UNE SEQUENCE D'ENSEIGNEMENT ARTICULANT LES LOIS DE PROBABILITE A DENSITE ET LE CALCUL INTEGRAL EN TERMINALE S

Charlotte DEROUET,
ESPE de l'académie de Strasbourg,
LISEC équipe AP2E
Sylvie ALORY,
Irem de Paris

Ce texte est également consultable en ligne sur le portail des IREM (onglet : Repères IREM) : http://www.univ-irem.fr/)

Résumé : Cet article présente une séquence d'enseignement articulant l'enseignement des lois à densité et le calcul intégral, conçue puis expérimentée en classe de terminale S, dans le cadre d'un travail collaboratif entre une enseignante et une chercheure en didactique des mathématiques. L'objectif de la séquence est de motiver l'apprentissage du calcul intégral, par le biais de l'étude de problèmes probabilistes. Nous insistons particulièrement sur les problèmes de modélisation probabilistes introductifs de la séquence, dont le but est de faire construire aux élèves la notion de fonction de densité et de faire naître le besoin de l'outil *intégrale*.

1. — **Introduction**

Avant la classe de terminale, les élèves ne rencontrent dans leur scolarité que des lois de probabilités discrètes. Dans le passage aux lois continues, on peut observer une rupture avec les probabilités discrètes. En effet, dans le cas continu, l'ensemble des valeurs prises par la variable aléatoire $X(\Omega)$ n'est plus constitué d'éléments isolés mais est un intervalle I de l'ensemble des réels ; pour décrire la loi de probabilité, on associe une probabilité à chaque événement $\{X \in [a,b]\}$, pour tout intervalle $[a,b] \subset I$. Dans le cas particulier des *lois à densité*, on arrive à une égalité entre une probabilité et une intégrale :

$$P(a \leq X \leq b) = \int_a^b f(x)dx \qquad (*),$$

où f est la fonction de densité associée à la variable aléatoire X. Cette égalité met en relation deux domaines mathématiques, les probabilités d'un côté et l'analyse de l'autre ; et plus particulièrement deux thèmes dans chacun de ces domaines, les probabilités à densité et le calcul intégral. Mais finalement, d'où vient cette égalité et surtout comment est-elle introduite en terminale ?

Plus généralement, *est-il possible de concevoir et de mettre en œuvre des problèmes d'introduction aux lois à densité qui permettent aux élèves de terminale S de donner du sens à l'égalité* (*) ?

En regardant le programme de mathématiques de terminale S (MEN, 2011), probabilité et intégrale sont en interaction par l'intermédiaire d'une troisième notion, l'aire sous la courbe. En effet, l'intégrale d'une fonction continue et positive sur un intervalle $[a\ ;\ b]$ est définie comme l'aire sous la courbe de f sur $[a\ ;b]$. Quant au lien entre probabilité et aire sous la courbe, il apparaît ainsi dans le programme : « *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* ». L'aire de ce domaine correspond à l'aire sous la courbe d'une certaine fonction, qui est la fonction de densité. La notion de *fonction de densité* est finalement l'objet mathématique central pour permettre de connecter les lois à densité et le calcul intégral. Nous focaliserons donc cet article essentiellement sur la question de l'introduction de la notion de fonction de densité en terminale S.

Nous allons présenter le fruit d'un travail collaboratif entre une enseignante de lycée (et notamment de terminale S) et une chercheure en didactique des mathématiques [1], qui a abouti à une proposition de séquence d'enseignement dont un des objectifs principaux est de donner du sens à la notion de fonction de densité et donc à l'égalité (*). Nous commencerons par présenter rapidement des propositions faites par des manuels pour introduire la notion de fonction de densité, ce qui nous permettra de mettre en évidence des manques voire des erreurs, qui ne permettent pas aux élèves d'appréhender correctement la notion de fonction de densité. Ensuite, nous présenterons en détail les séances d'introduction de la séquence originale que nous avons conçue et expérimentée chaque année depuis 2015. Nous détaillerons les problèmes conçus et nous donnerons des éléments de leur déroulement en classe. Nous passerons un peu plus rapidement sur les séances suivantes.

Dans tout l'article, quand nous parlerons d'intégrale, nous serons toujours dans le cas d'*intégrale de fonction continue positive* sur un intervalle, non nécessairement borné.

2. — Une analyse de manuels de terminale S

Avant d'aborder les propositions faites dans les manuels, nous souhaitons préciser qu'aucune préconisation n'est donnée dans le programme à propos de l'introduction des lois à densité et plus particulièrement de la notion de fonction de densité, mis à part le cas particulier de la loi normale centrée réduite pour laquelle il est demandé de l'introduire par « *observation du théorème de Moivre-Laplace* » (MEN, 2011). Ici, nous nous intéressons à l'introduction de la fonction de densité de façon générale, donc nous ne pouvons pas nous arrêter sur cette proposition. Le document Ressources dédié aux probabilités et à la statistique en terminale (MENJVA & DGESCO, 2012), lui non plus, ne fait aucune proposition [2]. Nous pouvons alors nous demander quels sont les choix faits dans les manuels.

2. 1 *Un premier bilan sur les manuels*

Tout d'abord, sur les huit manuels de terminale S édition 2012, tous considèrent que le calcul intégral est un prérequis pour l'étude des lois à densité. Un seul manuel ne propose pas d'activité d'introduction de la notion de fonction de densité et l'impose directement dans la partie cours : « *Dans le cas d'une variable*

1 Cette séquence d'enseignement co-construite a fait l'objet d'analyses fines dans la thèse de doctorat de Derouet (2016).

2 Il faut chercher dans le document d'accompagnement du programme précédent pour avoir une proposition (MJENR, 2002). Nous n'en parlerons pas ici mais elle a contribué à notre réflexion.

aléatoire continue, on a recours à une fonction définie sur $\mathbf{R}$ *appelée* ***densité*** » (*Odysée*, p. 395), sans aucune tentative de justification ou d'explication. Le manuel *Déclic* propose une activité d'introduction (p. 364) qui résume la fonction de densité au bord supérieur d'une cible de fléchette ! Cette introduction est très contextualisée à une expérience aléatoire d'un lancer de fléchette sur une cible avec des conditions particulières (sans que cela soit mis en évidence) et la généralisation semble aller de soi. Le fait que la cible ait une aire de 1 et donc le choix d'une fonction dont l'aire sous la courbe vaut 1 n'est pas questionné. Le manuel *Symbole* (pp. 386-387) propose une activité prenant appui sur un polygone de fréquences cumulées croissantes ; il cherche donc à introduire dans un premier temps la fonction de répartition (qui n'est pas un attendu du programme), et seulement ensuite la fonction de densité. Enfin, les cinq autres manuels appuient leur introduction de la fonction de densité sur la notion d'histogramme de fréquences. En nous basant sur des analyses épistémologiques et historiques de l'émergence de la notion de fonction de densité (Derouet, 2017), l'appui sur l'histogramme nous semble porteur de sens pour introduire la notion de fonction de densité. Nous allons détailler ces activités.

2. 2 *Des tentatives d'introduction, mais peu abouties et avec des erreurs*

Sur les cinq manuels qui proposent une introduction avec un appui sur l'histogramme, le travail de l'élève se résume essentiellement à un travail de visualisation. Sauf une exception, l'histogramme est donné et la courbe de densité est déjà tracée. Il reste à la charge de l'élève de repérer une similitude (visuelle) entre l'histogramme et la courbe et ensuite de faire un lien entre aire des rectangles de l'histogramme et aire sous la courbe (qu'ils rapprochent rapidement au calcul intégral, puisque le chapitre a été traité peu de temps avant). Nous remarquons la quasi absence de questionnements du type : Comment a été choisie cette courbe ? Quelles propriétés doit-elle vérifier ? Ces questions, qui nous semblent importantes et primordiales pour véritablement donner du sens à la fonction de densité, ne sont souvent pas abordées ou le sont de manière erronée.

Donnons l'exemple du manuel *Indice* qui justifie le passage de l'histogramme à la courbe de densité en disant : « *Si l'on relève les distances à* 0,01 *kilomètre près, on voit apparaître une courbe comme celle tracée sur le graphique* [figure 1] » (p. 322). En d'autres termes, si l'on diminue l'amplitude des classes de l'histogramme, on « voit apparaître » une courbe comme celle tracée. Implicitement, les auteurs du manuel ont dû vouloir signifier « les hauts des rectangles de l'histogramme laissent

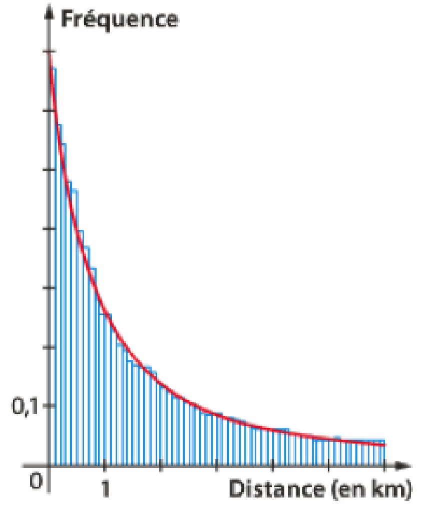


Figure 1. Exemple d'histogramme et courbe de densité proposés dans un manuel (Manuel *Indice*, Activité 1 p. 322)

deviner une courbe » (les rectangles seraient dans ce cas pratiquement des « segments accolés », du fait de la très petite amplitude). Cependant, même avec cette précision, cette remarque est fausse : si l'on diminue l'amplitude des classes, il apparaîtrait nécessairement des trous (des classes qui ne possèderaient pas d'éléments) et des pics. Si l'on prend l'exemple entre 0,4 et 0,5, le rectangle de l'histogramme est clairement au-dessus de la courbe rouge, il est donc impossible que les dix « petits » rectangles à intercaler dans cet intervalle aient tous une hauteur plus petite que celle du rectangle sur [0,4 ; 0,5], car l'aire totale doit rester la même.

Enfin, dans cet exemple, mais aussi dans les autres manuels, les connaissances des élèves sur l'histogramme, et notamment le fait que la fréquence d'une classe corresponde à l'aire du rectangle correspondant (et non à sa hauteur), ne sont pas considérées comme disponibles par les auteurs du manuel qui rappellent : « […] *on en déduit l'histogramme ci-contre, où chacun des* 60 *rectangles a pour base* 0,1 *et pour aire la fréquence de la classe correspondante* ». Pour autant, ces rappels écrits en langage courant n'empêchent pas les auteurs de proposer des graphiques erronés ; c'est le cas dans la figure 1. Les auteurs nomment l'axe des ordonnées « *fréquence* » alors que la fréquence n'est pas la hauteur des rectangles mais l'aire de ceux-ci. L'axe des ordonnées d'un histogramme de fréquences porte en fait la densité de fréquence (des compléments sur l'histogramme sont apportés dans Derouet (2018)). Bien entendu, ici, les classes étant d'amplitudes égales, fréquence et densité de fréquence sont proportionnelles mais tout de même ; cela pointe bien les difficultés de compréhension de la notion d'histogramme des auteurs de manuels et *a fortiori* des élèves (cf. Roditi, 2009) alors que c'est justement cette notion de densité de fréquences qui est importante pour introduire la notion de fonction de densité.

Pour finir, dans toutes les activités proposées, il n'y a pas de démarche de modélisation. D'une part, les problèmes sont souvent de « faux » problèmes (dans le sens où il s'agit de fausses contextualisations de données qui sont, la plupart du temps, des échantillons de simulations). D'autre part, le modèle est imposé car la courbe de densité est déjà tracée. De plus, les histogrammes donnés sont « trop parfaits » (ils ont été obtenus par simulation d'un échantillon de taille suffisamment grande, suivant la loi que les auteurs souhaitaient obtenir), ce qui empêche une véritable réflexion sur le choix de cette fonction de densité et donc sur les propriétés qu'elle doit vérifier. La notion de fonction de densité et plus généralement les lois à densité n'arrivent jamais comme une réponse à un problème, car finalement le problème est déjà résolu…

Les activités permettent de visualiser l'analogie entre histogramme et courbe de densité mais cet objet « fonction de densité » reste tout de même imposé aux élèves, qui n'ont pas véritablement construit ce nouveau savoir et notamment les caractéristiques de cette fonction (nécessité de l'aire totale sous la courbe égale à 1, positivité de la fonction) car ces propriétés n'émergent pas des activités proposées. De plus, comment introduire une notion (la fonction de densité) en s'appuyant sur une autre notion (l'histogramme) mal connue et mal maîtrisée par les élèves ?

2. 3 *Des points d'articulation possibles entre lois à densité et calcul intégral*

Pour mieux comprendre les liens qui peuvent être faits entre intégrale et probabilité, nous avons regardé les deux chapitres correspondants de façon plus globale. En analysant les manuels, dans la partie relative au calcul intégral, on peut repérer trois niveaux de calcul d'aires à mettre en lien avec le degré de complexité des fonctions en jeu :

— Le premier niveau : le calcul d'aires élémentaires. Il s'agit des calculs d'aires sous la courbe de fonctions affines par morceaux, qui reviennent à des calculs d'aires de rectangles, de triangles, de trapèzes.

— Le deuxième niveau : le calcul d'aires à l'aide de primitives. Il s'agit des calculs d'aires qui se font par utilisation du théorème fondamental de l'analyse :

$$\int_a^b f(x)dx = F(b) - F(a),$$

où F est une primitive de f. Ce type de calcul est possible lorsque l'expression explicite d'une primitive de f est connue des élèves. Ce niveau de calcul d'aire peut aussi être mobilisé pour des fonctions affines par morceaux bien entendu.

— Le troisième niveau : le calcul approché d'aires, essentiellement basé sur la méthode des rectangles qui est au programme de terminale S. Ce niveau concerne les fonctions dont les élèves ne connaissent pas d'expression explicite d'une primitive ou permet de retrouver un résultat obtenu par calcul avec une primitive.

Dans le déroulement du chapitre, quel que soit le manuel, on peut constater que, dans un premier temps, apparaissent des calculs d'aires de niveau 1, puis de niveau 3 pour enfin arriver au niveau 2, qui est le cœur du chapitre du calcul intégral. La présence de ces trois niveaux est pertinente dans notre cas, car nous pouvons remarquer que les calculs liés aux trois lois au programme (lois uniformes, exponentielles et normales) correspondent chacun prioritairement à l'un des niveaux de calculs d'aires (respectivement niveau 1, 2 et 3).

2. 4 *Bilan, suite à l'analyse de manuels*

A la suite de cette analyse de manuels mais aussi grâce à des analyses mathématique, historique (Derouet, 2017) et épistémologique, nous avons pu formuler des pistes à explorer pour concevoir une séquence et notamment des séances d'introduction à la notion de fonction de densité afin que les élèves construisent réellement ce nouvel objet mathématique et lui donnent du sens. Les points qui nous semblent importants sont les suivants :

— La fonction de densité doit être introduite comme réponse à un problème.

— Un appui sur la statistique descriptive (avec les histogrammes) et notamment sur des données statistiques réelles est porteur de sens.

— La notion d'histogramme doit être en amont retravaillée afin d'être disponible pour introduire la fonction de densité.

— Afin de relier plus solidement les probabilités à densité et le calcul intégral, il peut être intéressant de s'appuyer sur les trois niveaux de calculs d'aire.

3. — Une proposition de séquence

Nous avons ensuite travaillé de façon collaborative (une enseignante et une chercheure) avec pour objectif de proposer un (ou des) problème(s) qui invite(nt) les élèves à construire collectivement la notion de fonction de densité, pour leur permettre véritablement de donner du sens à ce concept mathématique. Ce travail collaboratif devait à la fois répondre à des critères didactiques mais aussi à des contraintes du terrain formulées par l'enseignante pour assurer sa viabilité dans les classes. Les contraintes posées par l'enseignante étaient de plusieurs ordres :

— Des contraintes de temps : la séquence et en particulier les séances d'introduction ne devaient pas avoir une durée supérieure à celle qu'elle consacrait pour la séquence les années précédentes (en particulier 2h pour l'introduction du calcul intégral et 2h pour l'introduction des lois à densité) ;

— Des contraintes de résultats : les élèves devaient être capables de résoudre au moins les mêmes tâches que les années précédentes ;

— Des contraintes matérielles : les séances d'introduction devaient se dérouler en classe entière, dans une salle de classe équipée d'un ordinateur pour l'enseignante et d'un vidéoprojecteur ; les élèves disposant de leur calculatrice.

Un autre point à prendre en compte était que la proposition faite ne devait pas trop s'éloigner des pratiques habituelles de l'enseignante pour que celle-ci puisse s'en emparer.

Précisons que l'enseignante travaille dans un lycée parisien du 16$^{\text{ème}}$ arrondissement, dans des classes de terminale S d'environ 35 élèves.

3. 1 *Point de départ et choix*

Une des originalités de cette séquence est d'articuler à la fois le chapitre sur les lois à densité (à l'exception du cas particulier des lois normales qui sont traitées plus tard dans l'année) et celui sur le calcul intégral (dans le cas des fonctions continues positives sur un intervalle) pour n'en faire qu'une seule et même séquence d'enseignement. Plus particulièrement, il s'agit de commencer par introduire la notion de fonction de densité en lui donnant du sens, pour ensuite créer le besoin de calculer des aires sous une courbe (pour déterminer des probabilités) et ainsi introduire l'intégrale comme réponse à ce problème. La suite de la séquence est assez « traditionnelle » dans le sens où elle diffère peu des propositions faites dans les manuels. Cela explique pourquoi ici nous nous attarderons plus particulièrement sur les séances qui ont pour double objectif de :

— Introduire la notion de fonction de densité

— Introduire le calcul d'aires sous une courbe et donc le calcul intégral.

3. 2 *Des anticipations avant la séquence*

Pour le bon déroulement de la séquence et plus précisément des séances d'introduction, il est primordial d'anticiper certains points. Avant le début à proprement parler de la séquence, des éléments indispensables sont à travailler en classe.

Un travail sur les fonctions du type $x \to e^{-kx}$ et $x \to e^{-x^2/2}$ doit être mené à l'occasion du chapitre sur la fonction exponentielle pour que les élèves les connaissent et qu'elles fassent partie de leurs fonctions de référence.

Ensuite, il est important d'anticiper le problème de l'aire finie d'un domaine infini. Il nous paraît essentiel que cette problématique soit abordée avec les élèves (et non esquivée comme c'est souvent le cas dans les manuels, par exemple). Cette problématique va évidemment susciter des discussions et surprendre les élèves, mais ce travail est indispensable pour éviter un éventuel blocage au début de la séquence qui nous intéresse. Ces trois dernières années, nous proposons un devoir maison sur la longueur finie d'une ligne brisée infinie, ce qui permet de travailler le même type de problématique. Dans ce devoir (en annexe 1), nous proposons d'étudier la longueur de deux spirales. Dans les deux cas, la longueur ajoutée à chaque étape de construction tend vers 0 mais dans le premier cas, la longueur de la spirale est finie, alors que dans le second, elle est infinie. Après une étape de conjecture, nous proposons aux élèves une démonstration guidée et une comparaison des deux phénomènes.

Pour finir, une nouvelle « première » rencontre avec la notion d'histogramme doit avoir lieu pour qu'elle puisse être un point d'appui à l'émergence de la notion de fonction de densité. Pour cela, nous proposons un devoir maison qui est corrigé en classe la séance précédant le début de la séquence.

Présentation du devoir maison préliminaire sur les histogrammes et la loi uniforme

Nous allons donner ici quelques éléments du devoir maison, présenté en annexe 2. Le premier exercice a pour objectif de réintroduire la notion d'histogramme. Nous nous appuyons sur des données statistiques sur la répartition des âges de la population parisienne en 2011 (cadre ci-contre).

Tout d'abord, il nous semble indispensable, avant de définir un histogramme, de définir ce qu'est la densité de fréquence. Une analogie qui nous semble adaptée pour les élèves et que nous reprenons de Bressoud & Kahané (2010) est celle avec la densité de population. Nous reprenons leur exemple :

> « *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 la densité de population, c'est-à-dire le nombre d'habitants au kilomètre carré. Avec une superficie de* 2,02 *km*^{2} *pour Monaco et de* 3 780 000 *km*^{2} *pour le Japon, les densités sont respectivement*

$$d_1 = \frac{32\,543}{2{,}02} = 16\,110{,}40\ h/km^2$$

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	2249974

Source : Insee, RP2011 exploitation principale.

Age	Fréquence
[0 ; 3[	0,03
[3 ; 6[	0,03
[6 ; 11[	0,05
[11 ; 18[	0,06
[18 ; 25[	0,11
[25 ; 40[	0,27
[40 ; 55[	0,19
[55 ; 65[	0,11
[65 ; 80[	0,10
[80 ; 105[	0,05
Ensemble	1

et $d_2 = \dfrac{127\,417\,244}{378\,000} = 337\ h/km^2$ *pour Monaco et pour le Japon. Autrement dit, alors que la population de Monaco est la moins importante en taille, sa densité est plus importante que celle du Japon* » (p. 11).

Ainsi, une représentation pertinente des populations de Monaco et du Japon doit rendre visible cette différence de densité. Il en est de même quand on cherche à représenter des fréquences de classes (notamment si elles sont d'amplitudes différentes) : il n'est pas pertinent, par exemple, de représenter de la même façon une fréquence égale à 0,05 de la classe [6 ; 11[et de la classe [80 ; 110[. Ces deux classes d'âge ont la même fréquence mais l'amplitude de la première classe est très inférieure à celle de la seconde, on pressent donc que la densité de la première classe est supérieure à celle de la seconde. Par analogie avec la densité de population utilisée en géographie, on calculera le quotient de la fréquence d'une classe d'âge par l'amplitude de cette classe, et on parlera de densité de fréquence. 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, qui n'est autre que l'histogramme, doit rendre visible ce phénomène. Nous proposons les définitions suivantes :

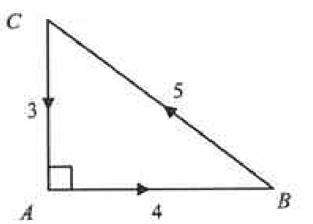
— Un histogramme de fréquences est une représentation graphique composée de rectangles « collés » dont les aires sont proportionnelles aux fréquences et dont les bases sont déterminées par les intervalles des classes.

— Dans le cas d'une variable quantitative continue [3], on définit la *densité de fréquence* d_i d'une classe de fréquence f_i et d'amplitude a_i par : $d_i = f_i / a_i$.

Nous trouvons donc que l'axe des ordonnées de l'histogramme représente la densité de fréquence (à constante multiplicative près), et non la fréquence comme on peut le voir très souvent dans les manuels. Bien entendu, fréquence et densité sont égales (ou tout du moins proportionnelles) lorsque les amplitudes des classes sont égales, mais l'histogramme n'a pas vocation à avoir toujours des classes d'amplitudes égales. Il pourrait être envisagé la définition (plus rigoureuse mathématiquement) de l'histogramme suivante : un histogramme est la représentation graphique de la fonction d qui, à tout nombre réel x [4], associe la densité de fréquence $d(x)$ de la classe à laquelle x appartient. La fonction d s'appelle la fonction de densité de fréquence de la série statistique étudiée [5].

Le deuxième exercice du devoir maison (Annexe 2) est une première rencontre avec une loi uniforme continue. Elle ne sera cependant pas institutionnalisée à ce moment-là. L'exercice est une adaptation d'une situation proposée par un groupe de l'Irem d'Aquitaine (Barbazo *et al.*, 2003) :

Un point mobile M se déplace à vitesse constante sur le « pourtour » d'un triangle, toujours dans le même sens (indiqué par la flèche). On suppose qu'il démarre au point C. Une panne se produit subitement et aléatoirement et le point mobile s'arrête instantanément.



3 Ou dans le cas d'un caractère quantitatif discret qui prend un grand nombre de valeurs et qu'il est alors plus intéressant de traiter les valeurs en classes comme pour un caractère quantitatif continu. C'est le cas dans l'exemple du devoir maison si l'on considère que l'âge de l'individu comptabilisé est celui à la date d'anniversaire de l'individu en 2011 (donc un nombre entier).

4 Appartenant aux classes concernées par le phénomène étudié.

5 Cette formulation n'a pas encore été testée en classe.

Le problème a ici été pensé en plusieurs étapes : une première phase de conjectures, où les élèves peuvent répondre de façon intuitive ; 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 » qui a été déduite des quatre histogrammes fournis. Cette courbe étant celle qui « lisse » les histogrammes (*la même* pour les quatre histogrammes). Introduire la notion de fonction de densité dans le cas de la loi uniforme continue n'est, selon nous, pas le plus pertinent car comme le montre la première partie de cet exercice, les calculs peuvent être faits par les élèves de façon intuitive et donc le besoin d'une courbe de densité n'est pas réel. Cependant, cet exercice permet aux élèves de rencontrer une certaine démarche de modélisation et de concevoir une « courbe de tendance » qui lisse les histogrammes, et de rencontrer la problématique de $P(X) = a$. Cela sera utile pour les premières séances de la séquence.

Retours sur le devoir maison

La correction de ce devoir est une étape importante. Le premier exercice ne pose pas de problème aux élèves. La correction de l'exercice 2 permet de revenir sur le fait qu'il est nécessaire de définir l'expérience aléatoire et la variable aléatoire définie sur cette expérience (X est la variable aléatoire correspondant à la distance parcourue par le mobile depuis son dernier passage au point C, au moment de la panne). En effet, aucun élève ne prend la peine de le faire. Nous faisons alors remarquer explicitement la nature différente de cette variable aléatoire : elle peut prendre *toutes* les valeurs de l'intervalle [0 ; 12]. On précise alors le vocabulaire : X est une *variable aléatoire continue*. Le calcul de la probabilité p_1 que le point mobile s'arrête sur l'hypoténuse est intuitif : la probabilité que la variable aléatoire X appartienne à un intervalle est proportionnelle à sa longueur de celui-ci car le point mobile se déplace à vitesse constante ; le modèle uniforme est sous-jacent.

Comme le montrent les extraits de copies (figures 2 et 3), les élèves rencontrent des difficultés avec l'évaluation de la probabilité p_2 que le point mobile (la puce électronique ici) s'arrête sur le sommet de l'angle droit. Il est inconcevable pour la majorité d'entre eux de dire que $p_2 = 0$. Malgré les différentes phases de l'exercice, peu d'élèves changent d'avis entre conjecture, simulation et calcul théorique.

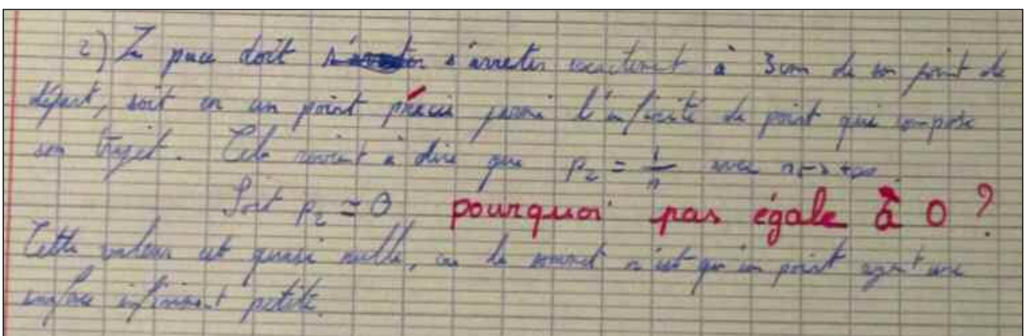


Figure 2. Réponse d'un élève à la question A2 de l'exercice 2 (Annexe 2)

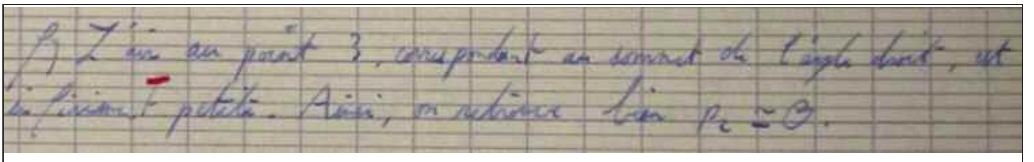


Figure 3. Réponse du même élève à la question C5 de l'exercice 2 (Annexe 2)

Le fait qu'un événement possible ait une probabilité nulle ne va pas de soi et constitue une rupture avec ce qu'ils ont rencontré auparavant. Il nous est donc apparu important de ne pas négliger cette difficulté. C'est l'occasion d'un débat dans la classe entre les tenants de $p_2 = 0$ et ceux de $p_2 \neq 0$, débat amenant systématiquement le besoin d'une preuve pour convaincre. Le débat est vif et même une fois une preuve produite, un petit nombre d'élèves demeure dubitatif. La correction permet de dégager différents points :

— Rencontre d'une variable aléatoire de nature différente de celles connues auparavant ;

— Nécessité de calculer la probabilité d'événements de la forme $\{X \in [a, b]\}$. Ce calcul est lié à un calcul d'aire ;

— $P(X = a) = 0$ alors que l'événement $\{X = a\}$ n'est pas impossible.

3. 3 *Le problème de la rencontre*

Nous allons maintenant présenter en détail les deux problèmes d'introduction de la séquence de terminale S articulant les lois à densité et le calcul intégral. Ces séances arrivent juste après la correction en classe du devoir maison préliminaire. Le premier problème, que nous appelons le problème de la rencontre, est une adaptation d'un exercice d'application proposé dans le document Ressources Probabilités et statistiques de la classe de terminale (MENJVA, 2012). L'énoncé est cependant beaucoup plus ouvert que dans l'exercice original.

Présentation du problème

L'énoncé du 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 ?

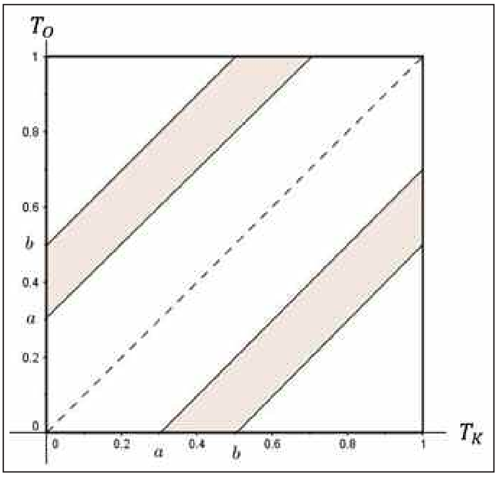
L'objectif de ce problème est de faire émerger la notion de fonction de densité et ses caractéristiques. La fonction de densité en question est une fonction affine ($f(x) = -2x + 2$ sur $[0 ; 1]$), il s'agira donc de calculs de probabilités correspondant à des calculs d'aires de niveau 1 (aires élémentaires de triangles ou de trapèzes).

La résolution de ce problème se fait par une réflexion collective des élèves. Il n'est pas prévu que les élèves s'emparent du problème de façon individuelle, c'est au contraire le collectif qui va mettre en place une démarche de modélisation pour arriver au choix d'une courbe de densité (que les élèves appelleront dans un premier temps « courbe de tendance »).

Une démarche de modélisation étant attendue, les élèves doivent se mettre d'accord pour faire des choix. Bien entendu, plusieurs modélisations sont possibles avec ce problème et plusieurs démarches aussi. Cependant, pour l'objectif d'enseignement que nous avons, nous espérons que les élèves partent dans une certaine direction, que nous détaillons ensuite.

Un des premiers choix à faire est de savoir si l'on considère le temps comme un phénomène discret ou continu. Ici, le problème amène assez naturellement à considérer le temps comme un phénomène continu. De plus, de manière assez immédiate les élèves considèrent que Karine et Olivier arrivent de façon indépendante et peuvent arriver de façon uniforme sur l'intervalle de temps considéré. Cela permet donc de considérer deux variables aléatoires T_K et T_O, représentant respectivement les moments d'arrivée de Karine et Olivier, et suivant toutes les deux une loi uniforme sur [7 ; 8] (cela revient au même de les considérer sur [0 ; 1]). Le temps d'attente du premier arrivé est alors la variable aléatoire $T_K - T_O$ à valeurs dans [0 ; 1] (on pourrait aussi faire le choix d'exprimer les temps en minutes et donc de considérer l'intervalle [0 ; 60]). On cherche alors la loi que suit cette variable aléatoire.

Dans le document Ressources, c'est une résolution géométrique qui est proposée. Ici, cela donnerait :



Cette solution permet de définir, pour tous a et b appartenant à [0 ; 1] tels que $a \leq b$, la probabilité $P(a \leq X \leq b)$ comme étant l'aire des parties colorées sur la figure. En effet, les parties colorées correspondent à l'ensemble des points de $[0 ; 1]^2$ tels que $a \leq |T_K - T_O| \leq b$. Cette démarche, correcte bien entendu, ne permet cependant pas d'arriver à la notion de fonction de densité, ce qui est l'objectif ici. Au cours des quatre années où ce problème a été expérimenté, elle n'a jamais été proposée [6] par les élèves.

La démarche espérée est que les élèves proposent de faire des simulations, sur tableur par exemple, pour avoir un (ou des) échantillon(s) de taille suffisamment grande (10 000 par exemple). La formule pour simuler l'expérience aléatoire à entrer est :

=ABS(ALEA()-ALEA())

sur tableur, ou

=Abs(Random()-Random())

sur le tableur du logiciel *GeoGebra* [7]. Inspirés par la démarche de l'exercice 2 du devoir maison, les élèves peuvent ensuite penser à représenter les données sous forme d'histogramme, avec des amplitudes plus ou moins grandes (fig. 4, page suivante). Contrairement au tableur, le logiciel *GeoGebra* permet de construire des histogrammes facilement et de façon assez transparente pour les élèves. De plus, dans un second temps, il permet de tester facilement les fonctions de densité. Nous présentons les étapes de construction d'un histogramme avec *GeoGebra* en annexe 3. Nous déplorons cependant le fait de ne pouvoir construire que des histogrammes avec des classes d'amplitudes constantes.

6 Sauf par un élève qui avait volontairement été isolé du reste de sa classe, pour ne pas influencer les autres.

7 Il est tout à fait pertinent d'utiliser le tableur de *GeoGebra* pour la suite ; cependant, il est vrai que le logiciel bugue assez souvent si l'on souhaite réaliser un échantillon de grande taille.

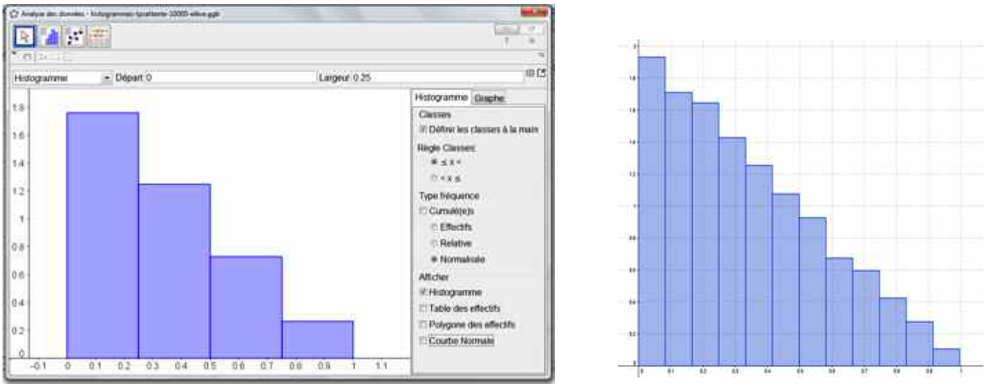


Fig. 4. Histogrammes avec des classes d'amplitudes 15 minutes (à gauche) et 5 minutes (à droite)

Une fois les histogrammes construits, les élèves doivent déterminer une « courbe de tendance » (comme dans le devoir maison). La contrainte de proximité avec le haut des rectangles ne suffit pas, car l'aire sous cette courbe va permettre de calculer des probabilités. Il faut pour cela que les élèves prennent conscience du fait que l'on a besoin de contraintes supplémentaires pour déterminer l'expression de la fonction affine décroissante qu'ils cherchent, notamment que l'aire sous la courbe sur [0 ; 1] doit valoir 1. De plus, au regard de l'histogramme, on peut imposer $f(1) = 0$. Ce choix est raisonnable car on observe que si l'on réduit de plus en plus l'amplitude des classes de l'histogramme, la densité de fréquence sur une classe de la forme $[1 - \varepsilon ; 1]$ (avec ε très petit) est presque nulle [8]. On arrive alors à la fonction $f(x) = -2x + 2$ sur [0 ; 1]. La connaissance de cette fonction permet de connaître la probabilité que le temps d'attente soit dans n'importe quel intervalle de [0 ; 1], ce qui répond au problème posé.

Nous poursuivons la séance par l'exercice d'application donné page ci-contre, qui permet aux élèves de prendre conscience que maintenant à l'aide de calculs d'aire sous la courbe qu'ils ont trouvée (aires de triangles et de trapèzes), ils peuvent déterminer toutes les probabilités qu'ils veulent. Cet exercice permet aussi de travailler sur l'écriture symbolique probabiliste d'événements.

Éléments du déroulement en classe

Cette recherche se déroule sur une séance de 2 heures. Tout le travail sur le logiciel *GeoGebra* se fait en classe entière avec un seul ordinateur vidéoprojeté qui est manipulé par l'enseignante ou par les élèves. Les élèves terminent l'exercice d'application à la maison.

Dans un premier temps, les élèves ont une phase de recherche individuelle. Le fait que la question soit très ouverte peut en bloquer quelques-uns, mais ils finissent tous par avoir quelques remarques à dire. Une première mise en commun permet d'engager des discussions dans la classe sur plusieurs points, que nous allons illustrer. Dans cette phase, l'ensei-

8 Par le calcul théorique, on trouve effectivement que la fonction de densité associée à la variable aléatoire $|X - Y|$, avec X et Y deux variables aléatoires suivant une loi uniforme sur [0 ; 1], est $x \to -2x + 2$, sur [0 ; 1].

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

a) Quelle est la probabilité que le premier arrivé attende plus de 40 minutes ?
b) Quelle est la probabilité que le premier arrivé attende moins de 40 minutes ?
c) Quelle est la probabilité que le premier arrivé attende entre 20 et 40 minutes ?
d) Quelle est la probabilité que Karine et Olivier arrivent en même temps ?
e) Quelle est la probabilité que le premier arrivé attende exactement 40 minutes ?
f) Finalement, ils ont décidé de ne pas attendre plus d'un quart d'heure. Quelle est la probabilité qu'ils se retrouvent ?
g) 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 ?

gnante écrit au tableau toutes les idées des élèves, qui sont invités à prendre des notes tout au long de la séance.

Au cours de la première année d'expérimentation (en 2015), toutes les séances ont été enregistrées, observées par la chercheure puis transcrites. Nous illustrons parfois nos propos par des extraits de *verbatim*. Les élèves discutent de la vraisemblance ou non du problème :

> Élève 1 : Je disais que c'est un peu bizarre comme rendez-vous… […] On se donne pas rendez-vous entre 7h et 8h…

L'enseignante peut au bout d'un moment recentrer le travail :

> P : Alors bon, peut-être, se donner rendez-vous entre 7h et 8h, c'est une plage horaire relativement grande… mais aussi pour que ce soit plus facilement traitable pour nous. On aurait pu se dire « ils se donnent rendez-vous entre 7h et 7h15 », est-ce que ça changeait ce qu'on va… est-ce que ça changeait le problème ?

Les élèves arrivent au caractère aléatoire de la situation et se questionnent sur le caractère discret ou continu du temps :

> Élève 2 : Si on considère qu'il y a autant de chance qu'ils arrivent à… chaque minute… entre 7h et 8h… donc ils ont une chance… en fait… Karine a 1 chance sur 60 d'arriver à… heure… donc ils ont un soixante… un sur soixante au carré…
> […]
> Élève 3 : Moi je dirais que c'est… on parle pas en minute, là on parle à tout moment donc… je dirais que même les secondes, il y a déjà beaucoup plus de… Par exemple, si on arrive à 7h28min37s, c'est pas pareil. En plus, il y a les millièmes de seconde et…
> Plusieurs élèves : Et les milliards de seconde.
> Élève 3 : Et les milliardièmes de seconde après.
> Élève 4 : C'est comme la puce hier [l'élève fait référence à l'exercice 2 du devoir maison].
> Élève 3 : Oui, c'est infini en fait.
> P : Ah, c'est comme la puce.
> Élève 3 : Oui, c'est comme la puce.
> Élève 5 : Sauf que…
> Élève 1 : Sauf que c'est des minutes…
> Élève 3 : Oui, mais c'est comme la puce.
> Élève 4 : Quand t'arrives… quand t'arrives à 7h31 et que l'autre arrive à 7h32…
> Élève 4 : Non à 7h31 virgule zéro cinq et

7h virgule trente-deux, t'es à la même heure.

Élève 1 : Oui, mais on est en maths.

P : Ah ! « Mais on est en maths ».

[…]

Élève 4 : C'est… Il y a une possibilité infinie de… d'arriver… enfin… […] Entre 8… entre 7h et 8h, on a une possibilité infinie d'arriver à une certaine heure.

P : Une infinité de moments tu veux dire ? d'instants ?

Les quatre années d'expérimentation, la modélisation par une variable aléatoire continue a toujours fini par être retenue.

Des considérations qualitatives peuvent alors suivre, comme par exemple : le temps d'attente est entre 0 et 1 heure, il y a plus de « chance » que le premier arrivé attende 5 minutes que 55 minutes… Progressivement des variables aléatoires T_O et T_K représentant respectivement les heures d'arrivée d'Olivier et Karine sont introduites par les élèves et la variable aléatoire représentant le temps d'attente X est définie par $X = |T_K - T_O|$. Une fois les remarques qualitatives épuisées, il faut souvent relancer la recherche. La plupart du temps, les élèves proposent de prendre un exemple : calculer la probabilité que le temps d'attente du premier arrivé soit inférieur à 10 minutes. Ayant rencontré, dans le devoir maison, des simulations ensuite mise sous forme d'histogramme, les élèves proposent assez naturellement (bien que cela puisse mettre du temps) de simuler des réalisations de la variable aléatoire X et ensuite de les représenter sous forme d'histogramme. Un temps est laissé aux élèves pour chercher la formule à entrer dans le tableur pour simuler des réalisations de la variable aléatoire X. L'utilisation du logiciel *GeoGebra* pour construire les histogrammes est amenée par l'enseignante. Suivant les classes, l'idée de chercher la « courbe de tendance » (vocabulaire employé dans le devoir maison) est plus ou moins immédiate :

P : Si je vous demandais de… avec ça, de… d'évaluer le… la probabilité que le temps d'attente soit entre 0,2 et 0,3. D'accord ? Imaginons. Comment vous feriez ?

P : Est-ce que c'est bien ça qu'on essaie d'évaluer ? La probabilité que le temps d'attente soit entre tant et tant ? Donc si je vous demandais ça… Si je vous posais cette question « évaluer le temps d'attente entre 0,2 et 0,3 », comment vous feriez avec cet… à l'aide de cet histogramme ?

[…]

Élève 5 : Entre 0,2 et 0,3 ?

P : Oui. Juste, comment tu ferais, pas ce que tu trouverais. Mais comment tu ferais ?

Élève 5 : On peut faire comme dans le DM, 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.

Ici, on remarque que l'élève a bien identifié les différentes étapes de la démarche (bien que cela ne soit pas nécessairement exprimé de la « bonne façon »). Les élèves sont tous d'accord qu'il s'agit d'une droite (plus exactement d'un segment). Un temps individuel est ensuite laissé aux élèves pour déterminer cette (ces ?) courbe(s) de tendance. Dans un premier temps, les élèves proposent des courbes en reliant « le haut des rectangles aux extrémités » ou en reliant le haut du rectangle sur l'axe des ordonnées et le point extrême d'abscisse 1 ou encore un segment « à vue d'œil » coupant les rectangles de façon équilibrée pour qu'il y ait « autant de vide en-dessous que de bout de rectangles au-dessus » (figure 5).

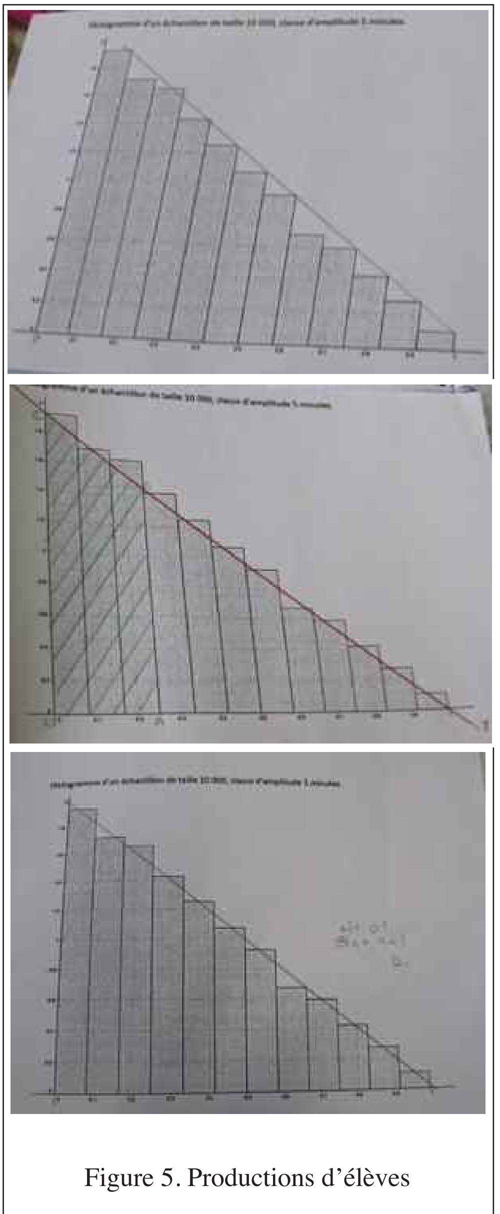


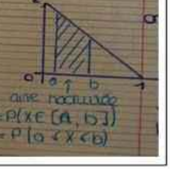
Figure 5. Productions d'élèves

Une mise en commun est nécessaire pour amener les élèves à prendre conscience des contraintes que doit vérifier cette fameuse courbe de tendance. Il émerge rapidement que l'aire sous la courbe doit être égal à 1. Ensuite, les élèves proposent parfois $f(0) = 2$, mais cette contrainte est moins « justifiable » que $f(1) = 0$ (cf. présentation du problème). On insiste alors avec les élèves sur le fait que nous faisons un choix de modèle.

Les élèves ont ensuite une synthèse de la séance à faire chez eux, pour faire ressortir les différentes étapes de la démarche qu'ils ont faite en classe. L'enseignante propose ensuite un résumé de leurs synthèses comme celle de l'annexe 4 réalisée en 2016. Au cours de la première année d'expérimentation (2015), la synthèse avait été faite en classe et avait abouti à l'institutionnalisation ci-dessous :

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.



Cela n'est pas visible dans l'écrit mais il est intéressant d'avoir conscience que toutes les propriétés de la fonction de densité qui apparaissent dans cette synthèse ont été justifiées par les élèves lors de la séance. L'enseignante introduit elle-même le vocabulaire « fonction de densité de probabilité », cependant elle demande ensuite aux élèves s'ils peuvent justifier ce choix de vocabulaire. Voici la réponse d'un élève :

> E1 : C'est parce que sur l'axe des ordonnées, c'est la densité.

Les autres élèves acquiescent en se remémorant le devoir maison sur l'histogramme, où il est fortement question du fait que l'axe des ordonnées pour un histogramme représente la densité de fréquence. Pour justifier la positivité de la fonction de densité puis le fait que l'aire sous la courbe est égale à 1, on retrouve à nouveau une justification prenant appui sur l'histogramme :

> E2 : Parce qu'en fait c'est par rapport à l'histogramme donc on peut pas avoir de rectangles en-dessous.
>
> E3 : Parce que ça correspond à la fréquence. L'aire correspond... [...] L'aire correspond à la somme de toutes les probabilités... [...] La somme des fréquences.
>
> P : Et donc on sait que ça fait...
>
> E3 : Un.

La construction de la nouvelle notion fonction de densité est ici possible grâce à la disponibilité des connaissances sur l'histogramme. Notamment, le fait que l'axe des ordonnées de l'histogramme soit véritablement défini (dans le devoir maison) et ait un sens pour les élèves permet de légitimer le vocabulaire « fonction de densité de probabilité », ce qui n'est jamais fait dans les manuels.

3. 4 *Le problème du volcan Aso*

Le second problème a lieu lors de la séance qui suit le problème de la rencontre. Ce problème s'appuie cette fois-ci sur des données réelles, que l'on trouve dans le manuel *Math'x* (2012). Elles ont cependant été complétées [9].

Présentation du problème

L'énoncé distribué aux élèves se trouve en annexe 5. Il s'agit notamment de l'ensemble des années des éruptions volcaniques du volcan Aso (Japon) depuis le XIIIème siècle et des temps d'attente entre deux éruptions consécutives.

Les questions posées aux élèves (qui évoluent chaque année) sont, pour l'année 2017, les suivantes :

> La dernière éruption du volcan Aso a eu lieu au cours du mois d'octobre 2016.
>
> a) Comment évaluer la probabilité que la prochaine éruption ait lieu avant octobre 2021 ?
>
> b) Comment évaluer la probabilité que la prochaine éruption ait lieu au cours de l'année 2031 ?

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. Dans ce problème, il s'agit, à partir de données réelles sur le passé, de « prédire » l'avenir. C'est une problématique de statistique inférentielle. Pour la première question, une première évaluation peut se faire à l'aide de la fréquence qui découle des données fournies. Pour répondre à la deuxième question, il est attendu des élèves qu'ils mettent en place une démarche similaire à celle du problème 1. Ils doivent rassembler les données sous forme d'un histogramme, puis chercher une fonction

9 Le site http://volcano.si.edu/volcano.cfm?vn=282110 recense l'ensemble des dates d'éruptions de ce volcan.

de densité. Tout l'enjeu de ce problème est la détermination de cette fonction de densité, qui cette fois-ci, n'est plus une fonction affine. On espère qu'après débat dans la classe, les élèves finissent par modéliser la situation par une loi exponentielle (sans la nommer ainsi bien sûr), mais d'autres lois pourraient être choisies. Cela amènera ensuite la question : comment calculer l'aire sous une telle courbe ? Dans ce cas, nous arrivons à un calcul d'aire de niveau 3 [10].

Plusieurs modèles du temps d'attente entre deux éruptions pourraient être envisageables. Il faut en tout cas garder en tête que le choix de modèle qui sera fait n'est autre qu'une conjecture de modèle. Il faudrait ensuite confronter cette proposition à la réalité pour la légitimer ou non, ce qui ne nous intéresse pas ici. D'ailleurs, le modèle exponentiel n'est en fait pas réellement adapté à une situation de temps d'attente dans des phénomènes naturels qui, d'après les spécialistes du domaine, ne sont pas des phénomènes sans mémoire.

Comme pour le problème de la rencontre, une démarche de modélisation est nécessaire. Ici, nous disposons de données réelles et nous pouvons remarquer que des choix ont déjà été faits, par exemple le fait que seules les années d'éruption soient précisées et donc que les temps d'attente entre deux éruptions consécutives sont donnés en années entières. Il faut donc choisir comment considérer la donnée « k années » ? Est-elle à voir comme une donnée isolée ou comme représentant un intervalle ? Lequel ? Il est assez naturel de considérer le temps d'attente entre deux éruptions comme continu. Ce qui est sûr, c'est qu'en réalité une donnée k peut représenter n'importe quelle valeur de l'intervalle $]k-1\ ;\ k+1[$, cependant cela entraîne des chevauchements. Il faut donc faire un choix arbitraire qui ne crée pas de chevauchements. Dans l'histogramme présenté, nous avons fait le choix de considérer une donnée k comme appartenant à l'intervalle $[k\ ;\ k+1[$ (étant donné que certaines valeurs valent 0). Ce questionnement n'est pas forcément abordé dans la classe, mais il est utile que l'enseignant ait conscience de cela.

Ce problème faisant suite à celui de la rencontre, il est attendu que les élèves se l'approprient facilement et proposent rapidement les différentes étapes de la démarche : représenter les données statistiques sous forme d'histogramme pour ensuite déterminer une « courbe de tendance » (courbe de densité) qui permettra de faire les calculs de probabilités attendues. Tout d'abord, il est nécessaire de laisser un temps individuel ou en binôme pour que les élèves commencent à manipuler les données. Il peut être intéressant de les laisser tracer eux-mêmes leur histogramme à la main. Ensuite, un histogramme doit être fait sur le logiciel *GeoGebra*. En effet, cela permet ensuite de tester rapidement les fonctions de densité candidates pour vérifier la cohérence visuelle avec l'histogramme, de vérifier le fait que la fonction est bien positive et de plus, la commande « Inspecteur de fonction » du logiciel permet de connaître l'aire sous la courbe entre deux bornes données. Les élèves, ne connaissant pas encore le calcul intégral, ont ce moyen pour obtenir l'aire sous la courbe. Cet histogramme sur *GeoGebra* (figure 6 de la page suivante) est donc le point d'appui pour la discussion sur le choix de la fonction de densité.

Nous avons déjà souligné que les histogrammes sous *GeoGebra* sont à pas constants, ce qui n'est pas le plus approprié ici. Le fait qu'il y ait des « trous » dans l'histogramme peut gêner certains élèves dans la recherche d'une fonction, qui soit continue. Toutefois, les élèves ayant bien assimilé la notion d'histo-

10 A ce stade, les élèves ne connaissent pas encore le calcul intégral.

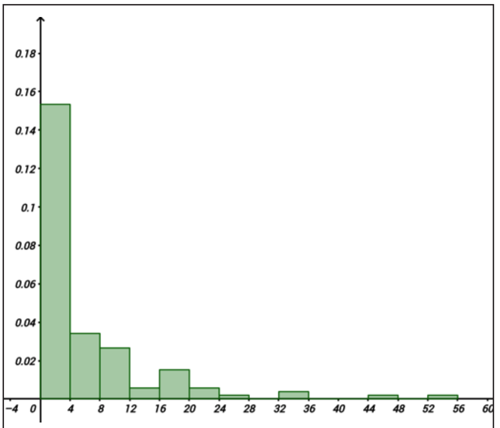


Figure 6. Histogramme des temps d'attente entre deux éruptions consécutives, avec des classes d'amplitude 4 ans

gramme grâce au devoir maison sont capables de passer au-delà :

> Elève : Je pense qu'il faudrait faire un histogramme avec des classes d'amplitudes différentes pour que ça puisse prendre toutes les valeurs… Enfin pour pas qu'il y ait de trous.

Une fois l'histogramme construit, les élèves doivent se mettre d'accord sur l'ensemble de définition de la fonction de densité qu'ils cherchent. Il est attendu qu'ils choisissent $[0\,;+\infty[$. Avec les données dont ils disposent, les temps d'attente ne dépassent pas 56 ans, or il n'est bien entendu pas raisonnable de penser que ce temps d'attente ne peut pas être supérieur. Ce choix d'un intervalle non borné amènera la question de la possibilité d'avoir une aire finie d'un domaine infini, d'où l'importance que cette problématique ait été anticipée au préalable dans l'année lors du devoir maison consacré à cela.

Nous attendons des élèves qu'ils proposent dans un premier temps une fonction de type fonction inverse (fonction de référence bien connue des élèves) de la forme : $\frac{a}{bx+c}$, qui peut être implémentée dans le logiciel *GeoGebra* avec des curseurs. Ces fonctions seront ensuite réfutées car elles ne peuvent pas vérifier toutes les conditions que l'on désire : l'aire sous la courbe sur $[0\,;+\infty[$ est supérieure à 1 (elle est infinie). Nous y reviendrons dans la présentation du déroulement. Cela permet aux élèves de prendre conscience de l'importance des caractéristiques des fonctions de densité, empêchant certaines fonctions d'être retenues pour ce rôle. Après avoir disqualifié ces fonctions, nous attendons que les élèves proposent une fonction du type exponentielle décroissante [11] (qui doit donc faire partie des fonctions de référence des élèves) de la forme : $\alpha\, e^{-\lambda x}$. Avec un jeu sur les paramètres à l'aide des curseurs de *GeoGebra*, les élèves peuvent se mettre d'accord sur l'expression d'une fonction. Il faut noter que plusieurs choix de la forme : $f(x) = \lambda\, e^{-\lambda x}$ peuvent être proposés, avec λ suffisamment proche de 0,2 (figure 7, ci-contre).

Il est possible que les élèves proposent des fonctions de type $x \rightarrow \frac{a}{bx^2+c}$; qui pourrait d'ailleurs convenir mais l'extrême difficulté à trouver les paramètres adéquats par tâtonnement pousse les élèves à abandonner cette possibilité. L'« Inspecteur de fonction » du logiciel *GegGebra* permet ensuite aux élèves de répondre aux deux questions du problème.

Après ce travail, la question devient : comment calculer nous-mêmes l'aire sous

11 Les fonctions exponentielles décroissantes ayant été travaillées à plusieurs reprises en classe de terminale, dans cette séquence il est attendu qu'elles fassent partie des fonctions de référence pour les élèves.

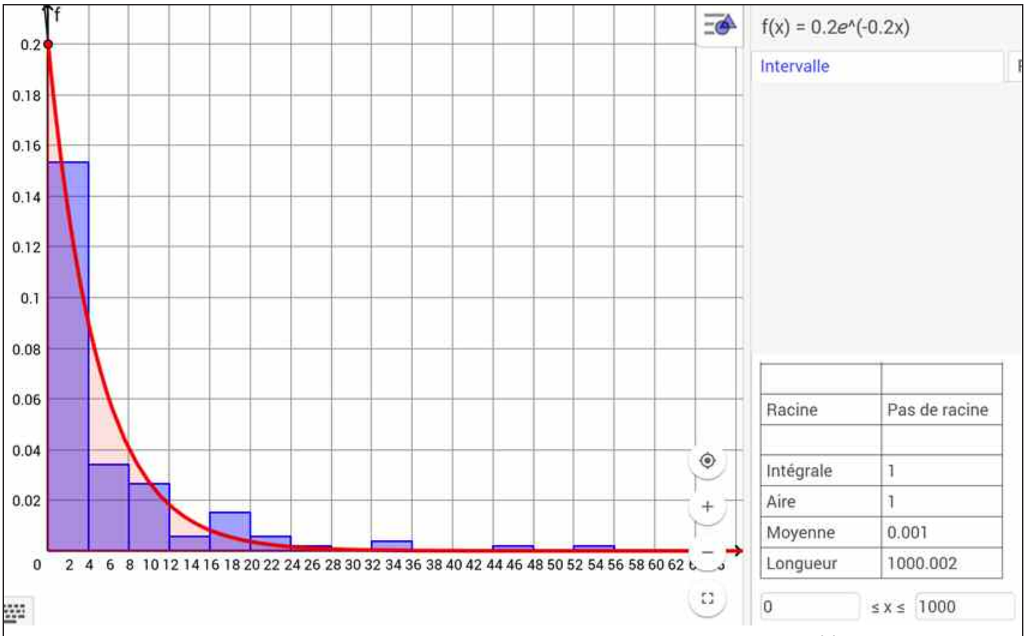


Figure 7. Histogramme et courbe de la fonction $x \to 0{,}2\, e^{-0{,}2x}$

une telle courbe (non affine) ? Il est demandé aux élèves d'évaluer par eux-mêmes (sans le logiciel) la valeur de la probabilité que le temps d'attente entre deux éruptions consécutives soit inférieur à 20 ans. Il s'agit ici d'introduire le calcul approché d'aire par une méthode choisie par les élèves. Il s'agit d'une activité classique d'introduction du calcul intégral, cependant ici ce sont les probabilités qui motivent ce travail.

Éléments du déroulement en classe

Ce problème se déroule sur une séance de 2h ou 2h30, suivant le temps laissé aux élèves dans les phases de travail individuel ou en binôme. Le travail se fait de nouveau en classe entière avec un seul ordinateur vidéoprojeté, pris en charge par l'enseignante ou un élève.

Nous reprenons ici le cheminement des élèves lors de l'expérimentation de 2015 [12], lors de la recherche d'une courbe de densité (cf. figure 8 de la page suivante). Une fois l'histogramme tracé sur le logiciel *GeoGebra*, les élèves doivent résoudre la tâche suivante : « *On cherche une fonction f définie sur* $[0 ; +\infty[$ *qui approche « au mieux » l'histogramme* ». La plupart des élèves proposent la fonction $x \to \frac{1}{x}$.
L'enseignante décide donc de faire une mise en commun en partant de cette fonction (phase A). Cette courbe ne convenant pas visiblement aux

12 En 2015, nous ne disposions pas des données des éruptions jusqu'à nos jours, ce qui explique que l'histogramme n'est pas exactement le même et donc que la fonction de densité choisie non plus. Cependant la démarche reste identique.

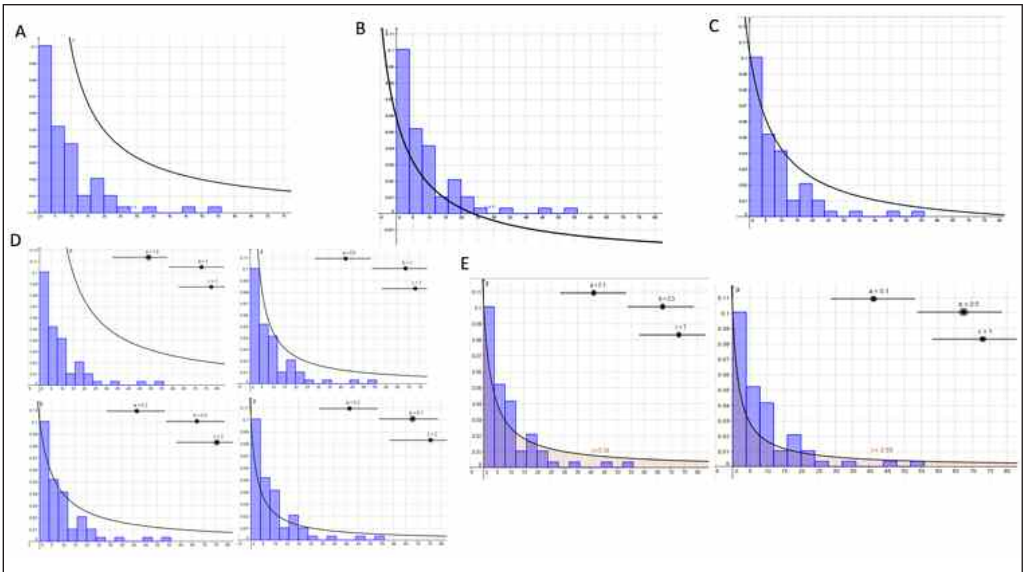


Figure 8. Evolution du travail de recherche de la courbe de tendance

élèves, ils décident de « la descendre ». Cependant, la fonction devient alors négative à partir d'un moment (phase B). Les élèves doivent mobiliser leurs connaissances sur la fonction de densité (la fonction doit être positive sur l'intervalle de définition) pour poursuivre leur recherche. La fonction représentée ensuite (phase C) paraît positive mais ne l'est plus à partir d'un moment. Cela entraîne le fait de chercher une fonction de la forme $f(x) = \dfrac{a}{bx + c}$, toujours positive sur $[0\ ; +\infty[$, d'où l'arrivée ensuite de curseurs. Le travail de recherche de la « meilleure » courbe se poursuit (phase D). Reste ensuite à prendre en compte le fait que l'aire sous la courbe doit valoir 1. Certains élèves disent que l'aire est nécessairement infinie, vu que l'on se place sur $[0\ ; +\infty[$, mais il y a toujours d'autres élèves qui contredisent cela en faisant référence au devoir maison qui les a fait travailler sur cette problématique. Ce point de blocage est donc vite écarté grâce à l'anticipation dans l'année.

L'« Inspecteur de fonction » permet de connaître l'aire sous la courbe sur $[0\ ; 1000[$ par exemple (nous ne pourrons bien entendu pas aller jusqu'à l'infini) et les élèves s'aperçoivent que l'aire est finie mais toujours supérieure à 1 (phase E). Pour que l'aire diminue, il faut choisir des fonctions dont l'allure finit par ne plus du tout correspondre à l'histogramme.

Ce long travail, indispensable pour bien consolider les propriétés de la fonction de densité, permet finalement d'invalider ces fonctions candidates. Les élèves proposent toujours ensuite une fonction du type $f(x) = \alpha\, e^{-\lambda x}$. A nouveau, un travail sur *GeoGebra*, avec les curseurs, permet de voir que α doit être égal à λ pour que l'aire soit égale à 1 et permet aux élèves de se mettre d'accord sur une expression pour la fonction. Une année où le matériel informatique n'a pas fonctionné, l'enseignante a dû imposer le paramètre λ et l'adhésion de la classe a été moindre. Il semble important que cette fonction soit le choix de la classe. Il peut être intéressant de mentionner aux élèves que les mathématiciens qui travaillent en statistique appliquée disposent de méthodes pour choisir les paramètres les plus adaptés à leur jeu de données (par exemple, la méthode du maximum de vraisemblance). Il est aussi indispensable de leur faire prendre conscience que nous ne savons pas si le modèle choisi est « le bon », comme précisé plus tôt. Il faudrait un retour vers la réalité pour que le modèle soit validé ou non.

A la fin de cette recherche, le nouveau problème est le suivant : comment calculer les probabilités nous-mêmes ? Une seconde phase de travail démarre : comment calculer l'aire sous cette courbe entre 0 et 20 ? La courbe choisie par les élèves[13] est distribuée à chaque binôme (le choix de donner un papier quadrillé guide les élèves vers des procédures de comptage de carreaux). Le déroulement est assez classique : un temps de recherche et une mise en commun des méthodes et des résultats trouvés. Les méthodes proposées par les élèves sont les suivantes :

- Comptage de carreaux (figure 9) souvent seulement sous la courbe ;
- Méthode de compensation : ils comptent les carreaux en dessous en découpant certains pour qu'il y ait compensation au-dessus et en-dessous ;
- Méthode des trapèzes ;
- Méthode des rectangles ;
- Méthodes des tangentes, de façon plus rare (figure 9).

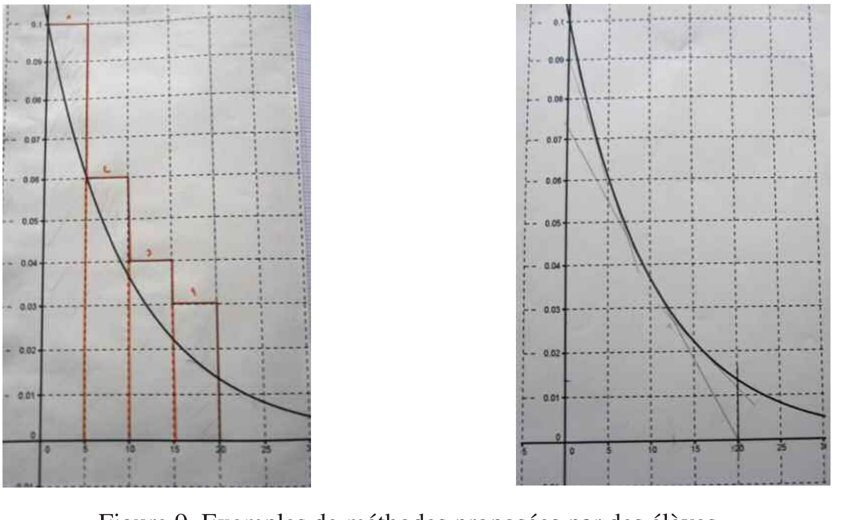


Figure 9. Exemples de méthodes proposées par des élèves

13 L'enseignante a préparé en amont les courbes de fonctions $x \to \lambda e^{-\lambda x}$ pour plusieurs λ proches de 0,2.

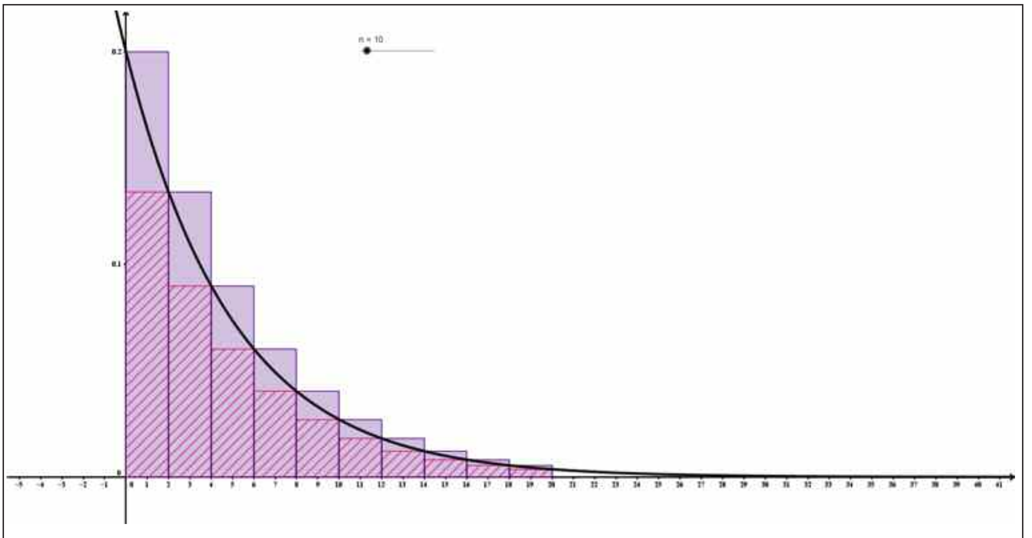


Figure 10. Illustration de la méthode des rectangles avec $f : x \to 0{,}2\, e^{-0{,}2x}$

Il est important de laisser le temps à chaque binôme de trouver une réponse. Il faut que chacun ait le temps d'expérimenter sa méthode afin d'accepter aussi celles des autres et de se convaincre que certaines démarches sont meilleures que d'autres et pourront être systématisées. La méthode des rectangles est reconnue par les élèves comme la plus facile et peut ensuite être institutionnalisée et programmée. Le logiciel *GeoGebra* permet aisément de montrer aux élèves ce qui se passe quand le nombre de rectangles tend vers $+\infty$ (figure 10).

3. 5 *La suite de la séquence*

Ces deux problèmes sont alors le point de départ pour toute une séquence articulant lois à densité et calcul intégral. Ils permettent d'introduire les probabilités continues et d'amener le besoin d'un nouvel outil mathématique pour calculer des aires sous une courbe : l'intégrale. Tous les éléments sont mis en place dans le cours, qui alterne probabilités continues et intégration (des fonctions continues et positives sur un intervalle $[a, b]$). Une première partie, assise sur les synthèses des problèmes de la rencontre et du volcan Aso et sur le devoir maison préparatoire, institutionnalise le vocabulaire : variable aléatoire continue, fonction de densité, calcul de probabilités et aire sous la courbe de la fonction de densité. La deuxième partie porte sur l'intégrale d'une fonction continue et positive sur un intervalle et suit un plan classique. En revanche, l'originalité de cette démarche tient au fait que tous les exemples sont déclinés systématiquement en analyse et en probabilités. Cela permet notamment de revenir sur le problème du volcan Aso et de faire alors le calcul exact de la probabilité $P(0 \leq X \leq 20)$ à ce moment-là (calcul d'aire de niveau 2 à ce stade du cours). La troisième partie traite des lois uniformes [14] et exponentielles.

14 La démonstration de l'espérance d'une loi uniforme sera faite ultérieurement lors de l'extension de l'intégration aux fonctions continues de signe quelconque.

Les exercices proposés dans cette séquence sont tantôt des exercices d'analyse pure, tantôt des exercices sur les lois à densité. L'idée n'est pas d'inventer de nouveaux exercices mais de les mélanger pour éviter de cloisonner des notions qui sont liées. Nous avons pu observer que cette façon de procéder ne perturbe pas les élèves, qui sont pleinement capables d'identifier quand ils sont dans le domaine de l'analyse ou dans le domaine des probabilités. Cependant, cela leur permet d'être plus en mesure de relier les deux domaines, ce qui n'est pas le cas pour beaucoup d'autres élèves n'ayant pas suivi cette progression (Derouet, Planchon, Hausberger & Hochmuth, 2018).

4. — Conclusions

Ce travail collaboratif a été l'occasion de questionner l'enseignement des lois à densité et du calcul intégral et de tester une séquence originale ayant pour objectif de donner du sens notamment à la notion de fonction de densité. Nous avons montré que ces séances sont effectivement réalisables en classe de terminale S dans des conditions ordinaires. Les contraintes imposées par l'enseignante ont été respectées, à la fois en termes de temps consacré aux séances d'introduction et de connaissances des élèves à la fin de la séquence. Nous avons même l'assurance que les élèves donnent plus de sens à la notion de fonction de densité et donc aux lois à densité.

D'autres points sont à prendre en considération. L'enseignante a depuis totalement adopté cette séquence dans ses pratiques et la met en place tous les ans depuis quatre ans. Les problèmes d'introduction de la séquence sont, selon elle, à la fois intéressants et stimulants pour l'enseignant mais aussi pour les élèves qui sont très motivés lors de ces séances.

Nous pourrions penser que le profil de classe de l'enseignante explique cette réussite, cependant depuis, d'autres enseignants dans d'autres contextes scolaires ont testé cette séquence, l'ont aussi appréciée et ont fait le même type de constat sur l'implication des élèves. Ces séances d'introduction demandent cependant beaucoup d'investissement de la part de l'enseignant, que ce soit en amont pour bien maîtriser tous les enjeux de ces problèmes, ou pendant les séances, car elles demandent de ne pas trop en dire tout en cadrant le déroulement, de laisser les élèves partir dans de mauvaises directions, de les laisser aller au bout de leurs démarches même si cela peut prendre du temps.

Notre objectif est maintenant de diffuser plus largement cette séquence afin que le plus grand nombre d'enseignants ait la possibilité de s'emparer de ce travail pour leur classe. Un projet de recherche, en collaboration avec des enseignant.e.s de l'Irem de Strasbourg et des formatrices de l'ESPE de Strasbourg, est en cours pour concevoir une ressource en ligne autour de cette séquence qui devrait être accessible en ligne courant 2019.

Références des manuels cités

Beltramone, J.-P. (dir.). (2012). *Déclic mathématiques TS. Enseignement spécifique et de spécialité.* Paris : Hachette Education.

Deschamps, C. (dir.). (2012). *Symbole maths Term S.* Paris : Belin.

Le Yaouanq, M.-H. (dir.). (2012). *Math'x. Term S. Enseignement spécifique.* Paris : Didier.

Poncy, M., Bonnafet, J.-L. & Russier M.-C. (dir.). (2012). *Indice. Maths. Term S. Enseignement spécifique.* Paris : Bordas.

Sigward, E. (dir.). (2011). *Odyssée. Mathématiques. Tle S. Enseignement spécifique.* Paris : Hatier.

Références bibliographiques

Barbazo, E., Bouscasse, J.-M., Pomès, R., Puyou, J., Puyou, M., & Terracher, P.-H. (2003). *L'esprit des lois continues ou Quelques aspects du calcul de probabilités au lycée.* Bordeaux : Irem d'Aquitaine.

Bressoud, E. & Kahané, J.-C. (2010), *Statistique descriptive*, 2è édition. Pearson France.

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 œuvre de tâches d'introduction articulant lois à densité et calcul intégral.* Université Paris Sorbonne Cité, Université Paris Diderot.

Derouet, C. (2017). Emergence historique des lois à densité. Des pistes pour l'enseignement. *Statistique et Enseignement*, *8*(2), 3–24.

Derouet, C. (2018, à paraître). L'histogramme sous une autre facette. *Au Fils Des Maths - Le Bulletin de l'APMEP*, *528*, 33-37.

Derouet, C., Planchon, G., Hausberger, T., & Hochmuth, R. (2018). Bridging probability and calculus: the case of continuous distributions and integrals at the secondary-tertiary transition. In V. Durand-Guerrier, R. Hochmuth, S. Goodchild, & N. M. Hogstad (Eds.), *Proceedings of the Second Conference of the International Network for Didactic Research in University Mathematics (INDRUM2018, 5-7 April 2018)* (p. 497-506). Kristiansand, Norway : University of Agder and INDRUM.

Ministère de l'Education nationale (MEN) (2011). Programme de l'enseignement spécifique et de spécialité de mathématiques. Classe terminale de la série scientifique. *Le B.O.*

Ministère de l'Éducation nationale de la Jeunesse et de la Vie associative (MENJVA), & DGESCO (2012). *Ressources pour la classe terminale générale et technologique. Probabilités et statistique.*

Ministère de la Jeunesse de l'Éducation nationale et de la Recherche (MJENR) (2002). *Mathématiques. Classe terminale. Série économique et sociale. Série scientifique.* Paris : Centre national de documentation pédagogique.

Roditi, E. (2009). L'histogramme : à la recherche du savoir à enseigner. *Spirale-Revue de Recherches en Éducation*, *43*, 129–138.

ANNEXE 1

Le devoir maison pour aborder la problématique d'une longueur finie d'une ligne brisée infinie

Question 1

On considère la figure ci-dessous construite par itération. On construit un carré $A_0B_0C_0D_0$ de 10 cm de côté. A l'intérieur du carré, on trace le quart de cercle de centre B_0 et de rayon B_0C_0 (étape 0). On appelle B_1 le milieu du segment $[B_0C_0]$ et on trace, à l'extérieur du carré $A_0B_0C_0D_0$, le carré $C_0B_1C_1D_1$ puis le quart de cercle de centre B_1 et de rayon B_1C_1. L'itération du processus donne les carrés : $C_1B_2C_2D_2$, $C_2B_3C_3D_3$... et la spirale $A_0C_0C_1C_2C_3$...

a) Compléter proprement la figure ci-dessous (étape 2 et 3).

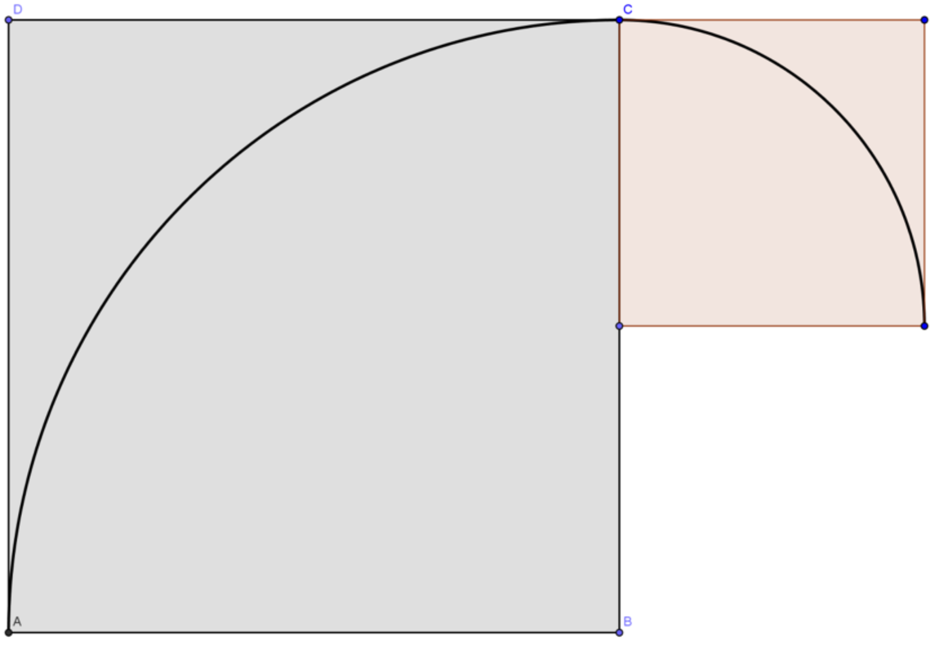
b) Calculer la longueur L_3 totale de la spirale obtenue à l'étape 3.

On souhaite connaître la longueur de la spirale quand on itère le processus une infinité de fois.

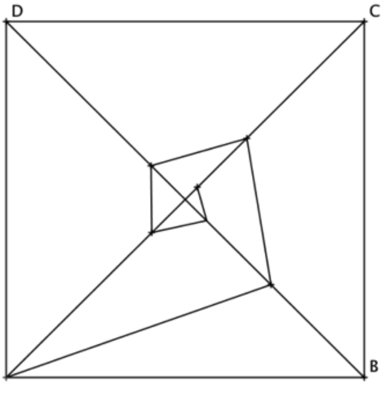
c) Emettre une conjecture sur la longueur de la spirale quand on itère le processus une infinité de fois. Expliquer votre réponse. On ne demande pas ici une preuve.

d) Exprimer en fonction de n, la longueur totale L_n de la spirale obtenue à l'étape n.

e) Etudier le comportement de la suite (L_n) quand n tend vers $+\infty$. Commenter.



Question 2 : On considère un carré ABCD de centre O, tel que OA = 1.
Soit n un entier naturel ; sur les diagonales [AC] et [BD] on place de différentes façons des points P_0, P_1, P_2, … , P_n tels que : P_0 = A, $P_1 \in$ [OB], $P_2 \in$ [OC], $P_3 \in$ [OD], $P_4 \in$ [OA], $P_5 \in$ [OB], et ainsi de suite. De plus, ces points sont tels que pour tout entier naturel $k < n$, $OP_{k+1} < OP_k$.



Selon la manière dont ces points sont définis, on obtient des spirales $P_0P_1P_2…P_n$ différentes, formées de n segments. Quand n vers $+\infty$, on obtient des spirales infinies et on s'interroge sur la longueur de la spirale infinie.

Soient $P_1 \in$ [OB] tel que $OP_1 = \frac{1}{2}$, $P_2 \in$ [OC] tel que $OP_2 = \frac{1}{3}$, $P_3 \in$ [OD] tel que $OP_3 = \frac{1}{4}$, et ainsi de suite de telle sorte que pour tout entier naturel n, $OP_n = \frac{1}{n+1}$.

1. Conjecturer (sans rien calculer) la longueur de la spirale infinie ainsi constituée. Vous expliquerez soigneusement votre conjecture.
2. Pour tout $n \geq 1$, on pose $v_n = P_{n-1}P_n$; montrer que $v_n \geq \frac{1}{n}$.
3. Pour tout $n \geq 1$, on pose $L_n = \sum_{k=1}^{k=n} v_k$ et $S_n = \sum_{k=1}^{k=n} \frac{1}{k}$. Que représente L_n ?
   Comparer L_n et S_n.
4. Pour tout $n \geq 1$, exprimer $S_{2n} - S_n$ en fonction de n ; quel est le plus petit terme de cette somme ? En déduire que $S_{2n} - S_n \geq \frac{1}{2}$.
5. En déduire, en raisonnant par l'absurde, que la suite (S_n) n'est pas convergente.
6. Montrer que la suite (S_n) est croissante.
7. Montrer que la suite (L_n) a pour limite $+\infty$ quand n tend vers $+\infty$. Que pouvez-vous en conclure ?

Question 3 : Comparer les résultats obtenus dans les exercices 1 et 2. Qu'en pensez-vous ?

ANNEXE 2

Le devoir maison préliminaire sur les histogrammes et la loi uniforme

DEVOIR A LA MAISON - TS

Attention ! Ce devoir constitue l'introduction d'un futur chapitre. Il doit faire l'objet d'une réflexion individuelle.

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	2249974

Source : Insee, RP2011 exploitation principale.

1. Quelle est la population étudiée ? Le caractère étudié ? Ce caractère est-il qualitatif ou quantitatif ? Discret ou continu ?

2. A l'aide du tableau ci-dessus, compléter le tableau de fréquences suivant (en arrondissant à 10^{-2} près) :

Tranche d'âge	$[0;3[$	$[3;6[$	$[6;11[$	$[11;18[$	$[18;25[$	$[25;40[$	$[40;55[$	$[55;65[$	$[65;80[$	$[80;105[$
Fréquence										

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 la Japon 127 417 244 (source ; Institut nationale 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 la densité de population, c'est-à-dire le nombre d'habitants au kilomètre carré. Avec une superficie de 2,02 km^2 *pour Monaco et de 378 000*km^2 *pour le Japon, les densités sont respectivement* $d_1 = \frac{32\,543}{2{,}02} = 16\,110{,}40 h/km^2$ *pour Monaco et* $d_2 = \frac{127\,417\,244}{378\,000} = 337 h/km^2$ *pour le Japon. Autrement dit, alors que la population de Monaco est la moins importante en taille, sa densité est plus importante que celle du Japon.*

Ainsi, une représentation pertinente des populations de Monaco et du Japon doit rendre visible cette différence de densité. De même dans notre exemple, il n'est pas pertinent de représenter de la même façon la fréquence égale à 0,05 de la tranche d'âge $[6\,;11[$ *et de la tranche d'âge* $[80\,;105[$. *Ces deux classes d'âge ont la même fréquence mais l'amplitude de la première classe est inférieure à celle de la seconde, on pressent que la densité de la première classe est supérieure à celle de la seconde. Par analogie avec la densité de population utilisée en géographie, on calculera le quotient de la fréquence d'une classe d'âge par l'amplitude de cette classe, et on parlera de densité de fréquence. Dans notre exemple, la densité de fréquence pour la tranche d'âge* $[6\,;11[$ *est* $\frac{0{,}05}{5} = 0{,}01$ *et la densité pour 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. *<u>La représentation graphique doit rendre visible ce phénomène.</u>*

Définitions

- *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 : $d_i = f_i/a_i$.*

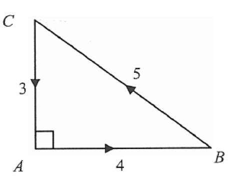
Rappel : L'amplitude a_i de la classe $[x_i ; x_{i+1}[$ est le nombre $a_i = x_{i+1} - x_i$.

3. a) Compléter l'histogramme de fréquences de la répartition de la population parisienne en 2011 par âge (sur la feuille annexe). L'aire d'un carreau correspond à une fréquence de 0,002.
b) Montrer que, dans un histogramme de fréquences, les hauteurs des rectangles sont proportionnelles à la densité. *Ainsi une classe de densité plus forte est représentée par un rectangle plus haut.* Que peut-on porter sur l'axe des ordonnées ? Graduer cet axe.

4. a) Dans quelle tranche d'âge la densité est-elle la plus importante ? A l'aide de vos connaissances en géographie, pouvez-vous justifier votre réponse ?
b) 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 :

Un point mobile M se déplace à vitesse constante sur le « pourtour » d'un triangle, toujours dans le même sens (indiqué par la flèche). On suppose qu'il démarre au point C. Une panne se produit subitement et aléatoirement et le point mobile s'arrête instantanément.

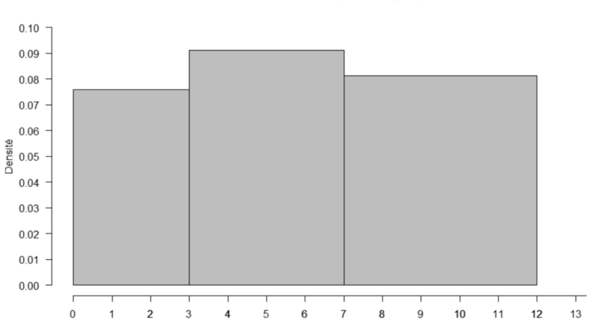


Partie A : Conjectures

1. Quelle conjecture faites-vous sur la probabilité p_1 que le point mobile s'arrête sur l'hypoténuse ?
2. Quelle conjecture faites-vous sur la probabilité p_2 que le point mobile s'arrête au sommet de l'angle droit ? Étayez votre conjecture sur des arguments raisonnés.

Partie B : Simulations

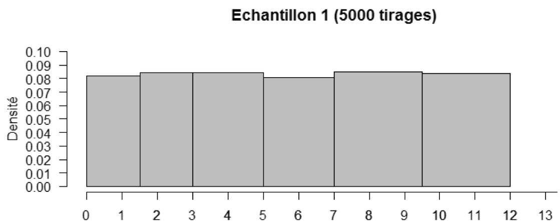
On peut remarquer que cette situation se ramène à l'expérience aléatoire qui consiste à choisir au hasard un nombre réel dans l'intervalle $[0 ; 12]$. A l'aide d'un tableur, on a simulé le tirage de 1000 nombres réels dans l'intervalle $[0 ; 12]$. Voici l'histogramme de fréquences obtenues :

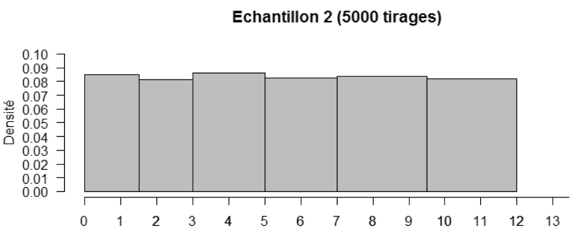


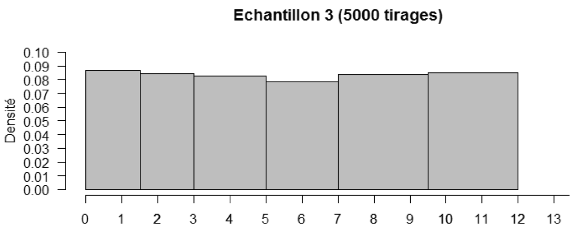
1. Pouvez-vous évaluer la probabilité que le point mobile s'arrête sur l'hypoténuse ? Vous détaillerez votre méthode et ferez figurer sur le dessin tout élément utile à la réponse.
2. Pouvez-vous évaluer la probabilité que le point mobile s'arrête au sommet de l'angle droit ? Expliquer votre réponse.

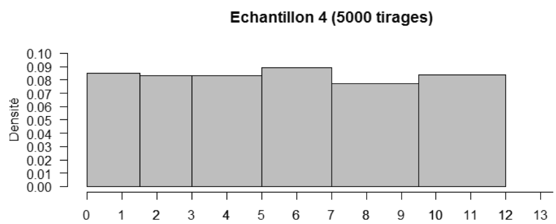
Partie C : Calculs théoriques

On a ensuite simulé, plusieurs fois, le tirage de 5 000 nombres réels dans l'intervalle $[0\ ;\ 12]$. Voici plusieurs histogrammes obtenus :





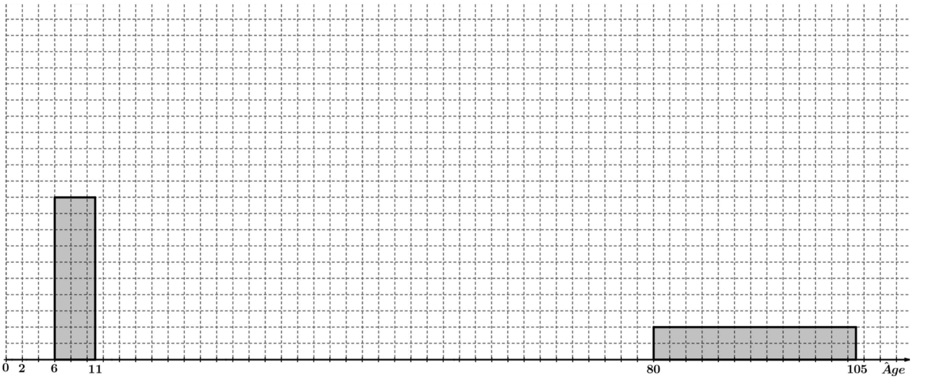




1. Les histogrammes ci-dessus sont différents tout en ayant la même allure. Justifier pourquoi.
2. Malgré ces différences, il semble y avoir une « courbe de tendance », c'est-à-dire une courbe qui « lisse » les histogrammes (**la même** pour les quatre). Tracer au jugé cette courbe, en vert, sur chacun des histogrammes ci-dessus.
3. On souhaite déterminer l'équation de cette courbe. Que doit valoir l'aire sous cette courbe ? En déduire l'équation de cette courbe. Tracer **sur votre copie** cette courbe.
4. Calculer, à l'aide de la courbe tracée sur votre copie, la probabilité p_1 que le point mobile s'arrête sur l'hypoténuse. Vous détaillerez votre méthode et ferez figurer sur le dessin tout élément utile à la réponse.
5. Calculer, à l'aide de cette courbe tracée sur votre copie, la probabilité p_2 que le point mobile s'arrête au sommet de l'angle droit. Cette réponse est-elle en cohérence avec votre conjecture de la partie A, question 2 ?

FEUILLE ANNEXE – Exercice 1 – 3.a)

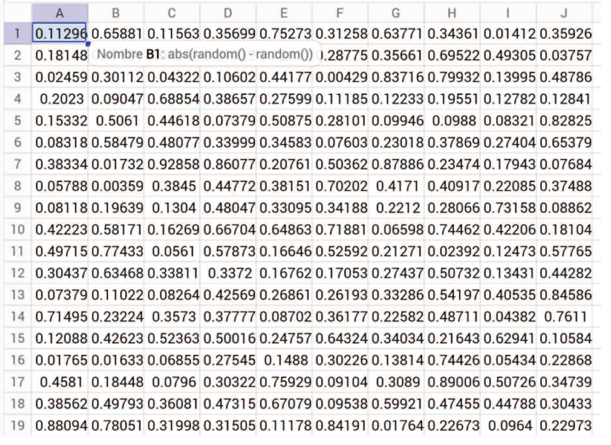
Tranche d'âge	$[0;3[$	$[3;6[$	$[6;11[$	$[11;18[$	$[18;25[$	$[25;40[$	$[40;55[$	$[55;65[$	$[65;80[$	$[80;105[$
Fréquence										



ANNEXE 3

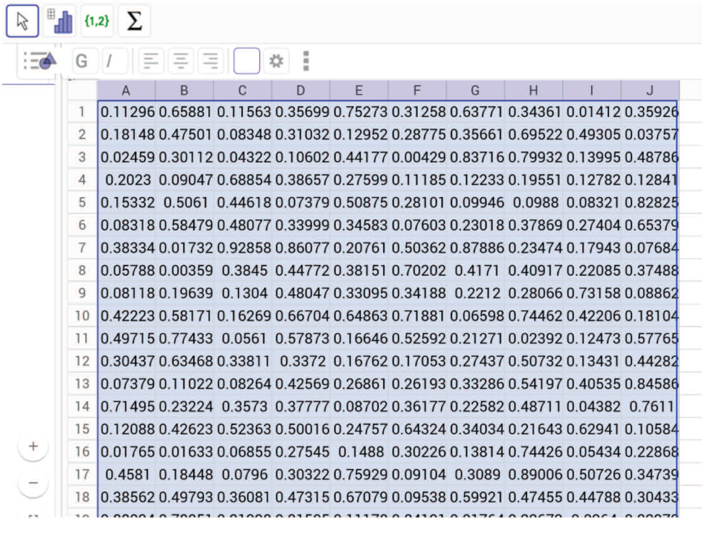
Tutoriel pour construire un histogramme avec le logiciel GeoGebra version 6. Exemple du problème de la rencontre.

Etape 1 : Rentrer vos données dans la fenêtre « Tableur » de *GeoGebra*. Dans le cas du problème de la rencontre, les données sont simulées avec la formule : =abs(random()-random()). Ici, nous avons un échantillon de 1000 réalisations (10 colonnes de 100 lignes).



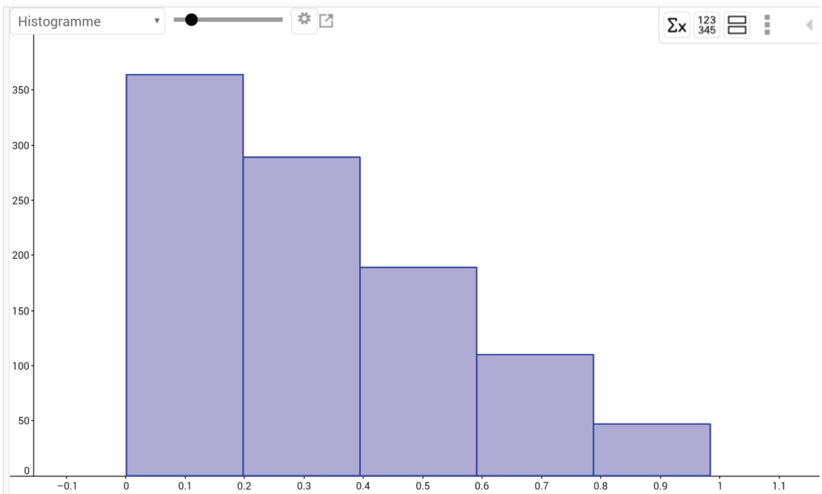
	A	B	C	D	E	F	G	H	I	J
1	0.11296	0.65881	0.11563	0.35699	0.75273	0.31258	0.63771	0.34361	0.01412	0.35926
2	0.18148	Nombre B1 : abs(random() - random())				).28775	0.35661	0.69522	0.49305	0.03757
3	0.02459	0.30112	0.04322	0.10602	0.44177	0.00429	0.83716	0.79932	0.13995	0.48786
4	0.2023	0.09047	0.68854	0.38657	0.27599	0.11185	0.12233	0.19551	0.12782	0.12841
5	0.15332	0.5061	0.44618	0.07379	0.50875	0.28101	0.09946	0.0988	0.08321	0.82825
6	0.08318	0.58479	0.48077	0.33999	0.34583	0.07603	0.23018	0.37869	0.27404	0.65379
7	0.38334	0.01732	0.92858	0.86077	0.20761	0.50362	0.87886	0.23474	0.17943	0.07684
8	0.05788	0.00359	0.3845	0.44772	0.38151	0.70202	0.4171	0.40917	0.22085	0.37488
9	0.08118	0.19639	0.1304	0.48047	0.33095	0.34188	0.2212	0.28066	0.73158	0.08862
10	0.42223	0.58171	0.16269	0.66704	0.64863	0.71881	0.06598	0.74462	0.42206	0.18104
11	0.49715	0.77433	0.0561	0.57873	0.16646	0.52592	0.21271	0.02392	0.12473	0.57765
12	0.30437	0.63468	0.33811	0.3372	0.16762	0.17053	0.27437	0.50732	0.13431	0.44282
13	0.07379	0.11022	0.08264	0.42569	0.26861	0.26193	0.33286	0.54197	0.40535	0.84586
14	0.71495	0.23224	0.3573	0.37777	0.08702	0.36177	0.22582	0.48711	0.04382	0.7611
15	0.12088	0.42623	0.52363	0.50016	0.24757	0.64324	0.34034	0.21643	0.62941	0.10584
16	0.01765	0.01633	0.06855	0.27545	0.1488	0.30226	0.13814	0.74426	0.05434	0.22868
17	0.4581	0.18448	0.0796	0.30322	0.75929	0.09104	0.3089	0.89006	0.50726	0.34739
18	0.38562	0.49793	0.36081	0.47315	0.67079	0.09538	0.59921	0.47455	0.44788	0.30433
19	0.88094	0.78051	0.31998	0.31505	0.11178	0.84191	0.01764	0.22673	0.0964	0.22973

Etape 2 : Sélectionner la ou les colonnes comportant les données (brutes) que vous voulez représenter sous forme d'histogramme (il n'y a pas besoin de regrouper les données par exemple). Plusieurs icônes apparaissent dont l'icône :  .

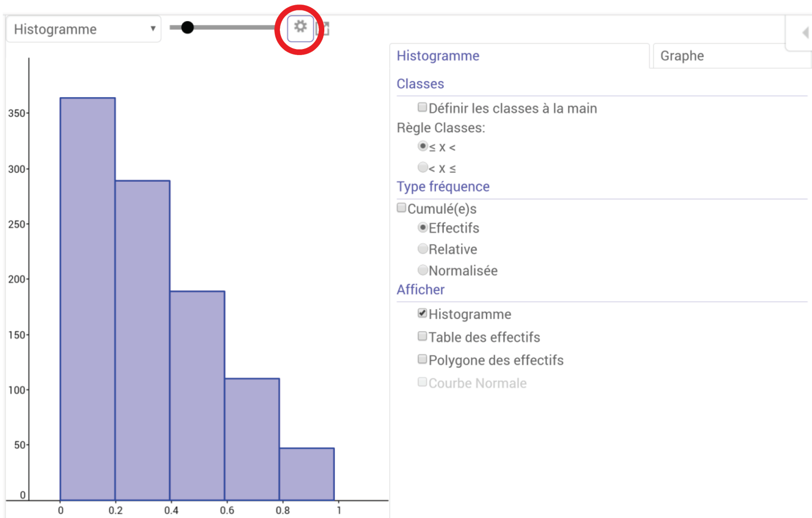


	A	B	C	D	E	F	G	H	I	J
1	0.11296	0.65881	0.11563	0.35699	0.75273	0.31258	0.63771	0.34361	0.01412	0.35926
2	0.18148	0.47501	0.08348	0.31032	0.12952	0.28775	0.35661	0.69522	0.49305	0.03757
3	0.02459	0.30112	0.04322	0.10602	0.44177	0.00429	0.83716	0.79932	0.13995	0.48786
4	0.2023	0.09047	0.68854	0.38657	0.27599	0.11185	0.12233	0.19551	0.12782	0.12841
5	0.15332	0.5061	0.44618	0.07379	0.50875	0.28101	0.09946	0.0988	0.08321	0.82825
6	0.08318	0.58479	0.48077	0.33999	0.34583	0.07603	0.23018	0.37869	0.27404	0.65379
7	0.38334	0.01732	0.92858	0.86077	0.20761	0.50362	0.87886	0.23474	0.17943	0.07684
8	0.05788	0.00359	0.3845	0.44772	0.38151	0.70202	0.4171	0.40917	0.22085	0.37488
9	0.08118	0.19639	0.1304	0.48047	0.33095	0.34188	0.2212	0.28066	0.73158	0.08862
10	0.42223	0.58171	0.16269	0.66704	0.64863	0.71881	0.06598	0.74462	0.42206	0.18104
11	0.49715	0.77433	0.0561	0.57873	0.16646	0.52592	0.21271	0.02392	0.12473	0.57765
12	0.30437	0.63468	0.33811	0.3372	0.16762	0.17053	0.27437	0.50732	0.13431	0.44282
13	0.07379	0.11022	0.08264	0.42569	0.26861	0.26193	0.33286	0.54197	0.40535	0.84586
14	0.71495	0.23224	0.3573	0.37777	0.08702	0.36177	0.22582	0.48711	0.04382	0.7611
15	0.12088	0.42623	0.52363	0.50016	0.24757	0.64324	0.34034	0.21643	0.62941	0.10584
16	0.01765	0.01633	0.06855	0.27545	0.1488	0.30226	0.13814	0.74426	0.05434	0.22868
17	0.4581	0.18448	0.0796	0.30322	0.75929	0.09104	0.3089	0.89006	0.50726	0.34739
18	0.38562	0.49793	0.36081	0.47315	0.67079	0.09538	0.59921	0.47455	0.44788	0.30433

Etape 3 : Cliquer sur l'icône , puis sur « Statistiques à une variable ». Une nouvelle fenêtre apparaît.



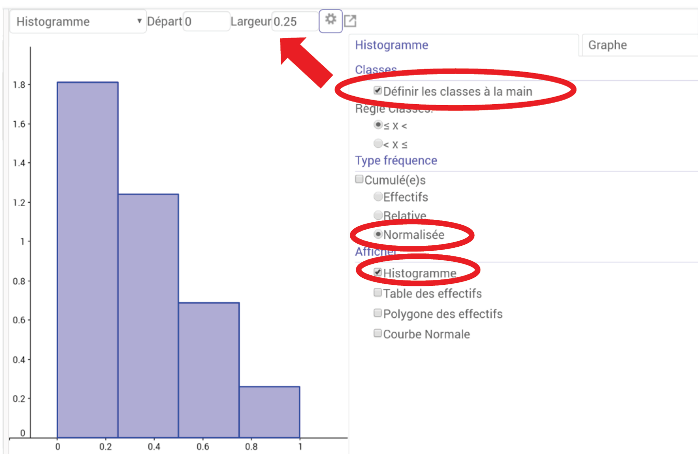
Etape 4 : Appuyer sur l'icône . Une nouvelle fenêtre apparaît.



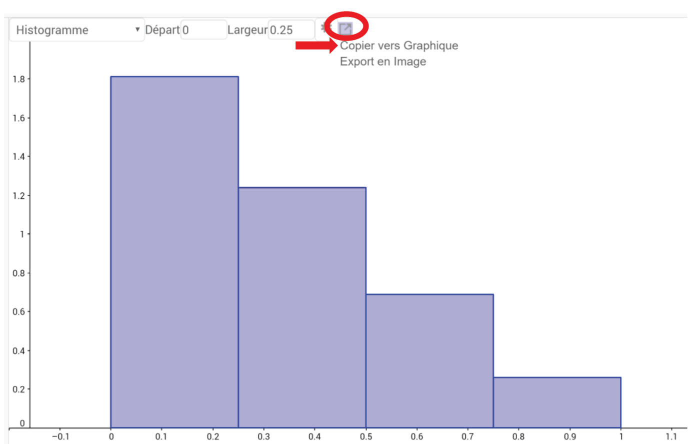
Etape 5 : Pour obtenir un histogramme de fréquences (ce qui nous intéresse pour passer à la fonction de densité) :

- dans la rubrique « Type fréquence », cliquer sur « Normalisée »
- dans la rubrique « Afficher », cliquer sur « Histogramme ».

Pour choisir vous-mêmes l'amplitude des classes de l'histogramme, cocher « Définir les classes à la main ». Puis, choisir le « départ » de vos classes et la largeur des classes (l'histogramme sera hélas à pas constant).



Etape 6 : Pour déplacer l'histogramme dans la fenêtre « Graphique », cliquer sur  , puis « Copier vers graphique ».



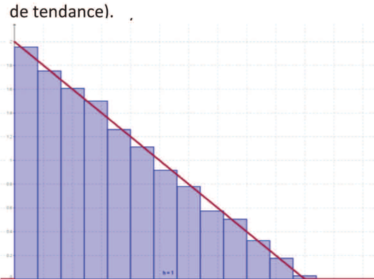
ANNEXE 4

Synthèse du problème de la rencontre, faite par l'enseignante à partir des synthèses des élèves (en 2016)

La première étape : La modélisation du problème.
Pour résoudre ce problème concret, nous devons d'abord modéliser la situation (choisir un modèle). Cela nécessite de faire des hypothèses. On fait abstraction d'événements extérieurs qui pourraient influer sur les instants d'arrivée de Karine et Olivier. Nous pouvons ainsi assimiler ces arrivées à un tirage au hasard de réels dans l'intervalle [0 ; 1]. Nous acceptons aussi de considérer que tous les réels de l'intervalle [0 ; 1] sont atteints et que les deux personnages puissent arriver au même instant.
Nous devons alors poser des notations communes pour faciliter les échanges. Nous notons Δt le temps d'attente, exprimé en heures, du premier arrivé. Δt prend au moins théoriquement toutes les valeurs de l'intervalle $[0; 1]$.
Nous reconnaissons une variable aléatoire continue comme dans le dm9.

Deuxième étape : Examen de remarques d'ordre qualitatif ou quantitatif.
Souvent en désaccord ou incapables de convaincre les autres, nous avons laissé en suspens nombres de ces remarques ou conjectures. Pour citer l'un de vous « seule la preuve nous apportera la certitude ».
Nous avons alors décidé d'évaluer $p\left(\Delta t \leq \frac{1}{6}\right)$ soit la probabilité que le premier arrivé attende moins de 10 min.
En effet, examiner un exemple permet souvent de déduire une méthode générale pourvu que l'exemple soit générique.

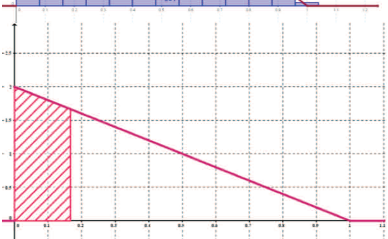
Troisième étape : Détermination du modèle.
Ne pouvant répéter l'expérience de nombreuses fois, nous avons simulé l'expérience et obtenu des échantillons de taille 10 000. Prenant exemple sur le dm9, nous avons alors construit avec un logiciel l'histogramme de fréquences d'amplitude 5 min. Si l'ordinateur n'avait pas cessé de fonctionner, nous aurions pu constater que pour différents échantillons de taille 10 000, les histogrammes ont même allure. Comme dans le dm 9, nous avons cherché à dessiner une courbe qui approche le mieux le haut des rectangles (appelée courbe de tendance).



La courbe est représentative d'une fonction f affine définie sur $[0; 1]$ et décroissante. f doit aussi vérifier des éléments indépendants des histogrammes représentés. f vérifie :

- l'aire sous la courbe sur $[0; 1]$ est égale à 1 car la somme des fréquences est égale à 1.
- $f(1) = 0$.

Cela a permis de déterminer l'expression de la fonction $f(x) = -2x + 2$.



Pour calculer $p\left(\Delta t \leq \frac{1}{6}\right)$, on calcule l'aire sous la courbe de tendance sur l'intervalle $\left[0; \frac{1}{6}\right]$ qui ici correspond à l'aire d'un trapèze.
On trouve $p\left(\Delta t \leq \frac{1}{6}\right) = \frac{11}{36} \approx 0{,}31$.

Etape 4 : Utilisation du modèle.
De la même façon, nous pourrons maintenant calculer la probabilité que la v.a. Δt prenne ses valeurs dans un intervalle quelconque $[a; b]$ de $[0; 1]$. Il suffira de calculer l'aire sous C_f entre a et b pour calculer $p(\Delta t \in [a; b])$. **La connaissance de la fonction f, nous permet ainsi de connaître de la v.a. Δt**. Voir exercice 1.

ANNEXE 5

Le problème du volcan Aso

Problème 2 : Le volcan Aso, situé sur l'île de Kyushu au Japon, est l'un des volcans les plus actifs au monde. On possède un relevé précis de ses éruptions, régulièrement tenues depuis le XIII^{e} siècle. Dans le tableau ci-dessous, nous disposons de ces données pour la période allant du XIII^{e} au XXI^{e} siècle. Nous nous intéressons au temps écoulé, en années, entre deux éruptions successives.

Années des 131 éruptions du volcan Aso du XIIIe au XXIe siècle

Année	Attente		Année	Attente		Année	Attente
1239			1612	1		1911	1
1240	1		1613	1		1914	3
1265	25		1620	7		1916	2
1269	4		1631	11		1918	2
1271	2		1637	6		1919	1
1272	1		1649	12		1920	1
1272	0		1668	19		1923	3
1273	1		1668	0		1925	2
1274	1		1675	7		1926	1
1281	7		1683	8		1928	2
1286	5		1691	8		1930	2
1305	19		1709	18		1932	2
1324	19		1765	56		1934	2
1331	7		1772	7		1935	1
1331	0		1781	9		1936	1
1335	4		1804	23		1936	0
1340	5		1806	2		1937	1
1343	3		1814	8		1937	0
1375	32		1815	1		1938	1
1376	1		1816	1		1940	2
1377	1		1826	10		1943	3
1387	10		1827	1		1943	0
1434	47		1827	0		1945	2
1438	4		1828	1		1946	1
1473	35		1829	1		1946	0
1485	12		1830	1		1947	1
1505	20		1830	0		1949	2
1506	1		1835	5		1950	1
1522	16		1837	2		1951	1
1533	11		1838	1		1953	2
1542	9		1854	16		1956	3
1558	16		1856	2		1956	0
1562	4		1872	16		1957	1
1564	2		1874	2		1960	3
1573	9		1884	10		1963	3
1574	1		1894	10		1964	1
1576	2		1897	3		1970	6
1582	6		1898	1		1973	3
1583	1		1906	8		1977	4
1584	1		1907	1		1979	2
1587	3		1908	1		1989	10
1592	5		1909	1		1992	3
1598	6		1910	1		1994	2
1611	13					2014	20
						2015	1
						2016	1

Attente	Effectif
0	10
1	40
2	20
3	11
4	5
5	4
6	4
7	5
8	4
9	3
10	5
11	2
12	2
13	1
16	4
18	1
19	3
20	2
23	1
25	1
32	1
35	1
47	1
56	1

La dernière éruption du volcan Aso a eu lieu au cours du mois d'octobre 2016.

a) Comment évaluer la probabilité que la prochaine éruption ait lieu avant octobre 2021 ?

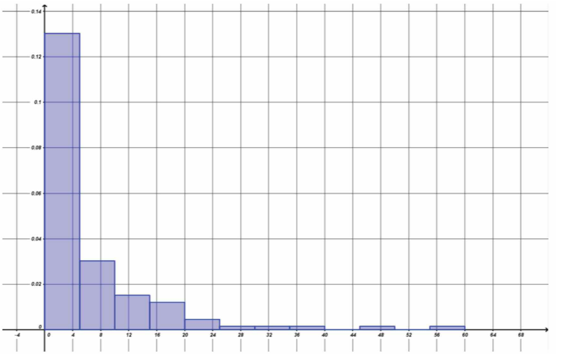
b) Comment évaluer la probabilité que la prochaine éruption ait lieu au cours de l'année 2031 ?

ANNEXE 6

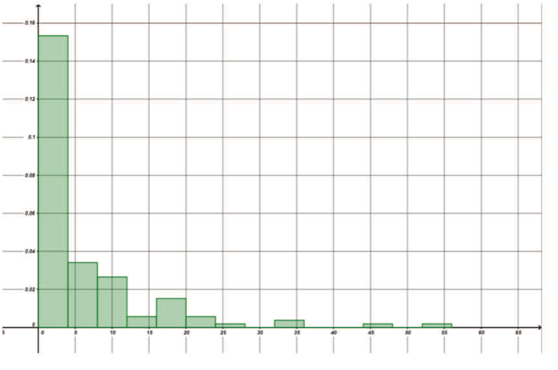
Document distribué aux élèves, après avoir introduits les histogrammes (problème du volcan Aso)

Ci-dessous sont tracés les histogrammes du temps écoulé en années entre deux éruptions successives, pour des amplitudes respectives de 5 ans, 4 ans et 2 ans.

Amplitude 5 ans :



Amplitude 4 ans :



Amplitude 2 ans :