Une approche fréquentiste des probabilités et statistiques en première année d'Université au Vietnam dans un cursus non mathématique

Jean-baptiste Lagrange[1] and Bui Ahn Kiet[2]

[1]LDAR, Université Paris-Diderot, France, jb.lagrange@casyopee.eu;
[2]Cantho University, VietNam

Au Vietnam, l'enseignement des probabilités donne une place importante à une approche « classique » basée sur la loi de Laplace, et les statistiques inférentielles sont vues comme une application. De plus, l'enseignement ne tient pas compte de l'utilisation des logiciels, génériques ou consacrés aux statistiques. Ceci entraîne des difficultés ayant une signification particulière pour les étudiants d'autres disciplines que les mathématiques, qui doivent être préparés à la compréhension de phénomènes aléatoires et à une approche statistique dans la vie réelle. Cet article vise à évaluer la possibilité d'introduire des innovations «viables» dans l'enseignement des probabilités et statistiques au Vietnam au niveau universitaire, ainsi que les améliorations que ces innovations apportent à la compréhension des étudiants.

Keywords: approche fréquentiste, probabilités et statistiques, R, tableur, modèles.

INTRODUCTION

Au Vietnam comme dans d'autres pays, les manuels d'enseignement et le programme en probabilités et statistiques à l'université sont élaborés selon cet ordre: l'analyse combinatoire, la formule de Laplace, les fréquences relatives dans des essais répétés, le calcul sur des événements, les variables aléatoires et puis les statistiques descriptives et inférentielles. La formule de Laplace s'applique à une loi de probabilité uniforme sur un univers fini. Elle permet de calculer la probabilité d'un événement comme le quotient du nombre d'éventualités dans cet événement par le nombre total d'éventualités dans l'univers. Dans le cas où la distribution modélise un phénomène aléatoire (par exemple un lancer de dé) la loi de probabilité est présupposée à partir du principe de raison suffisante[1]. L'étude des fréquences relatives dans des essais répétés débouche sur la loi des grands nombres selon laquelle les caractéristiques statistiques d'un échantillon aléatoire se rapprochent d'autant plus des caractéristiques de la population dont l'échantillon est issu, que la taille de l'échantillon augmente. Formalisée en théorème, la loi forte des grands nombres est le fondement théorique des statistiques inférentielles.

L'approche choisie par les manuels et les programmes, privilégie l'analyse combinatoire comme fondement du calcul de probabilité, donnant peu de place à l'aléatoire. Dans le chapitre sur les fréquences relatives, un nombre limité d'exemples d'expériences répétées est donné, mais il n'y a pas de tâches qui confronteraient les étudiants à l'aléatoire, telles que l'observation des fluctuations de façon à estimer la taille d’échantillon nécessaire pour l’estimation d’un caractère avec une incertitude donnée. Ceci entraîne des difficultés :

- Même pour des tâches «classiques», de nombreux étudiants ont des difficultés avec l'analyse combinatoire et ne peuvent donc calculer avec succès les probabilités théoriques. Ils n’ont aucun moyen pour confronter leur raisonnement et leurs calculs à la réalité, parce qu’il n’y a pas de lien réalisé avec des fréquences empiriques. La relation entre les fréquences empiriques et la probabilité théorique d'un événement reste superficielle.
- Le lien entre probabilités et statistiques inférentielles est fait sur un plan théorique très difficile pour ces étudiants, et dans la plupart des cas ils peuvent seulement appliquer des formules sans comprendre les théories sous-jacentes.
- Les étudiants ne sont pas vraiment confrontés à des phénomènes aléatoires, et manquent ainsi des occasions pour une compréhension profonde de la théorie des probabilités et de ses liens avec les questions statistiques de tous les jours, comme par exemple la taille de la population à retenir pour un sondage.
- Les manuels et les programmes ne tiennent pas compte de l'utilisation croissante des logiciels, génériques ou consacrés aux statistiques, les étudiants s’éloignent donc de la réalité pratique dans leur monde professionnel.

Ces difficultés ont une signification particulière pour les étudiants d’autres disciplines que les mathématiques, qui doivent être bien préparés pour la compréhension de phénomènes aléatoires et une approche statistique dans la vie réelle et le monde professionnel, plutôt que d'une approche purement mathématique. L’objectif général de l’étude est de remédier à ces difficultés et, plus précisément, d'évaluer la possibilité de la mise en œuvre de deux innovations, l'une liée à une "approche fréquentiste", et l'autre à la simulation, à l'intérieur d'un cours existant destiné aux étudiants d'autres disciplines que les mathématiques.

REVUE DE LITTERATURE ET BUTS DE L’ETUDE

Il existe des preuves de recherche solide que les conceptions erronées au sujet des probabilités ne disparaissent pas à la suite de l'enseignement traditionnel centré sur des définitions formelles, des règles et des procédures. Bien que les étudiants puissent apprendre des règles et procédures en probabilité et même dans le cas où ils obtiennent des réponses exactes aux tests mathématiques, ces mêmes étudiants se méprennent souvent à propos des idées et des concepts de base et ignorent souvent les règles lors de leur propre jugement sur les événements incertains. Surtout pour les étudiants universitaires, les conceptions erronées peuvent aussi apparaître comme le résultat de l’enseignement reçu sous la forme de théories explicites utilisées en dehors de leur champ d’application. Une intervention pédagogique conçue spécifiquement pour éliminer les conceptions erronées des étudiants sur les probabilités est nécessaire pour que des améliorations tangibles et stables dans les concepts des étudiants puissent être obtenues. La participation active des étudiants dans la construction des connaissances, la confrontation avec de grands échantillons, l'utilisation de la simulation informatisée sont des outils pour cet objectif[2].

Nous retenons aussi que l'idée de fréquence relative d'un événement et les modèles de comportement à long terme jouent un rôle très important. Nous souhaitons introduire une « approche fréquentiste » de la probabilité d'un événement basée sur l'observation de la convergence des fréquences relatives pour cet événement dans les essais aléatoires répétées de façon à sensibiliser les étudiants à la fois sur le caractère imprévisible à court terme des phénomènes aléatoires et la régularité à long terme que décrit la probabilité. De plus, cette approche fréquentiste sera utile en fournissant une approximation d'une probabilité réelle, grâce à un échantillon suffisamment grand : les étudiants seront en mesure de confronter des résultats empiriques obtenus par l'observation des fréquences, et les résultats théoriques obtenus par l'approche classique. L'observation d'une divergence pourra les aider à prendre conscience d'une conception erronée.

L'utilisation de la simulation est cohérente avec une approche fréquentiste, parce que l'observation de la convergence des fréquences relatives n'est généralement pas possible sur les données réelles. En outre, la simulation sert à construire un modèle d'un phénomène aléatoire et donc une profonde compréhension de la situation aléatoire sous-jacente, et prépare ainsi une «approche classique». Selon la littérature, par exemple Garfield, Chance & Snell (2000), un tableur comme Excel, et un langage de programmation comme R offrent des possibilités pour cela. Avec un tableur les étudiants peuvent construire simplement une simulation et en visualiser directement les résultats. Ils peuvent facilement produire et utiliser de nombreux échantillons. R est un logiciel professionnel et donc son usage aide les étudiants à se préparer pour de nouvelles utilisations dans leur vie professionnelle. Par rapport à un tableur, la simulation dans R peut être effectuée sur de très grands échantillons. Les questions de recherche découlent de ces considérations :

- Quelles sont ces tâches et les techniques liées à une "approche fréquentiste», et comment améliorent-elles l'enseignement/apprentissage des probabilités, en particulier en ce qui concerne les conceptions erronées et les modèles inadéquats des situations aléatoires ?
- Comment connecter cette approche fréquentiste et l'approche classique? Surtout, comment construire un milieu (Brousseau 1997) adéquat et mettre en œuvre des contrats didactiques appropriés à cette connexion?

MÉTHODOLOGIE

Notre étude est exploratoire, en ce sens que nous avons mis en place des séances expérimentales, conçues et menées par le second auteur de cet article dans le cadre de sa thèse, et que ces séances servent aussi bien à mettre en évidence des apports des innovations mises en place, qu'à repérer des occasions manquées qu'il s'agit d'analyser comme des points d'attention pour de futures implémentations.

Les séances expérimentales

Les séances ont été menées à l'Université de Can Tho dans une classe de 30 étudiants de différents cursus : économie, ingénierie, informatique, agriculture... Ces étudiants suivent le cours probabilités et statistiques de première année organisé sur une série de 45 leçons de 50 minutes sur 15 semaines. Trois leçons ont été regroupées pendant quatre semaines pour organiser quatre séances expérimentales de deux heures et demie. Les quatre séances expérimentales correspondent à quatre problèmes aléatoires. Dans chaque problème, nous considérons un ou plusieurs événements, ou une variable aléatoire et posons une question liée à la probabilité de ces événements ou à l'espérance de la variable. Nous choisissons ces problèmes parce que dans chacun d'entre eux, les données en jeu ne sont pas évidentes pour les étudiants et peuvent être un sujet de débat. Pour chaque problème, la construction d'une simulation doit apporter deux contributions différentes: d'abord pour obtenir des fréquences relatives ou des moyennes d'un événement pour un nombre donné d'essais, et ensuite pour développer un modèle qui sera aussi utile pour un calcul théorique de la probabilité. La mise en œuvre commune des quatre séances suit cette analyse : les étudiants font d'abord quelques expériences pratiques pour se familiariser avec le problème puis l'enseignant lance une discussion sur la question en jeu. Ensuite, les étudiants construisent une simulation et observent les fréquences ou les moyennes et leurs fluctuations en vue d'étudier la question. La dernière étape est le calcul mathématique classique pour confirmer la réponse.

Les problèmes

1. Le problème des deux dés : Soit X la somme des nombres apparus après le lancement de deux dés équilibrés. La question est de comparer $P(X = 7)$ et $P(X = 8)$. Ces chiffres ont été choisis parce un malentendu commun est que ces probabilités sont égales, la paire (4; 4) étant comptée deux fois dans la numérotation des événements élémentaires.

2. Le problème du lapin et de la tortue : Lancer un dé juste, si le nombre 6 apparaît alors c'est la victoire de lapin; si le numéro 6 n'apparaît pas alors la tortue fait un pas. Continuer à lancer le dé jusqu'à la victoire du lapin ou jusqu'à ce que la tortue fasse 6 pas et gagne. La question est de décider qui a plus de chance de gagner. La probabilité de gain de la tortue est proche de 1/3, mais ne peut être estimée sans un calcul ou une simulation. Le nombre de pas impose une simulation plus élaborée que dans le cas précédant.

3. Le problème des canards (Engel, 1990) : Il y a cinq chasseurs et cinq canards. Chaque tir d'un chasseur vise un des cinq canards au hasard et ne le rate jamais. La question est la moyenne du nombre de canards survivants quand les cinq chasseurs tirent simultanément. Le nombre 5 a été choisi afin que la taille de l'univers soit suffisamment grande pour que le simple comptage des éventualités soit irréaliste.

4. Le problème de Monty Hall : Il y a trois portes, derrière l'une d'elle se trouve une voiture et derrière les deux autres se trouve une chèvre. Le joueur est autorisé à choisir l'une des trois portes et celle-ci n'est pas ouverte. Puis l'animateur ouvre l'une des deux autres portes. Dans le cas où la voiture est derrière une des deux portes non choisies, l'animateur ouvre toujours la porte sans la voiture. Ensuite, le joueur peut rester sur son premier choix ou changer pour sélectionner l'autre porte qui n'est pas ouverte. La question est de décider si, pour gagner la voiture, le joueur a avantage à garder son choix original ou à passer à l'autre porte. Il s'agit d'une situation complexe mettant en jeu un raisonnement sous hypothèse. Les personnes non averties ont tendance à penser que changer de porte ne donne pas un avantage.

Logiciels et simulation

Pour présenter l'usage des logiciels dans les séances expérimentales, nous utilisons la notion de « fonctionnalités didactiques d'un outil numérique » (Cerulli et al., 2006) définie par trois éléments clés :

1) Un ensemble de fonctions de l'outil. Comme tous les tableurs, Excel est un logiciel basé sur des formules. Nous pouvons voir l'évaluation des formules dynamiquement mises à jour sur l'écran de l'ordinateur. La répétition se fait par la copie des formules dynamiques selon les lignes ou colonnes. Il n'y a pas de façon commode d'exprimer une condition d'arrêt et le nombre de répétitions est limité dans la pratique. Excel possède également des fonctions standard pour les statistiques descriptives et un générateur pseudo aléatoire. L'affichage graphique est utile pour la présentation des données recueillies à partir de simulations. Excel a également la touche F9 qui est utilisée pour recalculer toutes les formules, en donnant de nouvelles valeurs pour les données pseudo aléatoires. R est un langage de programmation fonctionnel. R peut être utilisé par des lignes de commande ou par programmation de fonctions structurées. Dans R la structure principale de la répétition est la boucle for. R dispose de fonctions pour les vecteurs (moyenne, unique) utiles pour les traitements statistiques. Le générateur pseudo aléatoire est appelé par l'intermédiaire de la fonction sample(a: b, n, repl). Le paramètre repl contrôle le «remplacement». Le rééchantillonnage peut être effectué en réexécutant.

(2) un objectif éducatif. Conformément à la loi des grands nombres, Excel et R peuvent être utilisés pour simuler une situation aléatoire et recueillir des données afin d'approcher la probabilité théorique d'un événement ou la moyenne théorique d'une variable aléatoire en augmentant la taille de l'échantillon. Les résultats de la simulation dans Excel ou R seront utilisés pour confirmer ou infirmer le calcul théorique de probabilité et aider les étudiants à mieux contrôler ces calculs. C'est la dimension pragmatique. La simulation dans le contexte de l'enseignement peut avoir aussi une dimension conceptuelle. Premièrement, les étudiants peuvent prendre conscience de la diminution de la fluctuation des fréquences relatives ou des moyennes statistiques lorsque l'échantillon augmente de taille et comprendre les probabilités ou les espérances théoriques comme des limites. Deuxièmement, pour

effectuer une simulation, les étudiants devront construire un modèle leur donnant un meilleur contrôle dans le calcul des probabilités théoriques.

(3) Les modalités d’utilisation de l'outil. Dans chaque séance expérimentale, les étudiants feront d'abord une simulation avec Excel, puis une simulation avec R, avant de faire un calcul théorique. Ils vont ainsi construire un premier modèle sur tableur et obtenir des résultats sur des échantillons limités, puis adapter le modèle ou en créer un nouveau avec R et obtenir des résultats sur des échantillons plus grands. Les modèles seront utiles pour le calcul théorique, et les résultats des simulations seront confrontés aux valeurs obtenues par calcul théorique.

EXPERIMENTATION

Pour chaque séance, nous donnons quelques indications sur le déroulement, puis une analyse résumée des observations réalisées par Bui (2015).

Séance expérimentale 1 : La somme des deux dés

L’enseignant amène les étudiants à se familiariser avec la réalisation d’une expérimentation concrète avec deux dés et d'une collecte des données. Il exploite les données recueillies afin que les étudiants voient la nécessité d'une plus grande taille de l'échantillon afin de pouvoir comparer les probabilités d’obtenir respectivement 7 et 8. Puis les étudiants effectuent des simulations en utilisant le tableur Excel sous la direction de l’enseignant. La fluctuation des fréquences conduit les étudiants à voir la nécessité d'une plus grande taille de l'échantillon. Puis, l'enseignant présente le logiciel R pour la simulation. A la fin de la séance, il fait discuter les étudiants sur l’intérêt de la simulation.

Les étudiants sont sensibles aux fluctuations et font le lien avec la taille de l'échantillon sans cependant les exploiter directement pour comparer les deux probabilités en jeu. Concernant la taille de l'échantillon lors des simulations, l’enseignant et les étudiants se limitent à deux tailles : 1000 (trop petite) et 100.000 (assez grande); l’occasion a été manquée de questionner la taille de l'échantillon, en vue de déterminer une taille optimale permettant de discriminer les deux probabilités en jeu ($P(S = 7) > P(S = 8)$ dans 95% des cas). Dans le tableur, les étudiants ont des difficultés avec la fonction *ALEA* qui appelle le générateur de nombre aléatoires : ils ne comprennent que chaque appel déclenche un nouvel appel et non une référence à la valeur antérieure. La fonction *sample* de R donnant directement un échantillon, ils ne rencontrent pas cette difficulté dans la seconde simulation.

Cette séance expérimentale témoigne des conceptions erronées des étudiants dans les différentes phases. Ces conceptions erronées sont déstabilisées par la confrontation aux fréquences obtenues par simulation sur un large échantillon, plutôt que par une réflexion sur le modèle développé pour cette simulation. En particulier, dans la dernière phase, lors du calcul de probabilités théoriques, les fréquences obtenues par la simulation sont utilisées pour invalider la conception erronée d’un espace constitué de paires (non ordonnées). Cependant, le modèle développé pour la simulation n'est pas exploité, alors qu'en considérant par exemple les données produites par le tableur, les étudiants pourraient voir que les sommes 4+3, 3+4 et 4+4 apparaissent avec la même fréquence. L'enseignant institutionnalise ce rôle pragmatique de la simulation, plutôt qu'une réflexion sur le modèle construit dans les phases de simulation.

Séance expérimentale 2 : Le lapin et la tortue

Cette séance marque une progression dans la complexité et n'est pas facile pour de nombreux étudiants. Après la première phase d'expérimentation avec un dé, l'enseignant et les étudiants ne reprennent plus la question initiale (le gagnant le plus probable) qui est ainsi oubliée au profit d'une centration sur la probabilité de chaque événement (le lapin gagne, la tortue gagne), d'abord approchée de façon fréquentiste, puis calculée de façon théorique. Ceci résulte d'un contrat didactique orienté vers la recherche de valeurs de probabilités et minimisant les questions statistiques.

Comme dans la séance précédente, les étudiants rencontrent des difficultés avec le tableur. Le modèle implémenté avec le tableur, est conforme à la situation dans le sens où le processus est arrêté après un gagnant, ce qui ne se réalise pas facilement dans l'organisation de la feuille de calcul. Cette condition d'arrêt n'est pas non plus commode à implémenter dans R, et un autre modèle est proposé qui revient à lancer le dé 6 fois, et à conclure que le lapin gagne s'il existe un 6 parmi les nombres obtenus. Ce modèle est à nouveau proposé par l'enseignant pour le calcul théorique de façon à guider les étudiants dans l'utilisation de la formule de Laplace, mais il est rapidement abandonné après qu'un étudiant propose d'utiliser la formule de multiplication des probabilités. Ainsi des modèles différents sont utilisés dans les différentes phases. Ce changement de modèle simplifie la programmation dans R et pourrait simplifier le calcul théorique, mais il ne fait pas l'objet d'une discussion en classe : le nouveau modèle n'est pas comparé au modèle précédent et l'équivalence des modèles n'est pas discutée[3]. C'est encore une occasion manquée, et comme dans la séance expérimentale précédente, le rôle pragmatique de la simulation est privilégié par l'enseignant aux dépens d'une réflexion sur les modèles utilisés dans la phase de simulation.

Séance expérimentale 3 : la chasse au canard.

Les étudiants ne rencontrent pas beaucoup de difficultés dans l'utilisation des fonctions Excel et R pour simuler. Comme dans les séances précédentes, l'enseignant souligne la nécessité d'augmenter la taille de l'échantillon appropriée à la simulation pour le passage à l'utilisation de R. De cette façon, les étudiants proposent une très grande taille d'échantillon sans qu'il y ait une discussion sur une taille appropriée. Cette analyse confirme que la simulation peut fournir un environnement riche pour la réflexion des étudiants. Cependant, il semble que l'enseignant puisse faire mieux en posant la question de la taille minimale de l'échantillon pour une précision donnée.

Les stratégies de calcul de la moyenne diffèrent dans les deux simulations : avec le tableur, la moyenne est calculée directement en globalisant les tirages (nombre total de canards tués divisé par le nombre de tirs successifs). Avec R, les étudiants calculent d'abord les fréquences empiriques de chacune des éventualités. Puis, après la présentation de la fonction moyenne par l'enseignant, les étudiants abandonnent les fréquences et calculent directement la moyenne comme avec le tableur. Pour le calcul théorique, l'enseignant conduit les étudiants à calculer la distribution de probabilité de la variable nombre de canards tués en confrontant aux fréquences empiriques obtenues avec R comme préalable au calcul de l'espérance. Le fait de ne pas discuter sur la possibilité de calcul direct de l'espérance sans passer par la distribution de probabilité est une autre occasion manquée : en effet, la discussion sur ce sujet pourrait porter l'attention des étudiants sur l'espérance mathématique dans des tirages répétés (somme des espérances dans chaque tirage).

Séance expérimentale 4 : Le problème de Monty Hall

A partir des expérimentations sans logiciels, les étudiants comprennent la situation et reconnaissent que le choix de changer de porte va apporter plus de chance de gagner. Les simulations qui suivent avec des tailles d'échantillon plus grandes aident les étudiants à éliminer les intuitions erronées et à quantifier plus précisément les probabilités de gain pour chaque choix (garder la même porte ou changer). Dans chaque cas, les simulations sont faites pour chacun des deux choix. Implicitement, le choix de la porte ouverte par l'animateur dans le cas où le joueur a choisi la porte gagnante, est équiprobable.

La première simulation avec Excel conduit à des formules assez complexes, mais les étudiants réussissent à présenter de nombreuses solutions différentes. Un étudiant remarque que dans le cas où le joueur garde la même porte, il gagne si et seulement s'il a choisi la porte gagnante au premier choix. Ceci simplifie notablement la simulation et est utilisé par la plupart des étudiants dans la simulation avec R. En revanche, pour le calcul théorique, l'enseignant engage les étudiants vers un calcul utilisant des probabilités conditionnelles conditionnées par le choix de l'animateur. Ces probabilités conditionnelles sont calculées en appliquant les connaissances académiques déjà enseignées dans le cours mais ne répondent pas en fait à la même question que celle posée et conduisent à la même valeur seulement parce que l'équiprobabilité est supposée pour le choix de l'animateur (Grinstead and Snell 2005, p.139). Une occasion de discuter de modèles, reliant la simulation et les probabilités théoriques a été manquée également dans cette situation.

DISCUSSION ET PERSPECTIVES

Tâches et techniques

La simulation apporte de nouvelles tâches associées à des questions statistiques. Le choix a été fait dans les séances expérimentales de proposer ces tâches avant les tâches classiques. Nous avons observé que les étudiants vérifient systématiquement leurs résultats par rapport aux résultats empiriques, en particulier dans la troisième séance où cinq probabilités théoriques sont calculées, et aussi pour déstabiliser les idées erronées ou les faux modèles. Cette vérification apparaît comme une technique intégrée dans des programmes d'action. Cela signifie que les étudiants articulent la simulation avec des techniques classiques déjà enseignées, afin d'obtenir un meilleur contrôle de ces techniques. La pratique de cette technique «mixte» semble avoir un effet positif relativement à des conceptions erronées : l'observation montre que les étudiants questionnent leurs modèles à partir de données de simulations. Elle a aussi des limites : un étudiant peut corriger un modèle faux afin d'obtenir une valeur théorique cohérente avec les données obtenues par la simulation sans que le modèle ainsi produit soit juste.

Milieu et contrats didactiques

Dans les séances expérimentales, la simulation peut être considérée comme un milieu à deux niveaux :

1. Un milieu d'action. Les fréquences et leurs fluctuations peuvent être considérées comme des rétroactions que les étudiants obtiennent lors de la construction et de l'exécution des simulations sur l'ordinateur, qui renforcent leur compréhension des relations entre fréquences et probabilités. Implicitement, la simulation permet aux étudiants de construire des modèles «en action» qui déstabilise les conceptions erronées. La simulation joue également un rôle en tant que milieu de l'action dans le calcul des probabilités théoriques lorsque les étudiants vérifient systématiquement leurs résultats par rapport aux fréquences empiriques.

2. Un milieu de réflexion. Nous avons vu qu'au cours de chaque séance, plusieurs modèles de la situation aléatoire ont été considérés, mais qu'ils ne sont pas discutés. Par ailleurs, nous avons vu que la fluctuation n'est pas étudiée de façon précise en regard des questions posées initialement : l'enseignant favorise en réalité une dimension pratique de la simulation, en insistant sur un échantillon de taille appropriée afin de motiver l'utilisation de R, plutôt que de discuter sur une taille pertinente à l'égard de la question statistique. Ceci montre que potentiellement, la simulation contribue à créer un milieu de réflexion, qui n'est pas exploité : l'existence de plusieurs modèles devrait être l'occasion de discuter leur équivalence ; discuter la taille de l'échantillon, en relation avec la question posée devrait être une préparation à la statistique inférentielle. Nous avons ainsi repéré plusieurs « occasions manquées » qui peuvent être interprétées comme une sous-estimation de cette dimension « réflexion » du milieu.

Les questions posés dans les 4 séances expérimentales sont problématiques pour les étudiants et motivent à construire des simulations. Cependant nous remarquons que dans ces séances, dès qu'ils passent à la simulation et aux calculs, les étudiants et l'enseignant ne se réfèrent plus à la question initiale. Ainsi, ces questions initiales génèrent une motivation afin que les étudiants cherchent des valeurs probabilistes, plutôt qu'une véritable enquête statistique. Nous interprétons cela comme la manifestation d'un contrat didactique orienté vers une utilisation pratique de la simulation pour approcher les valeurs probabilistes plutôt que pour enquêter sur la question en jeu, sous l'influence d'un contrat didactique « traditionnel » favorisant le calcul des probabilités théoriques. Cette analyse critique permet de proposer une reprise de chacune des 4 séances expérimentales de façon à mieux exploiter leur potentiel en mettant davantage l'accent sur les questions probabilistes et en prenant en compte davantage la simulation comme un milieu de réflexion sur les modèles de situations probabilistes.

REFERENCES

Batanero, C., & Sanchez, E. (2005). What is the nature of high school students' conceptions and misconceptions about probability? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 241–266). New York: Springer.

Bui, A. K. (2015) Apports de la simulation et de l'utilisation de logiciels pour l'enseignement /apprentissage des probabilités et des statistiques en première année d'Université auVietnam dans un cursus non mathématique. *Thèse de doctorat*. Université Paris-Diderot.

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht: Kluwer.

Cerulli, M., Pedemonte, B., & Robotti, E. (2006). An integrated perspective to approach technology in mathematics education, In Proceedings of CERME 4, Sant Feliu de Guíxols, Spain (pp. 1389-1399).

Engel, A. (1990). *Les certitudes du hasard*. Lyon : Aléas Editeur.

Garfield, J., Chance, B. L., & Snell, J. L. (2000). Technology in college statistics courses. In D. Holton et al. (Eds.), *The teaching and learning of mathematics at university level: An ICMI study* (pp. 357–370). Dordrecht: Kluwer.

Grinstead, C. M. & Snell, L. J. (2005). *Introduction to Probability* (2nd ed.). New York: Random House.

Parzysz, B. (2009). Des expériences au modèle, via la simulation (From experiences to a model, via simulation). *Repères-IREM, 74*, 91–103.

---

[1] « …puisqu'à cause de la similitude des faces et du poids absolument semblable du dé, il n'y a aucune raison (nulla sit ratio) qu'une des faces soit plus encline à tomber qu'une autre.. » Jacques Bernoulli, *Ars conjectandi*.

[2] Pour une synthèse sur les conceptions erronées, voir Batanero & Sanchez (2005).

[3] Pour une discussion sur l'équivalence de modèles dans la simulation d'une expérience aléatoire répétée avec loi d'arrêt, voir Parzysz (2009).