Connexions entre Espaces de Travail, à la croisée des probabilités et de l'algorithmique

Résumé : A la croisée des probabilités et de l'algorithmique, nous menons un travail correspondant au processus de découverte de l'intervalle de fluctuation qui est analysé dans ses genèses sémiotique, discursive et instrumentale, au sein de différents domaines disciplinaires.

Mots-clés : algorithmique, espace de travail, simulation, fréquence, modélisation.

Introduction

En l'absence d'un référentiel théorique, une approche expérimentale de la notion d'intervalle de confiance, en seconde (grade 10), peut être « facilitée » par l'informatique. Nous exposons ici une situation dans le domaine des probabilités, associée à une tâche de nature géométrique. Des considérations d'ordre didactique guident notre réflexion pour le choix de la tâche et les analyses du travail mathématique et algorithmique en jeu.

Une tâche pour aborder l'approche fréquentiste

Un objectif

Nous appuyant sur une étude des fluctuations d'échantillonnage, nous pensons conduire à des prises de décisions statistiques. Nous commençons par une situation des fréquences, plus simple parce qu'elles ne mettent en jeu qu'un seul paramètre. Les probabilités et les statistiques constituent des domaines où des situations extra-mathématiques jouent un rôle fondamental pour introduire l'aléatoire (Laval, 2018). Une approche « fréquentiste » à travers des observations de fréquences relatives de caractères représentatifs de ces situations, favorise la compréhension de la notion de fréquence théorique d'un événement. Parzysz (2009) observe aussi que l'introduction de l'intervalle de confiance au niveau de la 2nde se fait sans justification formelle. De plus, la simulation sur ordinateur de situations aléatoires est un moyen utile pour « obtenir des séries statistiques de taille suffisante pour observer la convergence des fréquences relatives et des moyennes statistiques » (Laval, 2018). Selon Laval, la simulation d'un phénomène aléatoire suppose l'utilisation d'une modélisation de la situation. En 2016, Nechache avait présenté un premier cycle de modélisation d'un modèle mathématique de type numérique. Tenant compte de l'apport des spécificités de l'algorithmique, nous complétons ce modèle pour obtenir des résultats numériques après avoir travaillé sur la conception et l'implémentation d'algorithmes de simulation dans un environnement numérique. Ainsi, nous nous intéressons à la construction par des enseignants stagiaires de tels modèles. Nous supposons que l'expérience réelle est simplifiée sous forme d'une expérience aléatoire permettant d'envisager un premier modèle probabiliste réel. Celui-ci est traduit en un modèle algorithmique associé à la simulation pour passer ensuite à un modèle probabiliste numérique. L'implémentation et l'exécution de l'algorithme de simulation donne des résultats numériques qu'interprète l'utilisateur. Ces résultats sont vus par rapport au modèle réel, puis interprétés en regard de l'expérience aléatoire. L'expérience aléatoire peut être ajustée suivant l'information véhiculée à

l'expérience réelle. Un nouveau modèle réel peut être envisagé pour affiner la compréhension de l'expérience réelle, correspondant ainsi à un nouveau cycle. Suite aux observations de Lagrange et Laval (2019) sur les connexions entre espaces de travail, nous supposons que l'introduction de l'algorithmique et d'un environnement numérique dans le cycle de modélisation laisse de la dévolution à l'apprenant n'ayant pas des connaissances théoriques suffisantes dans la construction d'un modèle probabiliste de type numérique (Nechache, 2016). Nous souhaitons aborder cette problématique spécifique de réalisation d'algorithmes de simulation qui, une fois implémentés dans un environnement numérique, est conjointe à une exploitation probabiliste des résultats obtenus. Nous formulons l'hypothèse que les connexions entre modèles algorithmique et probabiliste final simulant la fluctuation d'échantillonnage sont élaborées dans de meilleures conditions si on les soumet à plusieurs cycles de modélisation, dont nous cherchons à déduire la nature.

Vers le choix d'une tâche et son expérimentation avec les premières analyses

Le processus de découverte de l'intervalle de fluctuation reste au niveau seconde, expérimental et basé sur l'observation d'échantillons qui ont la même taille. Chaque valeur de fréquence correspond à un échantillon obtenu lui-même en répétant l'expérience aléatoire considérée. Il s'agit d'une approche empirique basée sur un raisonnement inductif. De plus, le « coût en temps » correspondant à la nécessité de répéter la même expérience un grand nombre de fois entraîne un besoin d'automatisation. Cette idée est « facilitée » par l'utilisation d'un calcul basé sur une opération de générateur pseudo-aléatoire. Par ailleurs, l'approche fréquentiste peut nécessiter de considérer plusieurs domaines, tant mathématiques qu'informatiques. En effet, un environnement numérique associé à l'implémentation d'un algorithme permet de réduire le temps passé à répéter l'expérience aléatoire. Cependant, des compétences spécifiques sont nécessaires au domaine algorithmique : connaissance des variables informatiques, des structures de boucles, syntaxe du langage machine, interface utilisateur, etc. Le manque d'expertise en informatique peut entraîner des difficultés supplémentaires pour les apprenants. L'utilisation du champ numérique permet d'automatiser un protocole expérimental qui nécessite l'utilisation de gestes, d'une syntaxe et des modes de pensée spécifiques. Le choix de s'intéresser aux algorithmes permet de mieux étudier les genèses instrumentales, sémiotiques et discursives qui sont impliquées lors des travaux mathématiques et algorithmique sur les fluctuations d'échantillonnage. Nous pensons que l'écriture d'algorithmes de simulation correspond à une étape de formalisation plus élevée pour les différents protocoles statistiques impliqués, et que le rôle du « langage » prend une importance particulière dans ce sens.
Cinq groupes de trois enseignants du lycée français d'Alger doivent résoudre le problème suivant :

> Quelle est la probabilité que la distance entre deux points choisis au hasard le long d'un S-segment de droite, soit supérieure à la moitié de la longueur du S-segment ?

Nous supposons que les enseignants vont proposer un modèle algorithmique de simulation numérisée avec la mise en place d'un générateur pseudo-aléatoire de type ALEA(), retournant un nombre aléatoire entre 0 et 1 simulant la position d'un point choisi au hasard sur le S-segment. Le travail de modélisation, mettant en évidence les fluctuations d'échantillonnage, trouve ainsi son expression formalisée, à travers un algorithme de simulation. L'expérience aléatoire fournissant la distance entre deux points sélectionnés est simulée par le calcul ABS(ALEA() – ALEA()). Nous considérons une telle instruction de calcul comme un algorithme de niveau A0. Il est appelé à chaque simulation de l'expérience aléatoire. Pour une plus grande automatisation, un deuxième algorithme, de niveau A1 (Figure 1), est construit à partir de l'algorithme précédent où nous avons une répétition automatisée par la machine de l'expérience aléatoire pour obtenir directement une fréquence de réussite simulée.

Une troisième étape est envisagée avec la construction et l'implémentation en machine d'un algorithme de niveau A2 (Figure 1) permettant l'automatisation de l'agencement du nuage de points de la liste des fréquences. Nous observons que les groupes parviennent plus ou moins facilement à réaliser un modèle algorithmique pour la simulation d'un calcul de fréquence. Le passage du niveau A1 au niveau A2 correspond à une étape délicate, en particulier une écriture algorithmique d'un *protocole statistique* qui consiste à exploiter les résultats fournis par l'algorithme de niveau A1. Nous supposons qu'il s'agit d'un obstacle épistémologique lié à l'encapsulation d'un algorithme dans un autre, au sens de sous-programme, afin d'automatiser l'exécution du premier. D'autres difficultés aux connections des champs des statistiques et de l'algorithmique sont identifiées au cours de la formation, comme l'introduction d'un compteur dans l'algorithme de niveau A1. Les observations montrent l'importance des connexions entre différents domaines : probabilités et statistiques, mais aussi algorithmique, lors de l'introduction d'une approche fréquentiste de l'intervalle de fluctuation.

```
DÉBUT
  N ← 0
  E ← 0
  TANTQUE N ≤ 400 FAIRE
    N ← N+1
    SI ABS(ALEA() – ALEA()) > 0,5 ALORS
      E ← E+1
    FINSI
  FINTANTQUE
  AFFICHER E/400
FIN
```

Algorithme de niveau A1

```
DÉBUT
  K ← 0
  TANTQUE k ≤ 50 FAIRE
    EXÉCUTER PROGRAMME NIVEAU A1
    k ← k + 1
    TRACER_POINT (k ; E/400)
  FINTANTQUE
FIN
```

Algorithme de niveau A2

fig. 1 : Algorithmes de niveaux respectifs A1 et A2, écrits en langage pseudo-code

Le choix d'un cadre théorique et la question de recherche

Les articulations, entre les *Espaces de Travail Mathématique Spécifique* (ETM_s) (Kuzniak & Richard, 2014) et les *Espaces de Travail Algorithmique* (ETA) (Laval, 2018), permettent d'étudier les différentes genèses des representamen et des processus, du point de vue des ETM_S et ETA idoines des apprenants, mais aussi d'aborder les interactions entre les probabilités/statistiques et l'algorithmique. Par ailleurs, des travaux de recherche récents, réalisés dans divers champs mathématiques et informatique sur les modèles des ETM_s/ETA, permettent de prendre conscience des connexions entre les *Espaces de Travail* (Laval (2018), Lagrange & Laval (2019)). La question de recherche est de savoir comment ce cadre permet d'analyser le travail d'un apprenant impliquant une activité algorithmique et de programmation dans une approche fréquentiste de l'intervalle de fluctuation.

Un cadre théorique – Les premières analyses

Les Espaces de Travails de référence

Dans la continuité des travaux de Laval (2018, 2021), nous voyons l'algorithmique comme une discipline à part entière qui peut être en interaction avec des ETM_s, d'où la nécessité de construire et de justifier l'existence d'un ETA comme *Espace de Travail* à part entière. Par ailleurs, afin d'étudier plus finement le travail des stagiaires dans des champs mathématiques et dans un domaine non nécessairement mathématique, nous nous référons aux *Espaces de Travail Connectés* (ETC) (Lagrange & Laval, 2019), « en tenant toujours compte de manière exhaustive des dimensions [genèses] sémiotique, discursive et instrumentale » (Laval, 2021).

Les Espaces de Travail (ETM et ETA)

Selon Kuzniak & Richard (2014), le travail dans un ETM se fait autour de 3 genèses : *Sémiotique*, avec utilisation de symboles, de graphismes, d'objets concrets compris comme des signes ;

Instrumental, avec des constructions faites à partir d'artefacts : tableaux, graphes, arbres, etc. ; *Discursif*, où se situe les aspects de preuve dans un référentiel théorique. Ces genèses génèrent chacune une relation dynamique entre plans épistémologique et cognitif. Les outils technologiques, sémiotiques et théoriques du plan épistémologique sont exploités à l'aide de schémas appropriés. On se réfère respectivement à des genèses instrumentale, sémiotique et discursive, produisant respectivement des construits, des visualisations et des preuves dans le plan cognitif. L'activation de deux genèses entraîne une circulation entre les genèses qui les portent. Cela est assimilé à une idée de modélisation (Laval, 2018). Nous nommons « projection » des ETM/ETA (Trunkenwald & Laval, 2019) la restriction à un champ spécifique analysée dans l'ETM/ETA considéré.

Les premières analyses : cas des algorithmes A0, A1 et A2 écrits en pseudo-code

Dans le poster, trois niveaux de représentations illustrent le schéma éclaté (vue de dessus des ETM_s et ETA) correspondant aux conceptions et exécutions de chaque niveau des algorithmes.

Conclusion

Les 3 niveaux algorithmiques permettent une construction par encapsulations successives de l'algorithme de simulation et entrent en résonance avec les phases expérimentales du protocole statistique. Cette congruence sémantique trouve son expression dans un certain aspect du travail. Mettant en évidence la dialectique des cycles des connexions évolutives entre les 3 domaines étudiés, notre analyse peut être utilisée pour construire une activité en 2^{nde}. L'objectif est d'éviter de confronter l'élève aux difficultés simultanées dans les différents domaines concernés.

Bibliographie

Kuzniak, A., & Richard, P. (2014). Spaces for mathematical work: Viewpoints and perspectives. *Relime, 17*(4.1), 17–26.

Lagrange. J.-B. & Laval, D. (2019) Connected Working Spaces: the case of computer programming in mathematics education. *Eleventh Congress of the European Society for Research in Mathematics Education*. Utrecht, Netherlands. ⟨hal-02418181⟩

Laval, D. (2018). L'algorithmique au lycée entre développement de savoirs spécifiques et usage dans différents domaines mathématiques. Thèse de Doctorat. Université Sorbonne Paris Cité.

Laval, D. (2021). Les Espaces de Travail Mathématique et Algorithmique connectés. Étude d'une activité algorithmique comme objet d'apprentissage de savoirs spécifiques et d'usage : Cas d'une ingénierie « dichotomie continue ». In A. Chesnais, H. Sabra (Eds.), *Actes du séminaire national de didactique des mathématiques 2020* (p.88-114). IREM de Paris - Université de Paris.

Nechache, A. (2016). La validation dans l'enseignement des probabilités au niveau secondaire. Thèse de doctorat. Université Denis Diderot Paris 7.

Parzysz, B. (2009). Des expériences au modèle, via la simulation. In *Repères-IREM*, (74), 91–103.

Trunkenwald, J. & Laval, D. (2019). Algorithms as a discovery process in frequentist approach to prediction interval. *Eleventh Congress of the European Society for Research in Mathematics Education (CERME11)*, Utrecht, Netherlands. ⟨hal-02412851⟩