



# **Introduire la notion de fonction de densité de probabilité : dynamiques entre trois domaines mathématiques**

Charlotte Derouet

## **► To cite this version:**

Charlotte Derouet. Introduire la notion de fonction de densité de probabilité : dynamiques entre trois domaines mathématiques. *Recherches en Didactique des Mathématiques*, 2019, 39 (2), pp.213-266. <hal-03196485>

**HAL Id: hal-03196485**

**<https://hal.science/hal-03196485v1>**

Submitted on 12 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# INTRODUIRE LA NOTION DE FONCTION DE DENSITÉ DE PROBABILITÉ : DYNAMIQUES ENTRE TROIS DOMAINES MATHÉMATIQUES

Charlotte Derouet\*

## INTRODUCING THE CONCEPT OF PROBABILITY DENSITY FUNCTION: DYNAMICS BETWEEN THREE MATHEMATICAL DOMAINS

**Abstract** – The aim of this article is to study the implementation of introductory tasks for the concept of probability density function in grade 12 course (17-18 years old's students) in France. The main goal of these tasks is to collectively construct this concept and the link between probability and integral, based on interactions between the domains of probability, analysis and statistics. Using theoretical and methodological tools of Activity Theory and of Mathematical Working Space Theory, we show the mathematical work in the class and the role of the students and the teacher in the work's progress. Finally, we conclude that students' activities effectively allow the construction of the new concept.

**Key words:** probability; analysis; statistics; probability density function; mathematical working space; activity theory.

## INTRODUCIR LA NOCIÓN DE FUNCIÓN DE DENSIDAD DE PROBABILIDAD: DINÁMICAS ENTRE TRES DOMINIOS MATEMÁTICOS

**Resumen** – Este artículo propone estudiar la implementación de tareas de introducción de la noción de función de densidad, concebidas en el marco de una ingeniería didáctica, en una clase de terminal científico (grado 12, estudiantes de 17-18 años) en Francia. Estas tareas tienen por objetivo permitir a los estudiantes construir colectivamente la noción de función de densidad y la relación entre probabilidad e integral, apoyándose en los dominios de probabilidad, del análisis y de la estadística. Con herramientas teóricas y metodológicas tomadas desde la Teoría de la Actividad en didáctica de las matemáticas y del modelo de los Espacios de Trabajo Matemático ponemos en evidencia el trabajo matemático y las dinámicas entre los dominios matemáticos puestos en juego en la clase,

---

\* ESPE de Strasbourg, LISEC équipe AP2E, Université de Strasbourg.  
charlotte.derouet@espe.unistra.fr

además el rol de los estudiantes y profesores en el avance del trabajo. Esto nos permite mostrar que las actividades de los estudiantes permiten mostrar que las actividades de los estudiantes permiten realmente la construcción de una nueva noción.

**Palabras-claves:** probabilidad; análisis; estadística; función de densidad de probabilidad; espacio de trabajo matemático; teoría de la actividad.

## RÉSUMÉ

Cet article se propose d'étudier la mise en place de tâches pour l'introduction de la notion de fonction de densité, conçues dans le cadre d'une ingénierie didactique, dans une classe de terminale scientifique (grade 12, élèves de 17-18 ans) en France. Ces tâches ont pour but de permettre aux élèves de construire collectivement la notion de fonction de densité et le lien entre probabilité et intégrale, en prenant appui sur les domaines des probabilités, de l'analyse et de la statistique. Avec des outils théoriques et méthodologiques empruntés à la théorie de l'activité telle qu'elle a été développée en didactique des mathématiques et au modèle des Espaces de Travail Mathématique, nous analysons le travail mathématique et nous mettons en évidence les dynamiques entre les domaines mathématiques en jeu dans la classe, ainsi que le rôle des élèves et de l'enseignante dans l'avancée du travail. Les résultats obtenus montrent que les activités des élèves permettent réellement la construction de la nouvelle notion.

**Mots-Clés :** probabilités ; analyse ; statistique ; fonction de densité ; espace de travail mathématique ; théorie de l'activité.

## INTRODUCTION

Depuis une vingtaine d'années, les probabilités et la statistique sont des domaines qui prennent de plus en plus d'importance dans les programmes de mathématiques, que ce soit en France ou dans d'autres pays, et de plus en plus tôt dans la scolarité. Depuis la rentrée 2012, en dernière année de lycée (grade 12, élèves de 17-18 ans), les élèves français de nombreuses filières étudient des lois de probabilité à densité, notamment la loi normale. Plus particulièrement, le programme de terminale (grade 12) de la filière scientifique (MEN, 2011), plus ambitieux, demande d'aborder trois lois à densité : la loi uniforme sur un segment, la loi exponentielle et la loi normale. La notion mathématique centrale dans l'enseignement des lois à densité à ce niveau scolaire est, nous

semble-t-il, la fonction de densité de probabilité. En effet, le programme (MEN, 2011) précise :

on définit [...] une variable aléatoire  $X$ , fonction de  $\Omega$  dans  $\mathbb{R}$ , qui associe à chaque issue un nombre réel d'un intervalle  $I$  de  $\mathbb{R}$ . On admet que  $X$  satisfait aux conditions qui permettent de définir la probabilité de l'événement  $\{X \in J\}$  comme aire du domaine :  $\{M(x; y); x \in J \text{ et } 0 \leq y \leq f(x)\}$  où  $f$  désigne la fonction de densité de la loi et  $J$  un intervalle inclus dans  $I$ . (p. 12)

La fonction de densité est donc la notion qui permet de mettre en relation les probabilités à densité avec le calcul intégral par l'égalité :  $P(a \leq X \leq b) = \int_a^b f(x)dx$ , l'intégrale d'une fonction continue sur un intervalle  $[a; b]$  étant définie comme l'aire sous la courbe sur cet intervalle.

Cet article se propose d'étudier l'introduction de la notion de fonction de densité de probabilité en dernière année du lycée français de la filière scientifique (terminale S), dans le cadre des programmes en vigueur de 2012 à 2020.

En didactique des mathématiques, encore peu de travaux sont consacrés à l'étude des probabilités continues. Nous pouvons citer essentiellement des recherches portant sur la loi normale, du point de vue de l'apprentissage des étudiants à l'université (Batanero, Tauber & Meyer, 1999 ; Batanero, Tauber & Sánchez, 2004 ; Bansilal, 2012) et du point de vue de la formation des enseignants du second degré (Batanero, Godino & Roa, 2004 ; Bansilal, 2014). Contrairement à ces travaux de recherche, notre travail de thèse (Derouet, 2016) s'est intéressé à la notion de fonction de densité, dans sa généralité, et non seulement en lien avec une loi particulière.

Dans le contexte français, un an après la réforme de 2002, Henry (2003) s'interroge sur la pertinence de l'enseignement des lois à densité et notamment pose la question « Comment donner un sens à la notion de densité ? » (p. 6). Il met en avant l'importance de la modélisation. Dantal et Henry (2003) ajoutent aussi que :

l'étude de certaines lois de probabilité et de leurs caractéristiques numériques permet de mettre en œuvre des connaissances essentielles des cours d'analyse et de probabilités, contribuant ainsi par leur utilisation à leur donner encore plus de sens. (p. 132)

Nous partageons cette idée que les probabilités et l'analyse peuvent être articulées plutôt que cloisonnées.

Dans Derouet et Parzysz (2016), nous avons étudié les propositions d'introduction de la notion de fonction de densité faites dans les différents manuels scolaires français de terminale S, édités en 2012. La majorité des manuels (cinq sur huit) propose une démarche s'appuyant sur l'histogramme de fréquences. Cet appui

sur la statistique descriptive nous semble justifié et porteur de sens. Cependant, nous avons identifié dans ces manuels un certain nombre d'erreurs mathématiques concernant la notion d'histogramme : erreurs dans les noms des axes, confusion entre fréquence et densité de fréquences<sup>1</sup>... Selon nous, ces erreurs constituent un frein à une introduction de la notion de fonction de densité pleine de sens pour les élèves. Nous avons aussi montré que les activités d'introduction des manuels limitent le travail des élèves à approuver visuellement une « correspondance » (non définie rigoureusement) entre un histogramme et une courbe déjà tracés sur un même graphique, sans que soit questionnée l'obtention de la courbe de densité ou, si c'est le cas, en le faisant de manière erronée. Les propriétés caractéristiques de la fonction de densité (fonction continue – propriété non nécessaire en général mais toujours vérifiée en terminale S – et positive sur son intervalle de définition  $I$  telle que l'aire sous la courbe sur  $I$  vaut 1) n'apparaissent pas à ce stade. La seule condition que cette courbe semble devoir vérifier est qu'elle « épouse » (ou « lisse ») l'histogramme. Il faut attendre la partie « Cours » des manuels pour que les auteurs présentent les différentes contraintes que doit vérifier la fonction. Finalement, le lien entre probabilités à densité et calcul intégral découle seulement de l'intuition, sans justification. Cette étude (Derouet & Parzysz, 2016) a permis de mettre en évidence une prise en charge très discutable de l'introduction de la notion de fonction de densité. Nous pouvons reprendre à notre compte la citation de Chevallard et Wozniak (2005) : « Le sentiment prévaut, ici, que les *raisons d'être* de [...] toute la *problématique de la statistique*, restent inconnues de qui a pourtant mission de l'enseigner » (p. 207).

Dans le cadre d'une méthodologie de type ingénierie didactique (Artigue, 1988), nous avons conçu et expérimenté une séquence d'enseignement articulant à la fois l'enseignement des lois à densité et celui du calcul intégral, et plus particulièrement des situations d'introduction de la notion de fonction de densité ayant pour but de permettre aux élèves de construire collectivement cette notion et le lien entre probabilité et intégrale, en prenant appui sur les probabilités, l'analyse et la statistique (Derouet & Alory, 2018).

Cet article se propose d'étudier la mise en place de ces séances d'introduction en terminale S en France, du point de vue du travail

---

<sup>1</sup> Ces résultats sur l'histogramme complètent ceux de la recherche menée par Roditi (2009) qui montre le manque de connaissances des enseignants sur cet objet mathématique et leur manque d'intérêt pour cette notion.

mathématique en jeu dans la classe, en se demandant notamment si l'objectif d'enseignement a réellement été atteint. Par travail, nous entendons une activité rationnelle ou instrumentée (Habermas, 1970) ayant une finalité insérée dans un contexte humain et social : ici, il s'agit de celle des élèves résolvant un problème mathématique.

Tout d'abord, nous allons présenter les outils théoriques qui guident notre recherche. Ensuite, nous détaillerons notre méthodologie. Après avoir présenté nos tâches d'introduction et donné quelques éléments d'analyse globale *a priori*, nous étudierons deux épisodes des séances d'introduction. À partir de ces analyses, nous dégagerons quelques résultats et proposerons des perspectives de recherche.

## CADRAGE THÉORIQUE

Pour étudier le travail mathématique effectué dans la classe, nous empruntons différents outils théoriques à la théorie de l'activité telle que développée en didactique des mathématiques (Vandebrouck & Robert, 2017) ainsi qu'au modèle des Espaces de Travail Mathématique (Kuzniak, Tanguay & Elia, 2016).

### Tâches et activités

La théorie de l'activité distingue la tâche de l'activité. En reprenant les mots de Rogalski (2008) :

La tâche est ce qui est à faire : le « *but qu'il s'agit d'atteindre sous certaines conditions* » (définition proposée par Leontiev en 1984). L'activité est ce que développe un sujet lors de la réalisation de la tâche : non seulement ses actes extériorisés, mais aussi les inférences, les hypothèses qu'il fait, les décisions qu'il prend, dans ce qu'il fait et ce qu'il se retient de faire ; l'activité comprend aussi la manière dont le sujet gère son temps, et également son état personnel, ainsi que ses interactions avec autrui dans la situation de travail. (p. 24)

Lorsque la tâche est complexe, le but à atteindre est composé de différents sous-buts. On parle dans ce cas de sous-tâches. En situation institutionnelle, l'activité est très orientée par la réalisation de la tâche. Dans cet article, nous nous intéressons aux activités des élèves (ou plus particulièrement du collectif classe) associées aux tâches d'introduction.

Les activités des élèves sont vues comme pouvant engendrer de nouvelles connaissances mathématiques. Cependant, les activités effectives des élèves ne sont accessibles que partiellement, nous ne

pouvons donc faire que des hypothèses sur ces activités. Nous parlons donc d'activités possibles (Robert, 2008b), qui sont à mettre en relation avec les activités attendues (*a priori*).

Robert et Vandebrouck (2014) proposent de distinguer trois sous-activités, liées aux connaissances en jeu et aux différentes adaptations (Robert, 2008a) des connaissances attendues :

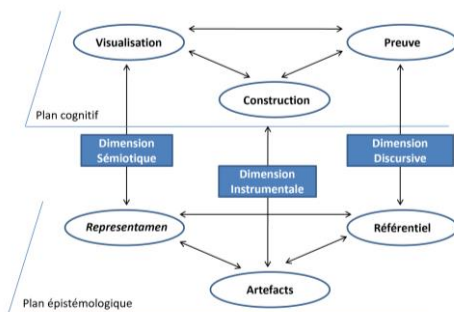
- Les (sous)-activités de reconnaissance d'outils ou d'objets mathématiques à mettre en fonctionnement : *« ce sont les théorèmes ou propriétés à mettre en fonctionnement. Cela peut comprendre des choix de connaissances, forcés ou non, selon les alternatives existantes »*. (Robert & Vandebrouck, 2014, p. 256)
- Les (sous)-activités d'organisation de raisonnement global : *« il s'agit de repérer les différents raisonnements précis à mener, avec les étapes éventuelles et leur ordre, les reprises de questions précédentes, les interprétations »*. (*ibid.*, p. 256)
- Les (sous)-activités de traitement interne : *« il s'agit des constructions de figure, des calculs à effectuer, du travail sur les formules, simple remplacement des données par leurs valeurs ou transformations, équivalences, implications, mais aussi de l'introduction d'intermédiaires, notations ou expressions, des changements de registres ou de points de vue (imposés ou choisis) et des mélanges de cadres éventuels (imposés ou choisis) »*. (*ibid.*, p. 256)

La tâche prescrite va donc engendrer des activités possibles des élèves, pouvant être rattachées aux sous-activités présentées ci-dessus. Dans notre cas, la tâche prescrite va être décomposée en sous-tâches attendues que l'on pourrait qualifier de prescrites, mais non explicitées. L'explicitation de ces sous-tâches est à la charge des élèves. Il s'agit d'une activité d'organisation de raisonnement.

### **Le modèle des Espaces de Travail Mathématique**

De façon plus globale, la tâche va activer un travail mathématique. De ce fait, nous inscrivons aussi notre travail dans le modèle des Espaces de Travail Mathématique (ETM) développé par Kuzniak (2011 ; Kuzniak & Richard, 2014 ; Kuzniak, Tanguay & Elia, 2016). Le modèle des ETM (figure 1) articule deux plans (Kuzniak, 2011), l'un de nature épistémologique, en lien avec les contenus mathématiques du domaine étudié, et l'autre de nature cognitive, qui concerne l'activité (dans le sens développé dans le paragraphe

précédent) et les moyens que met en œuvre l'élève résolvant des tâches mathématiques.



**Figure 1.** – Le modèle des ETM, inspiré de Kuzniak (2014)

Le plan épistémologique est composé de trois composantes : un ensemble de signes et de représentations sémiotiques (Duval, 1993), un ensemble d'artefacts (instruments) et un système de références théoriques, le référentiel théorique, c'est-à-dire un ensemble de définitions, de propriétés et de théorèmes. Le plan cognitif est constitué de trois processus : la reconnaissance (visualisation dans un sens large), la construction et la preuve. Ces deux plans sont articulés suivant trois dimensions<sup>2</sup> (ou genèses) :

- la dimension sémiotique, basée sur la production et l'interprétation de représentations, en relation avec les registres de représentation de Duval (1993) et la visualisation ;
- la dimension instrumentale, relative aux constructions déterminées par l'utilisation des instruments (par exemple, l'utilisation de logiciels) ;
- la dimension discursive, en lien avec le référentiel théorique et la preuve.

Des sous-activités de même type (de reconnaissance, d'organisation ou de traitement) peuvent être associées à des composantes et des dimensions différentes de l'ETM. Cependant, une activité de reconnaissance d'outils ou d'objets mathématiques

<sup>2</sup> Nous utilisons le terme « dimension » plutôt que « genèse » proposé par Kuzniak (2011) car il nous semble qu'il ne renvoie pas au même niveau d'analyse. Ici, nous voulons rendre compte de ce qui est en jeu dans les activités mathématiques de manière ponctuelle, nous ne cherchons pas à mettre en évidence un processus en évolution.



à mettre en fonctionnement est majoritairement rattachée à la dimension discursive de l'ETM, de manière plus ou moins explicite. Une activité de traitement interne est, la plupart du temps, en lien avec la dimension sémiotique mais ce peut être aussi avec la dimension instrumentale. Une activité d'organisation quant à elle, d'ordre plutôt « méthodologique », peut être associée à l'une ou l'autre des dimensions (ou plusieurs) suivant les cas.

Pour cette raison, la théorie de l'activité telle que développée en didactique des mathématiques et le modèle des Espaces de Travail Mathématique nous semblent complémentaires pour décrire le travail mathématique. Identifier des instants clés dans la circulation du travail mathématique dans l'ETM permet une relecture certes plus macroscopique mais éclairante des activités possibles des élèves.

### **Un ETM associé à un sous-domaine mathématique**

Dans un premier temps, un ETM peut être vu comme relatif à un domaine mathématique. Selon Kuzniak (2014), un domaine mathématique est une partie des mathématiques définie par des objets, des représentations de ces objets et un référentiel théorique sur ces objets. Cependant, on peut repérer plusieurs types d'ETM : des ETM associés à un domaine, plus localement à un sous-domaine ou encore des ETM globaux. Selon Montoya Delgadillo et Vivier (2014), « un ETM global regroupe l'ensemble des composantes cognitives et épistémologiques dans le travail mathématique de plusieurs domaines » (p. 78). Ils mettent en garde en précisant que la notion d'ETM global est à réserver au cas où les registres, domaines, notions, etc. en jeu sont suffisamment bien intégrés et le sont d'une manière cohérente. Ils ajoutent :

On peut d'ailleurs interpréter comme un enjeu majeur de l'enseignement des mathématiques le fait d'aboutir à un ETM global coordonnant les ETM associés à des domaines mathématiques. On peut en particulier penser que l'ETM personnel d'un enseignant de mathématiques est global ce qui lui permet d'élaborer un ETM idoine qui prend en charge plusieurs domaines, même si l'on peut interroger la coordination des ETM associés aux différents domaines au sein de cet ETM idoine. (Montoya Delgadillo & Vivier, 2014, p. 78)

Étant donné que nous nous intéressons, dans cet article, à l'articulation entre trois domaines mathématiques distincts (les probabilités, l'analyse et la statistique), il est assez naturel de considérer des ETM associés à chacun de ces trois domaines. Plus particulièrement, nous considérons trois ETM locaux, associés aux

sous-domaines en jeu : l'ETM des probabilités à densité (PaD), celui du calcul intégral (CI) et celui de la statistique descriptive (SD). Dans le sous-domaine du calcul intégral, nous incorporons ce qui concerne le calcul d'aire sous une courbe et aussi plus largement le sous-domaine des fonctions, indispensable au calcul intégral. Nous étudions l'ETM global à travers les changements de sous-domaine entre les probabilités à densité, le calcul intégral et la statistique descriptive. Il est à noter que les trois sous-domaines en jeu n'ont pas le même statut : l'un (les probabilités à densité) concerne la notion en construction tandis que les deux autres concernent des notions déjà abordées par les élèves, qui vont venir en appui à la construction du référentiel théorique associé à la notion de fonction de densité du premier sous-domaine.

### **Les différents niveaux des ETM**

Kuzniak, Tanguay et Elia (2016) distinguent trois niveaux d'ETM :

- l'ETM de référence, défini par rapport à des critères mathématiques sur les savoirs en question ;
- l'ETM idoine, qui décrit le travail planifié par un enseignant pour être effectif dans une classe, en fonction d'attendus institutionnels ;
- l'ETM personnel, relatif à chaque individu (élève ou enseignant).

Dans cet article, nous étudions l'ETM idoine proposé dans la classe de l'expérimentation. Nous parlons d'ETM idoine effectif, relatif à ce qui s'est effectivement passé dans la classe. L'ETM idoine effectif se construit petit à petit dans la classe à l'aide des ETM personnels de chaque élève. En observant une classe, notamment par le biais des mises en commun, nous n'avons pas accès aux ETM personnels des élèves (ou seulement de manière très ponctuelle) mais à un ETM collectif, où chaque élève peut apporter des éléments de son ETM personnel, qui permet de faire avancer le travail mathématique de la classe et, petit à petit, de façonner l'ETM idoine effectif. Bien entendu, ce n'est pas seulement l'ETM collectif de la classe qui permet de construire l'ETM idoine, l'enseignant aussi y contribue par ses interventions, ses aides, ses rappels...

### **Reformulation de notre question de recherche**

L'introduction d'une nouvelle notion mathématique vise à enrichir le référentiel théorique de l'ETM. Pour l'introduction de la notion

de fonction de densité, nous souhaitons que les élèves construisent les éléments du référentiel théorique suivants :

- R1 : l'idée que la courbe de densité « lisse » l'histogramme ;
- R2 : la fonction de densité  $f$  est définie et continue sur l'intervalle  $I$ , intervalle dans lequel la variable aléatoire prend ses valeurs ;
- R3 : la fonction de densité  $f$  est positive sur  $I$  ;
- R4 : l'aire sous la courbe de la fonction de densité  $f$  sur  $I$  est égale à 1 ;
- R5 : la courbe de densité permet de calculer des probabilités :  $P(a \leq X \leq b)$  correspond à l'aire sous la courbe de  $f$  entre  $a$  et  $b$ .

Nos questions de recherche sont les suivantes :

- Les activités des élèves associées aux tâches d'introduction conçues dans le cadre de notre ingénierie didactique permettent-elles de construire la notion de fonction de densité, à savoir tous les éléments du référentiel théorique cités ci-dessus ?
- Cette construction est-elle rendue possible grâce à des circulations entre les trois ETM en jeu ?
- L'ETM collectif de la classe permet-il seul de construire l'ETM idoine effectif ou l'enseignante joue-t-elle un rôle dans cette construction ?

## MÉTHODOLOGIE

Pour étudier le cheminement du travail mathématique dans la classe, nous cherchons à mettre en relation les sous-tâches, les différentes sous-activités possibles ou effectives des élèves, les sous-domaines et les composantes de l'ETM sollicités ainsi que les interventions de l'enseignant (notamment à travers ses aides, au sens de Pariès, Robert & Rogalski, 2009), à partir des données dont nous disposons.

### Les données

Les deux tâches d'introduction que nous étudions dans cet article ont été conçues en collaboration avec une enseignante de terminale S (Derouet & Alory, 2018). L'enseignante est très expérimentée

(plus de vingt d'ans d'enseignement) et a été formatrice à l'IUFM<sup>3</sup> pendant cinq ans. Cette même enseignante a testé la séquence dans sa classe de terminale S d'un lycée parisien, en 2015. La classe, composée de 34 élèves (17-18 ans) lors de l'expérimentation, est qualifiée de « bon niveau » par l'enseignante.

Les quatre séances d'introduction (environ 5 heures) ont été enregistrées avec des dictaphones puis transcrites dans leur intégralité. Les données sont complétées par des photos du tableau et des vidéoprojections, ainsi que par les documents distribués par l'enseignante et des productions d'élèves.

### Analyse globale des tâches et éléments de contexte

Les deux tâches d'introduction sont des problèmes de modélisation probabiliste. Dans ces deux problèmes, il ne s'agit pas pour les élèves de reconnaître le modèle probabiliste en jeu (ils ne le connaissent pas) mais de construire ce modèle (plusieurs sont possibles, même si un seul est attendu dans chaque cas pour répondre à notre objectif d'apprentissage). Pour définir ce modèle, il faut qu'émerge dans la classe le besoin de déterminer une fonction (qui s'avère être la fonction de densité) qui doit respecter des contraintes spécifiques (les éléments du référentiel théorique à construire), qui sont à prendre en considération pour la choisir.

Nous ne cherchons pas ici à étudier le processus de modélisation, mais bien la construction du nouveau référentiel théorique dans ce contexte. Pour cette raison, nous ne donnons ici que quelques éléments de l'analyse globale<sup>4</sup> (*a priori* et *a posteriori*) de ces deux tâches d'introduction. Nous ne présentons pas tous les possibles mais seulement le chemin attendu par la chercheure et donc par l'enseignante (pour l'objectif d'apprentissage fixé) et emprunté par la classe. Pour chaque problème, nous mettons en évidence l'objectif attendu de la tâche, les sous-tâches que nous souhaitons voir apparaître au cours de l'avancement du travail mathématique par rapport à notre objectif, afin de permettre ensuite de repérer les épisodes rattachés à ces sous-tâches dans le déroulement effectif. L'analyse globale nous permet d'indiquer, dans les grandes lignes, le cheminement global

---

<sup>3</sup> Institut Universitaire de Formation des Maîtres.

<sup>4</sup> Cette phase d'analyse a cependant été très importante pour la conception des tâches et est aussi primordiale pour guider l'analyse des déroulements, *a posteriori*.

attendu entre les sous-domaines en rapport avec les sous-tâches identifiées.

### **Analyse locale des déroulements, activités effectives et possibles**

Notre analyse plus locale se focalise sur les activités possibles des élèves (*a posteriori*) en lien avec la sous-tâche : *Déterminer l'expression de la fonction de densité* (sous-tâche qui n'est pas formulée de cette façon pour les élèves). Grâce à notre analyse globale, nous pouvons identifier rapidement les épisodes concernés, dont nous pensons qu'ils sont des moments clés dans la construction de la notion de fonction de densité.

Bien qu'une analyse locale *a priori* ait été faite, nous avons fait le choix dans cet article de nous centrer sur l'analyse des déroulements et des activités effectives et possibles de la classe.

Dans notre analyse locale des déroulements, en prenant appui sur le chronométrage, nous prenons en compte :

- les sous-domaines mathématiques en jeu et les éventuels changements de sous-domaines ;
- les sous-activités en jeu ;
- les registres de représentation mobilisés (Duval, 1993) ;
- les dimensions de l'ETM mobilisées et les articulations entre les différents ETM.

Nous identifions, d'une part, les activités des élèves et, d'autre part, celles de l'enseignante ainsi que les interventions représentatives correspondantes. Cela nous permettra notamment d'identifier qui, des élèves ou de l'enseignante, est à l'initiative ou prend en charge les différentes sous-activités ainsi que les changements de domaines ou de registres associés. Au niveau de l'enseignante, nous regardons un peu plus en détail ses interventions. Nous identifions tout particulièrement les aides qu'elle apporte : les aides à visée procédurale et les aides à visée constructive (Pariès *et al.*, 2009). Les aides « procédurales » sont les aides qui jouent directement sur la tâche prescrite. Elles peuvent modifier les activités mathématiques des élèves par rapport à celles prévisibles. Dans les aides procédurales, nous distinguons les aides directes, qui donnent une indication explicite dans la résolution, et les aides indirectes, par exemple lorsque l'enseignant pose une question sans induire une réponse. Les aides sont à visée constructive si elles ont une portée plus large que seulement celle de résoudre la tâche actuelle : ces aides « ajoutent quelque chose entre le travail de

l'élève et la construction (espérée) de la connaissance qui pourrait en résulter » (Pariès *et al.*, 2009, p. 107).

En conclusion de notre analyse, nous revisiterons l'épisode en adoptant une perspective globale sur le travail mathématique de la classe. À l'aide de l'enchaînement des diagrammes des ETM, nous mettrons en évidence les dynamiques entre les sous-domaines et nous reviendrons sur la question de la construction du nouveau référentiel théorique.

## PRÉSENTATION DES TÂCHES D'INTRODUCTION ET ANALYSE GLOBALE *A PRIORI*

L'idée générale de ces tâches d'introduction est que la fonction de densité arrive comme une Réponse À un Problème (notion RAP de Robert, 2008a), d'où le choix de proposer des problèmes de modélisation probabiliste, où les modèles ne sont pas connus des élèves. Une des originalités de la séquence entière est d'articuler à la fois la séquence sur les lois à densité (à l'exception du cas particulier des lois normales qui sont traitées plus tard dans l'année) et celle sur le calcul intégral (dans le cas des fonctions continues positives sur un intervalle) pour n'en faire qu'une seule (Derouet & Alory, 2018). Dans cette séquence, il s'agit de commencer par introduire la notion de fonction de densité en lui donnant du sens, pour créer le besoin de calculer des aires sous une courbe (afin de déterminer des probabilités) et ensuite introduire l'intégrale comme outil. Les séances d'introduction ont finalement un double objectif :

- Introduire la notion de fonction de densité (l'objectif de cet article) ;
- Introduire le calcul d'aire sous une courbe et donc le calcul intégral.

Du fait de ce deuxième objectif, les élèves n'ont pas de prérequis sur les primitives, les intégrales et le calcul d'aires sous une courbe quelconque. Nous considérons trois niveaux de calculs d'aires sous une courbe :

- Niveau A1 : calculs d'aires élémentaires (sous des courbes affines par morceaux) qui se ramènent à des calculs par des formules d'aires de triangles, de rectangles ou de trapèzes ;
- Niveau A2 : calculs d'aires sous des courbes de fonctions dont les primitives sont connues ;

- Niveau A3 : calculs approchés d'aires, par la méthode des rectangles, l'utilisation de logiciels ou autres.

Dans notre cas, les élèves n'ayant pas encore étudié les primitives, le niveau A2 n'est pas envisageable.

### **Le devoir maison**

Pour introduire la notion de fonction de densité, nous nous appuyons sur la statistique descriptive via l'histogramme de fréquences. La notion d'histogramme doit, selon nous, être disponible chez les élèves pour constituer un véritable point d'appui afin de faire émerger la notion de fonction de densité. Pour cette raison, il nous semble indispensable que la notion d'histogramme soit (re)travaillée par les élèves en amont de la séquence. Un devoir à la maison (annexe 1) a donc été donné aux élèves peu avant la séquence pour (ré)introduire cette notion, normalement connue des élèves à partir du cycle 4 (grade 7) mais en réalité très peu travaillée dans les classes (Roditi, 2009).

Dans l'exercice 1, l'objectif est de définir l'histogramme de fréquence comme la représentation graphique permettant de rendre compte de la densité de fréquence sur les classes considérées. Nous avons décidé de prendre appui sur la densité de population pour donner du sens à la densité de fréquence. Nous avons pris le parti, ce qui n'est jamais fait dans les manuels, de définir la densité de fréquence, qui permettra ensuite de faire le lien avec la densité de probabilité.

Ce devoir maison propose également, en exercice 2, une première rencontre avec la loi uniforme sur un segment. En appui sur des histogrammes de fréquences, construits à partir de différents échantillons simulés de taille 10 000, l'objectif est de concevoir une « courbe de tendance » qui lisse ces histogrammes. Le terme « courbe de tendance » et non « courbe de densité » est volontairement introduit lors du devoir maison pour ne pas imposer un vocabulaire qui ne « parle » pas aux élèves à ce moment-là. Ce problème est aussi l'occasion pour les élèves de rencontrer une certaine démarche de modélisation qui sera un point d'appui pour les séances d'introduction. Il a été pensé en plusieurs étapes : une première étape de conjectures, où les élèves répondent intuitivement ; une deuxième, qui s'appuie sur la simulation pour évaluer les probabilités à partir des fréquences observées sur l'histogramme donné ; et une troisième, basée sur des calculs théoriques prenant appui sur une « courbe de tendance », déduite des histogrammes fournis. Enfin, il permet aussi d'aborder la

problématique de  $P(X = a) = 0$  bien que l'événement  $\{X = a\}$  ne soit pas impossible.

Les séances suivant la correction en classe de ce devoir maison sont dédiées aux deux problèmes introductifs.

## Le problème de la rencontre

### Présentation de la tâche de modélisation

L'énoncé du premier problème proposé aux élèves est le suivant :

Karine et Olivier décident de se retrouver au café de l'Hôtel de Ville entre 7h et 8h. Ils peuvent arriver à tout moment entre 7h et 8h. Que peut-on dire du temps d'attente du premier arrivé ?

L'énoncé de ce problème est ouvert. La tâche en question est volontairement floue. Les élèves doivent finalement arriver à une reformulation (implicite ou explicite) de cette tâche du type : *Quelle est la probabilité que le premier arrivé attende entre  $t_1$  et  $t_2$  (avec  $t_1$  et  $t_2$  quelconques) ?* Pour cela, il est nécessaire de choisir un modèle. Plusieurs modèles seraient envisageables. La fonction de densité que nous attendons est la fonction affine  $f(x) = -2x + 2$  sur  $[0; 1]$  ; il s'agira donc de calculs de probabilités correspondant à des calculs d'aires élémentaires de triangles ou de trapèzes<sup>5</sup> (niveau A1).

### Eléments d'analyse globale a priori

Les sous-tâches attendues dans le cadre de notre expérimentation sont les suivantes :

- Sous-tâche 1 : Faire des hypothèses sur les temps d'arrivée de Karine et Olivier.
- Sous-tâche 2 : A partir de ces hypothèses, définir la variable aléatoire  $X$  correspondant au temps d'attente du premier arrivé.
- Sous-tâche 3 : Simuler des réalisations de la variable aléatoire  $X$ , à l'aide d'un tableur.
- Sous-tâche 4 : Construire un (ou des) histogramme(s) associé(s) à un (ou des) échantillon(s) d'un nombre suffisamment grand de données (avec des amplitudes plus ou moins grandes). Le logiciel qui sera privilégié dans la classe est *GeoGebra*.

---

<sup>5</sup> Pour une présentation plus détaillée de cette séquence d'enseignement, le lecteur peut se reporter à Derouet & Alory (2018).



- Sous-tâche 5 : Déterminer l'expression de la fonction de densité.
- Sous-tâche 6 : Déterminer des probabilités à l'aide de la fonction de densité (cette sous-tâche est associée à un exercice d'application, en annexe 2, qui est donné aux élèves une fois la fonction de densité trouvée).

Tout se joue dans les choix faits par les élèves dans la sous-tâche 1. Il est attendu que les élèves considèrent les temps d'arrivée de Karine et d'Olivier, respectivement  $T_K$  et  $T_O$ , comme étant des variables aléatoires continues (d'autres choix pourraient être faits) suivant des lois uniformes sur  $[7; 8]$  ou sur  $[0; 1]$ . Les élèves ont rencontré une loi uniforme sur un segment dans le devoir maison préliminaire. De là, le temps d'attente  $X$  est une variable aléatoire à valeurs dans  $[0; 1]$  définie par :  $X = |T_K - T_O|$ . La loi associée à une telle variable aléatoire (loi de Xenakis) n'est pas connue des élèves. Un passage par la simulation sur tableur, puis par un histogramme des données simulées, va ensuite amener les élèves à chercher une « courbe de tendance » (vocabulaire introduit dans le devoir maison), qui permettra ensuite de calculer des probabilités. Il est attendu que les élèves prennent appui sur la démarche rencontrée dans le devoir maison pour dégager et organiser les sous-tâches. Outre dans le devoir maison, dans lequel il est question de simulation sur tableur, les élèves ont déjà rencontré des simulations dans leur scolarité et précédemment dans l'année, essentiellement dans le cas discret (exemples : somme de deux dés, marche aléatoire...) et sont censés connaître la formule =ALEA() pour simuler le choix « au hasard » d'un nombre réel entre 0 et 1. Pour la construction de l'histogramme à l'aide de *GeoGebra*, les élèves ne connaissent pas cette fenêtre du logiciel. La sous-tâche doit donc être explicitée par les élèves mais sera prise en charge par l'enseignante. Les raisons du choix du logiciel *GeoGebra* sont multiples : tout d'abord, contrairement au tableur, *GeoGebra* construit de véritables histogrammes, de plus, bien que ce logiciel ne soit pas connu des élèves, sa manipulation est assez explicite pour être comprise et enfin, ce logiciel permet dans un second temps de tester des courbes de fonctions « sur » l'histogramme.

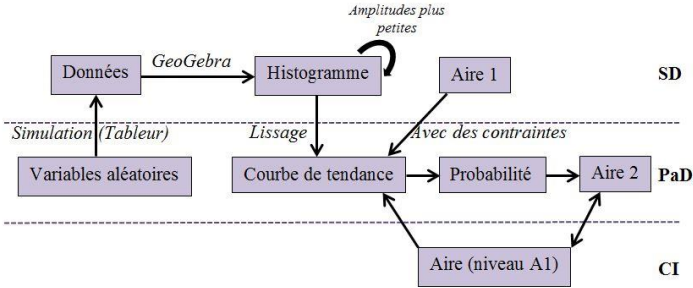
Pour déterminer l'expression de la fonction de densité, les élèves doivent s'appuyer sur l'histogramme : la courbe doit « lisser » l'histogramme (R1). L'histogramme (cf. figure 4) incite à chercher une courbe associée à une fonction affine décroissante sur  $[0; 1]$ . Cependant la contrainte de correspondance entre la courbe et l'histogramme ne suffit pas et c'est précisément ce qui

doit émerger dans la classe. Il faut notamment prendre en considération le fait que l'aire sous la courbe doit être égale à 1 (R4). De plus, il faut ajouter la contrainte  $f(1) = 0$ . C'est donc la sous-tâche 5 qui nous intéresse tout particulièrement. D'après le tableau 1, nous pouvons effectivement observer que c'est au niveau de cette sous-tâche que les trois sous-domaines qui nous intéressent sont présents simultanément.

Sous-tâche 1	PaD
Sous-tâche 2	PaD
Sous-tâche 3	SD
Sous-tâche 4	SD
Sous-tâche 5	PaD/CI/SD
Sous-tâche 6	PaD/CI

**Tableau 1.** – Sous-domaines mathématiques en jeu dans les différentes sous-tâches du problème de la rencontre

La figure 2 présente le cheminement attendu lors de la réalisation de la tâche et montre les dynamiques entre les trois sous-domaines.



**Figure 2.** – Cheminement attendu dans le problème de la rencontre

**Le problème du volcan Aso**

*Présentation de la tâche de modélisation*

Le second problème a lieu lors de la séance qui suit le problème de la rencontre. Il s'appuie cette fois-ci sur des données réelles, que l'on peut trouver dans le manuel *Math 'x* (2012). L'énoncé distribué aux élèves (en annexe 3) est un tableau regroupant les 73 années des différentes éruptions du volcan Aso (Japon) entre 1229 et 1897. La troisième colonne précise aussi le temps d'attente entre deux éruptions consécutives.

Les questions qui accompagnent ces données sont les suivantes :

<p>Le volcan Aso est actuellement en éruption. Comment évaluer la probabilité que la prochaine éruption :</p>
---

<p>a) Ait lieu dans les 5 ans ?</p>
-------------------------------------

<p>b) Au cours de l'année 2030 ?</p>
--------------------------------------

Les objectifs de ce problème sont, dans un premier temps, de réinvestir les propriétés de la fonction de densité pour continuer de donner du sens à cette notion et, dans un second temps, d'introduire le calcul approché d'aire. Ce deuxième objectif ne sera pas abordé dans cet article. Dans ce problème, il s'agit, à partir de données réelles sur le passé, de « prédire » l'avenir en faisant une estimation. C'est une problématique de statistique inférentielle. Pour la première question, une première évaluation peut se faire à l'aide de la fréquence qui s'obtient avec les données fournies. Cependant au vu de la deuxième question, il est attendu des élèves qu'ils mettent en place une démarche similaire à celle du problème de la rencontre. Cette fois-ci, la fonction de densité en jeu est une fonction exponentielle décroissante, dont les élèves ne savent pas calculer à la main les aires sous la courbe. Il s'agit d'un calcul d'aire de niveau A3 : ne connaissant pas le calcul intégral, les élèves devront faire du calcul approché d'aire.

### *Eléments d'analyse globale a priori*

Nous suggérons au lecteur de se reporter à Derouet et Parzysz (2016) pour une analyse *a priori* détaillée de ce problème. L'énoncé formule deux tâches (tâches a et b), mais pour y parvenir des sous-tâches intermédiaires sont attendues :

- Sous-tâche 1 : Regrouper les données sous forme d'un tableau de fréquences.
- Sous-tâche 2 : Tracer un histogramme (de fréquence) des données.
- Sous-tâche 3 : Déterminer l'expression de la fonction de densité.
- Sous-tâche 4 : Retour à la tâche a.
- Sous-tâche 5 : Retour à la tâche b.

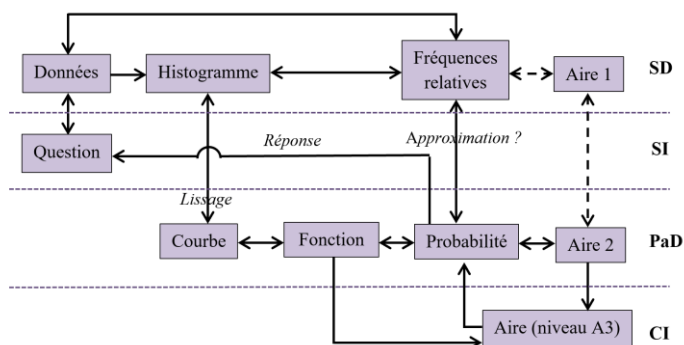
La problématique posée fait partie du sous-domaine de la statistique inférentielle (SI). Les données fournies avec l'énoncé du problème sont d'abord travaillées (sous-tâches 1 et 2) dans le domaine de la statistique descriptive (SD), puis il y a un passage au sous-domaine des probabilités à densité (PaD), jumelé à celui du calcul intégral (CI) (cf. tableau 2).

Enoncé	SI
--------	----

Sous-tâche 1	SD
Sous-tâche 2	SD
Sous-tâche 3	SD/PaD/CI
Sous-tâche 4	PaD/CI
Sous-tâche 5	PaD/CI

**Tableau 2.** – Sous-domaines mathématiques en jeu dans les différentes sous-tâches du problème du volcan Aso

Le figure 3 présente les dynamiques entre les trois sous-domaines et le cheminement attendu entre différentes notions lors de la réalisation de la tâche.



**Figure 3.** – Diagramme du cheminement du problème du volcan Aso (d'après Derouet & Parzysz, 2016)

Il est prévu que les tâches 2 et 3 s'appuient sur un travail sur le logiciel *GeoGebra*, avec la construction de l'histogramme des temps d'attente entre deux éruptions consécutives (cf. figure 15) puis la recherche d'une courbe de tendance « sur » l'histogramme. Effectivement, *Geogebra* permet de tester rapidement des courbes, en entrant des expressions de fonctions ou en déplaçant une courbe déjà tracée. Nous le détaillons dans l'analyse de déroulements, mais il est attendu que les élèves proposent, dans un premier temps, des courbes de tendance de la forme  $x \rightarrow \frac{a}{bx+c}$ , qui ne peuvent convenir car l'aire sous la courbe de ces fonctions sur  $[0; +\infty[$  ne peut être égale à 1, puis dans un second temps, qu'ils proposent des fonctions exponentielles décroissantes (qui ont été beaucoup travaillées plus tôt dans l'année, en anticipation de cette séquence). Dans cet article, il s'agit de se focaliser sur la sous-tâche 3 et plus spécifiquement sur un épisode clé rattaché à cette sous-tâche.

## ANALYSE DES DÉROULEMENTS DES TÂCHES D'INTRODUCTION

### Eléments de contexte de l'expérimentation

Avant la séance consacrée au problème de la rencontre, près d'une heure (séance 0) a été dédiée à la correction du devoir à la maison, portant sur les histogrammes et abordant une loi uniforme sur un segment (non institutionnalisé) dans lequel les élèves ont rencontré la démarche du type *simulation/tracé d'histogrammes/recherche d'une « courbe de tendance »*.

Le travail sur le problème de la rencontre, tenant compte de la synthèse et la correction de l'exercice d'application, a duré environ 2h25 en classe, sur deux séances (séance 1 et séance 2). La démarche de modélisation autour du problème (tâches 1 à 5) a eu lieu exclusivement lors de la séance 1 (un peu moins de 1h50). La salle de classe n'était équipée que d'un seul ordinateur dont l'affichage était projeté. C'est l'enseignante qui a pris à sa charge la manipulation des logiciels (le tableur et *GeoGebra*). Dans un premier temps, les élèves ont eu une phase de recherche individuelle. Une première mise en commun a permis d'engager des discussions dans la classe sur les premières idées. Ensuite il y a eu une alternance entre des phases collectives et des phases individuelles (sur des sous-tâches, voire des sous-sous-tâches, bien précises qui apparaissent dans la démarche de la classe). Ce premier problème a abouti à une première institutionnalisation sur la notion de fonction de densité (séance 2) dont la trace écrite est disponible en annexe 4.

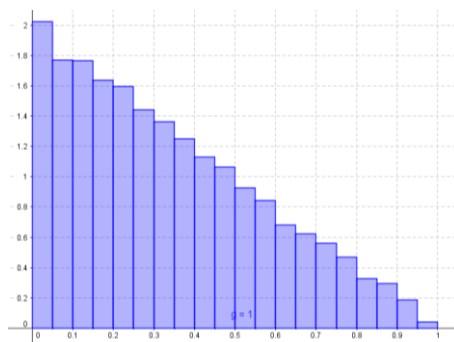
Le problème du volcan Aso a ensuite été travaillé sur deux séances (séances 3 et 4), sur une durée d'environ 1h20 en tout, avec un travail à la maison entre les deux séances. 18 minutes ont été consacrées à de la recherche individuelle (dont 14 minutes tout au début de la séance 3). La majorité du travail mathématique a été prise en charge par une réflexion collective. Comme pour le premier problème, le travail collectif prend appui sur un seul ordinateur vidéoprojeté, manipulé par l'enseignante.

### Analyse *a posteriori* du problème de la rencontre

Nous focalisons notre analyse sur la sous-tâche : *Déterminer l'expression de la fonction de densité*, sous-tâche 5 dans le problème de la rencontre. Nous considérons l'épisode relatif à cette sous-tâche, constitué de dix phases dont une phase préparatoire qui permet d'arriver à la recherche de l'expression de la fonction de

densité (phase 5.1) et des phases plus générales pour appuyer le sens de la notion de fonction de densité (phases 5.9 et 5.10).

A la fin de l'épisode précédent relatif à la sous-tâche 4, la classe aboutit à l'histogramme en figure 4 sur le logiciel *GeoGebra*, après avoir simulé un échantillon de taille 10 000 avec la formule  $=ABS(ALEA()-ALEA())$  sur tableur.



**Figure 4.** – Histogramme des temps d'attente du premier arrivé (amplitude 0,05h, soit 3min)

### *Analyse de l'épisode 5*

Cet épisode correspond au passage entre 1h11min44s et 1h40min25s de la séance 1 (un peu moins de 30 minutes).

Nous illustrons notre analyse par des diagrammes d'ETM, où l'on voit apparaître trois ETM (chacun relatif à un des trois sous-domaines PaD, CI, SD). Nous rappelons que dans le modèle des ETM (cf. figure 1) se trouve à gauche la dimension sémiotique, au milieu la dimension instrumentale et à droite la dimension discursive. Ensuite, pour rendre compte des dynamiques entre les différents ETM, nous entourons en gris foncé lorsque la dimension est sollicitée et travaillée par les élèves et en gris clair quand elle est prise en charge par l'enseignante. Les pointillés gris foncé sont utilisés quand la dimension est sollicitée par les élèves, mais grâce à une intervention de l'enseignante et inversement, en pointillés gris clair lorsque l'enseignante prend en charge le travail dans la dimension mais sous commande des élèves.

#### **Phase 5.1** [1:11:44] : Organisation du raisonnement

L'énoncé ouvert du problème de la rencontre est, à ce moment-là, reformulé dans le contexte d'un exemple générique de la façon suivante : *Évaluer la probabilité que le temps d'attente soit entre*

0,2 et 0,3<sup>6</sup>. Pour aider à organiser la suite de la démarche, l'enseignante dit : « *Comment feriez-vous avec cet histogramme ?* » (aide procédurale indirecte), en précisant que l'on ne cherche pas le résultat mais la méthode pour y arriver. Un élève propose de passer par plusieurs étapes :

E1 : On peut faire comme dans le DM<sup>7</sup>, c'est-à-dire... on peut tracer une courbe de tendance. On va pouvoir calculer l'aire de ce triangle... et puis prendre une amplitude qui correspond à 0,2 ~ 0,3.

Nous pouvons identifier trois étapes : 1. *Tracer une courbe de tendance* / 2. *Calculer l'aire du triangle* / 3. *Prendre une amplitude qui correspond à 0,2 ~ 0,3*. L'élève fait référence au devoir maison préliminaire et donc prend appui sur des connaissances en construction. L'enseignante demande ensuite à l'élève de répéter pour lui permettre d'insister sur certains points. « *Tracer la courbe de tendance* » est reformulé par l'enseignante par « *Essayer de lisser notre histogramme* » (R1). Le retour sur l'étape 2 permet à l'élève de compléter son propos :

E1 : Puis calculer l'aire de ce triangle-là.

P : Triangle. Quel triangle ?

E : Enfin, ce qui pourrait ressembler à un triangle.

P : Ah oui, tout le... J'avais pas compris. D'accord.

E1 : C'est égal à 1.

P : Ah oui, alors, on n'a pas besoin de le calculer puisque... ça fait 1. Pourquoi ça ferait 1 ?

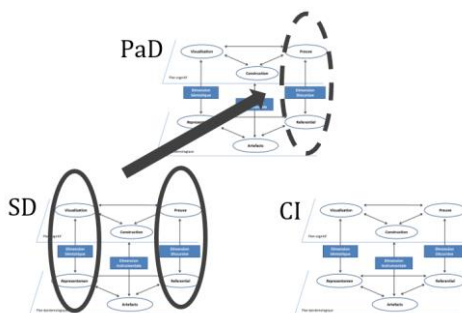
E1 : Parce que... ça indique la fréquence.

Cette phase d'activité d'organisation du raisonnement est prise en charge par un élève. La notion de courbe de tendance apparaît et des contraintes sur cette courbe sont mentionnées : courbe qui lisse l'histogramme (R1) par l'enseignante, et aire totale sous la courbe égale à 1 (R4) par l'élève, qui le justifie en s'appuyant sur la propriété suivante : l'aire totale des rectangles d'un histogramme représente la fréquence totale donc vaut 1. L'élément R4 du référentiel théorique relatif aux probabilités à densité est justifié avec un appui sur la statistique descriptive par la dimension sémiotique (visualisation de l'histogramme) et la dimension discursive liée à l'histogramme d'où la présence de la flèche qui va vers la dimension discursive PaD entourée en pointillé car c'est une intervention de l'enseignante qui a engendré cette justification (figure 5).

---

<sup>6</sup> Il n'est pas mentionné d'unité de durée.

<sup>7</sup> Abréviation utilisée par l'élève pour « devoir maison ».



**Figure 5.** – Dynamique entre les trois ETM dans la phase 5.1

**Phases 5.2 et 5.3 [1:13:14] :** Première approche de la courbe de tendance (R1) et aparté (apport théorique)

Au début de la phase 5.2, l'enseignante demande à l'élève précédemment interrogé :

P : Donc tu la vois comment cette courbe de tendance, toi ? [...]

E1 : Ça va être affine.

P : Une droite ?

E1 : Une droite.

Un autre élève interrompt le travail pour poser une question sur les graphiques :

E2 : En fait, je ne comprends pas ce qu'il y a en ordonnées.

P : Alors. Tu comprends pas ce qu'il y a en ordonnées.

E1 : C'est la densité.

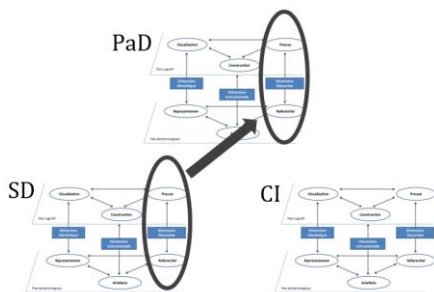
E2 : Ça va de zéro... Ah oui, c'est vrai.

Cet aparté (phase 5.3) permet une convocation du référentiel théorique lié aux histogrammes (SD) qui a été explicitement travaillé dans le devoir maison (d'où le « *Ah oui, c'est vrai* »), qui permettra ultérieurement de justifier le vocabulaire « fonction de densité » (PaD) (figure 6). En effet, bien que le terme « fonction de densité » ne soit pas introduit à ce moment-là, il est important que les élèves aient bien en tête que l'ordonnée représente la densité de fréquence pour l'histogramme et deviendra naturellement la densité de probabilité dans le domaine des probabilités.

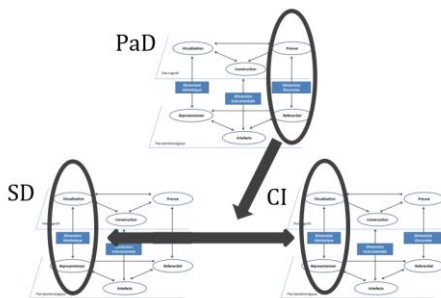
Ensuite, l'enseignante revient sur la proposition de l'élève pour la courbe de tendance (retour à la phase 5.2). L'élève « trace dans l'air » une droite au niveau du haut des rectangles de l'histogramme. Cette activité de traitement opère un changement de registre : il y a un passage du registre histogramme au registre graphique des fonctions par un changement de domaine (passage de SD à CI). Dans cette phase, il y a une première mobilisation en acte du référentiel théorique sur la « proximité » entre le haut des rectangles de l'histogramme et la courbe de densité (R1), en



construction. Seule cette contrainte de lissage de l'histogramme (implicite) est prise en compte dans ce travail de visualisation, qui cherche à mettre en correspondance visuelle l'histogramme et la courbe (figure 7) d'où l'activation des dimensions sémiotiques de SD et CI.



**Figure 6.** – Dynamique entre les trois ETM dans les phases 5.3



**Figure 7.** – Dynamique entre les trois ETM dans les phases 5.2

**Phase 5.4** [1:15:40] : Bilan sur l'organisation du raisonnement

Lors de cette phase de bilan de l'organisation du raisonnement, la notion de courbe de tendance est reprise par un élève et la notion d'aire apparaît mais c'est l'enseignante qui doit repréciser que l'aire totale sous la courbe doit être égale à 1 (R4).

**Phase 5.5** [1:16:28] : Mise en évidence de la contrainte de lissage de l'histogramme (R1)

Au début de cette phase, la sous-tâche explicite est : *Tracer votre courbe de tendance* [sur l'histogramme distribué aux élèves, en annexe 5]. Après une phase de recherche individuelle d'environ une minute, l'enseignante demande aux élèves s'ils ont tracé une courbe ; ils acquiescent. Elle leur demande s'ils ont tous la même courbe. Dans cette phase, l'enseignante cherche à faire expliciter le fait que l'on cherche une correspondance visuelle entre la courbe

et l'histogramme, d'où des aides procédurales indirectes : « *Vous ne l'avez pas tracée là, la droite ?* [en montrant une droite qui ne suit pas l'allure du haut des rectangles de l'histogramme] ? ».

P : Qu'est-ce qu'on cherche comme genre de fonction ?

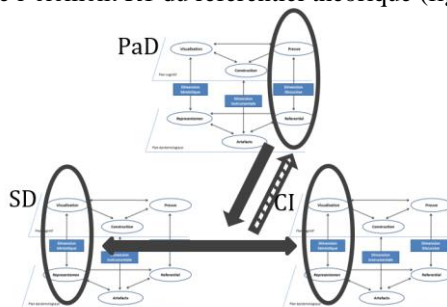
E3 : En fait, pour tracer faut qu'on essaye... de faire en sorte que... notre courbe quand on la trace qu'il y ait une partie qui soit en haut enfin...

P : Oui. A l'œil on va essayer de faire un truc...

E3 : Une moyenne.

P : qui passe à peu près voilà au milieu des hauts des rectangles. Qu'il y en ait autant au-dessus qu'en-dessous de vide.

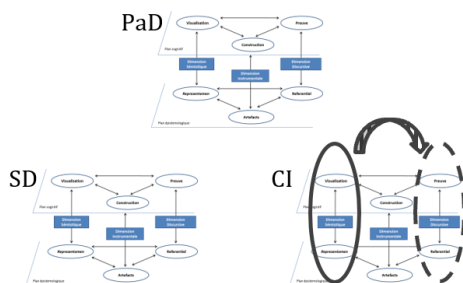
La recherche de la courbe de tendance s'appuie pour le moment seulement sur un travail de visualisation entre l'histogramme (SD) et la courbe cherchée (CI) pour permettre la construction et l'utilisation de l'élément R1 du référentiel théorique (figure 8).



**Figure 8.** – Dynamique entre les trois ETM dans la phase 5.5

#### **Phase 5.6 :** Changement de registre sur la courbe

L'enseignante amène à l'insuffisance de l'élément R1 du référentiel théorique en précisant que même avec les contraintes énoncées, « *on va se retrouver avec des droites tous différentes* ». Elle relance le travail en proposant l'aide procédurale indirecte suivante : « *Qu'est-ce qu'on cherche donc comme genre de chose ?* » (question ambiguë), pour les amener à changer de registre sur la courbe pour passer du registre graphique au registre algébrique ( $y = ax + b$ ). Cette activité de traitement est prise en charge par un élève, prenant appui sur ses connaissances (référentiel théorique sur les fonctions affines, compris dans l'ETM du CI ici) : il est capable de passer de la représentation graphique d'une fonction (dimension sémiotique) à son expression algébrique grâce à la dimension discursive. Ce changement de registres à l'intérieur d'un même domaine est visible en figure 9.



**Figure 9.** – Dynamique entre les trois ETM dans la phase 5.6

**Phase 5.7** [1:21:17] : Vers la deuxième contrainte sur la courbe de tendance (R4)

La sous-tâche explicitée au début de cette phase est : *Trouver l'équation de la droite. Trouver les conditions pour trouver a et b.* Après environ une minute de recherche individuelle, l'enseignante trace sur le logiciel une droite « au jugé » proche de l'histogramme. En appui sur le registre graphique, les élèves imposent que la droite passe par le point (1; 0). L'enseignante précise qu'il faut deux renseignements sur la courbe pour la déterminer. Avec l'aide de l'enseignante, un élève exprime que l'aire sous la courbe doit valoir 1 :

E4 : En fait, il faut... C'est le point (0 ; 2) parce que... on trace en fait le rectangle qui a une aire de 2 et on trace la diagonale du triangle et ça...

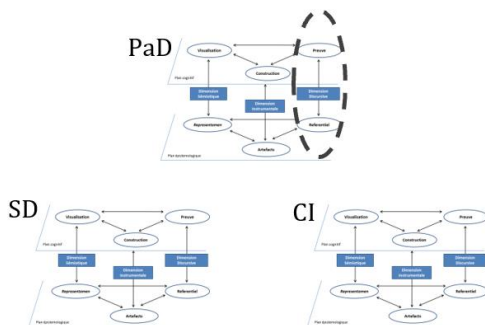
P : Donc c'est quel renseignement que tu utilises sans le dire ?

E4 : L'aire du rectangle.

P : Non. Dans l'autre sens, E4. Tu sais que l'aire sous la courbe doit valoir...

E4 : 1.

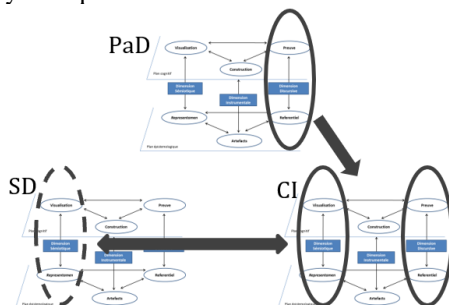
Dans cette phase, il y a émergence de l'élément R4 (l'aire sous la courbe qui doit être égale à 1) du référentiel théorique qui devient une contrainte pour le choix de la courbe, avec l'aide de l'enseignante (figure 10). Cette phase correspond à une activité d'organisation du raisonnement en amont de l'activité de traitement qui suit.



**Figure 10.** – Dynamique entre les trois ETM dans la phase 5.7

**Phase 5.8** [1 :25 :46] : Recherche individuelle – détermination de l'équation de la courbe de tendance

Cette activité de traitement se fait de manière individuelle pour chaque élève. Il y a une mobilisation pratique du référentiel théorique R4 (PaD) pour déterminer l'équation de la courbe de tendance : l'aire sous la courbe égale à 1. De plus, les connaissances sur la résolution d'équation sont mobilisées par les élèves pour trouver l'équation de la droite. La correspondance visuelle entre la courbe déterminée et l'histogramme reste un critère implicite, qui permet une vérification. La figure 11 rend visible cette dynamique.

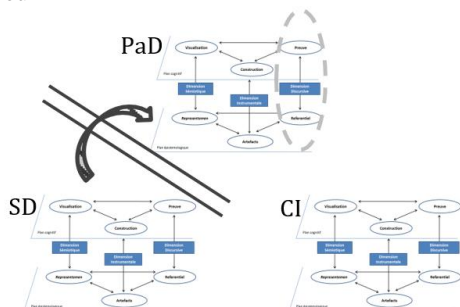


**Figure 11.** – Dynamique entre les trois ETM dans la phase 5.8

**Phase 5.9** [1:35:30] : Bilan : vers la rupture entre statistique descriptive et probabilités à densité

Dans cette phase, l'enseignante souhaite mettre en avant le fait que l'histogramme en jeu dans leur recherche de courbe de tendance est issu d'une simulation et donc que si l'on relance le générateur aléatoire, l'histogramme change. Les élèves précisent alors que les histogrammes auraient toujours la même « forme », « allure »... L'enseignante fait prendre conscience aux élèves que les

histogrammes ne permettent pas de calculer des probabilités, mais que justement la droite qui vient d'être trouvée va les aider à cela car « [cette] droite [...] modélise tous nos histogrammes en quelque sorte ». Un élève ajoute : « C'est le modèle théorique ». Ces deux phrases mettent en évidence le lien entre SD et PaD (implicitement, il s'agit de la loi des grands nombres) tout en créant une rupture entre les deux ETM, pour qu'ils ne soient pas assimilés l'un à l'autre (figure 12). Il apparaît en effet le caractère non spécifique de la courbe de tendance par rapport aux histogrammes, qui changent eux suivant l'échantillon choisi.



**Figure 12.** – Dynamique entre les trois ETM dans la phase 5.9

### Phase 5.10 : Utilité de la courbe de tendance (R5)

Dans cette phase, l'enjeu est de faire expliciter aux élèves ce à quoi va servir la courbe de tendance, c'est-à-dire l'élément R5 du référentiel théorique. Sur un exemple, un élève explique alors comment la courbe va permettre de calculer une probabilité :

P : On était en train d'essayer de calculer la probabilité que  $X$  soit entre 0 et 0,25.

P : Donc on aurait pu faire la somme des aires des petits rectangles de 0 à 0,25. Maintenant on fera quoi ?

E1 : On trace la droite  $x = 0,25$ .

[P complète la courbe tracée au tableau (figure 13)]

P : D'accord.

E1 : Maintenant on obtient un petit triangle.

[...]

P : Un triangle ?

E5 : A droite.

P : Ah, d'accord. Ok.

E1 : Là, c'est un rectangle donc on fait base fois hauteur... divisé par deux.

P : Je vois ton problème. Si, si je vois. Dis-moi, comment, à l'aide du dessin, tu vas calculer la probabilité qu'on souhaitait évaluer ?

E1 : Ah pardon [...]. Oui, oui. C'est un trapèze.

[...]

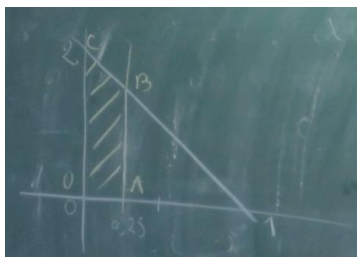
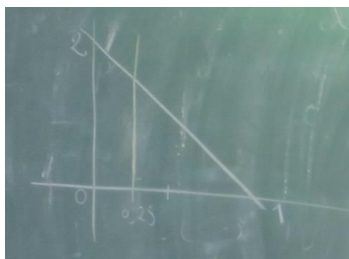
P : Comment est-ce que tu vas calculer cette probabilité, à l'aide de la courbe de tendance que tu... que tu as déterminée ?

E1 : [...] On se base sur l'aire de...

P : L'aire ! Voilà, ça sera... c'est l'aire. C'est l'aire de quoi ?

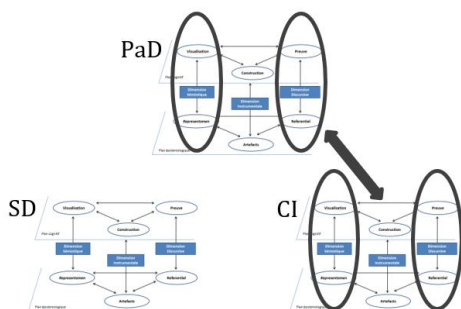
E12 : L'aire du trapèze.

P : [...] Bon on va peut-être le nommer ce trapèze, ça sera peut-être plus facile. Donc  $OABC$  pour vous... l'aire du trapèze  $OABC$ .



**Figure 13.** – Courbe tracée au tableau et complétée pendant la phase 5.10

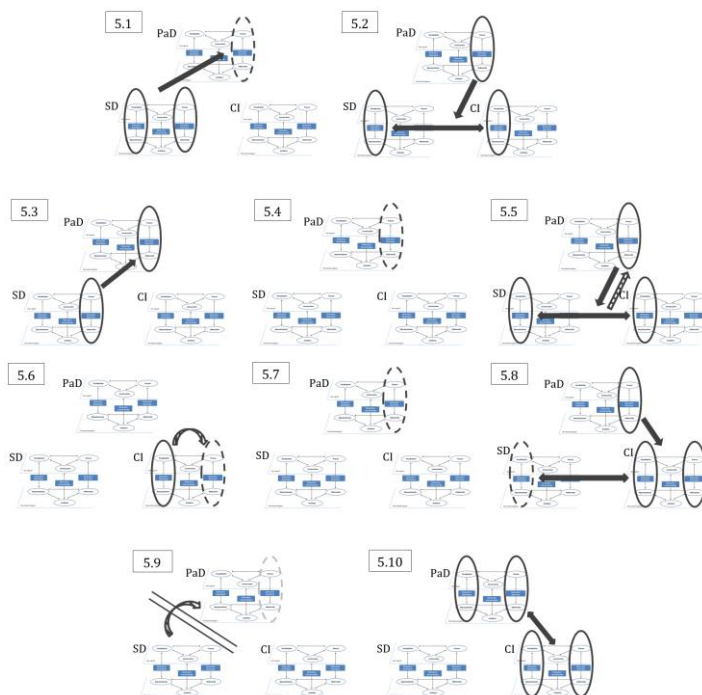
Cette phase permet de construire explicitement le lien entre PaD et CI par la construction de l'élément R5 du référentiel théorique « pour déterminer une probabilité, on calcule l'aire sous la courbe de tendance sur l'intervalle qui nous intéresse ». Dans ce problème, le calcul d'aire en question se ramène à des calculs d'aires élémentaires de niveau A1 (aires de triangles ou de trapèzes). La dynamique de cette phase, qui peut être vue comme une activité d'organisation du raisonnement, est illustrée par la figure 14. La courbe est à la fois la courbe d'une fonction et spécifiquement la courbe d'une fonction de densité (CI et PaD) d'où la dimension sémiotique activée dans les deux domaines et il y a alors un lien qui se crée entre probabilité (PaD) et aire sous la courbe (CI) au niveau de la dimension discursive. Pour le calcul effectif de l'aire, l'enseignante rappelle la formule de l'aire d'un trapèze.



**Figure 14.** – Dynamique entre les trois ETM dans la phase 5.10

### *Conclusion sur l'épisode 5*

L'enchaînement des diagrammes de cet épisode (figure 15) nous permet de repérer la présence d'un véritable jeu entre les trois sous-domaines (SD, PaD et CI). La statistique descriptive est le domaine crucial dans les sous-tâches précédant la sous-tâche 5. Elle est le point de départ de cet épisode car sans histogramme, ce travail mathématique de recherche d'une courbe de tendance ne serait pas possible. Il y a ensuite une construction au fur et à mesure de certains éléments du référentiel théorique autour de la notion de fonction de densité et du lien entre probabilités et calcul intégral, à savoir les éléments R1, R4 et R5. La nécessité de la positivité de la fonction (R3) n'apparaît pas dans ce problème. Chaque sous-domaine a son propre rôle dans la construction des connaissances. Les sous-domaines SD et PaD sont, au départ, travaillés en analogie (phase 5.1), mais à la fin une phase de bilan (phase 5.9) est proposée pour ne pas aller jusqu'à l'assimilation des deux sous-domaines et bien identifier la rupture entre les deux. Cette phase est prise essentiellement en charge par l'enseignante avec un appui sur les élèves. On peut l'expliquer par le fait qu'ici l'enseignante ajoute une nouvelle tâche aux élèves en revenant sur l'importance ou non du choix de l'histogramme sur lequel ils ont travaillé lors de cet épisode sur la courbe trouvée. L'enjeu ici est de mettre en évidence le rôle de la modélisation et le statut différent entre la statistique descriptive (réalité) et les probabilités (modèle).



**Figure 15.** – Enchaînement des dynamiques entre les ETM dans la sous-tâche 5

Nous repérons que le travail est en majorité à la charge du collectif classe (les dimensions entourées sont pratiquement toujours en gris foncé). Le rôle de l’enseignante n’est cependant pas à négliger dans les activités des élèves, car parfois elles sont rendues possibles grâce à une intervention de celle-ci (dimensions entourées en gris foncé en pointillé). C’est notamment le cas au niveau de la dimension discursive relative aux probabilités à densité, où l’enseignante met en place des aides procédurales indirectes (sous forme de question) pour aiguiller les élèves. Ce référentiel théorique étant en construction, des interventions de l’enseignante étaient prévisibles.

La dimension instrumentale n’est pas réellement sollicitée dans cet épisode, le logiciel est seulement un support pour visualiser l’histogramme et seule une courbe est tracée dessus à un moment de l’épisode. Cette dimension est sollicitée dans le travail mathématique relatif aux sous-tâches précédentes (non étudiées ici) pour la simulation ou le tracé des histogrammes. Les activités d’organisation du raisonnement sont le point de départ de cet



épisode et sont prises en charge par un seul élève. L'intervention de l'élève s'appuie sur la démarche rencontrée lors du devoir maison préliminaire. Ces étapes sont disponibles pour cet élève mais pas pour l'ensemble de la classe. L'enseignante prend du temps pour reformuler la démarche et la faire réexpliquer par un autre élève (phase 5.4). Les activités de traitement des élèves s'appuient la plupart du temps sur les dimensions sémiotiques entre SD et CI, avec un appui sur le référentiel théorique en construction de PaD.

Mises à part les aides procédurales indirectes, les autres interventions de l'enseignante sont de la reformulation, de la reprise, de la demande de précision, de la validation et une fois de la réfutation (phase 5.4).

Pour conclure, nous voyons émerger lors de cet épisode les éléments R1, R4 et R5 du référentiel théorique de la notion de fonction de densité en prenant appui sur les trois sous-domaines, grâce à de nombreux allers-retours entre chacun.

Lors de la séance suivante, une synthèse a été faite sur le problème (annexe 4) dans laquelle les éléments R1, R4 et R5 du référentiel théorique ont été rappelés, contextualisés au problème. L'enseignante a aussi amené les élèves vers l'élément R3 concernant la positivité de la fonction et a introduit le vocabulaire « fonction de densité ».

### **Analyse *a posteriori* du problème du volcan Aso**

Nous focalisons notre analyse sur la sous-tâche : *Déterminer l'expression de la fonction de densité*, sous-tâche 3 dans le problème du volcan.

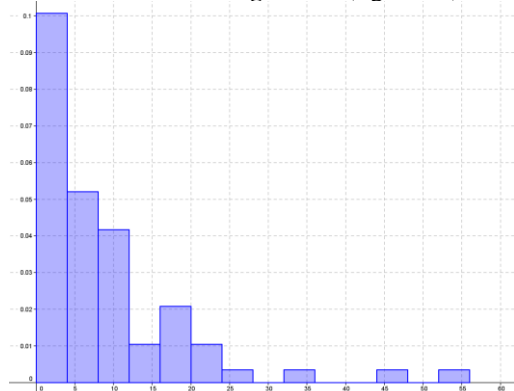
#### *Les épisodes de la sous-tâche 3*

Lors de la réalisation de la sous-tâche 3 (environ 27 minutes) du problème de la rencontre, plusieurs épisodes sont repérables dans le déroulement :

- 3.1 Phase collective : Rappel du vocabulaire « fonction de densité » ;
- 3.2 Phase collective : Détermination de l'ensemble de définition de la fonction de densité ;
- 3.3 Phase individuelle : Recherche de la fonction de densité ;
- 3.4 Phase collective : Réfutation des fonctions de la famille de  $f(x) = \frac{a}{bx+c}$  ;

- 3.5 Phase individuelle (à la maison) : Ajustement de la fonction exponentielle ;
- 3.6 Phase collective : Mise en commun.

À la fin de la phase 3.2, la classe se met d'accord et reformule la sous-tâche 3 par : *On cherche une fonction  $f$  définie sur  $[0; +\infty[$  qui approche « au mieux » l'histogramme (figure 16).*



**Figure 16.** – Histogramme de fréquence des temps d'attente entre deux éruptions du volcan Aso entre 1229 et 1897 avec une amplitude de 4 ans. Derrière la fonction en question dans l'énoncé de la tâche se cache la fonction de densité de la variable aléatoire du temps d'attente entre deux éruptions du volcan Aso. Cependant la manière dont est posée la question peut laisser penser que nous sommes simplement à la recherche d'une fonction (de son expression), et donc dans le domaine de l'analyse. Ici se mêlent en fait les deux sous-domaines PaD et CI.

Durant la phase de recherche individuelle (épisode 3.3), les élèves tracent à main levée une courbe qui semble leur convenir. L'enseignante leur demande de trouver l'expression de la fonction. Beaucoup d'élèves pensent à la fonction inverse. L'enseignante décide ensuite de faire une mise en commun. Nous allons ici étudier spécifiquement l'épisode 3.4, dans lequel les élèves sont véritablement confrontés aux propriétés que doit vérifier une fonction de densité.

#### *Analyse de l'épisode 3.4*

Cet épisode correspond au passage entre 39min30s et 48min50s de la séance 3 (un peu moins de 10 minutes). Il est découpé en six phases. Pour chaque phase, l'annexe 6 illustre les différentes courbes sur lesquelles porte le travail mathématique de la classe.

**Phase 3.4.1** [39:30] : Recherche de correspondance entre l'histogramme et la courbe (R1)

Dans cette phase, l'enseignante commence en traçant sur le logiciel *GeoGebra* la fonction  $x \rightarrow \frac{1}{x}$  dans la même fenêtre que l'histogramme. Les élèves entrent alors dans un travail de visualisation et demandent à l'enseignante de déplacer la courbe pour qu'elle approche au mieux l'histogramme :

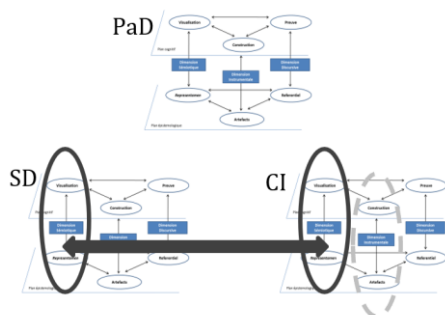
E3 : On la décale vers le bas.

[L'enseignante déplace la courbe sur *GeoGebra*]

P : Ah oui, c'est sûr, c'est plus facile que sur le papier.

P : Alors je la décale.

Ils recherchent effectivement une correspondance entre la courbe tracée et le haut de l'histogramme. Dans les diagrammes des ETM, nous signalons toujours en gris foncé les dimensions de l'ETM sollicitées par les élèves et en gris clair ceux par l'enseignante. Nous remarquons donc dans cette phase un travail de visualisation entre l'histogramme et la courbe, et donc entre les sous-domaines de la statistique (SD) et du calcul intégral (CI) (figure 17). La dimension instrumentale relative à la courbe (CI) est prise en charge par l'enseignante, car c'est elle qui manipule le logiciel, cependant elle agit sous commande des élèves (nous représentons donc cette action « sous commande des élèves » par des pointillés gris clair). Il s'agit ici d'une activité de traitement dans le registre graphique.



**Figure 17.** – Dynamique entre les trois ETM dans la phase 3.4.1

**Phase 3.4.2** [40:28] : Prise en compte de la contrainte de positivité (R3)

À la suite du déplacement de la courbe « vers le bas », l'enseignante dit :

P : Alors oui mais est-ce que je peux la décaler...? Regardez bien

ce qu'il se passe partout parce que...

E7 : Non, pas là, pas là.

P : Pourquoi je peux plus faire ça ?

Plusieurs élèves : Parce qu'elle est négative.

P : Parce qu'elle est négative. On a dit qu'on cherchait des fonctions... ?

Plusieurs élèves : Positives.

Cette aide procédurale indirecte de l'enseignante induit l'activité des élèves de reconnaissance d'une propriété (en construction) et donc il y a une mobilisation de l'élément R3 du référentiel théorique de la fonction de densité (PaD), qui relance le travail de visualisation (figure 18) entre SD et CI. Il y a à nouveau une activité de traitement dans le registre graphique :

E8 : Faut la décaler sur la gauche.

[...]

P : Là c'est bien ? E3, c'est toi qui avais proposé ça, tu me dis stop parce que là...

[P décale la courbe sur *GeoGebra*]

[...]

P : Alors je la reprends. Donc on la laisse au-dessus de l'axe des abscisses quand même et tu veux que je la décale vers la... par... enfin vous voyez ce que je veux dire.

[...]

P : Où est-ce que tu veux que je la décale, E3 ?

E3 : A peu près là.

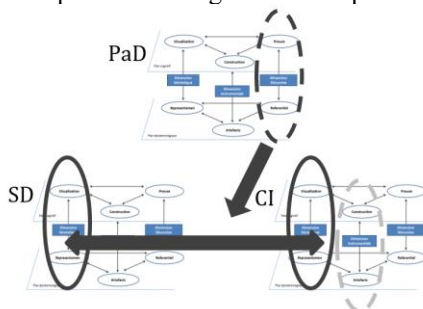
[...]

E3 : Oui là c'est bien.

E9 : Moi je dirais que faut la remonter.

P : Là, ça vous plaît ?

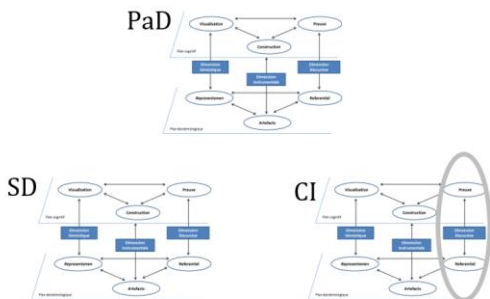
Les élèves ont l'air d'accord pour s'arrêter sur la fonction représentée en annexe 6 (phase 3.4.2, à gauche), mais finalement un élève propose de déplacer la fenêtre, et ils s'aperçoivent que la courbe coupe l'axe des abscisses et donc n'est pas positive sur  $[0; +\infty[$ . Dans cette phase, c'est la contrainte R3 qui guide le travail mathématique des élèves gardant à l'esprit la contrainte R1.



**Figure 18.** – Dynamique entre les trois ETM dans la phase 3.4.2

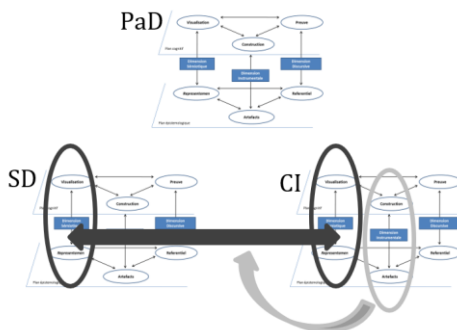
**Phase 3.4.3** [42:53] et **phase 3.4.4** [43 :33] : Passage au registre algébrique et travail sur les limites

L'enseignante propose ensuite aux élèves un changement de registre pour passer du registre graphique au registre algébrique des fonctions. Après une tentative d'aide procédurale indirecte de l'enseignante en demandant aux élèves quelle est la forme générique de telles fonctions, elle décide de leur donner directement (aide procédurale directe) la forme attendue :  $f(x) = d + \frac{a}{bx+c}$ . Ensuite l'enseignante leur demande : « *Quelle doit être la limite de la fonction en  $+\infty$  ?* » (aide procédurale indirecte) et finit à nouveau par leur donner la réponse et faire le raisonnement pour aboutir à  $d = 0$  et donc restreindre les fonctions à celles de la forme :  $f(x) = \frac{a}{bx+c}$ . Ces activités de traitement dans le registre algébrique des fonctions sont entièrement prises en charge par l'enseignante (figure 19).

**Figure 19.** – Dynamique entre les trois ETM dans les phases 3.4.3 et 3.4.4

**Phase 3.4.5** [44:24] : Nouvelle recherche de correspondance entre l'histogramme et la courbe

L'enseignante introduit dans le logiciel *GeoGebra* des curseurs  $a$ ,  $b$  et  $c$  (dont les valeurs peuvent varier) et entre la fonction  $f(x) = \frac{a}{bx+c}$  (aide procédurale directe). La dimension instrumentale, entièrement prise en charge par l'enseignante, relance à nouveau le travail de visualisation des élèves (figure 20). Il s'agit toujours d'une activité de traitement dans le registre graphique, prenant en compte les contraintes R1 et R3.



**Figure 20.** – Dynamique entre les trois ETM dans les phases 3.4.5

**Phase 3.4.6 [46:28] :** Prise en compte de la contrainte de l'aire (R4)

L'enseignante propose alors une aide procédurale indirecte :

P : Rappelez-moi, il y avait une contrainte sur la courbe qu'on cherche quand même. C'est quoi ?

E8 : L'aire c'est 1.

P : L'aire doit faire...

Plusieurs élèves : 1.

P : 1.

Plusieurs élèves font référence à R4 : l'aire sous la courbe doit être égale à 1. Cependant, cette propriété n'est pas tout de suite acceptée par l'ensemble des élèves, car il s'agit ici d'une aire sur un domaine non borné, sur  $[0; +\infty[$  :

E1 : Ah oui, oui. Du coup, ça marche pas si on met  $+\infty$ .

E10 : Faut que ça touche.

E1 : J'avais raison, alors.

P : Ah non, pourquoi tu dis ça ?

E1 : Parce que c'est une forme... parce que c'est une forme qui n'est pas finie.

P : [...] Alors tu veux dire que parce que...

E11 : On l'a vu, on l'a vu. L'aire elle grandit... Ah oui, c'est un DM.

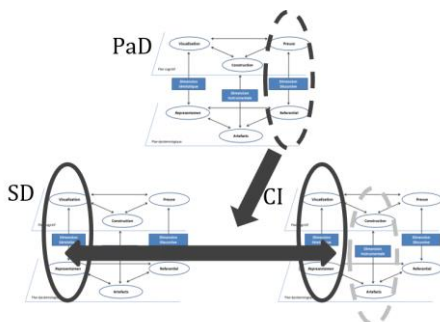
P : Oui. [...] On a vu, E1, que c'était pas un problème. Qu'on pouvait avoir un domaine infini et une aire...

E1 : C'est pas ça, c'est pas ça. C'est pas une forme parce qu'il n'y a pas de fermeture.

P : Oui mais non plus, dans les rectangles y avait pas de fin... On les continuait à l'infini. [référence à un devoir maison fait quelques semaines plus tôt]

En prenant appui sur un devoir maison travaillé plus tôt dans l'année (connaissances anciennes), tous les élèves acceptent ensuite cette possibilité d'aire finie d'un domaine infini. Cette activité de reconnaissance d'une propriété (en construction) de la fonction de densité relance à nouveau le travail de visualisation

(figure 21), avec une manipulation du logiciel pour déplacer les courbes. La commande *GeoGebra*, précisant la valeur de l'aire sous la courbe, va permettre de réfuter le choix d'une fonction de cette famille.



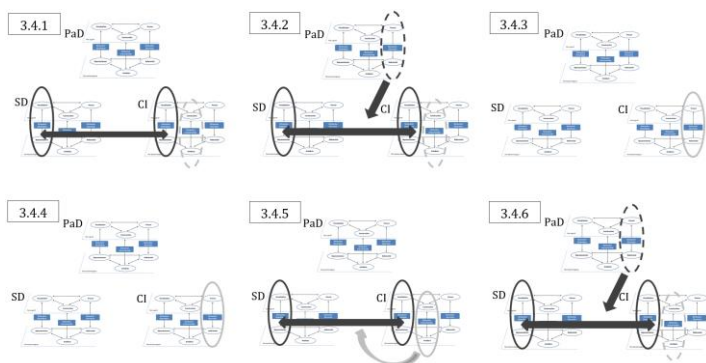
**Figure 21.** – Dynamique entre les trois ETM dans la phase 3.4.6

### Conclusion sur l'épisode 3.4

En résumé, comme pour le problème de la rencontre, nous remarquons qu'il y a véritablement une circulation entre les trois sous-domaines, où chaque sous-domaine a son propre rôle (figure 22). Pour le sous-domaine des probabilités à densité, il y a une véritable mobilisation de la dimension discursive avec les éléments théoriques R1, R3 et R4 (phases 3.4.2 et 3.4.6) pour relancer le travail de visualisation entre la statistique et l'analyse et permettre de réfuter les fonctions candidates : cela montre bien un réinvestissement du référentiel théorique de la fonction de densité, introduit dans les séances précédentes. Nous remarquons aussi que le travail mathématique est en majorité guidé par l'ETM collectif, et donc par certains élèves. Cependant, le rôle de l'enseignante n'est à nouveau pas à négliger au niveau des activités de traitement en lien avec la manipulation du logiciel (mais la plupart du temps sous commande des élèves) et les activités de traitement dans le registre algébrique des fonctions (phases 3.4.3 et 3.4.4). L'enseignante permet aussi régulièrement de relancer le travail de la classe et notamment de conduire les élèves à mobiliser les connaissances en construction grâce à des relances sous forme de questions (aides procédurales indirectes).

La dimension instrumentale est présente dans cet épisode. Elle n'est pas seulement mobilisée dans le domaine de la statistique, avec la construction de l'histogramme (sous-tâche précédente), mais essentiellement dans le domaine de l'analyse, pour la recherche d'une courbe adéquate. Ceci est dû au fait que

l'expression de la fonction ne peut pas ici se trouver à la main mais seulement par essai-erreur grâce au logiciel, du fait que l'aire sous la courbe est non calculable par les élèves pour le moment (aire de niveau A3).



**Figure 22.** - Enchaînement des dynamiques entre les ETM dans l'épisode 3.4

## CONCLUSIONS ET PERSPECTIVES

Revenons sur nos trois questions de recherche.

Les activités des élèves associées aux tâches d'introduction conçues dans le cadre de notre ingénierie didactique permettent-elles de construire la notion de fonction de densité, à savoir tous les éléments du référentiel théorique (R1 à R5) ? À travers les analyses des épisodes, nous avons pu repérer l'émergence (dans le problème de la rencontre) puis la consolidation (dans le problème du volcan Aso) des éléments R1, R3, R4 et R5 du référentiel théorique de la notion de fonction de densité que nous cherchions à construire. En réalité, l'élément R2 (définition et continuité de la fonction de densité) est aussi présent dans les deux problèmes mais de manière implicite. Par exemple, pour le problème du volcan Aso, les élèves cherchent *une fonction  $f$  définie sur  $[0; +\infty[$  qui approche « au mieux » l'histogramme* ; cela sous-entend que la fonction est définie et continue sur l'intervalle étant donné que toutes les fonctions étudiées en terminale sont continues sur leur ensemble de définition.

Cette construction du référentiel théorique est-elle rendue possible grâce à des circulations entre les trois ETM en jeu ? Nous avons mis en évidence que la construction du référentiel théorique du sous-domaine des probabilités à densité se fait progressivement



grâce à un appui sur les deux autres sous-domaines, le calcul intégral et la statistique descriptive. La statistique descriptive permet de faire émerger les propriétés de la fonction de densité à partir des histogrammes et le calcul intégral, avec un jeu sur les fonctions, permet de mobiliser concrètement ces propriétés. Ce sont ces dynamiques qui favorisent des activités permettant un enrichissement des connaissances des élèves sur la notion de fonction de densité.

L'ETM collectif de la classe permet-il seul de construire l'ETM idoine effectif ou l'enseignante joue-t-elle un rôle dans cette construction ? Le regard que nous avons porté sur le rôle respectif des élèves et de l'enseignante nous a permis de mettre en avant que l'ETM collectif de la classe est effectivement le moteur principal dans la construction de l'ETM idoine effectif de la classe. Cependant, il faut garder à l'esprit que ce ne sont pas tous les élèves qui sont nécessairement acteurs de l'ensemble des différentes activités. Nous pouvons seulement dire que ces activités sont principalement à l'initiative d'un (ou plusieurs) élève(s) et non de l'enseignante. Quoi qu'il en soit, le rôle central des élèves dans la prise en charge des activités mathématiques est possible grâce à plusieurs facteurs. Tout d'abord, les anticipations qui ont pu être faites avant la séquence (comme par exemple, la problématique de l'aire finie d'un domaine infini ou encore le devoir maison préliminaire) sont indispensables pour permettre aux élèves de faire avancer le travail mathématique de la classe. Ensuite, un point très important à ne pas négliger est le rôle de l'enseignante : les aides procédurales indirectes régulières de l'enseignante permettent de débloquer les élèves à des moments clés pour leur donner les moyens de construire de nouvelles connaissances. Il faut voir ces aides, non pas comme des constats négatifs sur le travail des élèves, mais comme des tremplins nécessaires pour que le nouveau savoir soit vraiment intégré par les élèves. D'ailleurs, dans nos analyses rattachées à une phase particulière du travail des élèves, nous avons décrit ces aides comme des aides à visée procédurale ; cependant, du fait que ces aides soient répétées entre les deux problèmes, elles pourraient être réinterprétées comme des aides à visée constructive. Nous pourrions notamment nous en convaincre en nous penchant sur les institutionnalisations qui suivent respectivement les deux problèmes, dans lesquelles nous retrouvons des points d'appui sur ce même type d'aides.

Pour poursuivre sur les institutionnalisations, nous ne l'avons pas mis en avant dans cet article mais elles constituent aussi des phases importantes dans la construction du nouveau savoir. Nous

donnons ici un rapide exemple. Le problème de la rencontre fait émerger la propriété que l'aire sous la courbe de la fonction de densité doit être égale à 1 (R4) et que la courbe doit « lisser » l'histogramme (R1). En revanche, la propriété de positivité de la fonction (R3) n'apparaît pas, car elle est implicite dans ce problème. C'est la phase d'institutionnalisation qui suit le problème de la rencontre qui permet d'exhiber cette propriété. Sur un jeu de contre-exemples, en proposant des courbes changeant de signes comme courbes de densité, l'enseignante amène les élèves à formuler le fait que la fonction de densité doit nécessairement être positive. Cette propriété est ensuite nécessaire et convoquée par les élèves pour le problème du volcan Aso.

Pour conclure, ces tâches d'introduction sont donc porteuses de sens sur la notion de fonction de densité, contrairement à celles que l'on peut trouver dans les manuels scolaires (Derouet & Parzysz, 2016). Ces résultats s'appuient sur une expérimentation dans des conditions particulières (enseignante impliquée dans la conception de la séquence, élèves d'un établissement favorisé...), il serait maintenant intéressant de questionner la reproductibilité de ces tâches d'introduction dans des conditions différentes (enseignants différents, public différent, modalités différentes...) et d'identifier les régularités et les variations suivant ces paramètres. Dans le cadre d'un projet en cours<sup>8</sup>, le problème de la rencontre sera analysé dans le contexte d'une expérimentation où les élèves ont travaillé cette fois-ci en groupe. Cela permettra notamment de voir si, dans ce cas de figure, les élèves réussissent à nouveau à construire les mêmes connaissances et si les types d'interventions de l'enseignant sont différents ou non.

## BIBLIOGRAPHIE

ARTIGUE, M. (1988). Ingénierie didactique. *Recherches en Didactique des Mathématiques*, 9(3), 281–308.

---

<sup>8</sup> Projet Co@Diff « De la conception d'une séquence d'enseignement à sa diffusion auprès des enseignants de mathématiques : développement d'une ressource en ligne sur l'enseignement des lois à densité en articulation avec le calcul intégral en terminale S », soutenu par le Groupement d'Intérêt Scientifique « Education et formation » de l'ESPE de Strasbourg.

- BANSILAL, S. (2012). Understanding student engagement with normal distribution problems. *Pythagoras*, 33(1), 37–49.
- BANSILAL, S. (2014). Using an APOS framework to understand teachers' responses to questions on the normal distribution. *Statistics Education Research Journal*, 13(2), 42–57.
- BATANERO, C., GODINO, C., & ROA, R. (2004). Training teachers to teach probability. *Journal of Statistics Education*, 12(1).
- BATANERO, C., TAUBER, L., & MEYER, R. (1999). From data analysis to inference: a research project on the teaching of normal distributions. In *Bulletin of the International Statistical Institute: Proceedings of the Fifty-Second Session of the International Statistical Institute (Tome LVIII, Book 1)* (pp. 57–58). Helsinki: International Statistical Institute.
- BATANERO, C., TAUBER, L. M., & SÁNCHEZ, V. (2004). Students' reasoning about the normal distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 257–276). Dordrecht: Springer.
- CHEVALLARD, Y., & WOZNIK, F. (2005). Enseigner la statistique au secondaire. Entre genre prochain et différence spécifique. In A. Mercier, et C. Margolinas (Ed.), *Balises pour la didactique des mathématiques* (pp. 195–218). Grenoble : La pensée sauvage.
- DANTAL, B., & HENRY, M. (2003). Variables aléatoires et lois de probabilité : utilisation de l'analyse en probabilités. In APMEP (Ed.), *Probabilités au lycée* (pp. 125–142). Paris : APMEP.
- DEROUET, C. (2016). *La fonction de densité au carrefour entre probabilités et analyse en terminale S. Etude de la conception et de la mise en oeuvre de tâches d'introduction articulant lois à densité et calcul intégral*. Université Paris Sorbonne Cité, Université Paris Diderot.
- DEROUET, C., & ALORY, S. (2018). Une séquence d'enseignement articulant les lois de probabilité à densité et le calcul intégral en terminale S. *Repères IREM*, 113, 45–80.
- DEROUET, C., & PARZYSZ, B. (2016). How can histograms be useful for introducing continuous probability distributions? *ZDM - Mathematics Education*, 48(6), 757–773.
- DUVAL, R. (1993). Registres de représentation sémiotique et fonctionnement cognitif de la pensée. *Annales de Didactique et de Sciences Cognitives*, 5, 37–65.
- HABERMAS, J. (1970). *Technology and Science as Ideology*. Boston : Beacon Press.
- HENRY, M. (2003). Des lois de probabilité continues en terminale S, pourquoi et pour quoi faire ? *Repères - IREM*, 51, 5–25.
- KUZNIAK, A. (2011). L'espace de Travail Mathématique et ses genèses. *Annales de Didactique et de Sciences Cognitives*, 16, 9–24.
- KUZNIAK, A. (2014). Travail mathématique et domaines mathématiques. *Revista Latinoamericana de Investigación En Matemática Educativa, Especial 2* (Tome II), 385–400.
- KUZNIAK, A., & RICHARD, P. R. (2014). Espaces de travail mathématique. Points de vue et perspectives. *Revista Latinoamericana de Investigación En Matemática Educativa, Especial 2* (Tome I), 29–40.

- KUZNIAK, A., TANGUAY, D., & ELIA, I. (2016). Mathematical Working Spaces in schooling: an introduction. *ZDM - Mathematics Education*, 48(6), 721-737.
- MINISTÈRE DE L'ÉDUCATION NATIONALE (MEN). (2011). Programme de l'enseignement spécifique et de spécialité de mathématiques. Classe terminale de la série scientifique. *Bulletin Officiel de l'Éducation Nationale*, [http://cache.media.education.gouv.fr/file/special\\_8\\_men/98/4/mathematiques\\_S\\_195984.pdf](http://cache.media.education.gouv.fr/file/special_8_men/98/4/mathematiques_S_195984.pdf)
- MONTOYA DELGADILLO, E., & VIVIER, L. (2014). Les changements de domaine dans le cadre des Espaces de Travail Mathématiques. *Annales de Didactique et de Sciences Cognitives*, 19, 73–101.
- PARIES, M., ROBERT, A., & ROGALSKI, J. (2009). Comment l'enseignant de mathématiques, en classe, met ses élèves sur le chemin des connaissances : un point de vue méthodologique en didactique des mathématiques. *Travail et Apprentissages*, 3, 95–123.
- ROBERT, A. (2008a). Une méthodologie pour analyser les activités (possibles) des élèves en classe. In F. Vandebrouck (Ed.), *La classe de mathématiques : activités des élèves et pratiques des enseignants* (pp. 45–58). Toulouse : Octarès Éditions.
- ROBERT, A. (2008b). Sur les apprentissages des élèves : une problématique inscrite dans les théories de l'activité et du développement. In F. Vandebrouck (Ed.), *La classe de mathématiques : activités des élèves et pratiques des enseignants* (pp. 33–43). Toulouse : Octarès Éditions.
- ROBERT, A., & VANDEBROUCK, F. (2014). Proximités-en-acte mises en jeu en classe par les enseignants du secondaire et ZDP des élèves : analyse de séances sur des tâches complexes. *Recherches En Didactique Des Mathématiques*, 34(2-3), 239–285.
- RODITI, E. (2009). L'histogramme : à la recherche du savoir à enseigner. *Spirale-Revue de Recherches En Éducation*, 43, 129–138.
- ROGALSKI, J. (2008). Le cadre général de la théorie de l'activité. Une perspective de psychologie ergonomique. In F. Vandebrouck (Ed.), *La classe de mathématiques : activités des élèves et pratiques des enseignants* (pp. 23–30). Toulouse : Octarès Éditions.
- VANDEBROUCK, F. & ROBERT, A. (2017). Activités mathématiques des élèves avec des technologies. *Recherches en Didactique des Mathématiques*, 37(2-3), 333-382.

Annexe 1 : Extrait du devoir maison traité en amont des séances d'introduction

### Exercice 1

Nous disposons des données statistiques suivantes :

Thème      Population  
Sous-thème      Évolution et structure de la population  
POP1A - Population par âge regroupé en 2011  
Commune de Paris  
© Insee

#### Population totale par âge regroupé

	Effectif
Moins de 3 ans	73946
3 à 5 ans	65679
6 à 10 ans	103810
11 à 17 ans	140886
18 à 24 ans	244026
25 à 39 ans	598565
40 à 54 ans	435966
55 à 64 ans	257600
65 à 79 ans	222118
80 ans ou plus	107378
Ensemble	2249975

Source : Insee, RP2011 exploitation principale.

1. Quelle est la population donnée ? Quelle est l'effectif total ?
2. Quel est le caractère étudié ?
3. Ce caractère statistique est-il qualitatif ou quantitatif ? Est-il discret ou continu ?
4. A l'aide du tableau ci-dessus, compléter le tableau de fréquences suivant (en arrondissant à  $10^{-2}$  près) :

Tranche d'âge	[0; 4[	[4; 6[	[6; 11[	[11; 18[	[18; 25[	[25; 40[	[40; 55[
Fréquence							

Tranche d'âge	[55; 65[	[65; 80[	[80; 105[
Fréquence			

5. Nous voulons représenter graphiquement ces données à l'aide d'un histogramme de fréquences. Nous allons d'abord faire quelques commentaires et utiliser une analogie avec la géographie.

En 2005, Monaco avait 32 543 habitants et le Japon, 127 417 244 (source : Institut national d'études démographiques). Bien sûr, les démographes diront que ces renseignements sont très largement insuffisants pour comparer la démographie des deux pays : il faut au minimum s'intéresser aux superficies de ces deux pays et calculer pour chacun d'entre eux la densité de population, c'est-à-dire le nombre d'habitants au kilomètre carré. Avec une superficie de 2,02 km<sup>2</sup> pour Monaco et de 378 000 km<sup>2</sup> pour le Japon, les densités sont respectivement  $d_1 = 32\,543 / 2,02 = 16\,110,40$  h/km<sup>2</sup> pour Monaco et  $d_2 = 127\,417\,244 / 378\,000 = 337$  h/km<sup>2</sup> pour le Japon. Autrement dit, alors que la population de Monaco est la moins importante en taille, sa densité de population est plus importante que celle du Japon.

Dans notre exemple, il n'est donc pas pensable de représenter de la même façon la fréquence 0,05 de la tranche d'âge [6; 11[ et de la tranche d'âge [80; 105[. Ces deux classes ont la même fréquence mais l'amplitude de la première classe est inférieure à celle de la seconde. Par analogie avec la densité de population, on comparera les rapports de fréquence d'une classe d'âge sur l'amplitude de cette classe. On parlera de densité.

Dans notre exemple, la densité de fréquence de la tranche d'âge [6; 11[ est  $\frac{0,05}{5} = 0,01$  et la densité de la tranche d'âge [80; 105[ est  $\frac{0,05}{25} = 0,002$ . Comme perçue, la densité de fréquence de la première tranche d'âge est plus grande que la densité de fréquence de la seconde tranche d'âge. La représentation graphique doit rendre visible ce phénomène.

### Définition

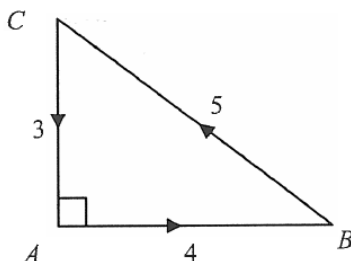
Un **histogramme de fréquences** est un diagramme composé de rectangles collés dont les **aires** sont proportionnelles aux fréquences et dont les bases sont déterminées par les intervalles des classes. Dans le cas d'une variable quantitative continue, on définit la **densité** de fréquence  $d_i$  d'une classe de fréquence  $f_i$  et d'amplitude  $a_i$  par :  $d_i = f_i / a_i$ .

- a) Justifier que, dans un histogramme, les hauteurs des rectangles sont proportionnelles à la densité. *Ainsi une classe de densité plus forte est représentée par un rectangle plus haut.*
- b) Compléter l'histogramme de fréquences de la répartition de la population parisienne en 2011 par âge (sur la feuille ci-après). On donne l'aire d'un carreau correspondant à une fréquence de 0,001. Que représente l'axe des ordonnées ? Graduer l'axe.
6. Dans quelle tranche d'âge la densité est-elle la plus importante ? Cela vous semble-t-il raisonnable ?
7. Dans un histogramme, si un rectangle est plus haut qu'un autre, peut-on affirmer que la fréquence associée est la plus grande ?

### Exercice 2

Une puce électronique se déplace à vitesse constante sur le « pourtour » d'un triangle, toujours dans le même sens (indiqué par la flèche). On suppose que la puce démarre au point C.

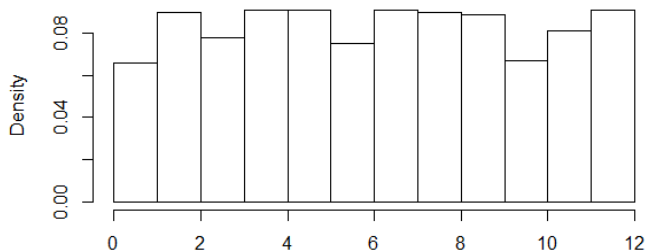
Une panne se produit subitement et la puce s'arrête instantanément.



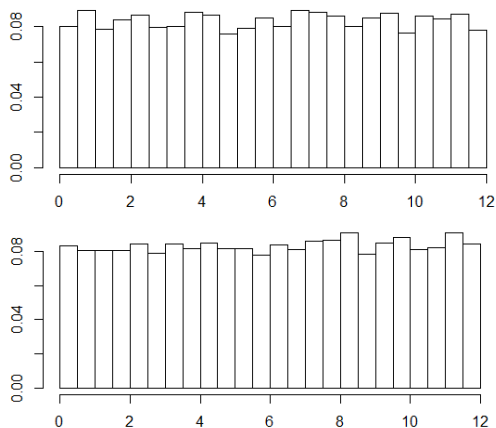
1. Quelle est la probabilité  $p_1$  que la puce s'arrête sur l'hypoténuse ?
2. Quelle est la probabilité  $p_2$  que la puce s'arrête au sommet de l'angle droit ? Commenter la valeur de  $p_2$  obtenue.
3. Quelle est la probabilité  $p_3$  que la puce s'arrête au milieu du segment  $[AB]$  ? Commenter la valeur de  $p_3$  obtenue.
4. On peut remarquer que cette situation se ramène à l'expérience aléatoire qui correspond à choisir un nombre réel au hasard dans l'intervalle  $[0 ; 12]$ .

A l'aide d'un tableur, on a simulé le tirage de 1000 nombres réels dans l'intervalle  $[0 ; 12]$ .

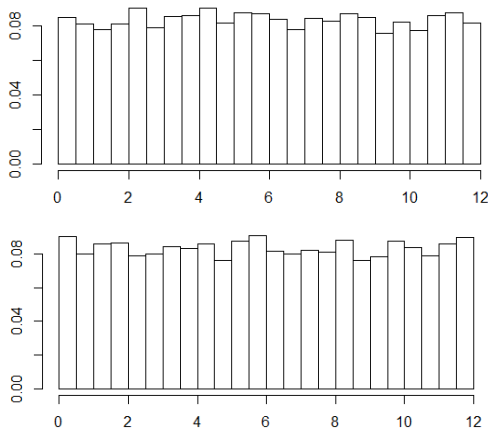
Voici l'histogramme de fréquences :



- a) Quel est l'amplitude de chaque classe ?
  - b) A quelle fréquence les puces s'arrêtent-elles sur l'hypoténuse ?
5. On a ensuite simulé, plusieurs fois, le tirage de 10 000 nombres réels dans l'intervalle  $[0 ; 12]$ . Voici plusieurs histogrammes correspondants :







- Les histogrammes ci-dessus sont différents. Justifier pourquoi.
- Malgré ces différences ne semble-t-il pas y avoir une « courbe de tendance », c'est-à-dire une courbe qui « lisse » l'histogramme (la même pour les quatre) ? Tracer la sur chacun des histogrammes ci-dessus.
- Que doit valoir l'aire sous cette courbe ?
- En déduire l'équation de cette courbe.
- A l'aide de cette courbe, retrouver la probabilité  $p_1$  que la puce s'arrête sur l'hypoténuse.
- De la même façon, retrouver la probabilité  $p_2$  que la puce s'arrête au sommet de l'angle droit.

## Annexe 2 : Exercice d'application du problème de la rencontre

**Exercice 1 :** En utilisant le modèle mis en place, répondez aux questions suivantes.

- Quelle est la probabilité que le premier arrivé attende plus de 40 minutes ?
- Quelle est la probabilité que le premier arrivé attende moins de 40 minutes ?
- Quelle est la probabilité que le premier arrivé attende entre 20 et 40 minutes ?
- Quelle est la probabilité que Karine et Olivier arrivent en même temps ?
- Quelle est la probabilité que le premier arrivé attende exactement 40 minutes ?
- Finalement, ils ont décidé de ne pas attendre plus d'un quart d'heure. Quelle est la probabilité qu'ils se retrouvent ?
- Le premier arrivé a attendu moins d'une demi-heure. Quelle est la probabilité pour qu'il ait attendu plus d'un quart d'heure ?

Annexe 3 : Énoncé du problème du volcan Aso distribué aux élèves (avec un extrait du tableau des données)

Le volcan Aso, situé sur l'île de Kyushu au Japon, est l'un des plus actifs au monde. On possède les statistiques de ses éruptions, régulièrement tenues depuis le XIII<sup>e</sup> siècle. Nous avons, dans le document tableur, les années d'éruptions jusqu'au XIX<sup>e</sup> siècle (à partir du XX<sup>e</sup> siècle, les données, d'une autre nature, ne sont pas compatibles(\*)).

**Années des 73 éruptions du volcan Aso du XIII<sup>e</sup> au XIX<sup>e</sup> siècle**

Donnée n°	Année	Temps d'attente	Donnée n°	Année	Temps d'attente
1	1229		34	1562	4
2	1239	10	35	1563	1
3	1240	1	36	1564	1
4	1265	25	37	1576	12
5	1269	4	38	1582	6
6	1270	1	39	1583	1
7	1272	2	40	1584	1
8	1273	1	41	1587	3
9	1274	1	42	1596	11
10	1281	7	43	1611	13
11	1286	5	44	1612	1
12	1305	19	45	1613	1
13	1324	19	46	1620	7
14	1331	7	47	1631	11
15	1335	4	48	1637	6
16	1340	5	49	1649	12
17	1345	5	50	1658	10

## Annexe 4 : Trace écrite à la suite du problème de la rencontre (séance 2)

## Synthèse

On note  $T_O$  le moment d'arrivée d'Olivier

$T_K$  le moment d'arrivée de Karine

$T_O$  et  $T_K$  sont des variables aléatoires qui prennent toutes les valeurs réelles de  $[7; 8]$  que l'on peut ramener à  $[0; 1]$ .

On dit que ce sont des variables aléatoires continues qui suivent le modèle de la puce.

On note  $X$  la variable aléatoire égale au temps d'attente du premier arrivé.

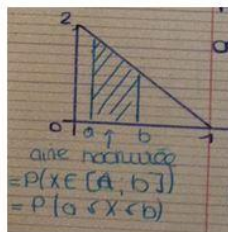
$X$  prend toutes les valeurs de  $[0; 1]$ .

Pour décrire la loi de probabilité de  $X$ , on introduit une fonction  $f$  définie sur  $[0; 1]$  par  $f(x) = -2x + 2$  telle que la courbe de  $f$  « lisse le haut des rectangles de l'histogramme ».

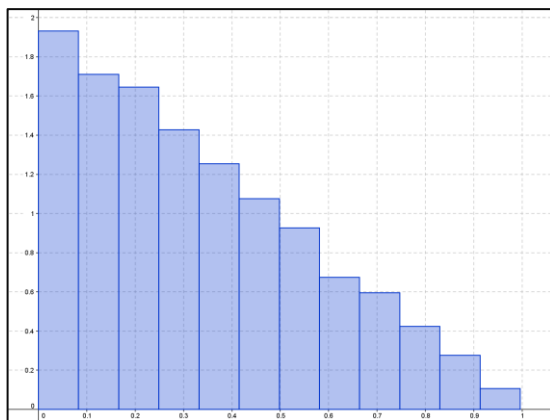
Ainsi pour calculer la proba  $P(X \in [a; b])$  où  $[a; b] \subset [0; 1]$ , on calcule l'aire sous la courbe  $C_f$  entre  $a$  et  $b$ .

$f$  s'appelle la fonction de densité de probabilité de  $X$ .

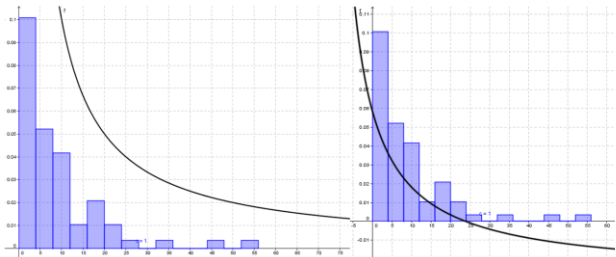
$f$  possède les propriétés suivantes :  $f$  est positive sur  $[0; 1]$ , l'aire sous  $C_f$  entre 0 et 1 vaut 1.



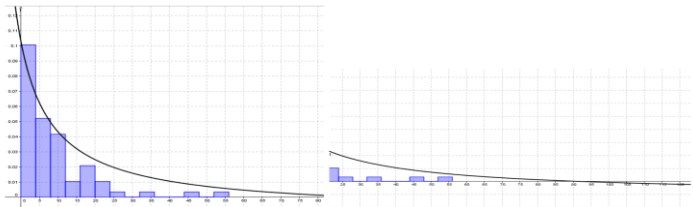
## Annexe 5 : Histogramme distribué aux élèves (problème de la rencontre)



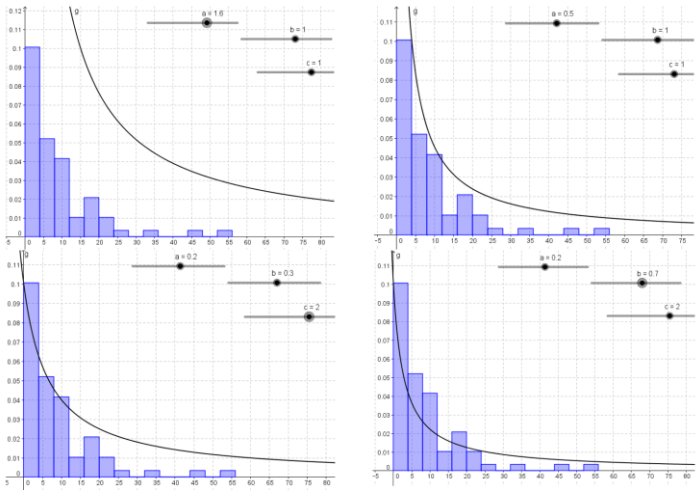
Annexe 6 : Évolution des courbes proposées lors de la phase 3.4 de la sous-tâche 3



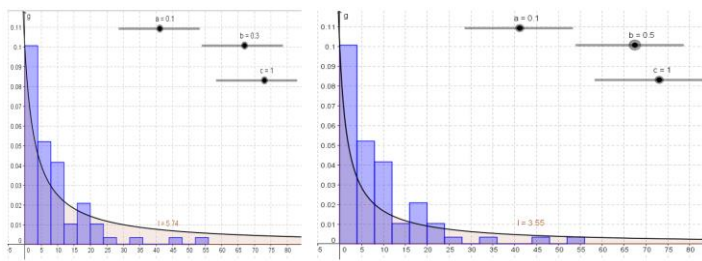
Phase 3.4.1



Phase 3.4.2



Phase 3.4.5



Phase 3.4.6