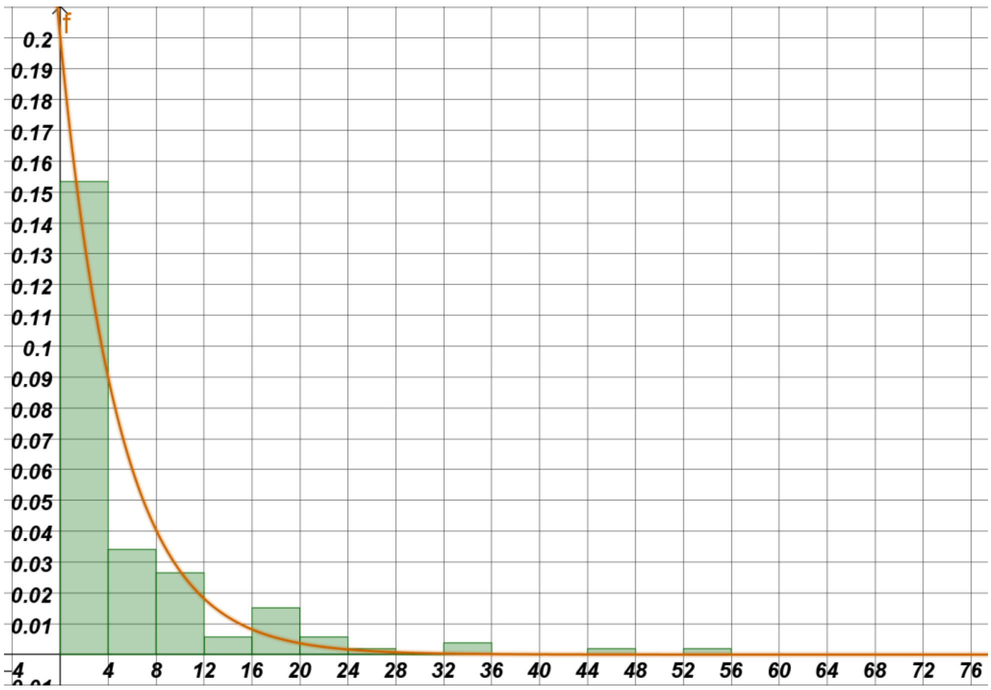

Figure 8. Choix de fonction de densité f associée à la variable aléatoire X (loi exponentielle de paramètre 0,2)

On peut alors remplacer la question par une autre comme par exemple :

> Quelle est la probabilité que le temps d'attente entre deux éruptions du volcan soit inférieur à 5 ans ?

Ou de façon plus générale, la question peut être :

> À partir de combien d'années A la probabilité que le temps d'attente soit inférieur à A est supérieure à 0,9 ?

En considérant que X suit une loi exponentielle de paramètre 0,2, on obtient :

$P(X \leq A) \geq 0{,}9$

$1 - e^{-0{,}2A} \geq 0{,}9$

$A \geq -\,ln(0{,}1) \,/\, 0{,}2$

Or $-\,ln(0{,}1) \,/\, 0{,}2 \approx 11{,}5$.

La probabilité que le temps d'attente entre deux éruptions du volcan Aso soit inférieur à 12 ans est supérieure à 0,9.

6. La place et le rôle de la statistique descriptive dans la modélisation probabiliste

Dans cette partie, nous allons aborder la question de recherche QR2, qui concerne la place de la statistique dans les trois catégories de démarches de modélisation probabiliste étudiées et le rôle qu'elle joue.

6.1. Les données statistiques

Au regard des trois exemples de problèmes de modélisation étudiés et des généralisations que l'on a pu en tirer, il est visible que les données statistiques en appui sur la réalité sont indispensables à la construction du modèle probabiliste dans le cas de l'approche fréquentiste et dans le cas de la statistique inférentielle.

Cependant dans le cas de l'approche fréquentiste, les données statistiques sont récoltées dans le respect du protocole expérimental. Il s'agit de données réelles obtenues dans le cadre du modèle pseudo-concret. On peut le voir comme un retour dans la réalité. On peut alors récolter un grand nombre de résultats de l'expérience, aussi grand que l'on veut dans la limite des possibilités temporelles et matérielles notamment. La distribution des fréquences va permettre de construire la distribution de probabilités. Un modèle probabiliste sera choisi en appui sur les fréquences. Cet enchaînement est illustré en figure 9.

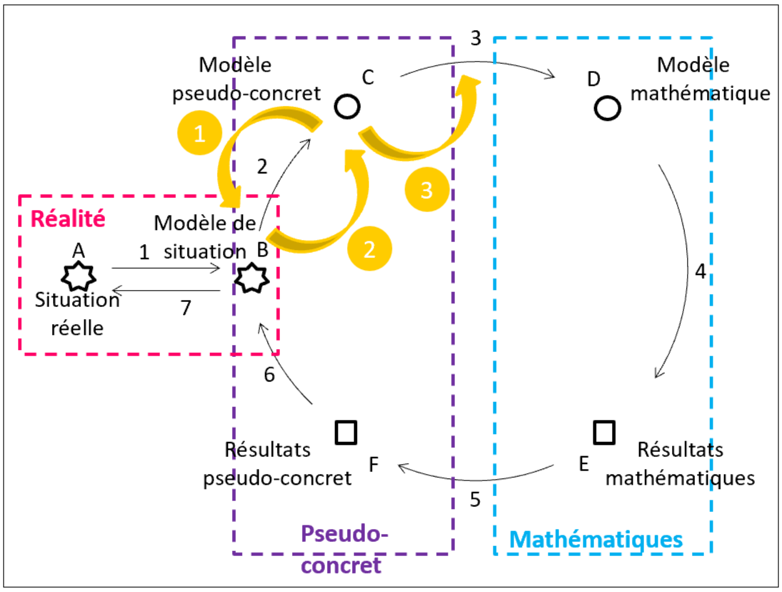

Figure 9. Approche fréquentiste : des données statistiques réelles contrôlées par le modèle pseudo-concret

Dans le cas de la statistique inférentielle, le rapport aux données est différent. En effet, les données sont issues de la situation réelle, ce n'est pas un retour à la réalité (tout du moins pas entendu dans le même sens que dans le cas précédent). Ces données sont le point de départ de la construction du modèle pseudo-concret, ce n'est pas le modèle pseudo-concret qui impose des données contrôlées (figure 10). Ces données ont tout de même été récoltées de manière rigoureuse. Elles sont en nombre fixé, qui peut être relativement assez limité. On ne peut pas en avoir autant que l'on veut comme précédemment. La distribution des fréquences va permettre de faire des inférences pour la construction du modèle. Cependant les fréquences ne vont pas directement être utilisées comme probabilités. Il y a un jeu de reconnaissance d'un modèle probabiliste connu ou de construction de celui-ci en appui sur des propriétés mathématiques.

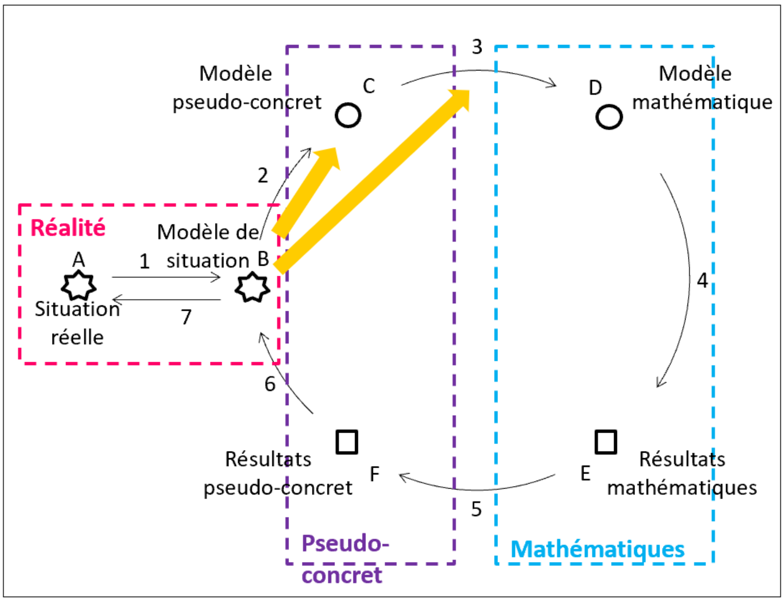

Figure 10. Statistique inférentielle informelle : les données statistiques comme point de départ

6.2. La simulation informatique

La statistique descriptive est aussi présente lors de simulations informatiques. En revanche, ce passage à la simulation informatique ne relève pas spécifiquement d'une des trois catégories. Elle peut en effet être présente dans chacune d'elles, mais non obligatoire. La simulation relève en fait du domaine des mathématiques (et bien entendu du domaine de l'informatique) et non de la réalité ou du modèle pseudo-concret. Elle s'appuie sur le modèle mathématique choisi et permet d'obtenir des résultats relatifs à des modèles plus complexes composés de modèles initiaux. Par exemple dans le problème de la rencontre, par l'intermédiaire de la simulation des instants d'arrivées de Karine et d'Olivier qui suivent des lois uniformes, on obtient des résultats sur le temps d'attente du premier arrivé T en considérant la valeur absolue de la différence des deux instants d'arrivées simulés.

Dans nos deux premiers exemples de situation de modélisation, il est bien visible que la simulation informatique a lieu après le choix du modèle. Comme la définit Dogme (1993) :

> La simulation est 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.

Parzysz (2009) explique aussi 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).

Cela permet de bien mettre en évidence la différence entre l'approche fréquentiste qui permet de construire un modèle (et des probabilités d'événements associés) et la simulation informatique qui permet d'estimer des probabilités d'événements dans un modèle choisi. En prenant appui sur le cycle de modélisation, on peut mettre en évidence que l'approche fréquentiste se situe à la liaison entre réalité et domaine pseudo-concret, tandis que la simulation est entièrement dans le domaine mathématique (au sein du cycle de modélisation), dépendant du modèle mathématique choisi. Cette différence est illustrée dans les figures 11 et 12. Cette place et ce rôle différents sont aussi à relier à la loi des grands nombres. En effet, la loi des grands nombres (dans sa version théorème de Bernoulli) n'est présente que dans la simulation, lorsqu'un modèle est arrêté. La simulation est d'ailleurs un outil très utilisé pour ensuite invalider ou non un modèle choisi, en comparant avec les données réelles (étape 6 du cycle de modélisation). En revanche, l'approche fréquentiste, permettant de construire un modèle mathématique, ne mobilise pas la loi des grands nombres. Cependant, la loi des grands nombres justifie (*a posteriori*) l'approche fréquentiste. Enfin, dans l'exemple simple du lancer d'un dé, comme dans toutes situations pouvant prétendre à l'équiprobabilité du fait notamment de symétries, l'approche laplacienne et l'approche fréquentiste donnent (fort heureusement) des modèles très proches (à hypothèses similaires).

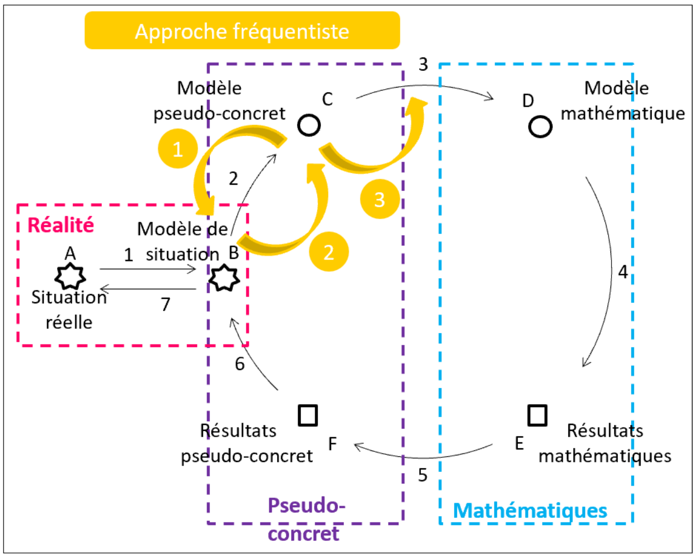

Figure 11. Place de l'approche fréquentiste dans le cycle de modélisation

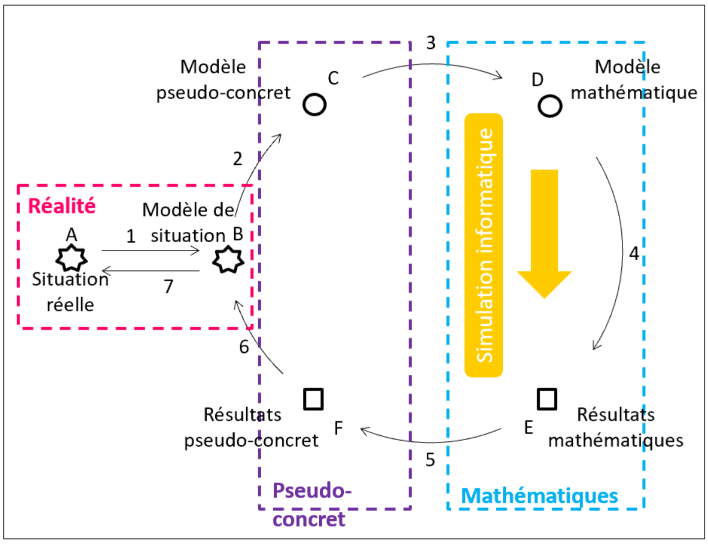

Figure 12. Place de la simulation informatique dans le cycle de modélisation

7. Conclusion et perspectives

En prenant appui sur trois exemples de problème de modélisation probabiliste, nous avons cherché à répondre aux deux questions suivantes :

- QR1 : Dans le cadre du cycle de modélisation, comment caractériser chacune des trois catégories de démarches de modélisation probabiliste considérées (approche laplacienne, approche fréquentiste, inférence statistique informelle) ?
- QR2 : Quelle place tient la statistique et quel rôle joue-t-elle dans les trois catégories de démarches de modélisation probabiliste considérées ?

En considérant les différentes étapes du cycle de modélisation, nous avons cherché à illustrer et à mettre en évidence les hypothèses de travail et de modèle sous-jacentes dans le contexte des trois problèmes de modélisation probabiliste « complexes » relevant chacun d'une des trois approches étudiées. Nous pouvons en extraire des informations généralisables :

- Une modélisation avec une approche laplacienne nécessite une hypothèse d'équiprobabilité à expliciter dès le modèle pseudo-concret ;
- Une modélisation avec une approche fréquentiste nécessite l'élaboration d'un protocole expérimental au niveau du modèle pseudo-concret pour permettre un retour dans la « réalité » (modèle de situation) pour récolter un grand nombre de réalisations de l'expérience aléatoire telle que décrite dans le protocole expérimental ;
- Une modélisation par inférence statistique informelle nécessite la mise à disposition de données réelles (réalité) qui sont dans un premier temps étudiées dans le domaine de la statistique descriptive (modèle pseudo-concret) et vont permettre de construire un modèle pseudo-concret afin d'être ensuite un point d'appui pour construire un modèle probabiliste compatible avec ces données (modèle mathématique).

Nous avons aussi mis en évidence que suivant l'approche mobilisée, la place et le rôle des données statistiques et de la statistique descriptive sont différents. En effet, les données statistiques réelles sont le point de départ d'une modélisation par inférence statistique, tandis que les données statistiques dans l'approche fréquentiste sont des réalisations d'une expérience aléatoire élaborée dans le modèle pseudo-concret. Il y a, dans ce second cas, un jeu d'aller-retour entre le modèle pseudo-concret et la *model situation*. Par une approche par inférence statistique, les données statistiques réelles sont ensuite simplifiées par exemple en regroupant des données, en éliminant des données considérées comme aberrantes… pour construire un modèle pseudo-concret, alors que dans l'approche fréquentiste la simplification se fait en amont de l'expérimentation, elle guide le recueil de données. Dans le cas de l'approche laplacienne, les données statistiques peuvent être complètement absentes.

Nous avons aussi essayé de clarifier la place des réalisations issues de la simulation informatique. En effet, la simulation informatique peut se retrouver dans les différentes approches, en revanche elle se situe à l'intérieur même du modèle mathématique, car simuler informatiquement une expérience aléatoire nécessite d'implémenter un modèle dans le tableur ou autre outil de simulation. Cependant, comme le signalait Parzysz (2014), « dans la pratique le modèle est souvent escamoté : on fait comme si on passait directement de la situation réelle à la simulation » (p. 71). La simulation informatique peut ensuite permettre d'estimer une probabilité d'un événement en appui sur la loi faible des grands nombres. Une confusion peut exister chez les enseignants et les élèves entre l'approche fréquentiste des probabilités et l'utilisation de la loi faible des grands nombres. Il est important de repérer que l'approche fréquentiste permet de construire un modèle mathématique, tandis que la loi des grands nombres permet d'estimer des probabilités dans le monde des mathématiques une fois un modèle choisi.

Ces distinctions apportées entre les différentes approches pouvant être en jeu dans la modélisation probabiliste ainsi que le couplage du cycle de modélisation avec les hypothèses de travail et de modèle nous apportent un cadre théorique pour étudier différentes questions d'enseignement-apprentissage dans le domaine des probabilités dans l'enseignement secondaire, voire aussi dans l'enseignement supérieur. Cela peut permettre de donner un outil d'analyse pour repérer les étapes de modélisation qui sont travaillées ou non en classe, et étudier les étapes pouvant être source de difficultés chez les élèves. Bien entendu, le détail des hypothèses de travail et de modèle formulées dans l'article ne l'est pas de façon aussi visible dans les classes, cependant elles peuvent tout de même être présentes plus ou moins explicitement ou encore totalement absentes, ce qui peut poser question sur le travail de modélisation des élèves. Pour étudier des déroulements en classe, il pourrait être intéressant de prendre en plus en compte les compétences spécifiques à la mathématisation horizontale (Yvain-Prébiski, 2018) pour compléter les analyses.

Outre un outil théorique et méthodologique pour la recherche, ces apports pourraient se révéler utiles pour la formation des enseignants. De plus, le protocole expérimental, les hypothèses de travail et de modèle mériteraient d'être explicités en classe, notamment lors des débuts de l'enseignement des probabilités (actuellement au collège), comme l'a déjà mentionné Parzysz (2009). Ces pratiques sont encore très rares (voire inexistantes), or elles permettraient de lever des ambiguïtés et des confusions entre réalité et modèle, fréquentes dans les classes (Girard, 2001b).

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 » (p. 33-34). Nous partageons cet avis. Cependant, force est de constater que vingt ans plus tard, la démarche de modélisation probabiliste est encore très peu présente dans les classes de collège et lycée, en France. La nouvelle réforme du lycée de 2019 ne va pas dans le sens d'une amélioration de ce phénomène dans la spécialité Mathématiques, en revanche l'option Mathématiques complémentaires en terminale pourrait être le lieu d'un véritable travail de modélisation probabiliste.

Bibliographie

ADJIAGE, R. ET RAUSCHER, J.-C. (2013). Résolution d'un problème de modélisation et pratique écrite de l'écrit. *Recherches en Didactique des Mathématiques*, 33(1), 9–43.

BEN-ZVI, D. (2006). Scaffolding students' informal inference and argumentation. Dans A. Rossman et B. Chance (Dir.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. International Statistical Institute.

BIEHLER, R. (1994). Probabilistic thinking, statistical reasoning, and the search for causes: Do we need a probabilistic revolution. Dans J. Garfield (Dir.), *Research Papers from ICOTS 4*. University of Minnesota.

BLUM, W., GALBRAITH, P. L., HENN, H.-W. ET NISS, M. (2007). *Modelling and applications in mathematics education: the 14th ICMI study*. Springer.

BLUM, W. et 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 et S. Khan (Dir.), *Mathematical modelling (ICTMA12) – Education, engineering and economics* (p. 222–231). Horwood.

BORROMEO FERRI, R. (2006). Theoretical and empirical differentiations of phasis in the modelling process. *ZDM - Mathematics Education*, 38(2), 86–95.

BURGHES, D. (1986). Mathematical modelling – are we heading in the right direction? Dans J. et al. Berry (Dir.), *Mathematical Modelling Methodology, Models and Micros* (p. 11–23). Horwood.

CABASSUT, R. et WAGNER, A. (2011). Modelling at Primary School Through a French–German Comparison of Curricula and Textbooks. Dans G. Kaiser, W. Blum, R. Borromeo Ferri et G. Stillman (Dir.), *Trends in Teaching and Learning of Mathematical Modelling. International Perspectives on the Teaching and Learning of Mathematical Modelling, vol. 1* (p. 559–568). Springer.

CHAPUT, B., GIRARD, J. C. et HENRY, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. Dans C. Batanero, G. Burrill et C. Reading (Dir.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (p. 85–95). Springer.

CHEVALLARD, Y. (1989). Le passage de l'arithmétique à l'algèbre dans l'enseignement des mathématiques au collège. *Petit x*, 19, 45–75.

COULANGE, L. (1998). Les problèmes "concrets" à "mettre en équation" dans l'enseignement. *Petit x*, 47, 33–58.

DEROUET, C. et 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.

DEROUET, C. (2019). Introduire la notion de fonction de densité de probabilité : dynamiques entre trois domaines mathématiques. *Recherches en Didactique des Mathématiques*, 39(2), 213–266.

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

DOUADY, R. (1986). Jeux de cadres et dialectique outil-objet. *Recherches En Didactique Des Mathématiques*, 7(2), 7–31.

DUVAL, R. (1993). Registres de représentation sémiotique et fonctionnement cognitif de la pensée. *Annales de Didactique et de Sciences Cognitives*, 5, 37–65.

FREUDENTHAL, H. (1991). *Revisiting mathematics education.* Kluwer Academic.

GARFIELD, J. et BEN-ZVI, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching*. Springer.

GIRARD, J.-C. (2001a). Qu'est-ce qu'une expérience aléatoire ? Dans M. Henry (Dir.), *Autour de la modélisation en probabilités* (p. 141–144). Presses universitaires Franc-Comtoises.

GIRARD, J.-C. (2001b). Un exemple de confusion modèle-réalité. Dans M. Henry (Dir.), *Autour de la modélisation en probabilités* (p. 145–148). Presses Universitaires Franc-Comtoises.

GREER, B. et MUKHOPADHYAY, S. (2005). Teaching and Learning the Mathematization of Uncertainty: Historical, Cultural, Social and Political Contexts. Dans G. A. Jones (Dir.), *Exploring Probability in School. Challenges for Teaching and Learning* (p. 297–324). Springer.

HANKELN, C. et HERSANT, M. (2020). Processus de modélisation et de problématisation en mathématiques à la fin du lycée. Une étude de cas dans une perspective de didactique comparée. *Education et Didactique*, 14(3), 39–67.

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* (p. 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 Franc-Comtoises.

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* (p. 149–159). Presses Universitaires Franc-Comtoises.

HENRY, M. (2001c). Notion d'expérience aléatoire. Vocabulaire et modèle probabiliste. Dans M. Henry (Dir.), *Autour de la modélisation en probabilités* (p. 161–171). Presses Universitaires Franc-Comtoises.

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

HENRY, M. (2010). Evolution de l'enseignement secondaire français en statistique et probabilités. *Statistique et Enseignement*, 1(1), 35–45.

KAISER, G. (1995). Realitätsbezüge im Mathematikunterricht – Ein Überblick über die aktuelle und historische Diskussion. Dans G. Graumann (Dir.), *Materialien für einen realitätsbezogenen Mathematikunterricht* (p. 66–84).

KAISER, G., BLUM, W., BORROMEO FERRI, R. et STILLMAN, G. (2011). *Trends in Teaching and Learning of Mathematical Modelling. ICTMA14*. Springer Netherlands.

KINTSCH, W. et GREENO, J. (1985). Understanding word arithmetic problems. *Psychological Review*, 92(1), 109–129.

KONOLD, C. et KAZAK, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, 2(1).

LEUNG, F. K. S., STILLMAN, G. A., KAISER, G. et WONG, K. L. (2021). *Mathematical Modelling Education in East and West. International Perspectives on the Teaching and Learning of Mathematical Modelling*. Springer.

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.

PFANNKUCH, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? Dans G. A. Jones (Dir.), *Exploring probability in school: Challenges for teaching and learning* (p. 267–294). Springer.

PFANNKUCH, M. (2006). Informal inferential reasoning. Dans A. Rossman et B. Chance (Dir.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. International Statistical Institute.

PFANNKUCH, M., BUDGETT, S., FEWSTER, R., FITCH, M., PATTENWISE, S., WILD, C. ET ZIEDINS, I. (2016). Probability modeling and thinking: what can we learn from practice? *Statistics Education Research Journal,* 15(2), 11–37.

PICHARD, J.-F. (2001). Les probabilités au tournant du XVIIIè siècle. Dans M. Henry (Dir.), *Autour de la modélisation en probabilités* (p. 13–45). Presses Universitaires Franc-Comtoises.

POLLAK, H. (1979). The Interaction between Mathematics and Other School Subjects. Dans UNESCO (Dir.), *New Trends in Mathematics Teaching IV* (p. 232–248).

PRATT, D. (2011). Re-connecting probability and reasoning from data in secondary school teaching. *Proceedings of the 58th International Statistical Institute World Statistical Congress*, 890–899.

RODRIGUEZ, R. (2007). *Les équations différentielles comme outil de modélisation mathématique en Classe de Physique et de Mathématiques au lycée : une étude de manuels et de processus de modélisation d'élèves en Terminale S*. [Thèse de doctorat, Université Joseph Fourier Grenoble I]. https://tel.archives-ouvertes.fr/tel-00292286/document

STAUB, F. C. et REUSSER, K. (1995). The role of presentational structures in understanding and solving mathematical word problems. Dans C. A. Weaver, S. Mannes et C. R. Fletcher (Dir.), *Discourse comprehension. Essays in honor of Walter Kintsch* (p. 285–305). Lawrence Erlbaum.

TREFFERS, A. (1986). *Three dimensions. A model of goal and theory description in mathematics education: The Wiskobas project*. Kluwer Academic.

WOZNIAK, F. (2012). Analyse didactique des praxéologies de modélisation mathématique à l'école : une étude de cas. *Education et Didactique*, 6(2), 63–86.

YVAIN-PRÉBISKI, S. (2018). *Etude 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 Montpellier, Montpellier]. https://tel.archives-ouvertes.fr/tel-01956661v1/document

ZIEFFLER, A., GARFIELD, J., DELMAS, R. et READING, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.

CHARLOTTE DEROUET

LISEC UR2310, Université de Strasbourg, UL, UHA

charlotte.derouet@inspe.unistra.fr