L'INTERVALLE DE FLUCTUATION D'ECHANTILLONNAGE : UNE FORMATION DIDACTIQUE A LA NOTION DE PREUVE

Jannick TRUNKENWALD, Mohamed ZORAI, Moulay BENMANSOUR

Résumé. Des expérimentations menées au lycée français d'Alger permettent d'envisager la conception d'une ressource exploitable en formation de formateurs. La thématique abordée dans cet atelier se focalise sur les aspects liés aux différents types de raisonnements pouvant être mobilisés pour justifier l'intervalle de fluctuation d'échantillonnage. Les rôles respectifs de l'informatique, et de la simulation sont au passage requestionnés en regard de questions liées à la modélisation probabiliste. L'analyse des séances exploite la théorie ETM afin de mieux identifier le processus de validation.

Introduction

L'approche dite combinatoire de la notion de probabilité peut correspondre à un travail de dénombrement permettant un emploi de la formule de Laplace pour évaluer une valeur de probabilité en appui sur une situation d'équiprobabilité. La valeur de probabilité est alors le quotient du nombre d'issues favorables par le nombre d'issues possibles. Un travail plus élaboré, faisant appel à des propriétés internes au domaine probabiliste peut aussi au besoin être mis en œuvre pour déterminer une valeur de probabilité.

L'approche fréquentiste de la notion de probabilité est quant à elle une démarche plutôt expérimentale, qui mène à une observation empirique du phénomène de fluctuation d'échantillonnage. Cette fréquence est obtenue en calculant le quotient du nombre de succès obtenus par le nombre de répétitions de l'expérience. La probabilité se manifeste alors comme une valeur limite des fréquences obtenues avec de grands échantillons. On s'appuie alors sur le référentiel de connaissances statistiques pour mettre en place un protocole permettant ce type d'observation portant sur les valeurs des fréquences.

Une bonne compréhension de la notion de probabilité semble passer par une mise en évidencede l'intervalle de fluctuation d'échantillonnage qui synthétise le lien entre ces deux approches combinatoire et fréquentiste. La mise en œuvre d'un tel objectif passe par un travail de simulation, qui est lui-même facilité par un emploi de l'informatique.

L'organisation de cet atelier CORFEM de formation de formateurs a pour objectif de requestionner la place du raisonnement et de la preuve lors de travaux menés en classe au niveau du lycée en lien avec la notion de probabilité. Notre réflexion s'appuie sur des expérimentations menées en classe à Alger au Lycée International Alexandre Dumas dans le cadre des activités du Laboratoire de Mathématiques Maurice Audin. Cette structure d'accueil, au sens de la mesure 16 du Plan Villani-Torossian, contribue au développement des pratiques professionnelles liées à l'enseignement des mathématiques dans un cadre de coopération éducative.

Nous commençons dans cet article par rappeler certains enjeux épistémologiques sous-jacents,puis nous présentons un cadre théorique permettant d'affiner notre analyse didactique, et enfinnous montrons différentes situations d'enseignement ayant été soumises à la réflexion des participants à l'atelier CORFEM, en résumant la nature des analyses et échanges qui en ont découlé.

Eléments d'une réflexion d'ordre épistémologique

Afin de décrire le lien mathématique entre les deux approches combinatoire et fréquentiste citées en introduction, nous commençons par exposer à ce qui correspond à une approche combinatoire de la notion de fréquence dans le cadre des notions enseignées au niveau du lycée.

La fréquence F_n des succès obtenus lors de n répétitions d'une expérience aléatoire peut être définie sur un univers formé de tous les échantillons possibles de taille n. Afin de nous rapprocher de la notion d'intervalle de fluctuation nous pouvons alors rechercher deux valeurs a et b telles que $P(a < F_n < b) \approx 0{,}95$ en effectuant une simple gestion des données combinatoires liées aux valeurs de probabilités de la distribution binomiale correspondant à la variable aléatoire fréquence.

Cela constitue une approche mathématiquement rigoureuse basée sur un raisonnement de typedéductif, mais qui peut donner lieu à des calculs trop laborieux pour être réalisés de manière raisonnable « à la main » (en raison du nombre de probabilités à calculer et cumuler qui peut être important). Un algorithme exploitant une fonction « loi binomiale » (rédigé ou à l'aide d'un outil tableur) peut alors faciliter le travail en dégageant les bornes a et b à l'issue d'un cumul progressif des probabilités associées à la loi de distribution binomiale en jeu. Cette approche combinatoire permet donc elle-aussi d'obtenir l'intervalle de fluctuation d'échantillonnage (et d'une manière exacte par utilisation du modèle probabiliste exact). L'usage de ce modèle exact était d'ailleurs entre les années 2012 et 2019 recommandé par l'institution française au niveau de la classe de première à travers une utilisation du tableur pour afficher le calcul des probabilités de toutes les valeurs prises par la variable X_n (ou ce qui est équivalent la variable $F_n=X_n/n$) exprimées de manière directe ou cumulée.

Par ailleurs, l'inégalité de Bienaymé-Tchebytchev peut être mise en évidence dans le cas général d'une loi finie (en minorant la variance avec le produit de $(\lambda\sigma)^2$ par la probabilité que l'écart entre X et E(X) dépasse $\lambda\sigma$). Puis cette inégalité peut être appliquée à la loi binomiale X_n , et on peut poser $\varepsilon = \lambda\sigma/n$. On obtient alors $P(|F_n-p|\geq\varepsilon) \leq p(1-p)/(n\varepsilon^2)$ qui permet d'établir un intervalle auquel la fréquence a plus de 95% de chances d'appartenir. Cela permet aussi de mettre en évidence la loi faible des grands nombres en faisant tendre n vers l'infini : $\lim P(|F_n-p|\geq\varepsilon)=0$. Cette approche présentée dans les documents d'accompagnement du programme de 1ère dès 2012, peut désormais s'avérer utile en tant que justification formelle générale pour l'idée de fluctuation d'échantillonnage au niveau de la classe de terminale. Cela peut se faire dans le cadre d'une situation de tâche probabiliste traitée en classe. Il s'agit de s'appuyer sur la loi binomiale à partir de la notion de variable aléatoire vue en 1ère, et des notions de dénombrement au programme de terminale. La loi de concentration probabiliste explicitée ci-dessus peut alors être introduite an appui sur l'inégalité de Bienaymé-Tchebytchev conformément au curriculum visé au niveau terminale. Mais en raison de la majoration intervenant au niveau de l'inégalité, cette approche fournit un intervalle moins précis, qui ne fait que contenir cet intervalle de fluctuation à 95% qui nous intéresse.

Dans son travail de thèse, Nechache (2016) avait mis en évidence l'absence au niveau 2nde d'un référentiel théorique probabiliste qui permettrait d'aborder cet intervalle de fluctuation par un raisonnement de type déductif. L'approche de cet intervalle de fluctuation reste en effetà ce stade, expérimentale et basée sur l'observation des fréquences d'échantillons de même taille. Chaque valeur de fréquence correspond à un échantillon obtenu lui-même en répétant uncertain nombre de fois l'expérience aléatoire considérée. Il s'agit bien d'une démarche empirique basée sur un raisonnement de type inductif. De plus, le coût temporel correspondantà la nécessité de répéter un grand nombre de fois la même expérience entraîne un besoin d'automatisation. Cette démarche est « facilitée » par un usage de l'informatique basé sur uneexploitation d'un générateur pseudo-aléatoire.

Le programme de seconde s'est adapté dès 2019 en inscrivant dans les attendus une approche algorithmique, programmée en langage Python, permettant d'évaluer le pourcentage de fréquences contenues dans cet intervalle de fluctuation.

Un environnement numérique (tableur, environnement de programmation, ou environnement de simulation préprogrammé) permet de réduire le temps dédié à la répétition de l'expérience aléatoire, en automatisant un protocole expérimental. Le savoir-faire spécifique à une utilisation d'un environnement numérique peut cependant être source de difficultés supplémentaires (algorithmique, syntaxe du langage machine) pour les élèves, et peut aussi poser un souci d'ordre métacognitif (voire d'ordre cognitif) aux enseignants. Tout cela nécessite l'usage de gestes, de syntaxes, et de modes de pensées spécifiques.

D'un point de vue purement algorithmique, nous pensons que l'approche fréquentiste est baséesur des protocoles expérimentaux permettant de pratiquer une observation empirique de phénomènes liés au hasard. L'enchainement de schèmes, directs ou procéduraux, imbriqués les uns dans les autres s'apparente en effet à ce qui est formalisé par écrit lors de la rédaction d'un programme de simulation.

D'un point de vue épistémologique, nous pouvons aussi considérer une « pensée algorithmique » (Modeste, 2012 et Laval, 2018) dans le cadre d'un travail mathématique portant sur les probabilités et les statistiques. Nous pensons aussi que le travail de rédaction de l'algorithme de simulation correspond à une étape supérieure de formalisation des différents protocoles statistiques en jeu : 1) recueil des résultats des expériences aléatoires simulées (série statistique), 2) interprétation et le pointage des succès (série de 1 ou de 0 dont la moyenne est la fréquence) et calcul de la somme puis de la fréquence (dont on perçoit clairement ici le lien avec la moyenne de réponses apportées par un variable aléatoire de loi de Bernoulli), 3) répétition de tout ceci pour obtenir une liste de fréquences simulées (constituant une nouvelle série statistique) qui est finalement représentée sous forme d'un nuage de points. Cette approche fréquentiste de la valeur de probabilité est basée sur un raisonnement de type inductif.

Cadre théorique

Espace de Travail Mathématique

Afin de cerner au-mieux les éléments du référentiel théorique opérant au cours du travail mathématique mené par les élèves nous allons devoir identifier leur rôle dans le processus de validation de ce travail par une preuve. Pour cela nous devrons prendre en considération les signes et symboles permettant un traitement mathématique au sein de chaque système de représentation, ainsi que la conversion du travail mathématique d'un registre sémiotiques mis en œuvre, à l'autre. Enfin, en lien avec l'ensemble de ce processus nous devrons identifier et distinguer la dimension instrumentale basée sur l'usage de schèmes d'actions, de schèmes mentaux, ou de boites noires, que l'on nommera artefacts. Cette dimension instrumentale n'appartient en effet ni à la discipline des mathématiques en elle-même, ni à son expression visuelle, mais participe du travail mathématique en contribuant à la connaissance. Noussommes de ce point de vue amenés à exploiter l'Espace de Travail Mathématique (ETM) tel qu'il a été défini par Kuzniak (2011). Les outils technologiques (*artefacts*), sémiotiques (*representamen*), et théoriques (*référentiel*) du plan épistémologique, peuvent alors être exploités à l'aide de schèmes appropriés. On parle dans ce cas respectivement de *genèses instrumentale*, *sémiotique*, et *discursive*. Ces genèses respectives produisent dans le plan cognitif des *constructions*, des *visualisations*, et des *preuves*. Enfin, l'activation de deux genèses peut entraîner une circulation entre les dimensions qui les portent, ce qui peut parfois être rapprochée de l'idée de modélisation.

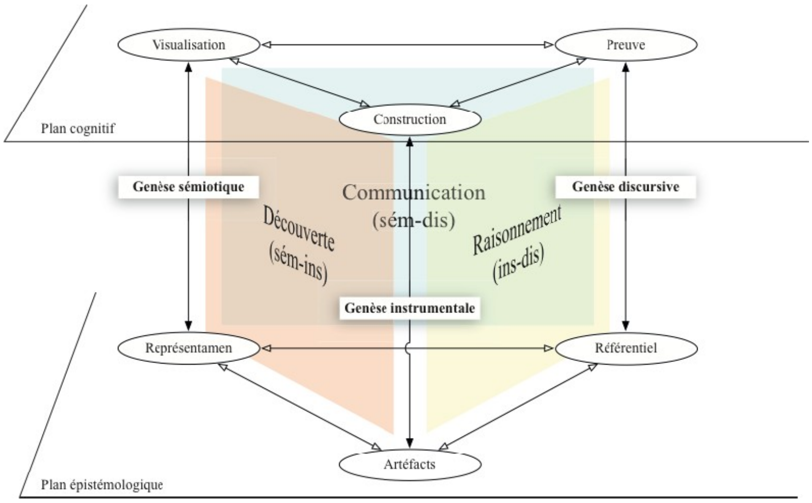

Figure 1. L'Espace de Travail Mathématique (Kuzniak & Richard, 2014)

Concernant la dimension discursive de l'ETM, un référentiel théorique issu d'un domaine mathématique particulier fournit un ensemble de propriétés, de théorèmes, de définitions, d'axiomes, qui peuvent être exploités pour établir une preuve mathématique. Lorsque c'est le cas on parle d'une genèse discursive. Un nouvel élément de connaissance peut alors éventuellement intégrer le référentiel théorique de l'ETM considéré. Cela enrichit alors le savoir de référence accessible au niveau du plan épistémologique pour ce sujet cognitif.

Concernant la dimension sémiotique de l'ETM, des signes visuels sont affectés à un rôle de signifiant correspondant à son signifié mathématique. Cette démarche intellectuelle fait partie intégrante de la culture du mathématicien. La pratique des mathématiques permet à une grande variété de tels signes de s'agréger sur le plan épistémologique en embarquant leurs propres propriétés opératoires. Il s'agit aussi de prendre en compte le processus cognitif de visualisation associé à un ensemble de manipulations de signes et outils de représentation pour le travail mathématique. Lorsque c'est le cas on parle de genèse sémiotique.

Concernant la dimension instrumentale certains objets physiques ou mathématiques, nommésartefacts, sont associés à des schèmes d'actions, et peuvent intervenir dans le travail mathématique en tant qu'outils, sans que le questionnement de l'élève porte sur la structure interne ou le fonctionnement de ces mêmes objets. Lorsqu'un tel artefact est instrumentalisé dans un but précis, la démarche de construction qui en ressort peut alors donner lieu à un automatisme qui s'agrégera au plan épistémologique en tant que nouvel artefact. D'un point de vue cognitif ces constructions sont parfois répétées afin de fournir des données empiriques, ou servent de la même manière dans différentes situations, à la manière de « recettes » automatisées. On parle alors d'une genèse instrumentale.

Exemple d'analyse didactique

Pour établir l'intervalle de fluctuation en 2$^{\text{nde}}$, l'absence de justification formelle disponible dans le référentiel théorique des élèves a été abordée par Parzysz (2009) :

> L'accès à la notion de modèle qui est une finalité visée à terme par le cycle terminal, peut être préparé par l'étude et la simulation d'expériences aléatoires diverses correspondants au même modèle probabiliste. La comparaison des procédures, des tableaux et des hypothèses sous-jacentes doit permettre aux élèves de se convaincre de l'analogie que présentent ces expériences sous leurs apparences diverses, et de déboucher sur la notion de schéma d'expérience, constituant en quelque sorte une classe d'équivalence d'expériences aléatoires. (Parzysz, 2009, p. 102).

Nechache (2016) aborde aussi ces questions en présentant le cycle de modélisation d'un modèlemathématique de type numérique, l'expérience réelle est d'abord simplifiée sous forme d'une expérience aléatoire permettant d'envisager un premier modèle probabiliste réel. Celui-ci est alors éventuellement traduit en modèle numérique pour la simulation. L'exécution de ce programme de simulation peut donner une réponse, qui est d'abord interprété par rapport au modèle réel, puis interprété en regard de l'expérience aléatoire. Nechache souligne cependant une difficulté liée à ce type de modèle numérique :

> Dans l'enseignement secondaire, la construction du modèle probabiliste de type numérique est basée sur des connaissances qui ne sont pas disponibles dans l'espace de travail personnel des élèves. C'est pourquoi, la construction de ce modèle est habituellement prise en charge par le professeur qui laisse aux élèves l'exécution de la simulation. (Nechache, 2016)

Nous souhaitons compléter cette réflexion liée à la problématique spécifique d'une réalisation du programme de simulation qui serait conjointe à son exploitation d'un point de vue probabiliste. Nous formulons l'hypothèse que le modèle final simulant la fluctuation d'échantillonnage est élaboré dans de meilleures conditions si on élabore en classe plusieurs cycles de modélisations abordant successivement l'expérience aléatoire, la construction d'un échantillon, puis la représentation des fréquences de succès associées à une liste d'échantillons de même taille.

En lien avec la phase pré-expérimentale associée au sujet de thèse de l'un des auteurs portant sur l'introduction de la notion de probabilité, nous souhaitons décrire le schéma d'expérience d'une approche fréquentiste suivant les trois cycles de modélisation précités. La situation mathématique considérée est très classique :

Enoncé :

JEU A : On lance un dé à 6 faces. On gagne si on obtient au-moins 5.
JEU B : On lance deux dés à 6 faces. On gagne si la somme des résultats vaut au-moins 9. Pour quel jeu est-il le plus facile de gagner ?

Remarque : Présentation de l'analyse a priori.

Concernant la tâche choisie, nous présentons donc ci-après notre analyse a priori de ce travail mathématique à l'aide d'un tel schéma éclaté de l'ETM suivant ses projections dans les domaines en jeu. Afin d'éviter une surcharge visuelle, nous représentons les trois projections de l'ETM par une vue du dessus. Sachant que le passage à une figure en deux dimensions nous permettra encore de présenter à l'aide de points et de traits mis en gras les différentes genèses Discursive (D), Instrumentale (I), et Semiotique (S) ainsi que les circulations dans les plans situés entre les axes de ces genèses.

Nous présentons cette analyse a priori suivant 3 niveaux de conception et exploitation, qui peuvent aussi être interprétés comme des cycles successifs de modélisation.

1er niveau : Modèle numérique simulant l'expérience aléatoire.

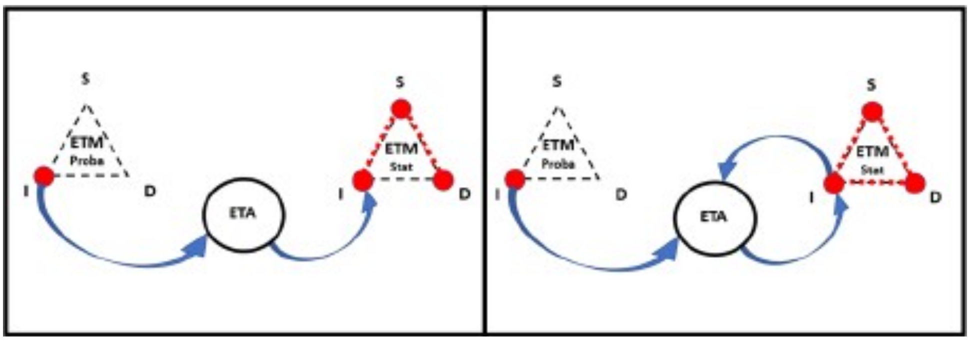

Figure 2 : Analyse par la théorie ETM du 1er niveau de conception/exploitation

- Lancement d'une genèse instrumentale dans l'ETM_{Prob} avec l'idée de tester l'expérience aléatoire. L' ETM_{Prob} entre en interaction avec l'ETA pour simuler l'expérience aléatoire à l'aide du générateur pseudo-aléatoire Alea(). On instrumentalise dans l'ETA l'artefact pour construire l'expérience aléatoire simulée : Alea(1,6)+Alea(1,6).
- Répéter l'expérience aléatoire en relevant les résultats amène à exécuter plusieurs fois l'instruction (interaction de l'ETA vers l'ETM_{Stat}) qui sert d'artefact dans une genèse instrumentale de l' ETM_{Stat} pour construire une série de résultats.
- Dans l'ETM_{Stat} la dimension sémiotique est activée pour organiser visuellement les données (à partir de signes du type « dépouillement » et/ou « tableau »), puis la dimension discursive pour l'idée de mobiliser à partir du référentiel la notion de fréquence des succès obtenus à l'issue du protocole expérimental. Ces outils sont mobilisés par le plan cognitifpour alimenter la genèse instrumentale (pas de genèse dans les dimensions S et D de l'ETM_{Stat}).
- L'idée d'automatiser l'ensemble du processus statistique réactive toutes les dimensions de l'ETM_{Stat} (expérience répétée, gérer les données, notion de fréquence) et engendre uneinteraction en retour de l'ETM_{Stat} vers l'ETA.

2ème niveau : Modèle numérique simulant une fréquence.

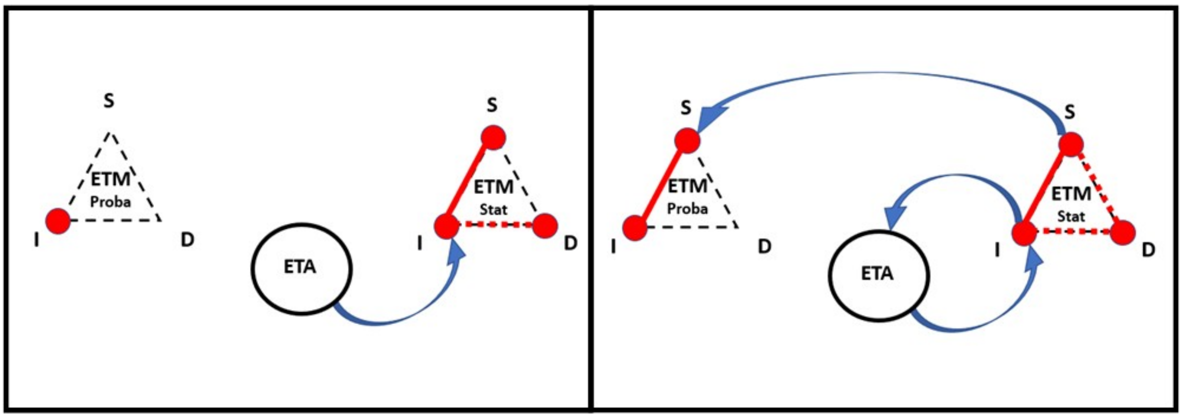

Figure 3 : Analyse par la théorie ETM du 2ème niveau de conception/exploitation

- L'objet « protocole statistique de simulation d'une fréquence » est repris dans l'ETA en tant qu'outil puis traduit sous la forme d'un objet du type « algorithme / feuille de calcul » confectionné pour simuler le calcul d'une fréquence.
- Plusieurs exécutions de l'objet « algorithme / feuille de calcul » mènent à collecter des valeurs de fréquences, et engendre une interaction de l'ETA vers l'ETM_{Stat} . L'objet « algorithme / feuille de calcul » devient un artefact qui engendre une genèse instrumentaledans l'ETM_{Stat} pour construire un nouveau protocole expérimental.
- Pour observer les résultats de fréquences on lance une représentation visuelle. La définition du repérage cartésien disponible dans le référentiel est mobilisée depuis du plancognitif, activant la dimension discursive dans l'ETM_{Stat}.
- Le registre graphique est exploité dans une genèse sémiotique de l'ETM_{Stat} pour visualiserle nuage des points fréquences. Ce lien renforce la genèse instrumentale, d'où une circulation I-S dans l'ETM_{Stat}.

- Cette approche du type « papier » (ou par saisie dans un éditeur de graphique), donne unaperçu du nuage de points et de la fluctuation d'échantillonnage. La visualisation statistique du phénomène correspond à un phénomène lié au hasard. Ce qui engendre uneinteraction entre l'ETM_{Stat} et l'ETM_{Prob} (fibration S-S).
- Ce lien entre deux domaines entraîne une genèse sémiotique dans l'ETM_{Prob}. Un lien estétabli avec la genèse instrumentale, en reliant graphiquement les idées d'expérience aléatoire et de fluctuation d'échantillonnage.
- L'idée d'automatiser l'ensemble du processus statistique réactive toutes les dimensions de l'ETM_{Stat} et engendre une interaction en retour de l'ETM_{Stat} vers l'ETA.

3ème niveau : Modèle numérique simulant la fluctuation.

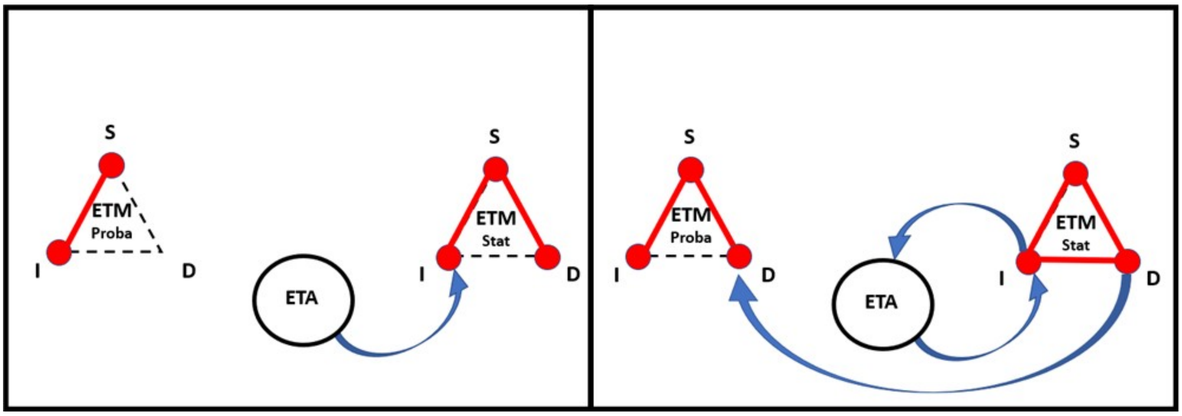

Figure 4 : Analyse par la théorie ETM du 3ème niveau de conception/exploitation

- L'objet « protocole statistique de construction du nuage de points » est repris dans l'ETA comme un objet du type « algorithme / feuille de calcul » pour construire automatiquement un nuage de points fréquences.
- L'exécution de cet « algorithme/feuille de calcul » consolide les genèses et la circulation I-S dans l'ETM_{Proba}.
- Les exécutions multiples de l'outil « algorithme / feuille de calcul » permettent d'observer l'amplitude du nuage de points (interaction de l'ETA vers l'ETM_{Stat}). On construit un protocole (genèse instrumentale dans l'ETM_{Stat}).
- Ecarter les valeurs extrêmes (repère cartésien) entraîne une genèse discursive de l'ETM_{Stat}(tracés horizontaux pour élaguer). Ces signes graphiques sont exploités dans une genèsesémiotique de l'ETM_{Stat} (localiser le nuage).
- Des modifications successives dans l'objet-outil « algorithme/feuille de calcul » (nombres de tests par échantillon, et d'échantillons) permettent de conjecturer une formule pour l'intervalle de fluctuation. Ces interactions entre l'ETM_{Stat} et l'ETA renforcent les genèses et la circulation dans dans l'ETM_{Stat} : I-S , S-D , et D-I .
- Conjecturer une formule pour les bornes de l'intervalle, nécessite un raisonnement inductif (genèse discursive dans l'ETM_{Stat}). Cette interaction entre l'ETM_{Stat} et l'ETM_{Prob}est du type fibration D-D.
- Cette genèse discursive dans l'ETM_{Prob}, prolonge la genèse sémiotique. Une circulation S-D se produit dans l'ETM_{Prob}. L'expérience aléatoire est reliée à la formule (modélisation formelle de l'intervalle de fluctuation).

Situations analysées

Les différentes situations qui suivent ont été soumises à la réflexion des participants à l'atelier CORFEM. Chacune correspond à une manière d'aborder en classe l'énoncé de problème présenté dans la partie précédente de cet article. Il s'agissait de réfléchir a priori au travail mathématique mis en œuvre en classe par les élèves pour aborder l'intervalle de fluctuation d'échantillonnage.

Exploitation statistique d'un nuage de points

Cette première situation a été expérimentée au niveau d'une classe de 2nde du Lycée International Alexandre Dumas à Alger. En lien avec l'énoncé précité, les élèves ont pu tester à l'aide de dés physiques l'expérience aléatoire. Puis à l'issue d'une mise en commun ayant permis de regrouper les données obtenues, l'outil tableur a été mis à leur disposition pour simuler des échantillons de taille 400.

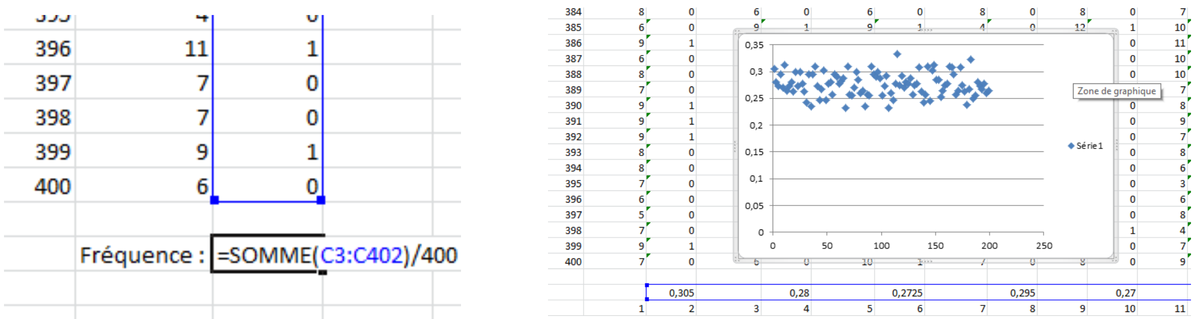

Figure 5 : Travaux menés sur tableur

Cette activité menée sur tableur correspond à l'approche fréquentiste analysée dans la partie précédente de cet article : simulation de l'expérience aléatoire, puis simulation d'un échantillon de taille 400 avec sa fréquence associée pour les succès, puis simulation d'une série d'échantillons de taille 400 permettant de construire un nuage de points représentant les fréquences obtenues. Chacune de ces trois étapes peut être assimilée à un nouveau cycle de modélisation permettant de préciser ce travail d'exploration du phénomène associé à l'expérience aléatoire visée.

Les élèves peuvent ensuite en déduire une conjecture de la valeur de probabilité de gagner à cejeu, en obtenant un aperçu clair du phénomène lié à l'intervalle de fluctuation.

Si le travail d'implémentation sur tableur de ce protocole de simulation se déroule plutôt dans le domaine algorithmique, il est aussi associé à une démarche statistique. Le fichier finalisé surtableur correspond alors du point de vue du domaine probabiliste à un artefact qui peut être instrumentalisé pour une construction plus précise du nuage de points. Dans l'exemple mis en illustration on remarque un mode de 0,27 correspondant à 13 fréquences, et une moyenne des extrema valant 0,2775 pour les fréquences élaguées à 95%.

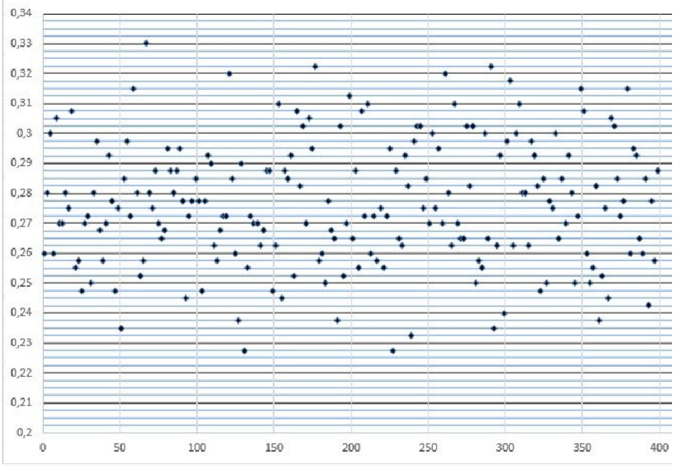

Figure 6 : 200 échantillons de taille 400 simulés au tableur (p=10/36)

Cet exemple a permis aux participants de l'atelier CORFEM de mieux appréhender l'analyse apriori de l'approche fréquentiste précédemment décrite. Les échanges ont principalement porté sur la question du raisonnement et de la preuve, qui est rattachée à la dimension discursive de l'ETM. Si en effet ce travail de conjecture relève plutôt de la dimension instrumentale d'un point de vue probabiliste, il s'agit cependant bien d'une mise en évidence rigoureuse de l'intervalle de fluctuation au sein du domaine des statistiques descriptives. L'importance de bien situer l'identification d'une démarche de raisonnement en regard d'un domaine mathématique particulier a donc été soulignée.

Exploitation de la loi binomiale

Cette deuxième situation est extraite d'une activité qui a été soumise à une classe de terminale du lycée El Malak (Alger) dans le cadre des travaux de thèse de l'un des auteurs. Il s'agit d'exploiter la loi binomiale pour dégager les bornes de l'intervalle de fluctuation.

Une première tâche, préliminaire, consiste à déterminer la probabilité que le nombre de succèssoit compris au sens large entre 3 et 7 lors de 20 parties jouées. Deux types de résolutions sontprésentées aux participants à l'atelier CORFEM : la première en utilisant sur tableur la fonctionnalité « loi binomiale », la deuxième en utilisant en mode papier crayon un raisonnement calculatoire dans le registre algébrique. Il s'agissait d'identifier les aspects de validation mathématique et de raisonnement associés à chacune de ces approches.

k	P(X=k)	CUMUL
0	0,00149	
1	0,01147	
2	0,0419	
3	0,09669	0,09669
4	0,15806	0,25475
5	0,19453	0,44929
6	0,18705	0,63634
7	0,14389	0,78022
8	0,08993	
9	0,04612	
10	0,01951	
11	0,00682	
12	0,00197	
13	0,00047	
14	9E-05	
15	1,4E-05	
16	1,7E-06	
17	1,5E-07	
18	9,6E-09	
19	3,9E-10	
20	7,5E-12	

Approximation numérique par le tableur

Expression exacte par une approche combinatoire

$$
\begin{aligned}
P(3 \leq X \leq 7) &= P(x=3) + P(x=4) + P(x=5) + P(x=6) + P(x=7) \\
&= \binom{20}{3}\left(\frac{10}{36}\right)^3\left(\frac{26}{36}\right)^{17} + \binom{20}{4}\left(\frac{10}{36}\right)^4\left(\frac{26}{36}\right)^{16} \\
&\quad + \binom{20}{5}\left(\frac{10}{36}\right)^5\left(\frac{26}{36}\right)^{15} + \binom{20}{6}\left(\frac{10}{36}\right)^6\left(\frac{26}{36}\right)^{14} \\
&\quad + \binom{20}{7}\left(\frac{10}{36}\right)^7\left(\frac{26}{36}\right)^{13}
\end{aligned}
$$

Figure 7 : Approche basée sur la loi binomiale

Les échanges lors de l'atelier mènent alors les participants à relativiser la nature des genèses enjeu au niveau de l'ETM. Cela confirme d'ailleurs les remarques associées à la situation précédentes qui donnent de l'importance au domaine mathématique considéré avant de conclure.

Dans l'approche tableur, le travail mathématique semble contenu dans la plan SEM-INS. Mais cela concerne le point de vue d'un travail mené dans le domaine des statistiques descriptives, en appui sur le tableur et ses fonctionnalités ainsi que le registre des tableaux. Cependant d'unpoint de vue probabiliste, le registre associé à ce tableau rempli de valeurs numériques masqueun travail de fond non apparent visuellement qui fait appel à différentes notions : lien avec l'expérience aléatoire, additivité des probabilités, incompatibilité entre les évènements associés à chaque ligne, variable aléatoire de loi binomiale, et lien avec la fréquence. Cela est d'autant plus exigeant que les élèves ne peuvent pas s'appuyer sur un registre plus explicite pour interpréter ces notions théoriques sous-jacentes.

Pour l'approche située dans le registre algébrique, on peut tout à fait envisager la formule de laloi binomiale comme un artefact automatisé correspondant au nombre de succès obtenus dès lors qu'une genèse sémiotique est activée en ce sens par l'écriture usuelle de calculs probabilistes citant les variables aléatoires. Tout dépend du niveau d'habitude des élèves sur cegenre de problème.

96	0,24	0,010770809	0,04987523
97	0,2425	0,012983052	0,06285828
98	0,245	0,01543903	0,07829731
99	0,2475	0,018114169	0,09641148
100	0,25	0,020970634	0,11738211
101	0,2525	0,023957312	0,14133942
102	0,255	0,027010695	0,16835012
103	0,2575	0,030056711	0,19840683
104	0,26	0,033013473	0,2314203
105	0,2625	0,035794828	0,26721513
106	0,265	0,038314493	0,30552963
107	0,2675	0,040490514	0,34602014
108	0,27	0,042249717	0,38826986
109	0,2725	0,043531819	0,43180168
110	0,275	0,044292865	0,47609454
111	0,2775	0,044507729	0,52060227
112	0,28	0,044171476	0,56477375
113	0,2825	0,043299473	0,60807322
114	0,285	0,041926278	0,6499995
115	0,2875	0,040103396	0,69010289
116	0,29	0,037896114	0,727999
117	0,2925	0,035379672	0,76337868
118	0,295	0,032635095	0,79601377
119	0,2975	0,02974498	0,82575875
120	0,3	0,026789549	0,8525483
121	0,3025	0,02384321	0,87639151

0,09641148

Par incompatibilité des évènements correspondant aux différentes valeurs possibles de nombres de succès :

P(100≤X≤120) = P(0,25 ≤ f ≤ 0,3) ≈ 0,8525483 – 0,09641148
≈ 0,75613682
≈ 75,6 %

0,8525483

Figure 8 : Exploitation d'une table binomiale pour obtenir l'intervalle de fluctuation

L'attention des participants a été attirée sur le fait qu'une adaptation de cette méthode apparait dans les anciens documents d'accompagnement officiel, de 2012, pour établir l'intervalle de fluctuation. On obtient ainsi : $P(94 \leq X \leq 129) = P(0{,}235 \leq f \leq 0{,}3225) \approx 95{,}6\%$.

Une question est cependant restée en suspens par rapport à cette approche combinatoire de l'intervalle de fluctuation, qui met en œuvre un raisonnement rigoureux basé sur une exploitation de la loi binomiale pour établir les bornes d'un intervalle de fluctuation, mais dont la réponse numérique obtenue après de nombreux calculs automatisés par le tableur, n'est qu'approchée (bien qu'extrêmement précise). Quel est le statut du point de vue de la preuve decette démarche ? Est-ce une démonstration mathématique ? Nous pouvons cependant être assurés que d'un point de vue probabiliste une genèse discursive a bien été activée (contrairement à la situation précédente donnant lieu à une approche fréquentiste et donc de nature empirique pour la valeur de probabilité).

Exploitation du principe d'équiprobabilité

En reprenant toujours avec une classe de terminale la même situation d'un jeu consistant à lancer deux dés, où l'on doit obtenir un total d'au-moins 9 pour gagner, il suffit de considérer un ensemble d'issues équiprobables puis d'appliquer la formule dite de Laplace pour conclure sur la valeur de probabilité qui sera de $10/36 \approx 0{,}2777$. Un dénombrement direct par les élèves des couples de résultats possibles pour chacun des deux dés permet d'arriver à ce résultat. L'exploitation d'un tableau cartésien 6×6, ou d'un arbre de dénombrement à 36 chemins, permettent aussi de décrire un tel univers équiprobable, en appui sur la visualisation de signes auxquels les élèves peuvent être habitués.

Dé 1 / Dé 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure 9 : Exploitation d'un tableau cartésien pour le lancer de deux dés

L'effet de la variable didactique correspondant au nombre de dés lancés, est alors soumis à la réflexion des participants à l'atelier CORFEM. On peut par exemple prendre en compte une variante de jeu où on lancerait 3 dés à six faces en considérant qu'une partie est gagnée lorsque la somme des résultats affichés sur ces 3 dés vaut au-moins 13. Pour trouver la probabilité de gagner on peut utiliser 6 tableaux cartésiens 6×6, ou un arbre de dénombrement à 216 chemins, ou encore raisonner sur les 3-uplets de résultats affichés sur chaque dé. Le nombre d'issues favorables est alors de 56 pour 216 issues possibles. Chacune des approches précitées présente déjà un coût important pour les élèves, qui s'aggraverait encore pour un jeu avec 4 dés voir davantage...

Un traitement algorithmique basé sur l'usage de boucles peut apporter une réponse efficace à ce problème posé du dénombrement des issues favorables. Un programme en Python rédigé ence sens est présenté aux participants de l'atelier (voir figure 10). Ceux-ci devaient réfléchir à lanature du travail mathématique attendu de la part d'élèves exploitant une telle approche.

La réaction des participants a pris en compte un élément de contexte essentiel : Si les élèves ne font qu'exploiter un tel programme de simulation, les élèves restent confinés dans la dimension instrumentale de l'ETM probabiliste. Si les élèves ont eux-mêmes réalisé le programme, une genèse discursive est identifiée a priori, et ceci tant du point de vue de domaine algorithmiqueque du point de vue du domaine probabiliste. La question de savoir si ce travail des élèves correspond à une véritable démonstration a cependant divisé les membres de l'atelier, car c'estplutôt l'algorithme qui effectue le travail de dénombrement proprement dit.

```
n=0
s=0

# Lancer de 3 dés à 6 faces

for i in range(1,7):
    for j in range(1,7):
        for k in range(1,7):
            n=n+1
            if i+j+k>=13:
                s=s+1

print("Nombre d'issues favorables : ",s)
print("Nomtre d'issues possibles : ",n)

print("Probabilité d'obtenir 13 ou plus :",s/n)
```

Figure 10 : Algorithme de dénombrement en Python pour le lancer de 3 dés

Approche fréquentiste sans recours à l'informatique

Une expérimentation réalisée au niveau de la classe de 1ère dans un lycée public de Brazzaville(Congo) est présentée aux participants de l'atelier. Ce travail mené en lien avec la thèse de l'un des auteurs a été réalisé en collaboration avec l'Inspection Générale de mathématiques. Il s'agitd'une approche fréquentiste menée à l'aide de dés physiques, en exploitant un support du typepapier-crayon. A ce niveau de classe, les statistiques descriptives ont déjà été abordés dès la classe de 2nde, et les élèves devront travailler sur un chapitre de dénombrement dont les applications donneront lieu à des calculs de probabilités. Une telle activité permet donc d'apporter une nouveauté aux pratiques en réinvestissant les statistiques tout en les reliant à cette notion émergente de probabilité.

Après quelques manipulations des dés permettant d'appréhender la règle du jeu, et de tester une série de parties, une série de tableaux à compléter est distribuée aux élèves, avec un graphiqueà réaliser sur papier millimétré. Cette approche permet de mettre en évidence, pour chacun des

15 binômes d'élèves impliqués dans l'activité, les fréquences de succès associées à 5 échantillons de taille 20. Une mise en commun permet alors à tous les élèves de bénéficier dans un tableau à deux entrées de l'ensemble des résultats des autres binômes pour chacun des 5

échantillons de taille 20. Un regroupement des données permet alors d'en déduire 5 échantillons de taille 300

DOCUMENT A

Tableau 1 :

Numéro du lancer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Fréquence
Résultat obtenu	6	7	9	5	8	10	8	3	3	5	10	9	10	7	7	9	10	8	4	8		
Succès (O ou N)	O	O	N	O	O	N	O	O	O	O	N	N	N	O	O	N	N	O	O	O	7	0,35

Tableau 2 :

Numéro du lancer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Fréquence
Résultat obtenu	5	5	8	9	4	7	6	8	5	6	6	5	12	3	9	5	8	11	5	5		
Succès (O ou N)	O	O	O	N	O	O	O	O	O	O	O	O	N	O	N	O	O	N	O	O	4	0,2

Tableau 3 :

Numéro du lancer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Fréquence
Résultat obtenu	5	9	5	5	3	10	4	10	9	6	12	7	7	7	11	7	11	5	3	10		
Succès (O ou N)	O	N	O	O	O	N	O	N	N	O	N	O	O	O	N	O	N	O	O	N	8	0,4

Tableau 4 :

Numéro du lancer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Fréquence
Résultat obtenu	3	9	6	5	8	11	7	5	2	7	7	6	2	9	11	6	12	7	9	10		
Succès (O ou N)	O	N	O	O	O	N	O	O	O	O	O	O	O	N	N	O	N	O	N	N	7	0,35

Tableau 5 :

Numéro du lancer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	Fréquence
Résultat obtenu	9	4	8	10	9	1	6	8	5	10	12	2	7	6	4	9	8	4	6	3		
Succès (O ou N)	N	O	O	N	N	O	O	O	O	N	N	O	O	O	O	N	O	O	O	O	6	0,3

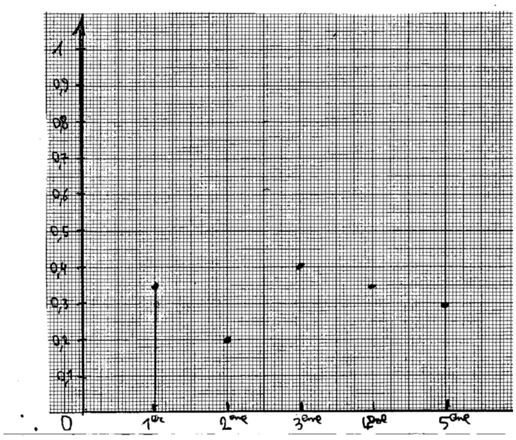

Figure 11 : Documents de travail pour les élèves (approche fréquentiste)

D'après les participants à l'atelier CORFEM c'est ce dernier travail demandé aux élèves qui permet de valider le travail mathématique mis en œuvre au cours de cette activité dans le domaine statistique. Enfin, toujours d'après les participants, une genèse discursive est également en œuvre dans le domaine probabiliste à la fin de l'activité lorsque la fluctuation d'échantillonnage due à l'augmentation de la taille des échantillons peut être observée à partir des deux graphiques.

DOCUMENT B

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Nombre total de succès	86	93	87	97	98
Fréquence	0,28	0,31	0,29	0,32	0,32

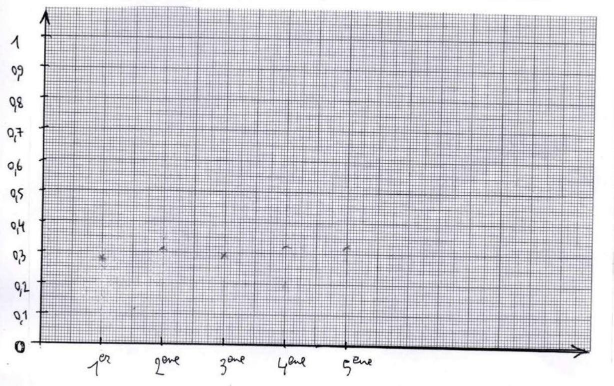

Figure 12 : Graphique représentant les fréquences des 5 échantillons de taille 300

Approche fréquentiste à l'aide de Python

Il s'agit de mettre en œuvre une approche fréquentiste à l'aide d'un programme de simulation en Python. Cette approche fréquentiste ne diffère que d'un point de vue informatique de la version tableur décrite précédemment. Cette version algorithmique amène les élèves à formaliser le protocole de simulation sous forme de consignes claires exprimées à travers un langage précis. De même que pour le tableur, trois étapes de simulation sont à mettre en œuvre :l'expérience aléatoire, la fréquence des succès pour un échantillon, et la production d'une sériede fréquence simulées. Cette liste de fréquences simulées obtenues permet, éventuellement après un classement par ordre croissant, de déterminer empiriquement un intervalle de fluctuation en déterminant le minimum et le maximum des valeurs obtenues (éventuellement après avoir élagué).

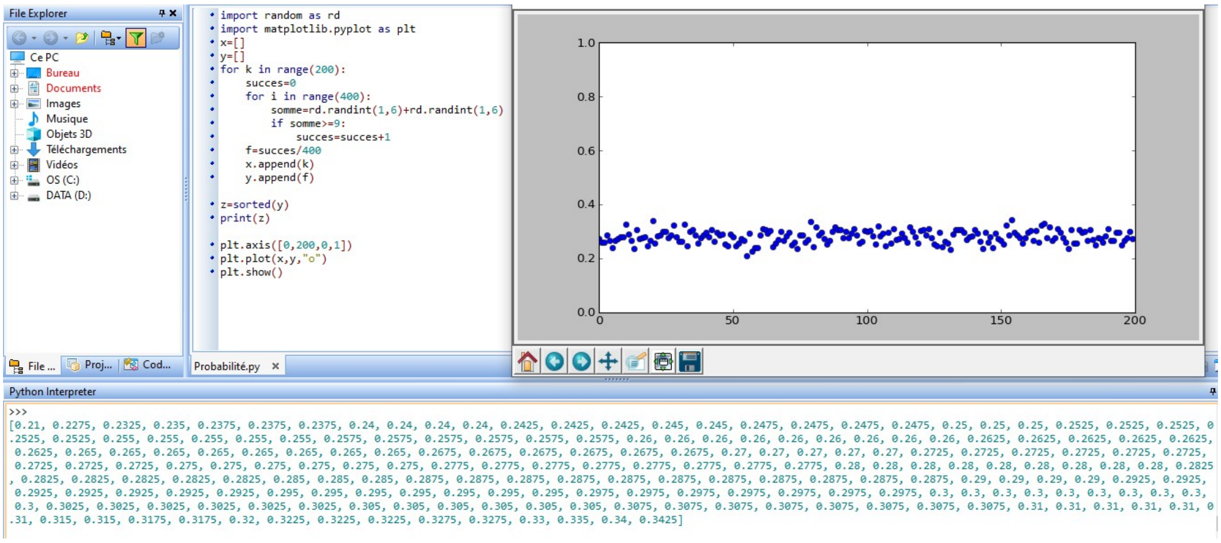

Figure 13 : Exemple de programme de simulation en Python

Les participants à l'atelier CORFEM doivent analyser des scripts en Python rédigés par des élèves. Cette réflexion permet alors de préciser la nature du travail mathématique dans le contexte spécifique d'une démarche de programmation algorithmique.

Au niveau des programmes de simulation trois niveaux d'erreurs sont identifiés : mathématiques, algorithmiques, syntaxiques.

Exemples d'erreurs mathématiques :

- randint(1,12) : cette instruction correspond à un biais d'équiprobabilité
- randint(2,12) : idem
- if n==9 : print(« gagné »)

Exemples d'erreurs algorithmiques :

- oubli d'initialisation de la variable compteur des succès
- lancers de dés en dehors de la boucle
- calcul de la fréquence à l'intérieur de la boucle
- pas de boucle

Exemple d'erreurs syntaxiques :

- import random : oubli de *
- fonction définie à l'intérieur de la boucle
- for i in range (1,2) : ne fait qu'une seule opération
- n=r+r : n'ajoute pas deux valeurs aléatoires distinctes

Les échanges menés avec les participants ont permis de dégager l'importance de la genèse discursive du point de vue du domaine algorithmique (rédaction du programme de simulation), et du domaine des statistiques descriptives (description du protocole expérimental). Par contre du point de vue probabiliste, l'approche fréquentiste reste essentiellement liée à une genèse instrumentale exploitant le programme de simulation comme artefact afin de conjecturer une valeur approchée de probabilité.

Exploitation de l'inégalité de Bienaymé-Tchebychev

Une activité permettant de faire découvrir par des élèves l'inégalité de Bienaymé-Tchebychevest présentée aux participant à l'atelier CORFEM (voir figure 14).

1) Démontrer que pour une variable aléatoire X de loi de probabilité finie :

$P(\mu\text{-}2\sigma \leq X \leq \mu+2\sigma) \geq 3/4$

2) Démontrer le cas général de l'inégalité de Bienaymé-Tchebytchev pour une variable aléatoire X de loi de probabilité finie :

$P(\mu\text{-}\lambda\sigma \leq X \leq \mu+\lambda\sigma) \geq 1 - 1/\lambda^2$ (pour tout entier $\lambda>1$)

C'est-à-dire : $P(\text{-}a \leq X\text{-}\mu \leq a) \geq 1 - \sigma^2/a^2$ (avec $a>\sigma$)

Figure 14 : Enoncé introduisant l'inégalité de Bienaymé-Tchebychev

Une deuxième activité peut exploiter cette inégalité pour établir la loi de concentration probabiliste pouvant être déduite de l'inégalité de Bienaymé-Tchebychev à partir d'une loi binomiale X_n de paramètres n et p, en posant $F_n = X_n/n$, nous donne pour tout $\varepsilon>0$:

$P(p\text{-}\varepsilon \leq F_n \leq p+\varepsilon) \geq 1 - p(1\text{-}p)/(n\varepsilon^2)$.

Pour un intervalle de fluctuation centré en p d'amplitude 2ε contenant la fréquence avec une probabilité d'au-moins 0,95 : $p(1\text{-}p)/(n\varepsilon^2)=0{,}05$ c'est-à-dire : $(10/36)(1\text{-}10/36)/(400\varepsilon^2)=0{,}05$

On obtient $\varepsilon^2=260/25920$ donc $\varepsilon=0{,}100154$. On en déduit un intervalle de fluctuation :

$J=[p\text{-}\varepsilon \; ; \; p+\varepsilon] \approx [0{,}177 \; ; \; 0{,}378]$. L'amplitude de cet intervalle (environ 0,2) est presque le doublede l'amplitude de l'intervalle directement rendu par la loi binomiale.

Les participants à l'atelier considèrent qu'une exploitation directe de cette loi de concentration,pour établir un intervalle de fluctuation, relève d'une genèse instrumentale d'un point de vue probabiliste. Tandis que le travail mathématique correspondant à la mise en place de cette loi de concentration correspond à une genèse discursive, en liaison avec le registre algébrique d'unpoint de vue sémiotique.

Conclusion

Cet atelier CORFEM permet de mettre en évidence la complexité associée aux notions de raisonnement et de validation du travail mathématique dans le contexte d'une mise en évidence de l'intervalle de fluctuation d'échantillonnage. La théorie ETM exploitée en lien avec l'identification des domaines mathématiques en jeu permet d'affiner le grain de l'analyse didactique.

Les échanges menés avec les participants à cet atelier de formation de formateur ont permis de constater la pertinence d'exploiter le domaine de recherche en didactique des mathématiques afin de préciser les concepts manipulés, et aussi de pouvoir les nommer. L'ensemble de ce travail pourra être complété afin de constituer une ressource de formation à destination des enseignants et de leurs formateurs.

Références bibliographiques

Artigue, M. (1992), Ingénierie didactique. *Recherches en Didactique des Mathématiques,* vol. 9-3,pp. 281-308.

Kuzniak, A. (2011a). L'espace de Travail Mathématique et ses genèses. *In Annales de didactique et de sciencescognitives, volume 16,* pp. 9-24.

Kuzniak, A. et Richard, P. (2014). Espace de travail mathématique. Points de vues et perspectives. *Relime,* 17(4.1), pp. 29-39.

Laval, D. (2018). L'algorithmique au lycéeentre développement de savoirs spécifiques et usage dans différents domaines mathématiques. Thèse de doctorat. Université Paris-Diderot.

Nechache A. (2016). *La validation dans l'enseignement des probabilités au secondaire.* Thèse de doctorat. Université Paris-Diderot, Paris, France.

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

Trunkenwald, J. (2018). La fluctuation d'échantillonnage : Une synergie entre les domaines probabiliste et statistique. *In Proceedings for the Sixth MWS Symposium.* Valparaiso, Chili.

Trunkenwald, J. & Laval, D., (2019). Algorithms as a discovery process in frequentist approach to prediction interval. *In Proceedings for the CERME 11, WG5a Probability and Statistics Education.* Utrecht, Nederland.

Vergnaud, G., (1989). La théorie des champs conceptuels. *In Publications de l'institut de recherche mathématiques de Rennes.* Fascicule S6. Vème Ecole d'Eté de Didactique des Mathématiques et de l'informatique. pp. 47-50.