REPERES - IREM. N° 98 - janvier 2015

UNE INITIATION A LA STATISTIQUE EN CLASSE DE SECONDE

Pascale BOULAIS, Marie DIUMENGE,
Martine VERGNAC, Claudine VERGNE
Irem de Montpellier

Les programmes de mathématiques à partir de 2009 donnent une place importante à l'initiation à la statistique inférentielle. Quelles sont les raisons d'être d'un tel choix ? Quels éclairages l'histoire peut-elle nous apporter sur la construction des notions de statistique et probabilité et sur le lien qu'elles entretiennent ? Comment permettre aux élèves de s'approprier une pensée statistique ? C'est dans cette analyse réflexive que nous invitons le lecteur à nous suivre avant de découvrir une proposition d'enseignement pour la Seconde.

1. — Les objectifs du programme de Seconde

Observons ce que propose le programme de Seconde (2009) pour l'enseignement de la statistique et des probabilités.

1. 1. *Pour la statistique descriptive*

> *L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées, pour synthétiser l'information et proposer des représentations pertinentes.*

Le programme précise que les capacités attendues sont :

— *Utiliser un logiciel (par exemple, un tableur) ou une calculatrice pour étudier une série statistique.*

— *Passer des effectifs aux fréquences, calculer les caractéristiques d'une série définie par effectifs ou fréquences.*

— *Calculer des effectifs cumulés, des fréquences cumulées.*

— *Représenter une série statistique graphiquement (nuage de points, histogramme, courbe des fréquences cumulées).*

On constate que la plupart des notions de statistique descriptive ont été abordées en classe de Troisième. Le programme de Seconde propose un enrichissement des techniques.

1. 2. *Pour l'échantillonnage*

> *L'objectif est d'amener les élèves à un questionnement lors des activités suivantes :*

— l'estimation d'une proportion inconnue à partir d'un échantillon ;

— la prise de décision à partir d'un échantillon.

et les capacités attendues

— *Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice.*

— *Exploiter et faire une analyse critique d'un résultat d'échantillonnage.*

1. 3. *Pour les probabilités*

Les capacités attendues sont :

— *Déterminer la probabilité d'événements dans des situations d'équiprobabilité.*

— *Utiliser des modèles définis à partir de fréquences observées.*

Le plus important restant l'objectif général défini par le programme (2009) :

L'objectif de ce programme est de former les élèves à la démarche scientifique sous toutes ses formes pour les rendre capables de :

— modéliser et s'engager dans une activité de recherche ;

— conduire un raisonnement, une démonstration ;

— pratiquer une activité expérimentale ou algorithmique ;

— faire une analyse critique d'un résultat, d'une démarche ;

— pratiquer une lecture active de l'information (critique, traitement), en privilégiant les changements de registre (graphique, numérique, algébrique, géométrique) ;

— utiliser les outils logiciels (ordinateur ou calculatrice) adaptés à la résolution d'un problème ;

— communiquer à l'écrit et à l'oral.

Dans la mesure du possible, les problèmes posés s'inspirent de situations liées à la vie courante ou à d'autres disciplines. Ils doivent pouvoir s'exprimer de façon simple et concise et laisser dans leur résolution une place à l'autonomie et à l'initiative des élèves. Au niveau d'une classe de seconde de détermination, les solutions attendues sont aussi en général simples et courtes.

L'enseignement de la statistique et des probabilités occupent une place importante au sein de ce programme, les objectifs sont ambitieux. En comparant avec les programmes des années antérieures, on constate que la place de ces enseignements augmente et l'analyse des programmes de Premières et Terminales montre que l'initiation à la statistique inférentielle absente jusqu'en 2000 semble prendre une place importante dans l'ensemble des cursus au lycée, y compris en série scientifique. Il est donc nécessaire que le travail fait en classe de Seconde construise des bases solides dans ce domaine ; en l'absence d'un travail approfondi en classe de Seconde, les enseignements de premières et Terminales sont compromis.

2. — Les raisons d'être de l'initiation à la statistique inférentielle.

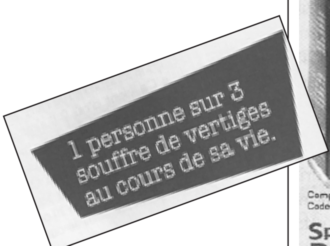
Pourquoi donner une telle place à cet enseignement, pourquoi aborder aujourd'hui en lycée la statistique inférentielle ? Les réponses à ces questions se situent à deux niveaux : d'une part dans les pratiques sociales modernes et d'autre part dans l'épistémologie du concept que l'on peut dégager d'une approche historique.

Aujourd'hui, la façon d'appréhender le monde n'est plus exclusivement déterministe.

Les statistiques produites sont presque toujours le résultat d'inférences à partir d'échantillons et ne sont que très exceptionnellement la description exhaustive d'une population. Cet état de fait légitime l'introduction de l'enseignement de l'échantillonnage et de la statistique inférentielle. Cherchons un peu plus finement à pointer des raisons d'être de cet enseignement au niveau secondaire relativement à la formation du citoyen, aux besoins des techniciens et enfin celles légitimées par les besoins des scientifiques de demain.

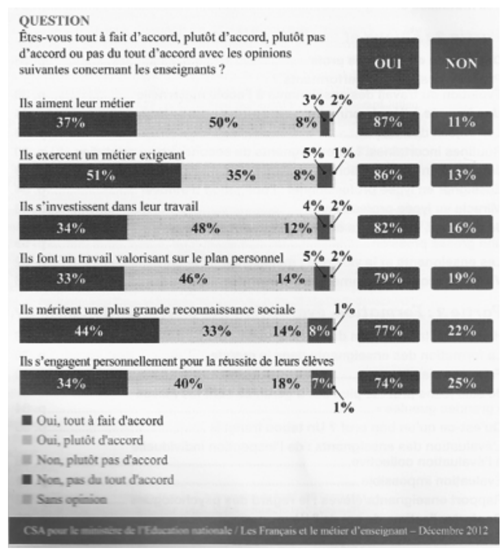
2. 1.[1] *Les besoins du citoyen*

Dans les médias foisonnent des informations exprimées dans un registre relevant de la statistique ; il semble clair que tous les citoyens ont besoin de décrypter ces informations. Mais les besoins dépassent ce simple cadre. En effet, tout citoyen a besoin :

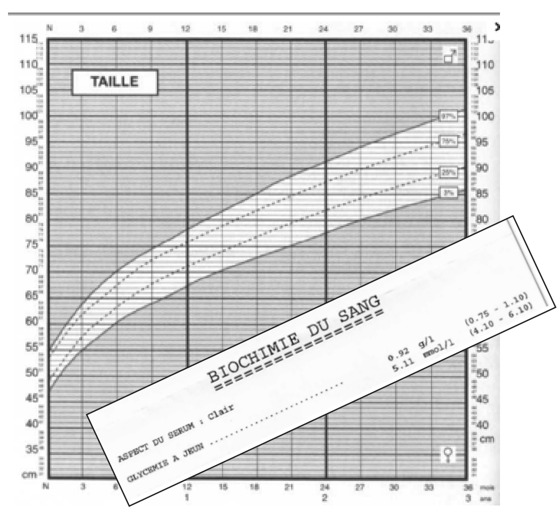


— *de comprendre une quelconque information résultant d'une étude statistique ;*

— *de comprendre l'information statistique issue d'un sondage*[2]*;*



— *de savoir interpréter les plages de normalité sur les analyses médicales, les carnets de santé, ... ;*



1 Avantages n°238 juillet 2008 ; page 96

2 Le nouvel Observateur Société – Hors-série : être enseignant aujourd'hui – janvier 2013 ; page 4

— *d'être capable d'évaluer les probabilités et les espérances de gain dans les jeux de hasard ; (cf. annexe 2)*

— *de relativiser l'exploitation médiatique de certains événements exceptionnels ;*

— *de comprendre pourquoi le recensement est maintenant pratiqué par sondage, et pourquoi le recensement exhaustif peut être moins efficace.*

Cette liste, loin d'être exhaustive, illustre l'importance des besoins dans ce domaine pour tout citoyen.

2. 2. *Les besoins des professionnels des médias*

L'information est relayée, traduite et interprétée par les journalistes de la presse spécialisée ou de la presse grand public. Tout journaliste se doit de maîtriser l'information statistique pour en rendre compte. Sur ce point la culture anglo-saxonne est très en avance sur la France, comme le montre Floriane Wozniak (2005) dans sa thèse.

2. 3. *Les besoins des techniciens*

L'enseignement de la statistique répond aussi à des besoins plus spécifiques de certains techniciens

— *Utilisations de cartes de contrôle dans l'industrie ;*

— *Prise en compte des incertitudes de mesure ;*

— *Etude de risque dans l'industrie, l'économie, la finance…*

— …

2. 4. *Les besoins des chercheurs du 21ème siècle*

Des connaissances en statistique sont nécessaires dans bien des métiers des sciences humaines, où la plupart des études reposent sur des données statistiques de type sondage, comme la psychologie, la sociologie et la démographie, entre autres. Elles sont aussi nécessaires aux biologistes pour l'estimation de populations animales, bactériennes, etc. De même, les études relatives aux effets des médicaments reposent en grande partie sur l'analyse de données statistiques.

La science du 20ème siècle a quitté progressivement sa quête d'un déterminisme exclusif et introduit de l'aléatoire dans ses modélisations avec notamment la naissance de deux grandes théories : la théorie du chaos et la mécanique quantique. Des phénomènes traités jusqu'ici de façon déterministe, sont interprétés en termes d'aléatoire ; cette modélisation permet de réduire les quantités de calculs nécessaires et limite les erreurs induites par ces calculs automatisés.

3. — Eclairage historique

Concernant la notion de probabilité, deux approches se juxtaposent : l'*approche clas-*

sique initiée par Pascal et Fermat en 1694 et l'*approche fréquentiste* proposée par J. Bernoulli en 1713. De ces deux approches, émergent deux tentatives de définition.

3. 1. *L'approche classique*

L'approche classique s'intéresse essentiellement aux jeux de hasard, qui ont été « conçus pour garantir l'équité » entre les joueurs. Elle s'exprime dans le langage des chances, le terme de « probabilité » reste longtemps absent. Une définition émerge de ces travaux : la définition, dite classique, que Laplace adopte comme premier principe dans son *essai philosophique sur les probabilités* en 1812. « *La probabilité est le rapport du nombre des cas favorables à celui de tous les cas possibles »* et qu'il complète dans son second principe *« mais cela suppose tous les divers cas également possibles… »*

Cette définition pose plusieurs problèmes

— *Le sens du nombre calculé est absent.* *Dans cette approche syntaxique, on dispose d'une technique de calcul qui ne construit pas de sens, si ce n'est une notion de proportion.*
— *Elle repose sur un postulat d'équiprobabilité, assumé par Laplace qui propose de choisir l'équiprobabilité en l'absence d'informations suffisantes. Cette définition est donc circulaire puisqu'elle repose sur la reconnaissance de l'équiprobabilité qui n'est autre qu'une probabilité, ce que l'on se proposait de définir.*

3. 2. *L'approche fréquentiste*

L'approche précédente, dite classique, ne s'applique guère que pour l'étude des jeux de hasard qui ont été conçus pour garantir l'équité entre les joueurs, c'est ce que souligne Bernoulli (1713, p.40), qui affirme que l'on ne pourra pas déterminer de cette façon l'évolution d'une épidémie ou les prévisions météorologiques. Puisqu'on ne peut l'obtenir par un calcul *a priori*, on l'estimera *a posteriori*. Il établit un théorème, maintenant appelé théorème de Bernoulli, qui n'est autre que la loi faible des grands nombres, dans le cas particulier de variables de Bernoulli.

Cette approche suggère une définition, dite fréquentiste, que Renyi (1966) cité par Henry (2001, p.81) énonce ainsi : « *Nous appellerons probabilité d'un événement le nombre autour duquel oscille la fréquence relative de l'événement considéré.* »

Si cette définition permet une meilleure approche sémantique, elle pose cependant de nouvelles difficultés :

— *La première est une confusion entre réalité et théorie. En effet la convergence presque sûre des fréquences est établie à l'intérieur d'un modèle, mais on ne peut pas dire, comme on l'entend souvent, que les fréquences expérimentales convergent vers cette valeur idéale produite par le modèle.*
— *La seconde tient à la formalisation de cette approche. Elle ne peut être donnée que par le théorème de Bernoulli : on obtient à nouveau une définition circulaire puisque ce théorème repose sur une convergence en probabilité.*

La synthèse de ce double visage de la probabilité est réalisée par Kolmogorov en 1933. Le concept de probabilité est défini par une axiomatique où la probabilité apparait comme une mesure positive de masse totale 1. Mais Kolmogorov, cité par Bordier(1991), dit aussi : *« La valeur épistémologique de la théorie des probabilités est basée sur ceci : dans leur action collective les phénomènes aléatoires, à large échelle créent des régularités strictes et non aléatoires. Le* *concept même de la probabilité mathématique serait sans utilité, s'il ne trouvait sa concrétisation dans la fréquence d'arrivée d'événements, suite à des expériences nombreuses réalisées dans des conditions uniformes* ».

Cette définition est trop abstraite et trop complexe pour être envisagée dans le secondaire. Comment permettre aux élèves du secondaire de faire la synthèse entre ces deux conceptions de la probabilité ? Michel Henry (2001) propose pour l'enseignement secondaire d'utiliser la modélisation comme outil didactique pour réaliser cette unification. C'est le point de vue que nous adopterons.

3. 3. *La modélisation*

Comment permettre une synthèse entre la vision de la probabilité comme rapport du nombre des cas favorables sur le nombre de cas possibles (l'équiprobabilité étant postulée) et la valeur limite des fréquences ? Michel Henry propose la modélisation pour réaliser la synthèse entre ces deux conceptions de la probabilité : ainsi, dans la conception laplacienne, l'équiprobabilité ne sera plus un postulat mais une hypothèse de modèle et dans l'approche fréquentiste, la fréquence observée sur un échantillon, si possible de grande taille, pourra servir d'hypothèse de modèle.

Pour Michel Henry (2001, p.149 à 159), la modélisation est un processus qui permet un passage progressif d'une réalité complexe, changeante que l'on souhaite étudier, à un univers mathématique théorique stable qui permet d'en étudier certains aspects.

Un objet ne possède pas en soi une probabilité, la probabilité est liée à une expérience, dont on dira qu'elle est aléatoire, parce qu'on ne peut pas établir de manière déterministe le résultat qu'elle va produire. L'étude repose donc sur la caractérisation de cette expérience.

Cette caractérisation doit assurer la reproductibilité de l'expérience, c'est pourquoi le processus de modélisation commence par l'explicitation d'un protocole expérimental qui décrit les conditions de réalisation de l'expérience. Elle explicite aussi ce que l'on va retenir ou pas d'observer : les *issues observables (*pour le lancer d'un dé, cela conduit à exclure le dé « cassé » des observations..). C'est une première simplification de la réalité.

Des hypothèses de travail sont exprimées dans le langage de l'expérience (qui, pour le dé amèneront à considérer que le dé est équilibré). Cette description qui idéalise l'objet, transforme le dé réel en un objet mathématique possédant des propriétés propres qui seront traduites en langage mathématique par l'explicitation du modèle mathématique retenu (pour le dé, la loi uniforme discrète sur l'ensemble des issues possibles).

Le travail dans le modèle choisi produira des résultats qu'il conviendra d'interpréter dans la situation réelle, cette interprétation pouvant conduire à rejeter le choix du modèle et à en définir un meilleur.

Ce travail sur la modélisation n'est pas spécifique à l'étude des probabilités, mais le champ des probabilités est un lieu privilégié pour faire vivre un processus de modélisation. Cette modélisation va permettre à l'élève de percevoir le caractère théorique d'une probabilité. Cette étape, difficile pour les élèves et délicate pour l'enseignant, est fondamentale.

Plus généralement, le travail sur la modélisation est étroitement lié au travail sur un problème concret. Le caractère concret du problème nécessite une phase de modélisation. Ce travail rarement pris en charge dans les classes de mathématiques est pourtant un attendu des programmes, comme le précisent les objectifs généraux du programme de Seconde :

> *L'objectif de ce programme est de former les élèves à la démarche scientifique sous toutes ses formes pour les rendre capables de :*
>
> — *Modéliser et s'engager dans une activité de recherche ;* [...]
>
> *Dans la mesure du possible, les problèmes posés s'inspirent de situations liées à la vie courante ou à d'autres disciplines.*

Quelles conclusions tirer de cette approche historique ?

— *Les jeux de hasard motivent l'étude de la notion de probabilité.*

— *L'estimation statistique ne nécessite pas de connaître la théorie des probabilités, ni même la définition d'une probabilité, une approche de la* ***notion*** *de probabilité est suffisante.*

— *La statistique inférentielle est intrinsèquement liée à la notion de probabilité dont elle éclaire le sens, ces deux thèmes gagneront à être menés conjointement.*

— *La modélisation doit apparaître comme processus dans les études proposées.*

4. — Réflexion sur les programmes de Seconde

Que retiennent les programmes comme capacités attendues sur le thème de la statistique inférentielle ?

4. 1. *Types de tâches visés et techniques*

Remarquons qu'un genre de tâches, la modélisation (au sens où nous venons de la définir) traverse ce travail sans lui être spécifique, mais revêt une importance didactique que nous avons soulignée.

On peut identifier certains types de tâches fondamentaux :

— Comparer des statistiques issues de divers échantillons aléatoires
  - *par un résumé statistique pertinent*
  - *par une représentation graphique adaptée*

— Produire des échantillons aléatoires par simulation

Ce type de tâche de la classe de Seconde a vocation à devenir une technique pour l'étude de modèles probabilistes et pour la résolution de problèmes relevant de l'aléatoire dans les classes ultérieures.

— Prendre une décision en situation de risque

On observe la fréquence f du caractère étudié sur un échantillon de taille n d'une population. L'affirmation : « la proportion du caractère dans la population est p » doit-elle être rejetée ou pas ?

On note IF un intervalle de fluctuation des fréquences au seuil de 95%. Une règle de décision est énoncée : Si la fréquence observée f, sur l'échantillon de taille n n'appartient pas à IF, alors on rejette l'affirmation. Sinon, n'ayant pas d'argument suffisant pour la rejeter, on garde p comme hypothèse de modèle.

L'IF prendra des formes différentes selon que l'on est en Seconde, Première ou Terminale. On utilisera en Seconde uniquement un seuil de 95%, cette valeur sera privilégiée au lycée, sans être exclusive.

— Interpréter des informations statistiques

La lecture de données statistiques n'est pertinente que si elle est éclairée par une pensée statistique où la fluctuation d'échantillonnage est mobilisée. En particulier, il convient d'être vigilant sur la taille des échantillons et sur les protocoles expérimentaux.

Un dernier type de tâche doit être envisagé tout en n'étant pas exigible en classe de seconde :

— Réaliser une estimation de p
  - par l'observation d'une fréquence *:* toute valeur de f est une estimation de p, plus la taille de l'échantillon est grande, plus l'estimation est pertinente.
  - par la donnée d'un intervalle de confiance, ce qui permet d'associer une précision à cette mesure : F la variable aléatoire qui associe à chaque échantillon de taille n, la fréquence du caractère étudié, alors au niveau de confiance de 95 % :

$$p \in \left[F - \frac{1}{\sqrt{n}}\ ;\ F + \frac{1}{\sqrt{n}}\right]$$

De façon plus réaliste pour la classe de Seconde par l'observation d'un intervalle de confiance (fourchette de sondage) : une valeur f de F étant observée sur échantillon de taille n, alors au niveau de confiance de 95 % on peut affirmer que

$$p \in \left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right].$$

Ce type de tâche n'est pas exigible en classe de Seconde mais le sera en Terminale. On remarquera que les théories mathématiques qui justifient des savoir-faire visés par ces études ne peuvent être abordées à ce niveau d'enseignement. Il s'agit de proposer aux élèves une approche heuristique, leur permettant de se familiariser avec l'aléatoire, notamment par des simulations informatiques, et de se construire des représentations efficaces pour traiter des problèmes relevant de la statistique inférentielle.

4. 2. *Difficultés et obstacles*

4. 2. 1. La définition de l'IF
proposée par le programme de Seconde

Le programme de Seconde (2009, page 8) propose la définition suivante :

> *L'intervalle de fluctuation au seuil de 95 % relatif aux échantillons de taille n est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation.*

Cet énoncé n'a de sens que si la fréquence évoquée est une variable aléatoire qui peut prendre aléatoirement, à l'issue de la réalisation d'un échantillon de taille n, une des valeurs

$$0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n-1}{n}, 1.$$

Il ne s'agit donc pas de la *fréquence observée*, dont la valeur est connue. Sans parler de variable aléatoire [3], avant de réaliser un échantillon de taille, on peut suggérer que l'on ne peut pas savoir quelle fréquence du caractère on observera dans l'échantillon en parlant par exemple de *fréquence observable* plutôt que de *fréquence observée*. Dans le programme de Terminale, la variable aléatoire fréquence du caractère dans

3 La notion de variable aléatoire n'est pas au programme de Seconde

un échantillon de taille n est notée F_n et f_n désigne la fréquence observée.

Il est énoncé que « *la fréquence* [...], *se situe, avec une probabilité égale à 0,95, dans un intervalle centré en p* ». Mais dans quel espace probabilisé cette probabilité égale à 0,95 est-elle calculée ? Si l'on considère que l'univers est l'ensemble des échantillons de taille n, qui sont au nombre de 2^n, à chaque échantillon est attachée la probabilité $p^k(1-p)^{n-k}$, où k désigne le nombre d'occurrences de la modalité dans l'échantillon et p désigne à la fois la proportion de la modalité dans la population et la probabilité qu'il apparaisse lors d'un tirage.

L'évènement « *dans l'échantillon, la fréquence du caractère est* $\frac{k}{n}$ » a pour probabilité la somme des probabilités des évènements élémentaires qui le constituent c'est-à-dire

$$\binom{n}{k} p^k(1-p)^{n-k}.$$

Ces justifications ne sont accessibles qu'en classe de Première, dans le cadre de la loi binomiale. En Seconde, il n'est pas raisonnable d'énoncer une définition dans laquelle se chevauchent implicitement deux probabilités définies dans deux espaces différents, probabilité d'apparition du caractère dans la réalisation d'un tirage, et probabilité que la fréquence prenne des valeurs appartenant à un certain intervalle, alors que l'un des objectifs essentiel est de construire la notion de probabilité.

La définition appelle plusieurs questions : cet intervalle existe-t-il ? Comment le déterminer ? Pour une variable aléatoire discrète, un intervalle I, centré en p tel que la probabilité que f appartienne à I soit *strictement égale* à 0,95, n'existe généralement pas. Mais le programme suggère de travailler avec des « intervalles approchés », ce qui nous amène à deux autres questions : qu'est-ce qu'un intervalle de fluctuation approché ? Comment le déterminer ?

Le programme envisage deux possibilités. La première est d'obtenir un intervalle approché par simulation. On proposera des situations où la simulation ne pose pas de difficultés excessives. Le programme offre une deuxième possibilité :

Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n \geq 25$ *et des proportions p du caractère comprises entre* 0,2 *et* 0,8 *:*

Si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}\right]$ *avec une probabilité d'au moins 0,95.*

Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais elle n'est pas exigible.

Ce résultat n'est pas exigible mais aucun travail de l'organisation mathématique relative aux intervalles de fluctuation et à la prise de décision n'est possible sans l'énoncé de cette propriété. Dans *non exigible*, sans doute faut-il comprendre qu'elle n'a pas à être connue par cœur et qu'elle doit être rappelée dans tout travail qui s'y rapporte.

De plus la propriété n'est pas toujours vérifiée : on trouve dans le document ressource de la classe de Terminale (2011, p. 27) des précisions à ce sujet ; lorsque $p = 0{,}5$, la variance est maximale et $P(0{,}5 - \frac{1}{\sqrt{n}} \leq f < 0{,}5 + \frac{1}{\sqrt{n}})$ est réalisé pour tout $n \geq 529$.

La « définition » de l'intervalle de fluctuation reposant sur l'unicité suggérée (***L****'intervalle de fluctuation*…) et l'égalité (*avec une probabilité* ***égale*** *à 0,95)* est une définition valide dans le cadre d'une loi continue. En classe de Seconde le travail repose sur des lois discrètes et donc sur un intervalle de fluctuation approché. Une formulation évoquant *une probabilité voisine de 0,95* serait préférable.

On peut comparer (encadré ci-desous) les présentations envisagées dans les différentes classes pour la notion d'intervalle de fluctuation au seuil 95 % en Seconde, Première, Terminale.

Remarque : seul le programme de Terminale S (2011) envisage des seuils autres que 95 %.

L'intervalle de fluctuation est une création didactique des nouveaux programmes de lycée (à compter de 2009). Lui donner un nom permet de l'identifier et de le distinguer de l'intervalle de confiance, avec lequel il ne doit pas être confondu. Les définitions successives, les incohérences (dont certaines sont évoquées dans les documents ressources de Première (2010) et de Terminale(2011)), risquent d'en faire, pour les professeurs, un objet difficile à enseigner et pour les élèves, un objet difficile à appréhender.

4. 2. 2. Proportion ou probabilité

On remarquera que l'intervalle de fluctuation dans le programme de Seconde fait référence à une proportion et non à une probabilité. Ce choix révèle une référence implicite à la théorie des sondages. On peut regretter ce choix :

— Ce choix est limitatif puisqu'il exclut la possibilité d'établir un lien entre l'observation des fréquences dans le cadre de jeux de hasard simples (dés, pile ou face, ...) et les probabilités des éventualités observées. Ce lien est pourtant essentiel dans la construction de la notion de probabilité.

— Ce choix crée des difficultés, puisqu'il va falloir admettre que les tirages sont toujours assimilables à des tirages avec remise. Bien sûr nous dirons que la taille de l'échantillon étant très petite par rapport à la population, la probabilité de chaque tirage reste sensiblement la même, mais nous savons que ce type » de discours, lorsqu'il n'est pas produit par l'élève lui-même a peu de chance d'être réellement assimilé. Des

	Conditions de validité	Propriétés de l'intervalle	Formulation
2nde	$n \geq 25$, $0{,}2 \leq p \leq 0{,}8$	approché, centré en p	$[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}]$
1ère	aucune	non centré	$[\frac{a}{n}, \frac{b}{n}]$ où $\begin{cases} a \text{ est le plus grand entier tel que } P(F < \frac{a}{n}) \leq 0{,}025 \\ b \text{ est le plus petit entier tel que } P(F > \frac{b}{n}) \leq 0{,}025 \end{cases}$
Terminale	$n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$: plus larges qu'en 2°	approché, asymptotique, centré en p	$[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}]$

activités permettant cela sont possibles en classe de Seconde mais le l'horaire de mathématique en Seconde ne le permet pas, sans compromettre des apprentissages plus essentiels.

— Ce choix risque de renforcer un obstacle didactique : l'assimilation entre probabilité et proportion ou fréquence.

C'est pourquoi nous proposerons comme définition de l'intervalle de fluctuation en Seconde :

Dans une population, la proportion du caractère étudié est p comprise entre 0,2 et 0,8. On réalise un échantillon aléatoire de taille $n \geq 25$ de la population, la fréquence du caractère observée dans l'échantillon est f. Si l'on considère tous les échantillons de taille n possibles, pour environ 95 % d'entre eux, la fréquence f appartient à l'intervalle

$$\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right].$$

Cet intervalle est appelé intervalle de fluctuation de la fréquence observée au seuil de 95 %.

4\. 2. 3. L'absence des diagrammes en boîtes

Les paramètres permettant la réalisation d'un diagramme en boîte sont connus dès le collège. Un diagramme en boîte permet de représenter très simplement, en situant quelques valeurs sur un axe, le résumé d'une série par le couple (médiane, écart interquartile). Ce diagramme focalise l'attention sur les valeurs les plus centrales de la série, en proportion d'au moins 50 %. Il fait apparaître un paramètre de position (la médiane) et deux paramètres de dispersion (l'étendue avec les moustaches et l'écart interquartile).

L'utilisation d'un seul axe permet de représenter plusieurs diagrammes en boîtes au dessus les uns des autres et facilite considérablement la comparaison de différentes séries. On peut regretter que l'introduction du diagramme en boîte n'apparaisse explicitement dans les programmes qu'en classe de Première.

4\. 3. *Le vocabulaire : « seuil et niveau »*

Auparavant, ces notions étaient abordées pour la première fois en BTS et il était d'usage d'appeler *seuil* ce que l'on ne peut pas dépasser, et *niveau* ce que l'on veut au moins atteindre. C'est cette convention qui est la plus usitée dans l'enseignement supérieur.

Pour un intervalle I, si $P(F \in I) \geq 0{,}95$, I est un intervalle de fluctuation au *niveau* 0,95 ou 95 %. Dans ce cas, $P(F \notin I) \leq 0{,}05$, le risque que F n'appartienne pas à l'intervalle ne dépasse pas le *seuil* de 0,05 ou de 5 % .

Plus généralement, soit α un réel de $]0, 1[$. $P(F \notin I) < \alpha$ exprime que le risque que n'appartienne pas à l'intervalle ne dépasse pas le *seuil* α et $P(F \in I) \geq 1 - \alpha$ exprime que la probabilité que F appartienne à I atteint le *niveau* $1 - \alpha$. De même, on veut pouvoir déterminer un intervalle de confiance avec un niveau de confiance au moins égal à, par exemple, 95 %.

Dans les textes du programme, un autre choix semble avoir été fait : celui de réserver le mot *seuil* à intervalle de fluctuation et le mot *niveau* à intervalle de confiance. Dans tout ce qui suit nous respecterons ce choix.

4\. 4. *Prise de décision à l'aide d'un test bilatéral*

L'intervalle de fluctuation utilisé en Seconde étant centré sur p, le rejet se fera sur l'observation d'une valeur de f trop petite ou trop grande pour appartenir à IF. Certains problèmes de prise de décision ne peuvent donc pas être traités. On peut regretter l'exemple fourni dans le document ressources de Seconde *Statistiques et Probabilités* (2009, page 17) où l'on s'interroge : « ... *sur les cas de leucémie chez les garçons de moins de quinze ans dans une petite ville des Etats-Unis* ». La question réelle étant doit-on s'inquiéter, il n'est d'aucune pertinence de rejeter une fréquence qui n'appartiendrait pas à IF parce qu'elle serait trop petite. Le document ressource de Première (2010) reprend cet exemple pour en montrer un traitement raisonnable basé sur une prise de décision à l'aide d'un intervalle unilatéral. On pourra approfondir cette question en se référant aussi à l'article de Ducel(2011).

4. 5. *Le hasard*

Pour les élèves du secondaire, le calcul des probabilités est intrinsèquement lié à l'étude de situations aléatoires, donc d'un point de vue purement didactique, à ce niveau d'enseignement : « *il n'y a pas de probabilité sans hasard* » comme le dit souvent Michel Henry. Le hasard est un concept qui n'est pas propre au champ mathématique. L'usage commun ne risque-t-il pas de se constituer en obstacle ? C'est ce que Brigitte Chaput et Claudine Vergne (2010, p. 6) précisent :

> *La notion de probabilité est indissociable de celle de hasard, car elle a pour objet de quantifier l'attente d'un évènement dont la réalisation est considérée comme dépendante du hasard. À leur arrivée en Troisième, les élèves ont déjà été confrontés à plusieurs conceptions du hasard dans leur vie quotidienne. Il convient, pour cette première rencontre avec les phénomènes aléatoires dans le cadre des mathématiques, de clarifier ceux qui, en mathématiques, permettront d'introduire des probabilités.* [...]
>
> *L'élève arrive avec des conceptions personnelles relatives au hasard et au tirage au sort, qu'il s'est forgées au quotidien dans son entourage, sous l'influence des médias, etc. Il est important de faire émerger ces conceptions de façon à lever les ambiguïtés, les malentendus qui pourraient faire obstacle à la compréhension de l'approche mathématique de la notion de probabilité.*

Dans la vie courante, le mot hasard est associé à la chance, à l'intervention divine, au destin, mais aussi aux coïncidences fortuites, le plus souvent le hasard est évoqué dans l'observation d'événements exceptionnels.

En classe de mathématique, le hasard étudié est un hasard construit, inhérent à un modèle mathématique. L'écart est important entre ces conceptions du hasard et la rationalité que les programmes proposent d'étudier. Ces conceptions culturellement répandues peuvent constituer des obstacles pour les apprentissages envisagés par les programmes.

Il semble nécessaire de clarifier dans nos classes, quel est le hasard dont l'étude est visée au lycée.

Cileda Couthiño (1999), propose d'amener les élèves à distinguer le hasard de contingence du hasard relevant de « situations potentiellement reproductibles », c'est-à-dire de situations modélisables. Sur le recensement de situations comportant « du hasard », les élèves évoqueront des situations de jeux de hasard, loto, machines à sous... Ces situations sont reconnues pour être reproductibles. Des situations issues de la vie quotidienne apparaitront aussi : sexe d'un bébé à la naissance, prévisions météorologiques, générateurs de musiques aléatoires, désignation de l'élève que le professeur va interroger... Le sexe des bébés est assez facilement classé dans les situations reproductibles par la référence à la quasi équiprobabilité des genres à la naissance, la désignation d'un élève sera rejetée des situations relevant du hasard, le professeur fait son choix en fonction de la connaissance des élèves et de ceux qu'il sait avoir déjà interrogés. Les phénomènes météorologiques sont étudiés comme des phénomènes aléatoires et modélisés, mais les modèles utilisés sont encore peu fiables. D'autres situations seront évoquées relevant de la coïncidence comme par exemple « je rencontre un copain dans la rue », cette dernière situation relève clairement de la contingence, en particulier parce que l'utilisation du pronom « je » singularise la situation. Cette première phase permet de distinguer le hasard modélisable du hasard contingent.

Nous proposerons en Seconde de classer des situations qui évoquent le hasard, afin de bien identifier celui qui sera l'objet de l'étude au lycée.

4. 6. *Générateurs aléatoires*

La génération de nombres aléatoires peut être réalisée avec deux fonctions, quel que soit le support utilisé, calculatrice, tableur, langage de programmation... Pour le tableur de Excel, cela sera soit avec la fonction ALEA, soit avec la fonction ALEA.ENTRE.BORNES ; la fonction ALEA fournit un nombre aléatoire de l'intervalle [0 ; 1[, elle modélise une loi continue, la loi uniforme, même si elle reste en toute rigueur une loi discrète portant sur les 10^n décimaux (comportant n décimales compris dans [0 ; 1[, 10^{14} pour le tableur Excel). La fonction ALEA.ENTRE.BORNES(n ; p) fournit un nombre « aléatoire » entier N, tel que $n \leq N \leq p$; elle modélise une loi discrète, la loi équirépartie sur $\{n ; n+1 ; ... ; p\}$.

Comparons les techniques et leurs justifications à mettre en œuvre pour ces deux fonctions sur deux exemples classiques : le lancer d'un dé, puis le tirage dans une urne contenant *une* boule rouge et *quatre* boules blanches. Pour le dé :

`=ENT(ALEA()*6+1)` ALEA() nécessite l'utilisation de la partie entière qui est au programme de Terminale et un travail sur des inégalités pour obtenir un nombre entier de l'intervalle [1 ; 7[.

`=ALEA.ENTRE.BORNES(6;1)` est plus simple.

Regardons pour le tirage dans une urne contenant 1 boule rouge et 4 boules blanches

`=SI(ALEA()<0,2;1;0` ALEA nécessite de se référer à la probabilité d'un intervalle dans le cadre de la loi uniforme : c'est le programme de Terminale S ou ES. Elle nécessite le recours à un test avec la fonction SI. On interprètera le 1 comme le tirage d'une boule rouge et le 0 comme celui d'une boule blanche.

`=ALEA.ENTRE.BORNES(1;5)` On interprétera le 1 comme le tirage d'une boule rouge et les autres chiffres comme celui d'une boule blanche.

Dans le cadre de la fonction ALEA(), la justification repose toujours sur la loi uniforme, pour ALEA.ENTRE.BORNES, la justification repose sur une loi discrète équirépartie. Nous utiliserons pour les classes de Secondes ALEA.ENTRE.BORNES.

Les nombres produits par cette fonction sont des nombres pseudo-aléatoires, produits par des algorithmes ayant des périodes extrêmement longues et indécelables à notre échelle. Bien qu'issues d'un processus déterministe, ces suites de nombres sont compatibles avec les modèles probabilistes qu'elles sont censées représenter. La première utilisation d'un générateur aléatoire sur des calculatrices neuves ou réinitialisées et du même modèle fournit des résultats identiques ! Le caractère pseudo-aléatoire ne peut être occulté au risque de créer des malentendus préjudiciables aux apprentissages visés. Nous pensons qu'il est important d'expliciter le caractère pseudo-aléatoire des nombres obtenus par de tels générateurs.

Sur tableur, l'utilisation de la touche de recalcul des formules F9 permet d'obtenir une nouvelle valeur aléatoire indépendante de la précédente. La fonction

ALEA.ENTRE.BORNES(n;p)

peut être représentée par le tirage dans une urne opaque contenant des boules indiscernables au toucher numérotées de n à p. L'utilisation de la touche F9 correspond alors à un tirage avec remise. Cette urne constitue un registre de représentation sémiotique qui favorise une construction sémantique de la fonction ALEA.ENTRE.BORNES. Cette urne constitue aussi un modèle pseudo-concret pour l'expérience aléatoire que l'on simule. Il semble donc important de permettre d'accéder à cette représentation.

5. — Parcours d'enseignement pour la classe de Seconde

L'étude que nous proposons en classe de Seconde se déroule en quatre temps. Le premier temps permet d'identifier et de caractériser le hasard dont l'étude est visée, puis elle s'articule autour de trois activités d'étude et de recherche : « *le problème d'Adrien* » qui permet de comprendre le caractère de modèle d'une probabilité et d'établir le lien entre fréquences et probabilités ; « *la parité* » qui permettent d'aborder la prise de décision avec un intervalle de fluctuation ; « *les anniversaires* » qui permet d'aborder l'estimation par intervalle de confiance.

5. 1. *Quel hasard étudions-nous ?*

Il s'agit dans ce premier temps de laisser s'exprimer la diversité des conceptions des élèves et d'établir une classification des situations selon qu'elles sont modélisables ou non. Cela permet de définir le hasard qui sera l'objet d'étude en cours de mathématiques.

Voici (encadré ci-contre) quelques exemples de productions, illustrant la diversité des conceptions que mobilisent spontanément des élèves.

On pourra institutionnaliser que :

— Certaines situations mettant en jeu des coïncidences fortuites ne seront pas étudiées.

— Dans d'autres situations, le hasard en jeu peut être modélisé. Ce sont de tels modèles que nous étudierons.

5. 2. *Comment choisir un modèle ? Le problème d'Adrien*

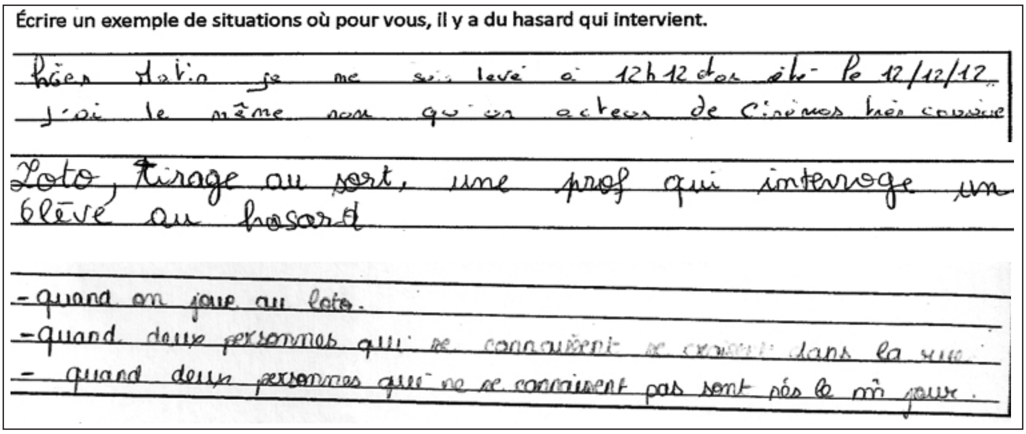
5. 2. 1. L'énoncé

Il s'agit d'une situation classique reposant sur la somme de deux dés en référence au problème du grand duc de Toscane.

Adrien est à la foire de la Saint Martin, une loterie lui propose de parier sur la somme de deux dés que le forain va lancer. Les deux dés sont dans un gobelet et vont être lancés dans un plateau de jeu. Si le nombre annoncé par Adrien correspond à la somme des deux faces des dés qui apparaissent alors il gagne un gros lot. Adrien doit payer pour participer à ce jeu, il souhaite faire un choix qui lui donne le plus de chance de gagner. Que conseillerais-tu à Adrien ?

5. 2. 2. Types de tâches mis en jeu

— *Choisir un modèle : sur quel univers faire porter l'équiprobabilité ?*


Écrire un exemple de situations où pour vous, il y a du hasard qui intervient.

hier matin je me suis levé à 12h12 ... le 12/12/12
j'ai le même nom qu'un acteur de cinéma très connu

Loto, tirage au sort, une prof qui interroge un élève au hasard

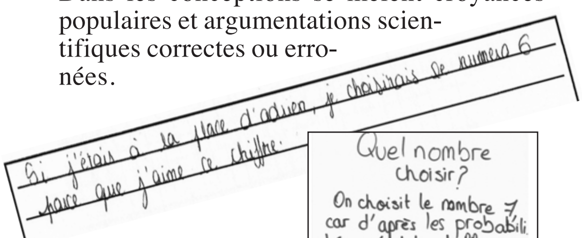
- quand on joue au loto.
- quand deux personnes qui ne connaissent se croisent dans la rue.
- quand deux personnes qui ne se connaissent pas sont nés le même jour.

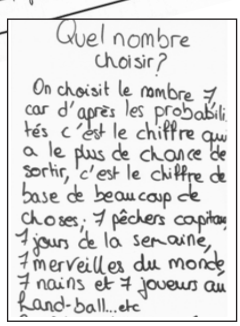
— *Simuler une situation aléatoire.*

— *Estimer une probabilité par la fréquence observée sur un échantillon de grande taille.*

5. 2. 3. Analyse de la situation

Une première phase de travail individuel permet l'expression des réponses spontanées. Cette phase permet d'amorcer le processus de dévolution et permet à l'élève de formuler ses conceptions du hasard. La diversité des conceptions initiales motive l'étude. Dans les conceptions se mêlent croyances populaires et argumentations scientifiques correctes ou erronées.


Si j'étais à la place d'adrien, je choisirais le numéro 6 parce que j'aime ce chiffre.


Quel nombre choisir?
On choisit le nombre 7 car d'après les probabilités c'est le chiffre qui a le plus de chance de sortir, c'est le chiffre de base de beaucoup de choses, 7 pêchers capitaux, 7 jours de la semaine, 7 merveilles du monde, 7 nains et 7 joueurs au hand-ball...etc

Trois modélisations doivent apparaître au cours de ce travail :

— *une faisant porter l'équiprobabilité sur les 11 sommes observables,*

— *une autre faisant porter l'équiprobabilité sur 21 façons d'obtenir ces sommes en ne différenciant pas les deux dés,*

— *la dernière faisant porter sur les 36 couples constituant l'univers.*

La situation proposée relate sur une expérience réalisée en classe et des simulations sur tableur. Elle va permettre d'articuler des raisonnements probabilistes et une approche fréquentiste des probabilités.

Elle doit permettre de mettre en jeu des registres de représentations sémiotiques variés :

— *distributions de fréquences sous formes de tableaux ou de graphiques statiques et dynamiques,*

— *tableaux à double entrée,*

— *liste de dénombrements,*

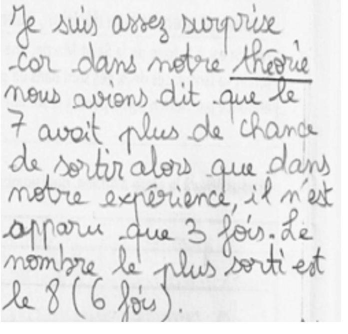
— *arbre de dénombrement*

— *langage naturel oral et écrit*

— *langage mathématique.*

Une large place au langage naturel est prévue : oral dans les différents travaux de groupes et écrit au travers des différents comptes-rendus à fournir.

Ce travail met en débat différents « candidats-modèles » selon les termes de Parzyzs (2007). Il doit permettre aux élèves de percevoir le caractère de modèle d'une loi de probabilité.


Je suis assez surprise
car dans notre théorie
nous avions dit que le
7 avait plus de chance
de sortir alors que dans
notre expérience, il n'est
apparu que 3 fois. Le
nombre le plus sorti est
le 8 (6 fois).

Le lien entre statistique et probabilités ne va pas de soi comme en témoigne la production ci-dessus d'une « bonne élève » au cours de ce travail :

Des entretiens menés en fin de parcours montrent que les avancées sur ce point sont significatives :

> Élève 1 : « *La probabilité c'est pour une infinité de choses c'est une formule précise, tandis que la fréquence c'est sur un exemple précis.* »

> Élève 2 : « *La probabilité c'est un calcul qui appartient à un monde idéal un monde qui n'existe pas, ce n'est pas la réalité, […] alors que la fréquence c'est sur un nombre fini, c'est la réalité, c'est ce qui se passe en vrai.* »

5. 2. 4. Indications sur l'organisation de l'activité d'étude et de recherche

Pour le déroulement de l'activité, des dés sont à disposition des élèves. Ils peuvent venir les chercher à tout moment du travail.

La première phase doit permettre à chaque élève de s'approprier la situation et d'exprimer ses représentations.

Une phase d'action en travail de groupe amène les élèves à confronter ces conceptions et à développer une argumentation pour en rejeter certaines et/ou en légitimer d'autres. La mise en commun fait apparaître des argumentations de type statistique obtenues par le lancer de dés, des argumentations probabilistes reposant sur le dénombrement des issues possibles. Deux modèles émergent à ce stade du travail : un modèle reposant sur 21 issues équiprobables, un autre sur 36 issues équiprobables. Mais les arguments ne permettent pas à la classe de décider entre ces deux modèles. Les élèves pensent que l'expérience doit permettre de décider.

Le protocole expérimental est décrit, les dés sont considérés comme équilibrés.

Lors de la seconde phase, les élèves recueillent des données statistiques par la réalisation de l'expérience et les résument. Le nombre de répétitions de l'expérience est de la responsabilité de chaque groupe, les nombres choisis sont souvent corrélés au nombre d'issues 11, 21 et 36 et ce choix témoigne de l'absence de perception de la fluctuation des fréquences.

La comparaison des résultats statistiques portant sur des effectifs différents conduit au calcul des fréquences. La comparaison de fréquences portant sur un même nombre d'observations est privilégiée par les élèves.

L'observation des fréquences expérimentales conduit beaucoup d'élèves à abandonner le modèle relatif aux 36 issues. Certains affirment que comme l'expérience relève du hasard, elle ne sert à rien. Certains élèves s'opposent à cette vision et affirment que si l'on en faisait beaucoup plus, on pourrait savoir. Un débat nait ; certains pensent : « *même si l'on en fait beaucoup, comme c'est du hasard, on ne pourra pas savoir* ». Ce débat témoigne d'une dévolution réussie et fait émerger une nouvelle question : « *Beaucoup de lancers peuvent-ils permettre de décider du modèle pertinent ?* ».

La troisième phase est une phase de simulation. En effet, la nécessité de faire beaucoup de lancers légitime le recours à la simulation sur tableur, même si le terme de « beaucoup » reste à préciser à ce stade du travail. Cette simulation est précédée d'une simulation avec un sac opaque contenant des papiers numérotés. Cela permet une bonne compréhension de la fonction ALEA.ENTRE.BORNES et de la touche F9 : tirage avec remise dans une urne.

Le nombre de répétitions réalisées est laissé à la responsabilité des élèves. Les choix faits sont variés. En fin d'activité, les élèves comparent les graphiques obtenus sur leurs écrans respectifs, observent que ceux qui ont utilisé des plus grands nombres ont des graphiques qui « bougent moins » quand ils utilisent le recalcul (F9) ils observent pour ceux qui ont dépassé 1 000 que la fréquence du 7 reste supérieure à celle du 6 et du 8. Cela permet que la stabilisation des fréquences simulées soit perçue et que le modèle à 36 issues soit retenu par tous. La démonstration est réalisée sans peine, puisque certains groupes l'ont effectuée lors de la première phase. Les rétroactions du milieu, données de l'expérience, modèles candidats pluriels, conceptions variées des élèves, données simulées avec le sac puis avec l'ordinateur, représentations graphiques ont permis que les élèves conservent la responsabilité de la résolution du problème. La plupart des changements de phases sont marqués par une évolution de la question étudiée.

5. 2. 5. Bilan et institutionnalisation

L'institutionnalisation se fera en deux temps : un bilan contextualisé sur la situation étudiée, puis une synthèse mathématique décontextualisée. Ces deux bilans prendront appui sur des propositions des élèves qui seront discutées. Le professeur apportera le vocabulaire spécifique correspondant à ces formulations des élèves.

5. 2. 5. 1. Bilan de l'activité

La synthèse de l'activité réalisée avec les élèves pourrait prendre la forme suivante :

— Pour des échantillons de taille 2 000, la fréquence du 7 reste toujours supérieure à celle du 6 et du 8, nous pensons donc que 6, 7 et 8 ne sont pas équiprobables. Il semble qu'Adrien aura plus de chance de gagner en jouant le 7.

— La probabilité qu'Adrien gagne est estimée par 0,16 (fréquence observée sur un échantillon de taille 1 000).

En travail à la maison, on demandera de chercher toutes les façons d'obtenir 6, 7 et 8 et éventuellement d'établir la probabilité de chacune des sommes

— Le modèle retenu sera explicité, il s'agit de la loi portant sur 36 issues équiprobables.

— La probabilité qu'Adrien gagne en jouant le 7 est calculée et simplifiée, on obtient 1/6.

5. 2. 5. 2. Institutionnalisation

— La notion d'échantillon sera définie conformément au programme (2009) :

« *Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience aléatoire*. »

— Fluctuation d'échantillonnage :

f_n désigne la fréquence de la valeur du caractère étudié au cours de l'observation d'un échantillon de taille n. On constate sur différentes observations d'échantillons de taille n, que les valeurs de f_n varient, on dit qu'il y a fluctuation des fréquences. On peut observer que l'amplitude des fluctuations est d'autant plus petite que n est grand (sauf dans des cas extrêmement rares).

— Stabilisation des fréquences :

Une probabilité est un nombre théorique relevant d'une expérience aléatoire simplifiée et idéalisée. Elle permet de décrire, d'anticiper les résultats de l'expérience réelle qu'elle modélise et de faire des prévisions.

La probabilité d'un événement A peut être estimée par la fréquence de la réalisation de A, observée sur un échantillon, avec d'autant plus de pertinence que la taille de l'échantillon est grande.

— Calcul de la probabilité d'un événement A de l'univers U :

Quand l'hypothèse de modèle repose sur l'équiprobabilité des issues, on a :

$$P(A) = \frac{\text{nombre d'éléments de } A}{\text{nombre d'éléments de } U}.$$

Quand les différentes éventualités ne sont pas équiprobables, on déterminera d'abord leurs probabilités respectives. Alors la probabilité de A sera la somme des probabilités des événements élémentaires qui le constituent.

— Simulation

Pour étudier certaines expériences aléatoires pour lesquelles un modèle a été choisi, on peut produire des échantillons par simulation de ce modèle. Une simulation peut être réalisée avec des objets (dés, pièces, …) ou des moyens informatiques.

Pour générer aléatoirement un nombre entier dans l'intervalle [n ; p] :

— avec le tableur, on utilise le plus souvent la fonction ALEA.ENTRE.BORNES(n;p)

— avec la calculatrice, on utilise **EntAleat(n,p)** qui se trouve dans $\boxed{\text{MATH}}$ PRB [4].

— avec Algobox, on utilise ALGOBOX _ALEA_ENT(n,p)

5. 2. 6. Exercices de travail sur la technique

— Comment simuler le lancer de trois pièces de monnaie pour déterminer la combinaison de Pile et Face la plus probable ?

— Comment simuler le tirage avec remise dans une urne contenant 50 boules : 30 blanches et 20 rouges afin de déterminer s'il est rare de n'obtenir aucune boule rouge quand on réalise 4 tirages d'une boule ?

— Comment peut-on simuler un sondage dans une population où il y a 40% de jeunes ?

— Pierre et Aline jouent : Pierre lance une pièce et s'il obtient « Face » il a gagné la partie ; sinon il relance la pièce et, s'il obtient « Face » il a gagné, tandis que s'il obtient « Pile » il a perdu et Aline a gagné. Qui a le plus de chances de gagner ? Comment simuler ce jeu ?

5. 3. *Comment prendre une décision dans une situation où le hasard intervient : la parité*

Avec la première activité d'étude et de recherche, les élèves ont découvert la notion de

4 Les calculatrices utilisées dans nos classes sont des TI.

modèle et celle de fluctuation des fréquences. La seconde activité a pour but de permettre aux élèves de percevoir un enjeu social important de la statistique : celui de la prise de décision dans une situation aléatoire. Il s'agit ici de faire comprendre aux élèves que dans une telle prise de décision, quelle que soit la décision prise, il y a un risque de se tromper.

5. 3. 1. L'énoncé.

Dans votre classe, la parité des sexes est-elle respectée ?

5. 3. 2. Types de tâches mis en jeu

— Modéliser une situation en termes d'aléatoire pour l'étudier.

— Produire des échantillons par simulation.

— Quantifier la fluctuation des fréquences.

— Prendre une décision dans le cadre d'une situation aléatoire.

5. 3. 3. Analyse de la situation

Quelle que soit la classe dans laquelle on pose cette question, la réaction des participants est toujours la même : ils se comptent, déterminent le nombre de filles, le nombre de garçons, les comparent et si ces nombres ne sont pas égaux, rejettent l'idée que la parité est respectée.

Ce comportement rend compte d'une conception arithmétique de la parité des sexes mais celle-ci n'est pas opérationnelle dès que l'effectif de la classe est impair (on peut alors l'aménager en disant qu'il y a parité des sexes si le nombre de filles et le nombre de garçons diffèrent d'une unité).

Toutefois, cette conception n'est pas la seule possible, on peut en envisager une autre, de type probabiliste : si pour constituer une classe, le sexe des individus n'est pas pris en compte et si la population dans laquelle les individus sont sélectionnés est composée d'à peu près autant de garçons que de filles, la classe peut être considérée comme un échantillon (aléatoire) d'une population où la proportion de filles est 0,5.

Cette valeur $p = 0{,}5$ est ici notre hypothèse de modèle. C'est avec ce nouveau point de vue que la question va être étudiée. On considère que le sexe des élèves n'a pas été pris en compte pour la composition des classes et que l'hypothèse de travail « il y a autant de filles que de garçons à l'entrée en Seconde » est raisonnable. On choisit donc de modéliser la situation étudiée par une loi équirépartie.

Remarquons que dans un autre cas, suivant la composition de la population, on aurait pu choisir $p = 0{,}51$ ou une autre valeur. Les élèves doivent être associés à la réflexion sur ce choix.

Le travail proposé amène à introduire l'intervalle de fluctuation au seuil de 95 % afin de prendre une décision quant au respect de la parité des sexes dans la classe.

Cette question d'actualité interpelle les élèves. Quand cette activité d'étude et de recherche est bien menée, on voit les élèves aller vérifier d'eux même s'il y a parité dans le gouvernement, à l'assemblée nationale, au sénat, etc.

Comme dans la précédente activité, la diversité des représentations graphiques et la simulation sur tableur jouent un rôle fondamental. Ils constituent une diversité de registres de représentations sémiotiques, cette diversité de registres graphiques articulés aux registres numériques dynamiques du tableur et au langage naturel des élèves sollicités contribuent à une construction efficace de la notion d'intervalle de fluctuation.

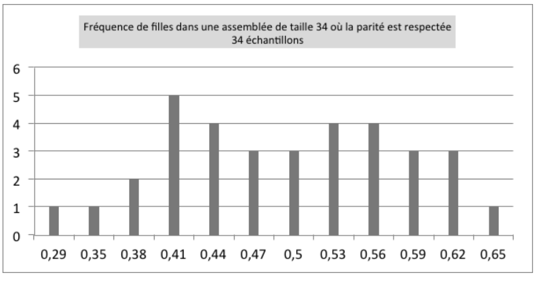
5. 3. 4. Indications sur l'organisation de l'activité d'étude et de recherche

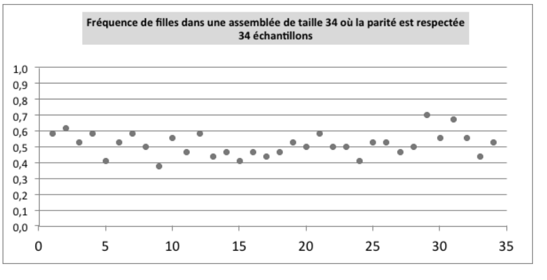
Dans la première phase, la réponse spontanée des élèves est suivie d'échanges destinés à faire changer le point de vue sur la conception de la parité homme-femme dans une assemblée.

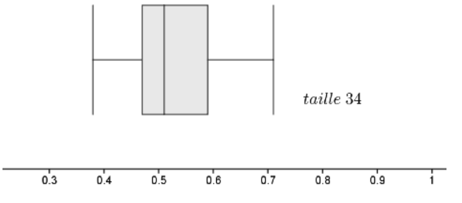
Lors de la deuxième phase, chaque élève écrit individuellement ce qu'il prévoit de faire pour répondre à la question en adoptant la conception aléatoire de la parité. Cette phase d'écriture permet aux élèves d'entrer dans la tâche proposée. Une mise en commun permet de faire apparaitre qu'avec une telle définition, on ne peut pas connaître à l'avance la composition d'une assemblée de 34 personnes (34 est le nombre d'élèves de la classe dans laquelle se déroule l'activité) ; on ne dispose pas de telles assemblées déjà constituées ; on peut en produire par simulation pour ensuite observer leur composition.

Lors de la troisième phase, les élèves travaillent en groupe de deux et élaborent une stratégie pour simuler une assemblée de 34 personnes. Cela peut-être à la main en tirant à Pile ou Face et codant, par exemple : Face-Fille, Pile-Garçon ; ou en lançant un dé et codant Pair-Fille, Impair-Garçon ; ou à la calculatrice en utilisant la fonction EntAlea(0,1).

En travail à la maison, chaque élève effectue la simulation d'un échantillon de taille 34 et calcule la fréquence de filles dans l'échantillon. Les résultats sont mis en commun et la série obtenue est étudiée. Cette phase permet de réactiver les connaissances de statistique descriptive étudiées en Troisième (moyenne, médiane, quartiles) et de construire puis observer plusieurs graphiques représentant la série (diagramme en bâtons, diagramme en boite, nuage de points). Cela permet de mettre en évidence et visualiser la fluctuation des fréquences (certains élèves risquent de proposer une fréquence qui leur paraît plausible sans avoir réalisé l'expérience ; il sera intéressant de confronter cette série aux résultats établis ultérieurement au cours de l'AER). Les élèves ne s'attendaient pas à de telles différences dans les fréquences obtenues, s'étonnent lorsque la moyenne ou la médiane n'est pas 0,5.



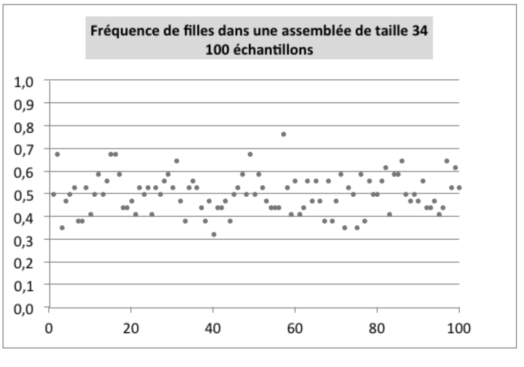




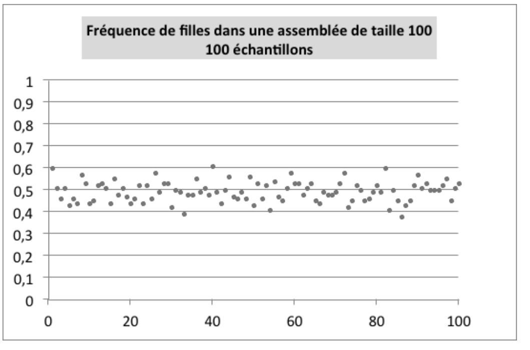
La question initiale est reposée : perplexité des élèves. Au vu de la variabilité des résultats obtenus, ils ont du mal à émettre une opinion et certains se disent que tout est possible (une élève déclare « *on pourrait avoir 99 % de garçons* ! »)

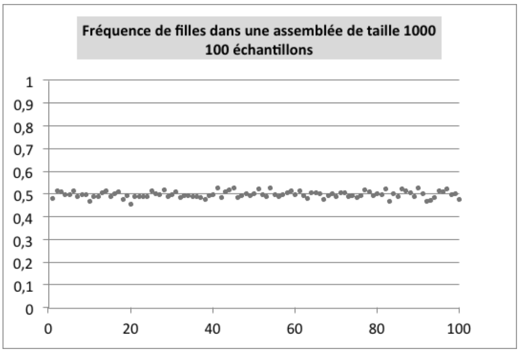
Le professeur propose de poursuivre l'observation en produisant davantage d'échantillons de taille 34 par une simulation sur tableur. En utilisant la fonction ALEA.ENTRE. BORNES(n;p), chaque élève produit une série de 100 fréquences de filles dans une assemblée de taille 34, détermine les caractéristiques de la série et fournit un nuage de points.

La simulation sur tableur est un moyen fiable et rapide d'obtenir des échantillons. L'utilisation de la touche F9 permet tout à la fois de produire de nouvelles séries d'échantillons mais aussi d'observer que d'une série à l'autre, certaines caractéristiques sont analogues (on notera qu'il est utile de bloquer l'axe des ordonnées du graphique entre 0 et 1 afin qu'il ne change pas à chaque production d'une nouvelle série).



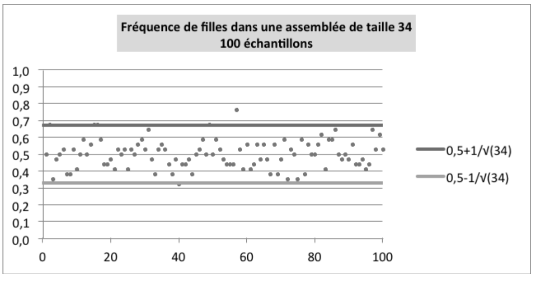
Devant la variabilité des fréquences, il est difficile de voir apparaître une propriété. Toutefois la première activité d'étude et de recherche a permis de voir que la fluctuation diminue lorsque la taille des échantillons augmente. Le professeur relance alors l'activité en proposant d'observer cette fois des assemblées de plus grande taille. Chaque élève produit une série de 100 fréquences de filles dans une assemblée de taille 100, puis de taille 1000 ainsi que les nuages de points correspondants.

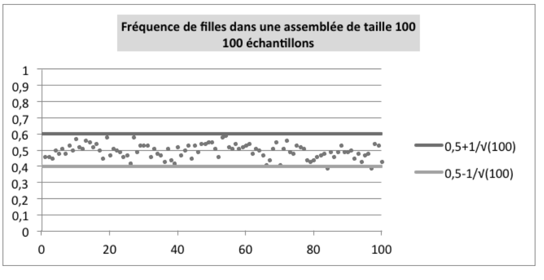


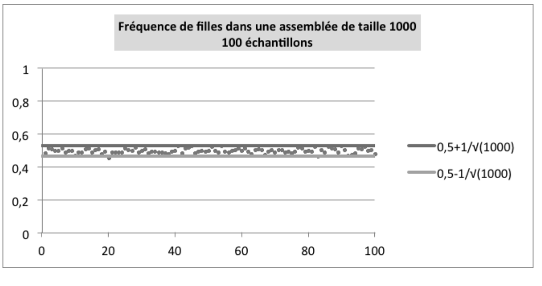


L'observation simultanée des trois graphiques est assez suggestive pour permettre aux élèves d'énoncer des conjectures qu'ils sont invités à rédiger. Dans ces conjectures, les élèves décrivent souvent, pour chaque graphique, l'intervalle dans lequel se situent les fréquences. Le professeur amorce la construction des éléments à institutionnaliser : pour 95 % environ de tous les échantillons de taille n possibles, la fréquence f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}} \; ; \; p + \frac{1}{\sqrt{n}}\right]$, appelé intervalle de fluctuation au seuil 95 %.

Une feuille comportant trois graphiques de ce type est distribuée aux élèves. Ils sont invités à écrire les intervalles de fluctuation et à tracer sur chacun des trois graphiques les droites d'équation $y = p - \frac{1}{\sqrt{n}}$ et $y = p + \frac{1}{\sqrt{n}}$ pour donner sens à cette propriété.







En se basant sur l'intervalle de fluctuation $\left[0{,}5 - \frac{1}{\sqrt{34}}\ ;\ 0{,}5 + \frac{1}{\sqrt{34}}\right]$, le professeur énonce la règle permettant de décider si, dans la classe, la parité homme-femme est respectée.

5. 3. 5. Bilan et institutionnalisation

5. 3. 5. 1. Bilan de l'activité

Si l'on considère que le caractère homme-femme est aléatoire, la composition d'une classe de 34 élèves peut être très variable. Dans environ 95 % des cas, la fréquence de filles appartient à l'intervalle $\left[0{,}5 - \frac{1}{\sqrt{34}}\ ;\ 0{,}5 + \frac{1}{\sqrt{34}}\right]$.

Pour décider si dans une classe donnée, la parité homme-femme est respectée, on adopte la règle suivante :

— *Si la fréquence de filles dans la classe n'appartient pas à* $\left[0{,}5 - \frac{1}{\sqrt{34}}\ ;\ 0{,}5 + \frac{1}{\sqrt{34}}\right]$, *on considère que la parité n'est pas respectée,*

— *Si la fréquence de filles dans la classe appartient à* $\left[0{,}5 - \frac{1}{\sqrt{34}}\ ;\ 0{,}5 + \frac{1}{\sqrt{34}}\right]$ *on n'a pas de raison de considérer que la parité n'est pas respectée.*

5. 3. 5. 2. Institutionnalisation

Intervalle de fluctuation

Dans une population, la proportion du caractère étudié est p comprise entre 0,2 et 0,8. On réalise un échantillon de taille $n \geq 25$ de cette population, la fréquence du caractère dans l'échantillon est f. Si l'on considère tous les échantillons de taille n possibles, pour environ 95 %

d'entre eux, la fréquence *f* appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$, appelé intervalle de fluctuation au seuil de 95 %. Il est souvent noté IF.

Prise de décision

Soit un échantillon de taille *n* dans lequel la fréquence observée est *f*. L'affirmation « l'échantillon provient d'une population où la proportion est *p* » doit-elle être rejetée ou pas ?

— *Si la fréquence f n'appartient pas à IF, alors, au seuil de 95 %, on rejette l'affirmation.*

— *Si la fréquence f appartient à IF, alors, au seuil de 95 %, on n'a aucune raison de rejeter l'affirmation.*

5. 3. 6. Exercices de travail de la technique

1. Deux entreprises recrutent leur personnel parmi des candidats comportant autant d'hommes que de femmes. La répartition entre hommes et femmes est la suivante :

	Hommes	Femmes
Entreprise A	57	43
Entreprise B	570	430

Peut-on penser que, dans l'une de ces entreprises, la parité homme-femme à l'embauche n'est pas respectée ?[5]

2. Contrôle qualité. Un grossiste en fourniture de bureau revend des rouleaux de ruban adhésif transparent et affirme que seulement 0,8 % des rouleaux présente un défaut de jaunissement du papier. Un client achète 500 rouleaux et constate que 6 rouleaux, soit 1,2% des rouleaux, présentent un jaunissement du papier. Le client peut-il faire une réclamation auprès du grossiste ?

3. On considère que la proportion de femmes dans la population française est 1;2. A l'Assemblée Nationale, il y a 577 députés, dont 108 femmes. Peut-on considérer que cette répartition est un effet de la fluctuation d'échantillonnage ou bien dire que la parité des sexes n'est pas respectée à l'Assemblée Nationale? (On assimilera l'ensemble des députés à un échantillon de taille 577 de la population française).[6]

4. Où l'on retrouve la parité des sexes (Ressources pour faire la classe en mathématiques pour la voie professionnelle, mise à jour 2010).
Dans la réserve indienne d'Aamjiwnaag (Canada) on recense, parmi les 132 enfants de moins de 3 ans, 46 garçons. Ces observations sont-elles le fruit du hasard ?

5. 4. *Comment estimer une probabilité : Le problème des anniversaires*

Les élèves savent maintenant quantifier la fluctuation des fréquences en fonction de la taille d'un échantillon.

La situation proposée doit les amener à s'interroger sur le problème réciproque : comment estimer la probabilité d'un évènement ? Peut-on le faire à partir des fréquences observées dans une expérimentation ou une simulation ? Et si oui, comment ? L'objectif est d'introduire l'estimation par intervalle de confiance.

La situation retenue est une adaptation du paradoxe des anniversaires énoncé par Von Mises au début du 20ème siècle.

5 D'après le document ressource de Seconde, 2008, page 17
6 Maths 2de, Collection Pixel, édition Bordas 2010, page 277

5. 4. 1. L'énoncé

« Est-il rare que dans une classe deux élèves au moins aient la même date anniversaire ? »

5. 4. 2. Types de tâches mis en jeu

La situation permet d'aborder les types de tâches suivants :

— Interpréter des informations statistiques.

— Modéliser une situation aléatoire.

— Simuler une expérience aléatoire sur tableur.

— Réaliser une estimation par intervalle de confiance.

Et éventuellement, si le niveau de la classe le permet :

— Calculer la probabilité par l'utilisation de l'évènement contraire et si A désigne l'événement étudié et U l'univers, p(A) = $\dfrac{Card(A)}{Card(U)}$. Ici, si chaque jour est repéré par son rang dans l'année (année non bissextile) et si le nombre d'élèves de la classe est 35, alors les 35-uplets constitués des entiers de 1 à 365 forment l'univers U.

5. 4. 3. Analyse de la situation

La situation choisie s'appuie sur un thème familier aux élèves, l'organisation doit permettre l'expression de conceptions initiales et leur éventuelle remise en cause.

Oui c'est possible chez le cas des frères jumeaux ou des soeurs jumelles qui soient nés au même jour et qu'ils sont dans la même classe. Ou sinon dans chaque naissance, il y a 1 chance sur 365 d'avoir la même date d'anniversaire

En particulier il est probable que certains penseront naturellement qu'il faut environ 365 personnes pour avoir beaucoup de chances que deux personnes aient la même date anniversaire.

On peut également s'attendre à ce que certains pensent que c'est très rare et d'autres non.

Oui cela peut-être fréquent, sachant que dans notre classe nous sommes 32.
La probabilité est donc de $\frac{1}{32}$.
Cela me parraît pas très fréquent.

Non, car il y a $\frac{1}{365}$ chances que 2 élèves aie la même.

$32 \times \frac{1}{365} = 0{,}08 = 0{,}08\ \%$ de chances

De façon très spontanée, l'élève peut partir de sa propre expérience et déclarer : *« Dans les classes que j'ai fréquentées, j'ai souvent/rarement observé que deux élèves au moins fêtaient leur anniversaire le même jour »*.

Cela est plutôt rare mais ça arrive : dans la classe on en a 2 : Majid et Ikrame, tous les 2 naissent le 18 juin.

La nécessité de modéliser la situation apparaît déjà :

Ça dépend du nombre d'élèves dans la classe, si certain élèves sont redoublant, s'il y a eu un grand nombre de naissance au cour d'une période dans l'année.

La diversité des réponses permet la mise en place d'un débat scientifique et motive la poursuite de l'étude. La subjectivité des réponses et la diversité des vécus aiguisent la curiosité des élèves qui expriment leur souhait de consulter de vraies données. La nécessité de recueillir des données objectives permet d'évoquer les problèmes fondamentaux dans la société, de faisabilité d'une enquête, de coûts. L'impossibilité d'un recensement exhaustif sur toute la France par exemple amène les élèves à mettre en œuvre une démarche statistique.

En proposant un fichier constitué de l'ensemble des classes du lycée, cette situation permet d'aborder un objectif du programme : faire travailler les élèves de Seconde sur des données réelles à partir de gros fichiers informatiques. Ce sera l'occasion de mobiliser une partie des statistiques descriptives.

Les résultats obtenus sur l'étude des classes dans nos lycées paraissent surprenants aux élèves, en effet on obtient une fréquence d'environ 75%. Beaucoup d'élèves pensent que leur lycée est exceptionnel ; cela relance la motivation de cette étude. Un travail de simulation sera proposé pour observer davantage de données.

Ce travail de simulation nécessite une réflexion riche du point de vue de la modélisation :

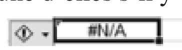
— on fixe le nombre d'élève par classe ;
— on décide qu'il y a 365 dates possibles d'anniversaires (on omet le 29 février) ;
— on décide qu'il y a autant de naissance chaque jour de l'année (cette hypothèse de travail est raisonnable au vu des données disponibles sur l'INSEE) et donc le modèle mathématique reposant sur l'équirépartition des issues peut permettre d'étudier cette expérience.

L'observation des données simulées amène les élèves à proposer une estimation de la probabilité cherchée. Le professeur invitera à faire le lien avec l'intervalle de fluctuation et proposera la mise en place de l'intervalle de confiance. Dans des classes à profil scientifique, on pourra envisager le calcul probabiliste pour clore l'étude mathématique du problème.

5. 4. 4 Indications sur l'organisation

Lors de la première phase de travail individuel, chaque élève exprime son point de vue puis travaille sur les listes photocopiées des dates anniversaires de deux classes. Ce travail « à la main » est fastidieux. Il ne permet pas de conclure quant à la question posée. Les élèves suggèrent qu'il faudrait étudier davantage de classes.

Le professeur propose d'étudier l'ensemble des classes du lycée : le fichier permettant cette étude a été obtenu à partir d'une extraction de la base de données du lycée, seuls les noms des classes et les dates anniversaires des élèves de chaque classe ont été conservées, les années de naissance ont été remplacées par 2000, puis cachées, afin que seuls le jour et le mois de l'anniversaire apparaissent. Les dates anniversaires des élèves de chaque classe ont été disposées en colonne, par classe, la première ligne contenant le nom de la classe.

Pour la seconde phase, les élèves ont en charge l'étude sur l'un des niveaux du lycée : Secondes, Premières, Terminales ou Post-bac. Chaque élève doit, à l'aide de la fonction MODE, déterminer pour les classes du niveau étudié, si oui ou non deux élèves au moins ont la même date anniversaire. Cette fonction MODE renvoie la valeur la plus fréquente de la série s'il y en a une, l'une d'elles s'il y en a plusieurs ou l'indication 

si chaque date n'apparaît qu'une fois.

L'élève doit calculer la fréquence, parmi les classes qu'il observe, de celles dans lesquelles deux élèves au moins ont la même date anniversaire. Ce travail peut être fait « à la main », en comptant les résultats obtenus avec la fonction MODE ou en utilisant la fonction NB qui compte combien de cellules d'une plage contiennent un nombre, ce qui n'est pas le cas des cellules contenant #N/A.

La mise en commun des résultats conduit à un calcul de moyenne pondérée pour établir la fréquence de l'événement étudié sur le lycée. Le résultat obtenu sur le lycée, voisin de 0,8, parait exceptionnel. Le fait que les effectifs sont très différents d'une classe à l'autre amènent les élèves à interroger la pertinence de l'étude réalisée. Ils expriment le souhait de disposer de données plus nombreuses et plus homogènes.

La troisième phase consiste à simuler des classes d'effectif constant. Le professeur relance : « Comment modéliser et simuler l'expérience aléatoire qui consiste à étudier si, dans une classe, deux élèves au moins ont la même date anniversaire ? ». Les différents éléments de la modélisation sont discutés et notés. Pour la simulation, on peut s'attendre à ce que les élèves proposent de mettre toutes les dates possibles de l'année dans un sac et de tirer au hasard avec remise une trentaine de fois (n tirages au sort successifs avec remise dans une urne contenant une fois chaque date anniversaire). Les différentes propositions font l'objet d'un débat qui aboutit à l'utilisation la fonction ALEA.ENTRE.BORNE(1;365).

Les élèves simulent alors sur tableur un grand nombre de classes. Ce nombre de classes simulées est laissé à la responsabilité de l'élève, il permet de remettre en débat la notion de grand nombre pour une étude statistique, certains choisissent 10 ! L'étude des fréquences observées sur ces échantillons obtenus par simulation (et l'utilisation de la fonction F9) permet alors aux élèves d'émettre une hypothèse sur la probabilité de l'évènement étudié. La confrontation avec les résultats de leurs voisins amènent les élèves à augmenter la taille de l'échantillon pour gagner en précision dans l'estimation.

Lors de la mise en commun des résultats, la fluctuation d'échantillonnage est évoquée pour expliquer les variations des résultats obtenus. Pour obtenir la précision de l'estimation obtenue, le professeur propose d'établir l'équivalence entre l'intervalle de fluctuation et la fourchette de sondage. C'est l'élaboration d'un début de justification.

Pour un seuil de 95 %, on sait que

$$f \in \left[p - \frac{1}{\sqrt{n}} \; ; \; p + \frac{1}{\sqrt{n}}\right].$$

Or $f \in \left[p - \frac{1}{\sqrt{n}} \; ; \; p + \frac{1}{\sqrt{n}}\right]$ si, et seulement si $p \in \left[f - \frac{1}{\sqrt{n}} \; ; \; f + \frac{1}{\sqrt{n}}\right]$. On dira que, au niveau de confiance de 95 %, $p \in \left[f - \frac{1}{\sqrt{n}} \; ; \; f + \frac{1}{\sqrt{n}}\right]$.

5. 4. 5. Bilan de l'activité et institutionnalisation

5. 4. 5. 1. Bilan de l'activité d'étude et de recherche

Des informations statistiques obtenues à partir de données simulées ont permis d'émettre une conjecture sur une probabilité non connue, pour une classe de 36 élèves, $p \approx 0{,}8$ avec d'autant plus de précision que la taille de l'échantillon simulé est grande. Par exemple, si on a obtenu $f = 0{,}8$ sur un échantillon de taille 100, pour un niveau de confiance de 95 %, on aura : $p \in \left[0{,}8 - \frac{1}{\sqrt{100}} \; ; \; 0{,}8 + \frac{1}{\sqrt{100}}\right]$

c'est-à dire $p = 0{,}8$ avec une précision de 1/10.

Si l'on a obtenu $f = 0{,}8$ sur un échantillon de taille 10 000, on aura :

$p \in \left[0{,}8 - \frac{1}{\sqrt{10\,000}}\ ;\ 0{,}8 + \frac{1}{\sqrt{10\,000}}\right]$, c'est-à dire $p = 0{,}8$ avec une précision de 1/100.

5. 4. 5. 2. Institutionnalisation

Toute fréquence observée sur un échantillon permet d'estimer la probabilité cherchée. Cette estimation est d'autant plus précise que la taille de l'échantillon est grande. Pour un niveau de confiance de 95 %, la précision est de $\frac{1}{\sqrt{n}}$.

Autrement dit : Si f est la fréquence observée sur un échantillon de taille n, alors pour un niveau de confiance de 95 %,

$$p \in \left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right].$$

5. 4. 6. Exercices de travail de la technique

1. Une personne est chargée de contrôler la qualité des pièces produites par une machine. Sur 1 500 pièces prises au hasard dans la production, elle en trouve 60 défectueuses. Elle doit fournir une estimation de la proportion de pièces défectueuses dans la production par intervalle de confiance pour un niveau de confiance de 95%. Quelle réponse communiquera-t-elle ?

2. Lors d'un referendum, un sondage aléatoire simple pratiqué sur 1 000 personnes a donné 55 % pour le Oui et 45 % pour le Non. Peut-on prévoir le résultat du referendum ?

3. Une enquête préliminaire à l'implantation d'un supermarché dans une petite ville a montré que, sur 820 familles interrogées au hasard, 270 envisageraient d'y aller le samedi après-midi. Sachant que la ville compte 2 000 familles, quel est le nombre de places de parking à prévoir pour que dans 95 % des cas il ne soit pas complet le samedi après-midi ?

Références bibliographiques

Bernoulli, Jakob (1713). *Ars Conjectandi*. Traduit du latin par Norbert Meusnier, *Jacques Bernoulli et l'Ars Conjectandi,* 4ème partie*,* IREM de Rouen, 1986.

Bordier, J. (1991). *Un modèle didactique utilisant la simulation sur ordinateur pour l'enseignement de la probabilité*, Thèse de doctorat, Paris, Université Paris 7.

Chaput, B. & Vergne, C. (2010). *A propos de l'introduction aux probabilités en troisième*. Commission Inter-IREM Statistiques et probabilités. http://www.univ-irem.fr/IMG/pdf/A_propos_de_l_introduction_aux_probabilites-2.pdf

Corpart, A. Lassalle, N. (2012) Un biberon comme outil de simulation au lycée*, Bulletin vert n°500 de l'APMEP*, *477-493*.

Courtebras, B. (2001). « Sur quelques conceptions du hasard », *Autour de la modélisation en probabilités*, 95-132. Besançon : Presses Universitaires de Franche-Comté.

Couthiño, C. (2001). *Introduction aux situations aléatoires dès le Collège : de la modélisation à la simulation d'expériences de Bernoulli dans l'environnement informatique Cabri-géomètre II*. Thèse Université Joseph Fourier, Grenoble I. Fichier pdf en ligne : http://www-diam.imag.fr/thesesequipe.htm

Ducel, Y. & Saussereau, B. (2011). « La prise de décision de la Seconde à la Première ». *Repères-IREM 85,* 31-49.

Henry, M. (1994). « Enseignement du calcul des probabilités dans le second degré, perspectives historiques, épistémologiques et didactiques ». *Repères-IREM 14,* 69-104.

Henry, M. (2001). « Notion de modèle et modélisation dans l'enseignement », *Autour de la modélisation en probabilités*, 149-160. Besançon : Presses Universitaires de Franche-Comté.

Henry, M. (2009). « Émergence de la probabilité et enseignement ». *Repères-IREM 74,* 67-89.

Laplace, P. S. (1825). *Essai philosophique sur les probabilités*. Rééd. de la 5e édition, Paris : Bourgois, 1986. Première éd. dans *Traité analytique sur les probabilités*, 1812.

Parzysz, B. (2007). « Expériences aléatoires et simulation, le jeu de croix ou pile ». *Repères-IREM 66*, 27-44.

Renyi, A.(1966). C*alcul des probabilités*, Rééd. J. Gabay, Sceaux, 1992.

Wozniak, F.(2005). *Conditions et contraintes de l'enseignement de la statistique en classe de Seconde générale et technologique*. Thèse, pdf en ligne :

et les documents officiels disponibles en ligne sur Éduscol :

Ressources pour la classe de Seconde, juin 2009, *Probabilités et Statistiques*, Éduscol.

Ressource pour la classe de Première, juin 2011, *Statistiques et probabilités*, Éduscol.

Ressource pour la classe de Terminale, février 2012, *Probabilités et statistique*, Éduscol.

Ressources pour faire la classe en mathématiques pour la voie professionnelle, mise à jour novembre 2010, Éduscol.

B.O. n°30 du 23 juillet 2009, *Programme de Seconde,* Éduscol.

B.O. n°9 du 30 septembre 2010, *Programme de Première S,* Éduscol.

B.O. n°8 du 13 octobre 2011, *Programme de Terminale S,* Éduscol.

Cadre de l'étude

ANNEXE 1

Le travail que nous avons présenté est issu d'une recherche proposée par l'Institut Français de l'Education (IFE), dans le réseau PERMES (Parcours d'Etude et de Recherche en Mathématiques dans l'Enseignement Secondaire). Cette recherche menée de 2010 à 2013 a pour objectif de produire des exemples de séquences d'enseignement des mathématiques illustrant le concept de « Parcours d'Etude et de Recherche » (PER) introduit par Yves Chevallard[7]. Nous avons choisi de travailler sur l'enseignement des statistiques et probabilité en lycée.

Explicitons les caractéristiques d'un PER que nous avons retenues pour cette étude.

Des questions porteuses de sens

Un constat établi par plusieurs institutions, telles que l'inspection générale des mathématiques, atteste de la perte de sens de l'enseignement des mathématiques. D'après Chevallard « les objets enseignés condensent des réponses à des questions que nous avons perdues » (2005). L'étude épistémologique et les usages sociaux doivent nous permettre de retrouver la raison d'être des enseignements mathématiques. C'est en donnant aux élèves des questions problématiques et authentiques que l'élève « rencontre en acte les raisons d'être » (2005) de ce qu'il apprend. Ces questions peuvent se poser en dehors mais aussi depuis l'intérieur des mathématiques. Un PER doit permettre de faire revivre aux élèves une activité mathématique riche permettant la construction d'un sens mathématiquement authentique.

Un PER doit s'appuyer sur une Question à Fort Pouvoir Générateur d'étude (QFPG). Cette question permet de mettre en évidence l'intérêt du thème à étudier. Elle motive les apprentissages visés par les programmes. Cette question doit permettre à l'enseignant de concevoir des activités pertinentes à tous les niveaux d'enseignement. Nous pensons que, le plus souvent, cette question n'a pas à être transmise aux élèves sous sa forme générale. Enoncée dans le contexte de la situation à étudier, elle motive l'activité des élèves et favorise la dévolution du problème. Par exemple, une QFPG d'étude pour la géométrie pourrait être « comment mesurer une distance inatteignable », pour les élèves d'une classe de quatrième, cela pourrait devenir « comment mesurer la distance entre le bureau du professeur et l'arbre de la cour, sans sortir de la classe ? »

Pour un niveau scolaire donné, pour un thème donné, un PER permet d'aborder la plupart des apprentissages visés par le programme. Il amène souvent à un décloisonnement des chapitres du programme.

Un PER intègre tous les moments de l'apprentissage, au sens de l'organisation didactique de Chevallard, Activité d'Etude et de Recherche (AER), institutionnalisation, travail sur l'Organisation Mathématique (travail sur la technique, étude de la portée de la technique,…) et évaluation. Un Parcours d'Etude et de Recherche est un enchainement cohérent de ces différents temps d'apprentissage répartis sur l'ensemble d'une année scolaire.

7 Chevallard, Y., 2005, Onisep, Les sciences aujourd'hui. La recherche, les métiers, les formations, « *Étudier et apprendre en mathématiques : vers un renouveau* », p.55-56.

Un professeur directeur d'études :

Selon Chevallard (2008) « *Le travail engagé pour l'étude d'une question engendre d'autres questions et c'est ce qui fait la dynamique de l'étude. Pendant la recherche, à l'aide d' un jeu de questions cruciales formulées par la classe sous l'impulsion du professeur qui dirige l'étude, on avance vers les connaissances visées ce qui sous-entend une analyse a priori bien fouillée pour pouvoir orienter le travail de façon pertinente et disqualifier les techniques non souhaitées qui apparaissent.* »

Les Activités d'Etude et de Recherche (AER) doivent permettre une dévolution des questions mathématiques aux élèves, afin que l'activité mathématique soit menée par l'élève. Les élèves au cours de ce travail développent des compétences fondamentales en mathématique : prendre des initiatives, communiquer à l'écrit et à l'oral, poser un regard critique sur une production, dégager et étudier un cas particulier, établir une conjecture, argumenter... Le professeur joue un rôle de directeur de l'étude.

Notre étude a porté sur l'enseignement de la statistique inférentielle et du calcul des probabilités au lycée. Cet article présente nos propositions pour la classe de Seconde. Nous avons choisi de poursuivre le travail réalisé par un autre groupe de l'IREM de Montpellier sur le même thème pour la classe de Troisième [8]. Nous avons repris la même Question a fort pouvoir Générateur d'Etude : « *Comment prendre une décision dans une situation où le hasard intervient ?* ».

8 Irem de Montpellier – Michel Roche- publication à venir.

... le loto

ANNEXE 2

Palmarès des numéros (depuis octobre 2008)

N'oubliez jamais que le hasard ne se contrôle pas.
Dernière mise à jour : **04/02/2013**
Ce tableau prend en compte tous les tirages depuis le 6 octobre 2008.

Numéross	Nombres de sorties	% de sorties*	Dernière sortie	Ecart**
1	78	11.29	22/12/2012	19
2	72	10.42	26/01/2013	4
3	78	11.29	26/12/2012	17
4	74	10.71	31/12/2012	15
5	65	9.41	26/12/2012	17
6	64	9.26	04/02/2013	0
7	73	10.56	02/02/2013	1
8	57	8.25	17/11/2012	34
9	71	10.27	23/01/2013	5
10	68	9.84	14/01/2013	9
11	84	12.16	30/01/2013	2
12	67	9.7	24/12/2012	18
13	84	12.16	12/01/2013	10
14	62	8.97	05/11/2012	39
15	84	12.16	07/01/2013	12
16	74	10.71	16/01/2013	8
17	69	9.99	26/12/2012	17