Sommaire

Introduction

Les probabilités, régulièrement traitées au cycle 4, permettent un recours aisé à l'intuition grâce à une contextualisation sous forme ludique. Pourtant, la définition même de la notion de probabilité par l'approche fréquentiste est souvent plus difficile d'accès pour les élèves. L'introduction de la fluctuation d'échantillonnage en classe de seconde présente les mêmes difficultés. Ce chapitre, alors qu'il offre des possibilités d'applications à la fois très concrètes, contemporaines, et riches du point de vue du raisonnement, est souvent passé en revue un peu rapidement dans une progression dense. Il est pourtant d'une grande importance pour cimenter les notions de probabilité, et son étude à l'aide d'outils plus complexes lors du cycle terminal nécessite une compréhension solide des bases vues en seconde.

Nous nous sommes donc posé la question de savoir comment introduire ces chapitres en tentant d'en exploiter le potentiel ludique et concret pour en faciliter l'appropriation par les élèves. A la lecture d'articles de littérature didactiques et au gré des échanges instructifs avec nos formateurs, la notion de situation-problème s'est imposée comme une piste de réponse à cette question. Ces situations-problèmes ont été formalisées en 1987 par Philippe Meirieu mais sont inspirées de la branche des problèmes ouverts en mathématiques.

Voici à titre d'information la définition que donne Wikipedia de cette notion : "On appelle situation-problème une activité pédagogique consistant en l'aménagement d'une tâche de travail destiné à faire découvrir, par l'élève lui-même, des solutions à un problème. La résolution de ce problème doit permettre à l'élève l'acquisition de nouvelles connaissances (savoir, savoir-faire…)."

Les situations-problèmes se doivent d'interpeller les élèves, de les plonger dans un problème concret et de les mettre au centre de sa résolution, et se faisant de réinvestir ou découvrir les notions en jeu. Il s'agit d'un mode d'introduction plus actif et autonome que les activités d'introduction traditionnelles. Elles s'appuient souvent sur des manipulations concrètes, et dans notre cas, sur des simulations informatiques.

L'objet de ce mémoire professionnel est donc d'analyser et d'expérimenter l'apport de deux situations-problèmes bien choisies pour introduire certaines notions relatives aux probabilités et à la fluctuation d'échantillonnage en classe de seconde.

En particulier, nous présenterons l'expérience dite de "la bouteille de Brousseau" (du nom du célèbre didacticien) pour les probabilités puis le problème du sex-ratio dans une réserve indienne pour l'échantillonnage. Nous les analyserons à partir de la littérature existante et des expérimentations que nous avons réalisées en classe de seconde, avant d'en effectuer un bilan qui pourrait être utile à ceux qui voudraient à leur tour mettre en pratique ces activités.

Il est à noter que ces deux activités sont indépendantes, mais peuvent s'enrichir mutuellement. Nous les présentons ici dans l'ordre logique (il est d'usage d'aborder d'abord les probabilités) et chronologique dans lequel nous les avons expérimentées.

I Probabilités : la bouteille de Brousseau

1) Motivations

Les probabilités ne constituent pas un chapitre nouveau à l'arrivée en classe de seconde. Fort d'une intuition souvent sollicitée et d'un travail au collège, les élèves se disent souvent à l'aise avec les probabilités et récitent les situations des exercices traditionnels qu'ils ont déjà rencontrés : lancer de dé, d'une pièce, etc. Les probabilités sont désormais introduites dès la classe de 5ème et l'approche fréquentiste est préconisée en classe de 4ème. Pourtant la définition générale d'une probabilité, hors de tout exemple particulier, est beaucoup plus difficile à obtenir. Il est vrai que la question est difficile : il n'y a qu'à faire l'expérience sur des adultes, les mots manquent rapidement et les références deviennent circulaires. Elle contient d'ailleurs dans l'approche fréquentiste une notion implicite de limite qui ne sera formalisée que bien plus tard.

Le document IREM (Groupe IREM "Ressources TICE pour la formation et l'enseignement", 2015) analyse les questions soulevées par l'introduction des probabilités. L'extrait ci-dessous porte à l'origine sur les anciens programmes et le traitement des probabilités en classe de 3ème mais les questions d'ordre général nous semblent toujours pertinentes :

"Les programmes […] préconisent une double approche de la notion de probabilité : fréquentiste et théorique. Mais laquelle choisir pour introduire la notion de probabilité ? A quel moment, en classe, poser une définition d'une probabilité ? Quelle définition ? Les programmes préconisent également de recourir à l'expérimentation à travers des situations familières pour les élèves. Expérimenter ou manipuler ? Dans ce cadre institutionnel, il s'agit donc de faire entrer les élèves dans un processus de modélisation à un moment d'un processus d'enseignement, choisi par l'enseignant. "

Dans l'aménagement du programme de 2nde, il est écrit que les "objectifs visés par l'enseignement des probabilités à l'occasion de résolution de problèmes" sont notamment de "rendre les élèves capables d'étudier et modéliser des expériences relevant de l'équiprobabilité" et de "proposer un modèle probabiliste à partir de l'observation de fréquences dans des situations simples".

Le but recherché par l'activité d'introduction aux probabilités que nous proposons est similaire à ce qui est fait en troisième (elle peut d'ailleurs tout à fait être proposée à ce niveau). Il s'agit avant tout de remobiliser ces connaissances et d'asseoir, si ce n'est dans un langage formel au mois dans l'intuition, la définition d'une probabilité par l'approche fréquentiste.

2) Présentation et analyse a priori

L'expérience de la bouteille de Brousseau porte le nom de son créateur le didacticien Guy Brousseau qui l'a décrite pour la première fois en 1974 dans l'article "Une expérience de premier enseignement des statistiques et des probabilités", après l'avoir testée en classe de CM2. L'énoncé original est le suivant : "Une bouteille contient 5 jetons d'au plus deux couleurs différentes, noire et blanche. L'enseignant ne connait pas la composition. Quelle est la composition de la bouteille ? Ouvrir la bouteille n'est pas autorisé. "

Il faut préciser que la bouteille est opaque et que l'on ne peut voir qu'un jeton à la fois, au niveau du goulot. Cette expérience nous a aussi été présentée lors de l'UE Probabilités de l'ESPE pendant notre année de formation, et nous avons décidé de la proposer à notre classe de 2nde, après quelques modifications et prolongements que nous détaillerons plus loin. La méthode attendue est de tirer de nombreuses fois un jeton, de noter les couleurs obtenues puis de calculer la fréquence associée à chaque couleur, que l'on ramène alors au total de billes (cinq) pour obtenir une prédiction d'autant plus fiable que le nombre de tirages est important.

Cette situation-problème présente à première vue de nombreux avantages. L’analyse qui suit est personnelle mais inspirée principalement de ce qui nous a été présenté à l’ESPE.

Tout d'abord la formulation du problème est simple, épurée de tout formalisme ou vocabulaire mathématique, si bien qu’elle permet à tous les élèves de se saisir du problème. D'autre part, la manipulation concrète, qui joue un rôle fondamental dans la construction du savoir dans le cycle primaire, est ici remise au goût du jour pour une meilleure appropriation des notions sous-jacentes. Le fait que l'exercice s'incarne dans une bouteille concrète permet également une auto-validation objective, indépendante de l'enseignant qui ne connaît d'ailleurs pas lui-même le résultat dans cet énoncé original. Il est interdit d'ouvrir la bouteille pendant l'expérience, mais on sait que le résultat attendu n'est qu'à portée de bouchon, et qu'il se vérifiera en dernier recours en ouvrant la bouteille. Enfin la question "Quelle est la composition de la bouteille ?" est surprenante. Il ne s'agit pas simplement de donner son avis mais bien de trouver un résultat, de construire une certitude. Le premier réflexe est d'ailleurs d'affirmer qu'on ne peut pas savoir.

A priori, la difficulté réside justement dans le passage de ce cap, dans la transformation du jugement et de l'intuition en un raisonnement et une méthode. L'énoncé est sur ce point un peu trompeur puisqu'il paraît clair, y compris aux élèves il nous semble, qu'en pareille occasion il ne peut y avoir de certitude absolue : on peut malencontreusement tirer toujours un jeton noir par exemple. Il faut alors ne pas renier le problème en bloc, et chercher plutôt une prédiction éclairée par un raisonnement solide. C'est cette incertitude sur un tirage particulier qui rend les probabilités ludiques mais qui dans le même temps peut ne pas inciter les élèves à chercher plus loin.

L'approche fréquentiste n'est ici pas évidente, puisque les données du problème font penser à un calcul de probabilité dans le cas de l'équiprobabilité. De la même manière, si l'on demandait à un élève de vérifier que la probabilité de tirer un 1 au dé est bien de 1/6, il invoquerait beaucoup plus facilement la définition théorique qui rend la question évidente plutôt que de se lancer dans de nombreux tirages fastidieux. D'autres difficultés peuvent apparaître dans cette grande autonomie laissée aux élèves, comme le nombre de tirages à effectuer (quand s'arrêter ?), et le passage du nombre de jetons noirs tirés à la fréquence de jeton noirs tirés. Même si le chapitre de statistiques a été vu plus tôt dans l'année, cela n'est pas toujours évident.

Enfin, comme préconisé par (Groupe IREM "Ressources TICE pour la formation et l'enseignement", 2015), cette activité a un prolongement naturel grâce à l'outil informatique qui permet de simuler un grand nombre de tirages, et ainsi d'observer directement sur le graphique la convergence des fréquences vers une valeur limite qu'on appellera alors probabilité. On peut également simuler des bouteilles avec un grand nombre de jetons et comprendre qu'il faut un nombre d'essais beaucoup plus important (trop fastidieux pour être mis en œuvre en pratique avec des bouteilles) pour obtenir une prédiction stable et à peu près certaine. Il faut donc amener les élèves à se rendre compte des limites de la manipulation physique pour mieux s'en affranchir par la simulation informatique (qui nécessite néanmoins que le modèle de la bouteille, et donc sa composition, aient étés programmés au préalable).

3) Mise en place

Nous avons donc mis en place cette activité en classe de seconde, dans le but d'introduire ou de remobiliser l'approche fréquentiste des probabilités et dans un souci de pouvoir y faire référence à la fois dans le cours, les exemples et les chapitres ultérieurs comme l'échantillonnage. Notre classe de 2^{nde} a 27 élèves, et cette activité a été proposée en demi-groupe. Les élèves ont été répartis au volontariat dans des groupes de 3 ou 4 élèves, organisés en ilots. Chaque groupe possédait une bouteille opaque dans lesquelles ont été placées des billes de couleurs bleue et blanche. Un trou dans le bouchon permettait de voir la couleur d'une bille à la fois.

Il était précisé que le nombre total de billes dans chaque bouteille est de cinq, afin de ne pas compliquer l'expérience dans un premier temps, et il fallait donc retrouver la composition de chaque bouteille. Le contenu de chaque bouteille était identique mais les élèves ne le savaient pas, et nous nous sommes servis de cette information au cours de l'activité pour confronter des résultats contradictoires.

Le document qui nous a servi pour guider le déroulement de cette séance, ainsi que des photos du dispositif se trouvent en annexe 1. Pour les détails pratiques, nous avons utilisé des petites briques de jus de fruit qui ont l'avantage d'avoir un goulot suffisamment petit pour ne laisser passer qu'une bille et sont déjà opaques. Voici des photos du dispositif proposé aux élèves :

Figure 1: une bouteille et son contenu

Figure 2 : une bouteille par groupe

Figure 3 : Tirage blanc

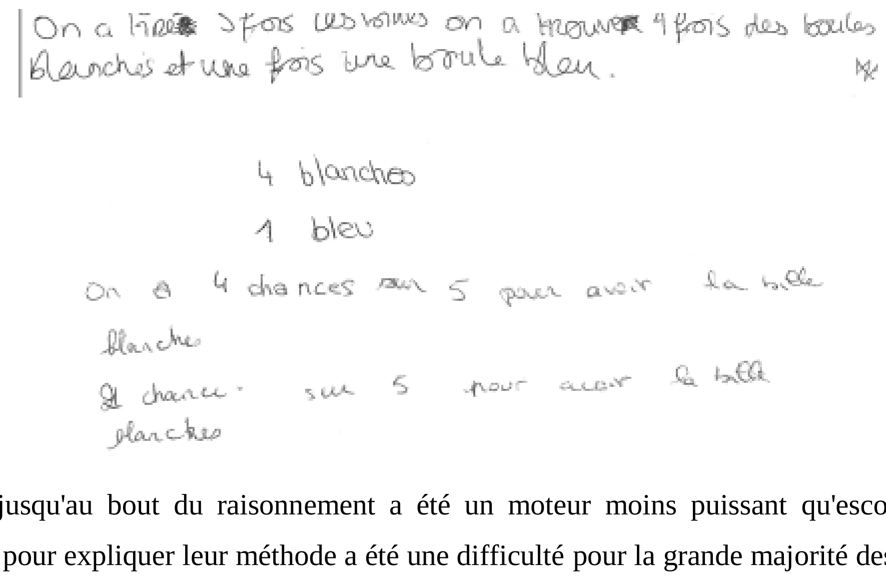
Figure 4 : Tirage bleu

Il se trouve que l'expérience avec la bouteille et son exploitation ont pris plus de temps qu'estimé, et la simulation informatique a donc fait l'objet d'une séance complète, également en demi-groupe. L'énoncé de ce TP informatique qui a donc pour but de constater la convergence des fréquences avec différents paramètres tout en familiarisant les élèves avec les commandes usuelles du tableur se trouve en annexe 2. La simulation des tirages bleu et blanc se fait à l'aide d'une "macro" que nous avons implémentée et qui permet de reconstituer l'effet séquentiel des tirages. A chaque nouveau tirage, le graphique "fréquence observée en fonction du nombre de tirage effectués" s'enrichit d'un nouveau point. Chaque partie de l'activité, physique et informatique, a duré une heure en tout, pour chaque demi-groupe.

4) Analyse des productions d'élèves et retour critique

Dans l'ensemble, les élèves ont bien accepté la situation proposée malgré un temps à expliquer qu'il ne fallait ni dévisser le bouchon ni essayer de regarder à l'intérieur. Certains élèves ont d'ailleurs proposé des contournements intelligents que nous n'avions pas anticipés, comme celui de faire une marque distinctive au stylo sur chacune des billes à travers le bouchon percé, pour en compter plus facilement le nombre. Il a fallu aussi pour un nombre non négligeable d'entre eux leur montrer le geste pour effectuer un tirage élémentaire (retourner la bouteille au-dessus de ses yeux), et j'ai pu constater y compris sur des adultes que ce n'était pas toujours intuitif. Dans le deuxième demi-groupe, alors que certains groupes sont partis beaucoup plus rapidement sur la bonne méthode, il a paradoxalement fallu sans cesse recadrer pour éviter que l'activité ne dérive en un jeu sans intérêt mathématique.

De manière générale, l'enjeu a été bien compris mais la plupart se sont contentés de peu de tirages et donc de conclusions peu probantes. Voici par exemple la production d'un groupe qui s'est arrêté tôt :


On a tiré 5 fois les boules on a trouvé 4 fois des boules Blanchés et une fois une boule bleu.

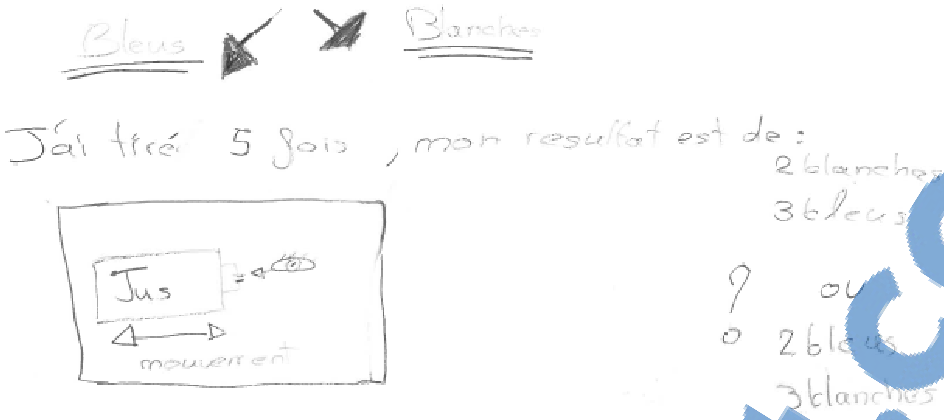
4 blanches

1 bleu

On a 4 chances sur 5 pour avoir la bille blanches

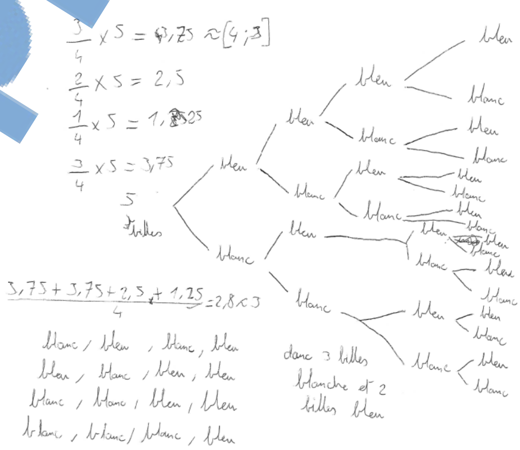
1 chance sur 5 pour avoir la bille blanches

L'envie d'aller jusqu'au bout du raisonnement a été un moteur moins puissant qu'escompté. Surtout, le passage à l'écrit pour expliquer leur méthode a été une difficulté pour la grande majorité des groupes, comme on peut le constater sur l'ensemble des productions d'élèves en Annexe 3 dont voici un exemple :



Nous voyons plusieurs raisons à cela. Nous pensons qu'ils n'ont pas assez l'habitude de narrer leurs recherches (et il est vrai que c'était le premier exercice de ce type avec eux depuis le début de l'année) et il y a une réticence à coucher sur papier des raisonnements ou des bribes d'idées qu'ils considèrent comme fausses ou non abouties – une réticence que l'on retrouve parfois dans les ratures de réponses pourtant intéressantes dans des productions plus classiques. De ce point de vue, cette activité a peut-être permis de vaincre en partie cette réticence. Souvent, en leur demandant de nous expliquer l'état de leur réflexion, il fallait insister pour qu'ils écrivent ce qu'ils venaient de dire, et seul cet aval donné et répété par l'enseignant pouvait les convaincre de se lancer. Ensuite, peut-être que le contraste entre cette activité d'aspect très libre et la contrainte de la rédaction était trop net, qu'il aurait fallu attendre davantage que leurs réflexions se structurent. Enfin, la maîtrise de la langue française est un blocage pour certains d'entre eux. En tous cas, il a fallu passer du temps pour faire émerger l'intérêt de beaucoup de tirages, de celui de calculer la fréquence (en rappelant ce que c'était) et d'en déduire le nombre de billes de chaque couleur.

Parmi les pistes intéressantes, un groupe a dessiné des diagrammes en arbre parce qu'il a reconnu des probabilités et que cela leur semblait une méthode traditionnelle de résolution de ces problèmes. L'idée de la fréquence est tout à fait présente, mais curieusement (peut-être pour des questions de place pour l'arbre qui pourtant ne joue pas de rôle), le groupe a réalisé plusieurs séries de 4 tirages puis a calculé la moyenne des résultats obtenus. Le résultat est équivalent à calculer la fréquence sur l'ensemble des tirages mais présente en plus l'avantage de donner une idée du résultat final après chaque série.



D'autres sont restés très attachés au nombre cinq (le total de billes), n'effectuant que des séries de cinq tirages successifs. Ainsi, ils ont listé les différents contenus possibles pour la bouteille (quatre blanches et une bleue ; trois blanches et deux bleues ; …) et à chaque série de cinq tirages, notaient quelle configuration ils obtenaient.

4 blanche 1 bleu pas possible - X 5
3 blanche 2 bleu X 3
3 blanche 2 bleu
2 blanche 3 bleu X 2

Leur proposition de résultat était le contenu possible qui était advenu le plus souvent. Il a alors fallu prendre un cas extrême, comme souvent en mathématiques, pour leur montrer la faille dans leur raisonnement : "si on obtient deux fois que des blanches, et trois fois que des bleues, va-t-on conclure qu'il n'y a que des bleues dans la bouteille ?" Comme dans le groupe précédemment cité, l'utilisation d'une moyenne aurait permis de boucler astucieusement le raisonnement mais le groupe d'élèves a préféré reprendre la réflexion du début.

Comme cela avait été anticipé dans l'analyse a priori, certains élèves s'arrêtent rapidement au « on ne peut pas savoir ». Certains répondent aussi que «tirer 10 ou 20 fois, ça ne change rien », au sens que cela ne modifie pas la façon dont sont tirées les billes, sans comprendre que ce qui change, c'est le degré de certitude de notre prédiction.

Voici par exemple la production d'un groupe qui se méprend sur la qualité de démonstration de leur méthode et n'a pas encore compris l'intérêt d'un grand nombre de tirages :

On a fait le geste et on a obtenue 4 fois sur 6 fois de billes blanche donc on pense qu'il y'a plus de billes blanche que de bille bleu. on va vous le démontrer !:

nombre de geste	1	2	3	4	5	6
couleur de billes	blanc	blanc	bleu	bleu	blanc	bleu

↑ Donc il y a vraiment plus de bille blanche que de bleu.

J'ai trouvé que le recours à une situation de pile ou face les interpellait davantage, par des questions comme : « Si on tire 3 fois et que l'on obtient 3 fois pile, peut-on dire que le jeu résulte toujours par un pile ? Comment se convaincre qu'on a autant de chance de faire pile que face ? ». L'intérêt de tirer un grand nombre de fois émerge lentement. Certains pensent alors (peut-être par paresse), qu'au-delà de 20 tirages, "ça ne sert à rien".

Le recours à la fréquence (ou au "pourcentage") n'est pas très intuitif et les raisonnements, sans aide extérieure, restent en général qualitatifs, c'est-à-dire très souvent "il y a plus de blanc que de bleu donc plus de billes blanches". Il faut alors interroger le sens de "plus" : est-ce beaucoup plus ? Un peu plus ? Comment quantifier ? etc.

Bien sûr, certains groupes sont parvenus au résultat par la méthode attendue sans qu'il n'y ait besoin de trop les guider. Par exemple :

5billes Combien de billes blanche combien de billes bleu ?

Nous allons faire des essaie avec la bouteille pour voire laquelle couleur revien le plus souvent pour etablir une hypothèse.

Essaie :
1 : blanche
2 : blanche
3 : blanche
4 : blanche
5 : bleu
6 : bleu
7 : blanche
8 : bleu
9 : bleu
10 : blanche

Totale : blanche : $\frac{6}{10}$
bleu : $\frac{4}{10}$

On peut etablir une frequence et une hypothèse sur le faite qu'il y a plus de billes blanche que de bleu, même un nombre 2 billes bleu pour 3 billes blanche.
Mais pour être plus sûre nous allons faire plus d'essaie. (100)
Après 100 essaie nous avons eu comme resultat :
bleu : 40
blanche : 60
Comme notre 1er essaie est fait
40% de chance de tomber sur une bille bleu et 60% de chance sur une bille blanche ce qui fait 2 billes bleu et 3 billes blanche car 40% de 5 est egale à 2. Notre hypothèse et donc exact.

Et surtout, tous les groupes ont fini par rédiger une solution satisfaisante en fin d'heure, même si l'aspect rédactionnel est inégal. Par exemple :

en applicant une autre méthode : × 50 $\begin{array}{r}34\\+16\\\hline 50\end{array}$

blanc : |||||||||||||||||||||||||||||||||| = 34
bleu : |||||||||||||||| = 16
Donc : • $34 \div 50 = 0,68 \rightarrow \frac{68 \times 5}{100} = 3,4 \simeq 3$
• $16 \div 50 = 0,32 \rightarrow \frac{32 \times 5}{100} = 1,6 \simeq 2$

Donc on pense qu'il y a 2 billes bleu et 3 billes blanche.

En fin de séance, nous avons vérifié le contenu des bouteilles et modifié la composition de chaque bouteille (une différente dans chaque groupe cette fois) pour vérifier si les élèves avaient assimilé la méthode. Lorsque le "un peu plus de blanc" revenait, ils ne pouvaient cette fois pas conclure 3 blanches et 2 bleues car ils ne connaissaient pas le nombre de billes total contenues dans la bouteille. L'intérêt de la fréquence est alors devenu plus clair, sans que l'on ait véritablement le temps de pousser jusqu'au bout. La fréquence a aussi montré ses limites, car en cas de parité (c'était le cas dans l'une des bouteilles), le groupe en question a bien compris qu'on ne pouvait pas (et surtout ne pourrait pas) distinguer la solution 2/2 de la solution 3/3 par exemple.

Nous avons alors dégagé ensemble l'intérêt de passer à une simulation informatique, et nous avons montré la convergence des fréquences de billes bleues vers la valeur de 40%, en insistant qu'en l'absence d'autres informations cela pouvait constituer une définition de la probabilité de tirer une bille bleue.

En ce qui concerne le TP informatique, cette fois-ci réalisé de façon individuelle, les tâches des élèves étaient davantage guidées. Les fonctions de base d'Excel, même avec l'aide en ligne, ne sont pas vraiment maitrisées et il faut souvent décoder pour eux ce qui est attendu dans chaque argument. Cela donne parfois l'impression qu'ils se perdent dans la technicité et ne voient pas ce qu'on est en train de traiter. Par contre le cadre du TP comme prolongement de la manipulation précédente n'a posé aucun problème, et l'intérêt a été bien perçu. Tout le monde a néanmoins eu le temps de tracer la courbe de convergence des fréquences, ce qui était l'objectif principal. Une des difficultés du fichier que nous avons réalisé, c'est qu'il doit se lire de façon séquentielle, ligne par ligne, et que la fréquence obtenue à une certaine ligne doit ignorer tous les tirages obtenus à une ligne ultérieure, même s'ils sont affichés à l'écran, et ce n'est pas toujours facile ni à concevoir ni à expliquer. Un travail sur les couleurs (grisé pour les fréquences des lignes antérieures, en gras pour celles de la dernière ligne) pourrait peut-être améliorer cet aspect du fichier.

5) Bilan

En conclusion, il nous semble que cette situation-problème est tout à fait adaptée à l'introduction des probabilités en classe de seconde. Sous une forme ludique qui a intéressé les élèves, y compris ceux en difficulté, elle a permis d'asseoir l'approche fréquentiste des probabilités par la manipulation et la simulation informatique.

Son cadre éloigné de la forme scolaire traditionnelle est bénéfique pour la motivation, mais n'est pas garant ni de la rigueur ni de l'envie de rigueur, surtout en probabilités où plusieurs réponses sont parfois possibles, et où les élèves ont parfois l'impression que les avis se valent. Cette activité entretient une relation subtile et bénéfique avec l'intuition, car les groupes ont souvent rapidement la bonne réponse mais doivent s'en convaincre ou éventuellement convaincre les autres. C'est cette partie-là qu'il conviendrait d'améliorer, en incitant par exemple à une rédaction plus structurée. Un groupe ayant des résultats dissonants et questionnant la méthode des autres aurait pu jouer le rôle de catalyseur mais il n'y en a pas eu dans cette expérimentation. Les groupes étaient soit sur la bonne voie, soit trop peu sûrs d'eux pour alimenter le débat. Le travail en groupes et la mise en commun des idées entre les groupes occupe quand même une place prépondérante et valorisante dans cette activité. Il convient donc d'y consacrer un temps raisonnable, d'une séance à chaque fois pour la manipulation et la simulation informatique.

Pour améliorer cette activité, nous pensons qu'il faudrait la rendre plus adaptable à l'avancement et aux difficultés éventuelles des différents groupes. Si cela est possible, on peut par exemple faire passer les élèves qui ont bien compris plus rapidement sur les ordinateurs, ou au contraire insister sur la rédaction claire d'une méthode de résolution. En préparant davantage de prolongements ou de questions annexes, on renforce ainsi son pouvoir de différentiation. On peut aussi proposer deux bouteilles aux contenus inconnus et leur demander la probabilité de tirer deux boules blanches. Avec un peu de chance ou d'astuce certains groupes étudieraient chaque bouteille séparément, et d'autres groupes la fréquence d'apparition de l'événement sur les deux bouteilles vues comme un ensemble. Cela serait une bonne manière de justifier la multiplication des probabilités sur un chemin d'un arbre pondéré.

Cette activité a en tous cas permis de créer une représentation mentale commune du concept de probabilité, à laquelle nous avons pu faire référence régulièrement. Dans le cours, nous avons ainsi fait noter le paragraphe ci-contre :

Intuition : (voir Activité)

- On retourne 10 fois la bouteille, on obtient 5 « bleu » et 5 « blanc »
- On retourne 100 fois la bouteille, on obtient 35 « bleu » et 65 « blanc »
- On retourne 1000 fois la bouteille, on obtient 382 « bleu » et 618 « blanc »

Plus on répète l'expérience un grand nombre de fois, plus la fréquence d'apparition des "bleu" tend vers 40% (50% → 35% → 38,2%). C'est ce qu'on appelle la loi des grands nombres.

Définition : A chaque évènement E, on peut associer une probabilité, notée p(E) : c'est la fréquence théorique obtenue en répétant un nombre infini de fois l'expérience.

II Echantillonnage : un problème de sex-ratio

1) Motivations

L'échantillonnage est un chapitre au statut quelque peu particulier dans le programme de la classe de seconde. A la fois nouveau pour les élèves, qui n'ont pas rencontré directement cette notion au collège, il l'est parfois aussi pour les enseignants puisqu'il est d'une introduction récente (2009) dans les programmes de mathématiques – les générations précédentes, nous y compris, n'ont donc pas été particulièrement formées sur ces aspects. Il est également difficile à situer puisqu'il se situe au carrefour de deux grands domaines : les statistiques et les probabilités. Il est d'ailleurs souvent nommé "statistique inférentielle" dans les articles de la littérature comme (Sotura, Janvier 2013) ou (Parnaudeau, Octobre 2013), car il s'agit d'inférer à partir d'une étude statistique sur un échantillon le degré de validité d'une hypothèse générale. Enfin, son contenu théorique, et donc souvent sa trace écrite sur le cahier des élèves, se limite presque à une unique formule pour l'intervalle de fluctuation – formule qui est en plus le produit d'une double approximation (Cerclé, Avril 2013) qui ne fait pas directement sens en classe de seconde. Il n'est donc pas exceptionnel d'entendre de façon plus ou moins informelle des collègues avouer en fin d'année qu'ils traiteront le sujet en guère plus d'une heure pour terminer le programme.

Il est naturel qu'un tel chapitre soulève de nombreuses questions de la part des enseignants, dont certaines sont relayées dans (Cerclé, Avril 2013). En outre, le chapitre est abordé à chaque niveau du lycée, mais avec une approche différente qui conduit à des définitions également différentes d'un intervalle de fluctuation. Il n'est pas toujours aisé de se rendre compte que "la problématique reste la même sur les trois niveaux et que seuls les moyens disponibles pour y répondre diffèrent : la simulation en classe de seconde, la loi binomiale en première, la loi normale en terminale." (Sotura, Janvier 2013)

C'est à la lumière de ce léger discrédit qu'il nous est apparu essentiel de mettre en œuvre une activité d'introduction particulièrement porteuse de sens, pour les élèves comme pour les enseignants, d'autant que l'échantillonnage ouvre justement un vaste champ d'applications concrètes et ludiques qui nécessite un raisonnement fin.

C'est dans cet esprit que nous avons été guidé par les recommandations de l'inspecteur Philippe Dutarte dans (Dutarte, Octobre 2013) : "L'enseignement au lycée de la notion d'intervalle de fluctuation est essentiel, tout d'abord dans le cadre de la prise de décision, ensuite comme préliminaire à l'investigation de la notion d'intervalle de confiance. Il faut faire comprendre « que toute décision s'accompagne d'un risque, mais que ce risque peut être évalué ». Dans cet enseignement, l'apport principal du professeur de mathématiques doit être de dégager la logique qui préside à la démarche statistique, les raisonnements mis en œuvre. On pourrait confier les détails des calculs à un logiciel (ou programmer un algorithme), là n'est pas l'essentiel. Partir de situations-problèmes « concrètes » est indispensable pour mettre en perspective les principes introduits, comprendre les définitions et les précautions d'utilisation (rôle de la zone de rejet, distinction entre la règle de décision et la prise d'échantillon, signification des pourcentages 95%, 5%, …). Il ne s'agit pas d'élaborer des théories mathématiques « à vide », dans un cadre purement probabiliste : à l'arbitraire de cette introduction s'ajouterait l'incapacité d'appliquer à bon escient ces concepts."

La notion de situation-problème, déjà rencontrée avec la bouteille de Brousseau, trouve donc dans ce chapitre une application naturelle qu'il convient de bien choisir.

Le programme de la classe de seconde et son document d'aménagement datant de 2017 précisent que les élèves doivent savoir :

- Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide d'un tableur et d'une calculatrice
- Exercer un regard critique sur l'information obtenue à partir d'un échantillon, notamment en faisant le lien entre la taille de l'échantillon et la largeur de l'intervalle de fluctuation

Plus loin, il est également précisé que "l'objectif est d'amener les élèves à un questionnement lors des activités suivantes :

- L'estimation d'une proportion p inconnue à partir d'un échantillon ;
- La prise de décision à partir d'un échantillon"

C'est ce dernier aspect – prise de décision – qui semble avoir le plus la capacité d'interpeller les élèves, de formuler un problème concret qui va nécessiter un recours au raisonnement mathématique. Après un rapide tour d'horizon de la littérature existante et des activités proposées dans les manuels, nous avons donc choisi de mettre en place l'activité autour du sex-ratio proposée dans (Sotura, Janvier 2013), qui précise que "cette activité est proposée dans les documents ressources de lycée professionnel (disponible sur le site Eduscol) et est issue des travaux de l'IREM Paris Nord (statistique et citoyenneté)."

2) Présentation et mise en place

Voici donc l'énoncé original de l'activité choisie : "Dans la réserve indienne d'Aamjiwnaang, située au Canada à proximité d'industries chimiques, il est né, entre 1999 et 2003, 132 enfants dont 46 garçons. Cette situation doit-elle être considérée comme anormale et interroger les autorités sanitaires ? (Le sexe-ratio canadien est d'environ 105 garçons pour 100 filles.)"

Il y aurait peu d'intérêt à détailler ici la résolution attendue par des élèves si cet énoncé était celui d'un exercice d'évaluation de fin de chapitre. Sans simulation ou explication préalable, il est d'ailleurs tout à fait impossible qu'un élève de seconde qui n'est pas déjà familier avec l'échantillonnage puisse répondre seul à cette question de façon parfaite, le seuil de 95% étant ici implicite et arbitraire. L'objectif de cette activité est bien d'amener les élèves à développer une démarche statistique, et ce faisant d'interroger leur notion de fréquence, d'échantillon, de "normalité".

Nous avons légèrement modifié l'énoncé, comme on peut le constater dans les différents documents que nous avons utilisés dans la séquence et présentés dans l'annexe 5. Nous avons fait notamment passer le nombre de garçons de 46 à 50, parce qu'après avoir simulé informatiquement ces échantillons, nous avons pensé qu'il était trop rare d'obtenir 46 garçons et que cela risquait d'inciter les élèves à répondre "ce n'est pas possible", et donc de contourner le questionnement central autour de la fréquence d'apparition. 50 reste en dehors de l'intervalle de fluctuation au seuil de 95% mais s'obtient au moins quelques fois sur 1000 échantillons par exemple.

Nous avons conçu la séquence autour de l'échantillonnage en quatre temps, en posant cette situation-problème non seulement comme une introduction à la séquence mais aussi comme un fil rouge qui a permis de dérouler les différents aspects du chapitre en faisant constamment référence à cette activité. Les quatre temps sont, dans l'ordre chronologique et celui des documents de l'annexe 5 :

- Un débat en classe pendant environ une demi-heure sur cette situation-problème, en essayant de reproduire au moins en partie l'expérience relatée dans (Parnaudeau, Octobre 2013).
- Une séance en demi-groupe et en salle informatique pour répondre à la question posée en s'appuyant sur les données de la simulation et en observant la fluctuation d'échantillonnage
- Une séance de travail sur l'influence de la taille d'un échantillon en modifiant la situation proposée, avec une vérification empirique de la formule proposée dans les programmes. La séance se conclut par une synthèse du cours reprenant la définition préconisée dans (Cerclé, Avril 2013)
- Une ou plusieurs séances d'exercices du manuel sur des situations similaires de prise de décision.

On remarquera également plusieurs différences notables avec l'activité proposée par (Sotura, Janvier 2013), notamment dans le TP qui constitue le cœur de la mise en œuvre de la démarche statistique. Nous n'avons par exemple pas introduit de raisonnement sur la fréquence de naissances de garçons avant la toute dernière question, qui de fait n'a pas été traitée par tous les groupes. Après l'expérience de la bouteille de Brousseau qui avait montré une difficulté sur le passage à la fréquence, et au vu du retour d'expérience de l'article en question, nous avons préféré ne pas rajouter cette étape, et se concentrer sur le principe de l'échantillonnage et de l'intervalle de fluctuation en raisonnant directement sur le nombre (et non la fréquence) de garçons nés dans cette réserve. Nous avons aussi laissé tomber, pour les mêmes raisons, l'information finalement assez technique concernant le sexe-ratio canadien.

L'auteure avait demandé aux élèves de produire des échantillons de taille 132 grâce à des bouteilles de Brousseau bien choisies vues comme des machines à produire du hasard. Cela aurait été l'occasion rêvée de reprendre la manipulation déjà effectuée pour l'appliquer dans un autre contexte, et c'est d'ailleurs ce qui avait aussi conduit à ce choix d'activité dans un premier temps. Cependant, étant donné qu'amener les élèves à faire ne serait-ce que 20 tirages lors de l'activité sur cette bouteille avait été déjà compliqué, il nous paraissait utopique, y compris pour des questions de gestion de classe, de leur demander de réaliser chacun 132 tirages. Nous sommes donc directement passés sur tableur pour la simulation d'échantillons. La vision fréquentiste enracinée par l'expérience de la bouteille a quand même été d'une grande utilité dans ces activités, notamment dans celle où l'on constate que la taille de l'échantillon (le nombre de tirages) réduit l'amplitude des fluctuations (convergence vers une valeur limite).

3) Analyse de la séquence

L'analyse qui suit s'appuie sur les documents de l'annexe 5.

Introduction du problème

La présentation du problème a été suivie comme prévu d'une discussion en classe entière pendant une demi-séance. Elle fut très fructueuse même si un peu décousue et qu’il fallait parfois répéter les avis pertinents pour que toute la classe entende. Il y a tout de suite eu une tension intéressante entre le « c’est le hasard » intuitif et la mention d'industries chimiques dans l'énoncé. Au passage, il a fallu expliquer que des substances présentes dans notre alimentation ou notre environnement, et notamment les perturbateurs endocriniens, étaient susceptibles d'influer sur le sex-ratio. Certains élèves sont restés longtemps convaincus, indépendamment du problème mathématique, que l'occurrence de filles et de garçons ne peut être que le fruit du hasard. Peut-être que cette discussion annexe aurait pu être évitée en choisissant par exemple d'étudier un taux de maladie à proximité d'une centrale nucléaire. Il faudrait dans ce cas mentionner le taux normal d'apparition de la maladie, ce qui rend l'énoncé moins élégant.

La classe était à peu près partagée en deux lorsqu'il leur a été demandé intuitivement s'ils pensaient que c'était normal, c'est-à-dire que le hasard expliquait bien cette situation. Il y a eu consensus sur le fait que le hasard était en mesure d'expliquer cette situation, mais que ce n'est pas forcément très probable. Le débat s'est ensuite enlisé quelque temps dans la confrontation d'avis qualitatifs pas du tout étayés dont la force de persuasion résidait uniquement dans la conviction et la réputation de celui qui l'émettait. Le calcul d'une fréquence a fini par venir naturellement d'un élève, qui a dit que 37% (valeur approchée de $\frac{50}{132}$) "c'était pas loin de 50%", ce qui a eu l'air de convaincre plusieurs personnes. Nous avons demandé si "du coup 37% de garçons à l'échelle de la France c'était normal ?", ce à quoi la classe a répondu en cœur "non". Ils ont ainsi dégagé le fait que le nombre de naissances étudiées avait aussi son importance.

Quand nous avons demandé comment s'appuyer sur des données chiffrées pour trancher les différents avis, un élève a naturellement pensé à comparer les chiffres dans les autres réserves. Plus précisément, il a fallu faire apparaître le fait qu'il fallait comparer avec des populations non exposées aux industries en question. Quand nous avons demandé quels outils on pouvait utiliser si on ne disposait pas de telles données, la simulation informatique est apparue, puis, en imposant la contrainte "et sans ordinateur ?" le pile ou face a été mentionné.

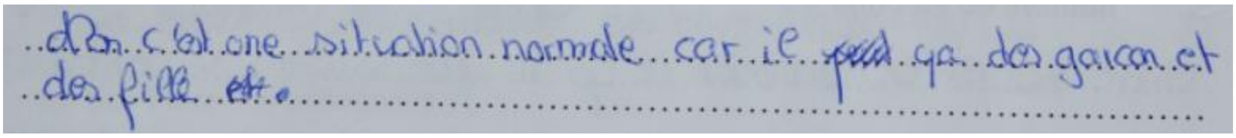
Cette activité introductive a donc été une très bonne manière de construire avec les élèves les différentes intuitions de la statistique inférentielle : écart à une valeur de référence, influence de la taille de l'échantillon considéré, recours à la simulation, etc. Le constat sur le passage à l'écrit est cependant resté le même que pour la bouteille de Brousseau : beaucoup d'élèves se sont contentés de réfléchir à l'oral, et après avoir bien insisté, ont fini par noter quelques explications orales de l'enseignant qu'ils demandaient de répéter mot à mot comme s'ils en reportaient l'étude du sens à plus tard. Malgré le manque de rédaction, la feuille distribuée a été collée dans le cahier de cours (c'est la raison pour laquelle il n'y en a pas d'extraits dans ce document), car nous souhaitions qu'ils gardent trace, même maladroite, de la démarche utilisée.

TP informatique

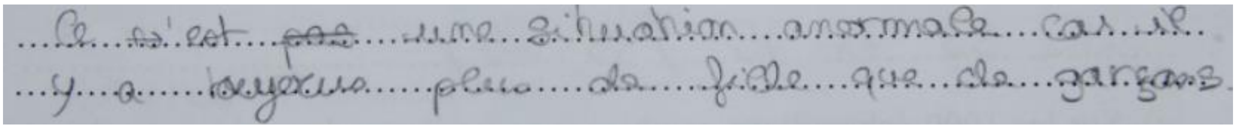
Le TP informatique a joué dans cette séquence un rôle plus important encore que pour les probabilités où il avait servi de validation à une intuition construite grâce à la manipulation. Ici, c'est véritablement la simulation informatique qui permet de répondre à la question posée, et même de justifier l'existence et la formule donnée dans le cours de l'intervalle de fluctuation. Nous y avons donc accordé une importance particulière.

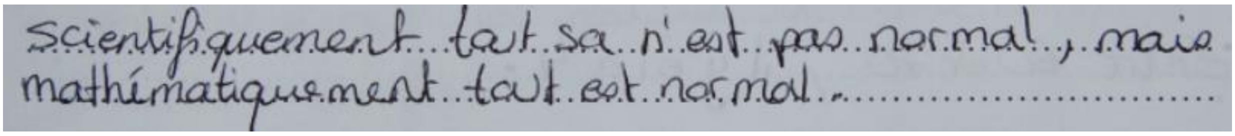
Ce TP, réalisé à chaque fois en demi-groupe pendant une séance, s'est globalement très bien déroulé. Nous avions repris la trame utilisée dans (Sotura, Janvier 2013), en la modifiant un peu et en redessinant les graphiques. Nous remercions d'ailleurs une idée passée inaperçue au premier abord, celle d'inclure les graphiques résultats dans l'énoncé de sorte qu'un élève qui n'arriverait pas à obtenir ces graphiques puisse répondre quand même aux questions mathématiques qui s'y rapportent. En effet, réaliser 1000 échantillons de 132 nombres aléatoires est à la portée d'un ordinateur d'enseignant stagiaire, mais pas toujours à celle d'un poste élève dans l'établissement. Les fonctionnalités Excel comme SI et NB.SI ont nécessité un rappel conséquent mais n'ont pas posé de problème particulier. L'idée d'une situation anormale car produite trop rarement par la simulation aléatoire a été bien comprise mais encore une fois formulée de manière très inégale.

Concentrons-nous ainsi sur la question 6, qui est celle où il faut répondre au problème une fois la simulation effectuée, et balayons le spectre des réponses des élèves. Tout d'abord, pour des raisons qui ne tiennent peut-être pas uniquement au cours, quelques élèves ne sont pas rentrés dans le problème :


donc c'est une situation normale car il ~~peut~~ ya des garcon et des fille ~~et~~.

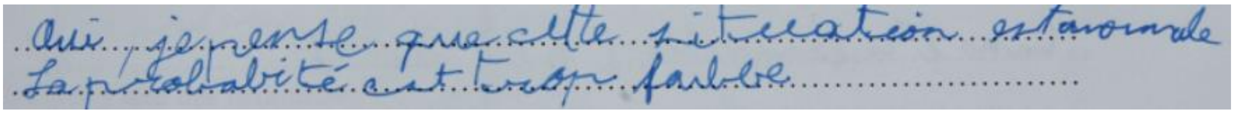
Ou lorsque l'on corrige pour mettre la réponse des autres sans que cela fasse sens :

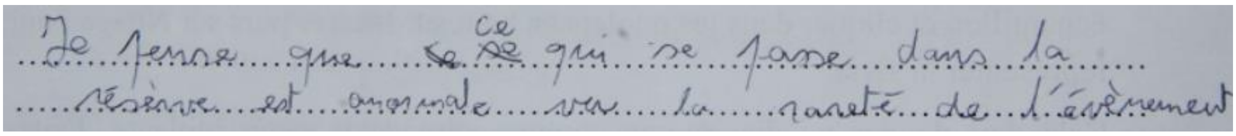

Ce ~~n'~~est ~~pas~~ une situation anormale car il y a toujours plus de fille que de garçons


Scientifiquement tout sa n'est pas normal, mais mathématiquement tout est normal.

Malgré son aspect contradictoire, la phrase précédente est éclairante sur la conception des mathématiques par certains élèves en difficulté : les mathématiques ne seraient pas une science, ou plutôt ils seraient une science si confuse que l'arbitraire y est parfois légitime. Peut-être est-ce à rapprocher de l'expression que l'on peut entendre dans les analyses de classements sportifs par exemple : "mathématiquement, c'est encore possible", alors que l'enjeu c'est précisément de faire comprendre que "mathématiquement, c'est très peu probable".

D'autres élèves, plutôt à l'aise en mathématiques, ne prennent pas la peine de justifier leur assertion par la simulation menée :


Oui je pense que cette situation est anormale
La probabilité est trop faible


Je pense que ~~se~~ ce qui se passe dans la
réserve est anormale vu la rareté de l'événement

Pour beaucoup d'élèves, 50 garçons n'est pas apparu dans les premiers échantillons (c'est avant de produire et d'analyser le résultat des 1000 échantillons), leur réponse ne contient donc pas de référence à une probabilité ou une rareté mais sous-entend plutôt une impossibilité. C'est ce que nous avions voulu éviter en ramenant la valeur de 46 à 50 mais ce n'était peut-être pas suffisant :

~~Non~~ Oui c'est anormale car 50 garçon n'appart pas à notre intervale de naissance suite à notre simulation.

Cette situation parait anormal car les valeur trouver lors des simulation ne dessendent pas a 50 garçons sur 132 naissance

Oui c'est une situation anormale car les valeurs n'atteignent pas 50 elle varie entre 58 et 81

Enfin, des élèves ont une justification tout à fait correcte même si elle pourrait contenir plus de détail :

Cela peut être du hasard mais sur 1000 il y a tres peu de cas alors nous pouvons en deduire que le hasard ne fait pas tout car cette situation est anormal

C'est anormal car le nombre de garçons (50) arrive rarement avec le hasard

Ce n'est pas une situation normale car en relançant plusieurs fois le calcul on tombe rarement sur le nombre 50

Dans l'ensemble, le constat est satisfaisant.

Nous avons par contre constaté les mêmes difficultés que (Sotura, Janvier 2013), prévisibles, quant au lien entre les deux représentations graphiques. Le premier graphique est non structuré, il montre le nombre de naissances de garçons pour chaque échantillon pris dans l'ordre d'apparition dans la simulation, il ne pose en général aucun problème. Le deuxième graphique a regroupé tous les échantillons qui correspondent à un même nombre de garçons. Alors qu'il apparaît sous la forme valeur/effectif familière aux élèves depuis le chapitre de statistiques, c'est pourtant le plus difficile à comprendre puisqu'il est le produit d'un post-traitement de la simulation, une fois tous les échantillons obtenus. Nous avons fini par pencher la feuille, pour faire apparaître comme des bâtons la série de points du premier graphique, et cette technique a permis de mieux faire comprendre le lien.

Il est à noter que ce type de représentation n'est pas fondamental en classe de seconde, mais prépare le terrain pour l'étude de la loi binomiale qui sert de support à l'échantillonnage en classe de première. Le nombre de garçons d'un échantillon est alors considéré comme une variable aléatoire obtenue par la répétition de n expériences identiques et indépendantes (une naissance) qui ont toutes une probabilité p (le sex-ratio) de donner un garçon. Il est donc naturel de ne le traiter que si le reste a été bien compris.

Taille de l'échantillon

Quant à l'activité sur l'influence de la taille de l'échantillon, que nous avons nous-même créée de toutes pièces, elle a particulièrement bien marché lorsque nous l'avons proposée en classe entière. Les deux premières questions, formulées de manière identique, amènent naturellement des réponses intuitives (normal pour 10 naissances, anormale pour un million) et incitent rapidement à désigner comme responsable le seul paramètre qui diffère, la taille de l'échantillon. Néanmoins, si les élèves se sont approprié ce constat sans difficulté, le "on voit bien pourquoi" a eu du mal à se transformer en une explication convaincante. A moins de tracer l'amplitude de l'intervalle de fluctuation en fonction de la taille de l'échantillon, ce qui donne un graphique complexe à comprendre, on en est réduit à juxtaposer des graphiques pour plusieurs tailles d'échantillons, et c'est ce qui est fait dans l'activité. Le lien avec la convergence des fréquences vues avec la bouteille de Brousseau est peu apparu naturellement mais il a été bien compris une fois l'explication appuyée sur les graphiques.

Les questions 6 et 7 sous la forme "peut-on affirmer que c'est uniquement à cause du hasard ?" avaient pour but de faire comprendre le sens du seuil de 95% (pas encore explicité d'ailleurs). En effet, cette question est particulièrement subtile, comme cela avait été souligné dans (Cerclé, Avril 2013) et (Dutarte, Octobre 2013). Le cas qui nous occupe de la prise de décision est en réalité un test d'hypothèse, celle de l'hypothèse nulle : on teste si le hasard seul explique raisonnablement la valeur observée. Ce qui est important, c'est que 95% ne correspond pas à la probabilité que l'hypothèse testée soit vraie si la valeur observée est dans l'intervalle de fluctuation mais bien qu'on prend un risque (de première espèce) de seulement 5% de rejeter l'hypothèse lorsque la valeur observée est en dehors de cet intervalle. Sans le formuler de la sorte aux élèves, il faut leur expliquer qu'il ne s'agit pas de répondre de manière manichéenne "oui, le hasard est seul responsable", ou "non, il n'est pas seul responsable", mais plutôt s'il est raisonnable de considérer que cela peut être vrai. Cela donne lieu à des formulations à double négation comme "sur la base de cet échantillon, il n'y a pas lieu de considérer que le hasard ne peut expliquer seul cette situation", ou au contraire "il est raisonnable de penser que le hasard n'est pas la seule cause à cette situation". Elles ne sont pas toujours très élégantes mais sont les seules à refléter fidèlement les conclusions que l'on a tirées par l'étude statistique. Nous avons ensuite veillé à conserver cette subtilité et ces formulations ("il faut/ne faut pas remettre en cause…") dans les exercices et les évaluations. Les questions 6 et 7 de cette activité sont donc l'occasion de déclencher ce débat important.

Enfin, aux questions 8 et 9, lorsqu'il faut constater à la calculatrice que les valeurs $p-\frac{1}{\sqrt{n}}$ et $p+\frac{1}{\sqrt{n}}$ correspondent à peu près à l'intervalle trouvé graphiquement aux questions précédentes, les élèves ne voient pas vraiment – et à juste titre – ce qu'il faut constater, quels liens logiques ces nombres entretiennent. Et pour cause, c'est le moment où l'enseignant fait accepter, essentiellement par un argument d'autorité qui ne peut être expliqué en classe de seconde, que la formule proposée convient et peut être généralisée. Nous ne savons pas s'il existe un moyen simple de contourner cette rupture dans la démarche, qui du raisonnement à la simulation, se condense brutalement dans une formule au sens un peu mystérieux. Peut-être faudrait-il insister auprès des élèves sur la simplification d'usage qu'elle apporte ou sur l'universalité de son écriture pour mieux la justifier. Nous constatons cependant qu'une fois cette étape franchie et la formule synthétisée dans le cours, elle ne pose pas de problème aux élèves lors de la résolution d'exercices. Certains s'y raccrochent d'ailleurs comme une vérité immuable dans un univers où le raisonnement vous mène à des conclusions changeantes. C'est pour cette raison qu'il ne faut introduire cette formule, et le seuil arbitraire des 95%, qu'à la toute fin de la démarche pour qu'elle ne cache pas l'essentiel.

4) Bilan

Dans l'ensemble, l'activité choisie, modifiée et déclinée en plusieurs activités et sur plusieurs supports, a parfaitement répondu aux attentes que nous avions placées en elle, à la fois quant à sa capacité à "parler" aux élèves et quant à sa capacité à faire émerger les questions cruciales qui sont relatives à l'échantillonnage, à construire la démarche idoine.

Nous avons également montré que l'utilisation de la fréquence peut très bien se faire dans un deuxième temps, après avoir raisonné sur une quantité plus facilement accessible à la représentation. Elle intervient alors comme un moyen de comparer des quantités relatives à des tailles différentes d'échantillon. C'est d'ailleurs la fréquence qui déclenche le débat sur l'influence de la taille de l'échantillon dans notre activité du même nom, en ce qu'elle permet de normaliser le calcul et donc la formulation (40% pour 10 naissances / 40% pour 1 000 000 de naissances). Il serait intéressant d'essayer dans cette activité d'expliciter l'utilité de ce passage ultérieur à la fréquence, voire de l'intégrer à la réflexion autonome des élèves.

Reprenons dans le tableau suivant les critères listés dans (Dutarte, Octobre 2013) (voir aussi fiche de lecture en Annexe 4) pour juger a posteriori de la pertinence de ces activités :

Critères de P. Dutarte	Adéquation avec l'activité proposée
"– Le premier critère nous semble devoir être de poser un problème, une situation, ayant quelque écho avec le monde réel. Cette situation doit être telle que la proportion *p* dans la population est supposée connue. Cela exclut les situations où l'on n'a aucun *a priori* sur la valeur de *p* (type « sondage ») […]"	Ce critère est parfaitement vérifié, avec l'élégance supplémentaire que la proportion p est implicite et connue de tous (50% de filles et de garçons).
"– Il faut ensuite que cette situation soit de type « bilatérale » : par rapport à la valeur de p « visée » (celle de l'hypothèse) on s'intéresse à un écart trop important à gauche et à droite."	Le problème n'est pas posé de manière bilatérale, c'est l'écart trop important à gauche que l'on doit juger, mais il est clairement symétrique, et les élèves comprennent que la démarche serait similaire s'il fallait qualifier un nombre trop élevé de garçons.
– La règle de décision doit être élaborée, autant que possible, avant la prise d'échantillon. Il est plus honnête de décider d'une règle avant de jouer, qu'après la partie. […]	Ce critère n'est pas respecté mais il est ambitieux pour une activité d'introduction : imposer une règle (seuil de 95% par exemple) serait inscrire l'activité dans un cadre plus contraint, moins propice à l'intuition et au tâtonnement. Il faudrait idéalement avoir un débat

	préalable avec les élèves sur la notion de "normalité" et se fixer une règle dans ce cadre mais il nous semble que cela alourdirait le déroulement de l'activité sans grand bénéfice. Il nous semble plus opportun de discuter cette notion a posterio, une fois avoir fait appel à l'intuition normal/anormal, en dégageant alors une règle de décision (un seuil) qui pourrait s'appliquer à toutes les situations de ce type. Ce critère est peut-être plutôt un prolongement possible de l'activité.
– L'échantillon doit être prélevé par tirage au hasard avec remise (équiprobabilité garantie par randomisation).	Ce sont les échantillons de référence (tirage au hasard dans la population canadienne) qui vérifient cette propriété, on cherche si l'échantillon de la réserve est conforme à ce tirage.
– En cas de rejet de l'hypothèse, il faut savoir que l'on peut interpréter le 5% comme la probabilité de commettre une erreur de décision (probabilité de rejet de l'hypothèse sachant qu'elle est vraie). En cas d'acceptation de l'hypothèse, il faut savoir que l'estimation de la probabilité d'erreur de décision est plus compliquée.	Critère vérifié dans l'activité sur la taille de l'échantillon, comme cela a été discuté dans la partie précédente.

L'activité n'était pas rigoureusement une situation-problème puisque les élèves n'ont pas forcément tous les outils en main pour y répondre dès le début, et que la validation se fait surtout par l'enseignant, mais elle en reproduit les avantages : concret, ouvert, surprenant, sujet à débat, etc…

Nous remarquons également qu'un simple problème de ce type, et il en existe beaucoup d'autres dans le genre, est suffisamment riche pour être traité sur de multiples supports (y compris informatique) et pour servir de trame à la séquence entière. C'est aussi efficace, et c'est peut-être le plus important, puisque la bonne compréhension acquise s'est transposée en une bonne aisance sur les exercices type du manuel et en bons résultats d'ensemble sur l'évaluation qui a suivi.

Conclusion

Permettons-nous de citer le rapport intitulé "21 mesures pour l'enseignement des mathématiques" remis au ministre en février 2018 par une mission pilotée par le célèbre mathématicien Cédric Villani et l'inspecteur général Charles Torossian :

"Notre société abonde d'informations chiffrées. Les mathématiques apportent des outils essentiels à l'exercice d'une citoyenneté active. L'absence ponctuelle de regard critique et le manque de temps pour analyser une donnée numérique (pourcentages, sondages, lectures graphiques, statistiques, etc. dont les médias sont friands) nous rendent particulièrement vulnérables aux tentatives de manipulation.

Les mathématiques sont nécessaires à la démocratie parce qu'elles favorisent l'autonomie et la capacité d'innovation. Il est ainsi indispensable de fournir aux enfants tous les outils de logique, de calcul, de développer chez eux l'intuition et la démarche scientifique, la rigueur et sa nécessité, et enfin de leur permettre de mener des raisonnements et d'élaborer des preuves."

Puis un peu plus loin : "La lecture et l'analyse de graphiques et la compréhension des statistiques ont souvent été présentées à la mission comme essentiel à la formation du citoyen, notamment à travers l'usage que l'on fait des proportions, du sens que l'on donne aux indicateurs de position (médiane, moyenne) ou de dispersion, pour justifier une décision sociale ou politique."

C'est pour ces raisons – la formation de citoyens, la construction de repères intellectuels dans un monde qui se présente de plus en plus sous forme de données chiffrées – que nous pensons également que des chapitres comme les probabilités et peut-être à plus forte raison l'échantillonnage sont essentiels à l'enseignement des mathématiques.

Paradoxalement c'est leur côté ludique et concret qui désarçonnent parfois les enseignants, en particulier les stagiaires puisqu'ils sortent en général d'études où formalisation, abstraction et rigueur sont les maîtres mots. Pourtant ces aspects sont tout à fait conciliables, au moins ponctuellement. On pourrait même soutenir que c'est en assumant complètement leur côté ludique et concret, en laissant place à la manipulation, à l'expérimentation, à la simulation, que l'on peut ensuite dégager une abstraction rigoureuse et utile.

Nous espérons en tous cas que les activités présentées dans ce mémoire apporteront leur eau à ce moulin par leur valeur d'exemple. Bien sûr, ces activités ne sont ni parfaites ni incontournables – c'est l'essence même de notre métier que de veiller à les améliorer ou à les renouveler chaque année – et ne sont qu'un outil parmi d'autres au service de la compréhension des élèves, mais nous avons été séduit par leur philosophie de situation-problème, par leur méthode de recherche active judicieusement appuyée sur les outils informatiques et par leur efficacité à introduire puis consolider des concepts importants.

Il nous semble qu'il faudrait chercher à généraliser cette approche autant que possible, y compris sur des chapitres d'apparence plus formels, et ce sera l'une des pistes que nous poursuivrons au cours de cette enthousiasmante carrière que nous venons d'embrasser.

Bibliographie

Cerclé, V. (Avril 2013). Quelques interrogations du professeur de lycée autour des intervalles de fluctuation.

Dutarte, P. (Octobre 2013). Du bon usage d'un intervalle de fluctuation. (APMEP n°505).

Groupe IREM "Ressources TICE pour la formation et l'enseignement". (2015). L'utilisation du tableur dans une situation d'introduction aux probabilités au collège - La bouteille de Brousseau revisitée pour la formation auxx TICE. *IREM Paris VII*.

Parnaudeau, J.-M. (Octobre 2013). Statistiques inférentielles : un débat scientifique en classe de seconde. (APMEP n°505).

Sotura, B. (Janvier 2013). Une activité pour initier à la statistique inférentielle en classe de seconde. (APMEP n°502).

Annexe 1 : Expérience de la bouteille de Brousseau –
Document personnel pour en guider le déroulement et photos du dispositif

---

Modalités : En demi-groupe, donc 13 ou 14 élèves en tout, répartis en groupes de 3 ou 4, dans une salle qui contient des ordinateurs. Un rapporteur est désigné dans chaque groupe.

Consigne à l'oral : il y a 5 billes dans chaque bouteille. Combien y-a-t-il de billes de chaque couleur dans la bouteille ? Expliquer votre démarche par écrit.

Réflexion en autonomie (15 minutes). Le professeur vient guider.

Mise en commun (10 minutes)

- Tout le monde est-il d'accord sur le résultat ? Si non, discussion. Est-ce la même méthode ? Peut-on savoir qui a raison ?
- Chaque groupe explique sa démarche.
- Expliciter si nécessaire le rapport entre la proportion (ou fréquence) et le nombre de billes.
- Réflexion autour du degré de certitude.

Passage à l'ordinateur (restant de la séance)

Annexe 2 : Expérience de la bouteille de Brousseau –
Enoncé du TP informatique distribué aux élèves

Nom : …………………   Prénom : ……………………

Attention : votre feuille de calcul est à sauvegarder à la fin de l'heure dans le dossier 2GT1/Rendu/Maths

- Ouvrez le fichier "Activité Simulation Bouteille v2". Cette feuille de tableur va permettre de simuler et d'analyser les résultats obtenus avec l'expérience de la bouteille.

La colonne A correspond au nombre de tirages effectués.
La colonne B enregistre le résultat obtenu : Blanc ou Bleu

Les colonnes D et E correspondent au nombre de fois qu'on a tiré bleu et blanc depuis le début de l'expérience.

Les colonnes G et H correspondent à la fréquence d'apparition du bleu et du blanc, c'est-à-dire la proportion de bleu tirés parmi l'ensemble des tirages. On l'exprimera en pourcentage.

Les colonnes I et J permettront d'estimer et prédire le nombre de billes bleues dans la bouteille

	A	B	D	E	F	G	H	I	J
1	Nb Tirage	Tirage	Nombre Bleus	Nombre Blanc		Fréquence Bleu	Fréquence Blanc	Estimation Nombre de boules bleues	Prédiction Nombre de boules bleues
2	1	Blanc	0	1					
3	2	Bleu	1	1					
4	3	Blanc	1	2					
5	4	Bleu							
6	5	Bleu							

- Complétez la cellule G2 par une formule qui permet de calculer la fréquence d'apparition du bleu d'après les résultats des autres colonnes. Complétez de la même façon la cellule H2 avec la fréquence d'apparition du blanc.

- Utilisez la poignée de glissement en bas à droite de chaque cellule sélectionnée pour copier vos formules aux trois premières cellules des colonnes G et H.

On souhaite que le tableur complète automatiquement les colonnes D et E pour prendre en compte les nouveaux résultats. Pour cela, on aimerait compter le nombre de fois que le mot "Bleu" apparaît dans la colonne B.

Voici un extrait de l'aide du tableur à propos de la fonction NB.SI :

> NB.SI, l'une des fonctions Statistiques, permet de compter le nombre de cellules qui répondent à un critère ; par exemple, pour compter le nombre de fois où le nom d'une ville apparaît dans une liste de clients.
>
> Dans sa forme la plus simple, la fonction NB.SI se décompose ainsi :
>
> =NB.SI(où voulez-vous rechercher ?;que voulez-vous rechercher ?)
>
> Par exemple :
>
> =NB.SI(A2:A5;"Londres")

- A l'aide de cette fonction NB.SI, compléter la cellule D5 pour qu'elle indique le nombre de "Bleu" obtenu depuis le début. Faire de même avec "Blanc" en cellule E5.

- Sélectionnez les cellules D5 et E5 et copiez leurs formules sur la ligne suivante avec la poignée de déplacement. Vérifiez que le résultat est cohérent.
  N'oubliez pas que le symbole $ permet de figer une cellule dans une formule qu'on copie en glissant.

- Remplissez la cellule I2 par une formule qui permet d'estimer à chaque ligne le nombre de billes bleues dans la bouteille. On utilisera pour cela la fréquence des boules bleues en colonne G et le nombre total de billes en cellule N1

- Complétez la cellule J2 par une formule qui arrondit le calcul effectué en I2 : ici, « =ARRONDI(I2 ;0) »

- Faites glisser les formules de I2 et J2 aux colonnes suivantes et vérifier que les résultats sont cohérents.

En sélectionnant la dernière cellule de la première colonne et en appuyant sur les touches "**Ctrl + B**", le tableur simule un tirage supplémentaire.

	A	B	D	E	F	G	H	I	J
1	Nb Tirage	Tirage	Nombre Bleus	Nombre Blanc		Fréquence Bleu	Fréquence Blanc		Prédiction Nombre de boules bleues
2	1	Blanc	0	1		0,00%	100,00%	0	0
3	2	Blanc	0	2		0,00%	100,00%	0	0
4	3	Blanc	0	3		0,00%	100,00%	0	0
5	4	Bleu	1	3		25,00%	75,00%	1,25	1
6	5	Bleu	2	3		40,00%	60,00%	2	2

- En restant appuyé sur ces touches, réaliser 100 tirages. Quelle est la fréquence bleue obtenue ? Combien peut-on prédire de boules bleues dans la bouteille ? Cette prédiction est-elle stable ? Et si oui, à partir de combien de tirages ?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

- Appuyer sur le bouton "Recommencer les tirages" en haut de la page, et réaliser à nouveau 100 tirages. Les résultats sont-ils identiques ? Les résultats sont-ils proches ?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

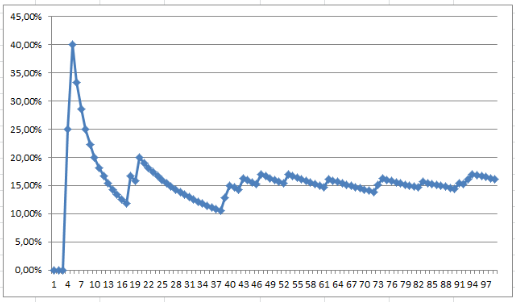
……………………………………………………………………………………On souhaite tracer la courbe des fréquences obtenues après chaque tirage.

	A	B	D	E	F	G	H	I	J
1	Nb Tirage	Tirage	Nombre Bleus	Nombre Blanc		Fréquence Bleu	Fréquence Blanc		Prédiction Nombre de boules bleues
2	1	Blanc	0	1		0,00%	100,00%	0	0
3	2	Blanc	0	2		0,00%	100,00%	0	0
4	3	Blanc	0	3		0,00%	100,00%	0	0
5	4	Bleu	1	3		25,00%	75,00%	1,25	1
6	5	Bleu	2	3		40,00%	60,00%	2	2
7	6	Blanc	2	4		33,33%	66,67%	1,66666667	2
8	7	Blanc	2	5		28,57%	71,43%	1,42857143	1
9	8	Blanc	2	6		25,00%	75,00%	1,25	1
10	9	Blanc	2	7		22,22%	77,78%	1,11111111	1
11	10	Blanc	2	8		20,00%	80,00%	1	1
12	11	Blanc	2	9		18,18%	81,82%	0,90909091	1
13	12	Blanc	2	10		16,67%	83,33%	0,83333333	1
14	13	Blanc	2	11		15,38%	84,62%	0,76923077	1
15	14	Blanc	2	12		14,29%	85,71%	0,71428571	1
16	15	Blanc	2	13		13,33%	86,67%	0,66666667	1
17	16	Blanc	2	14		12,50%	87,50%	0,625	1
18	17	Blanc	2	15		11,76%	88,24%	0,58823529	1
19	18	Bleu	3	15		16,67%	83,33%	0,83333333	1
20	19	Blanc	3	16		15,79%	84,21%	0,78947368	1
21	20	Bleu	4	16		20,00%	80,00%	1	1
22	21	Blanc	4	17		19,05%	80,95%	0,95238095	1
23	22	Blanc	4	18		18,18%	81,82%	0,90909091	1
24	23	Blanc	4	19		17,39%	82,61%	0,86956522	1
25	24	Blanc	4	20		16,67%	83,33%	0,83333333	1
26	25	Blanc	4	21		16,00%	84,00%	0,8	1
27	26	Blanc	4	22		15,38%	84,62%	0,76923077	1

K	L	M	N	O	P	Q
		Nombre total de billes :	5			

Recommencer les tirages

Changer la répartition des billes et recommencer



- Sélectionnez les résultats de fréquences "Bleu" en colonne G et cliquez sur l'onglet "Insertion" dans le menu du haut puis "Ligne". Que constate-t-on ?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

………………………………………………………………………………
……………………………………………………………………………..

On souhaite maintenant changer le nombre de billes dans la bouteille.

- Changer la valeur de la cellule N1 pour qu'il y ait 25 billes dans la bouteille.
  Cliquez sur le bouton « changer la répartition des billes et recommencer »
  Effectuer à nouveau un grand nombre de tirages (toujours avec « les touches Ctrl + B »). A partir de combien de tirages la prédiction semble-t-elle se stabiliser ? Expliquer la différence

………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
……………………………………………………………………………..

- Changer maintenant la valeur de la cellule N1 pour qu'il y ait 100 billes dans la bouteille.
  Cliquez sur le bouton « changer la répartition des billes et recommencer »
  Effectuer à nouveau un grand nombre de tirages (toujours avec « les touches ctrl + b »). A partir de combien de tirages la prédiction semble-t-elle se stabiliser ?

………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………

Annexe 3 : Productions des élèves pour l'expérience de la bouteille de Brousseau

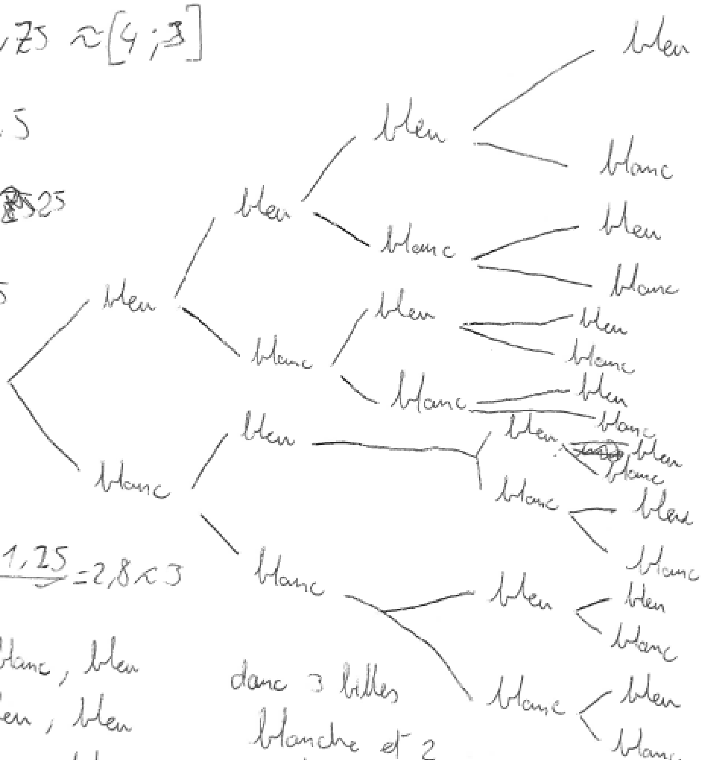
$\frac{3}{4} \times 5 = 3{,}75 \approx [4;3]$

$\frac{2}{4} \times 5 = 2{,}5$

$\frac{1}{4} \times 5 = 1{,}25$

$\frac{3}{4} \times 5 = 3{,}75$

5 billes



$\frac{3{,}75 + 3{,}75 + 2{,}5 + 1{,}25}{4} = 2{,}8 \approx 3$

blanc / bleu, blanc, bleu
bleu / blanc / bleu / bleu
blanc / blanc / bleu / bleu
blanc / blanc / blanc / bleu

donc 3 billes blanche et 2 billes bleu

1 bleu, blanc, bleu, bleu, blanc
2 blanc, blanc, bleu, blanc, bleu
3 blanc, bleu, blanc, blanc, bleu
4 blanc, bleu, bleu, blanc, bleu
5 blanc, blanc, blanc, blanc, bleu
6 bleu, blanc, blanc, bleu, blanc

13 bleu
17 blanc

Fréq bleu $= \frac{13}{30} \approx 0{,}4$

Fréq blanc $= \frac{17}{30} \approx 0{,}6$

2GT1 Mariame
Laetitia
Soraya

On a fait le geste et on a obtenue 4 fois sur 6 fois de billes blanche donc on pense qu'il y'a plus de billes blanche que de bille bleu. on va vous le démontrer!:

nombre de geste	1	2	3	4	5	6
couleur de billes	blanc	blanc	bleu	bleu	blanc	bleu

↑ Donc il y a vraiment plus de bille blanche que de bleu.

en applicant une autre méthode ↓:

× 50

$$\begin{array}{r} 34 \\ +16 \\ \hline 50 \end{array}$$

blanc : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | = 34

bleu : | | | | | | | | | | | | | | | | = 16

Donc : • $34 \div 50 = 0,68 \rightarrow \frac{68 \times 5}{100} = 3,4 \simeq 3$

• $16 \div 50 = 0,32 \rightarrow \frac{32 \times 5}{100} = 1,6 \simeq 2$

Donc on pense qu'il y a 2 billes bleu et 3 billes blanche.

Souleymane meite.
Ryann MBIAKOP

4 blanche 1 bleu. par possible. X 5

3 blanche 2 bleu X 3

3 blanche 2 bleu

2 blanche 3 bleu X. 2

~~X4 blanche 1 bleu~~

la boule blanche

on a tire 50 fois.
on trouve 33 boule blanche
on trouve 17 boule bleu.

$\frac{33}{50} = 0{,}66 \times 100 = 66\%$

$\frac{17}{50} = 0{,}34 \times 100 = 34\%$

$66\% \times 5 = 3{,}3 \simeq 3$

$34\% \times 5 = 1{,}7 \simeq 2$

on peut en deduire que la nombre de boule de blanche est superieur a celui de boule bleu.

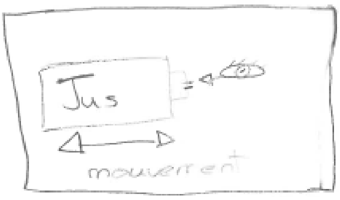
il y a 3 blanche et 2 bleu

Dabkcia - Jennifer 2nd 1

5B

Bleus Blanches

J'ai tiré 5 fois, mon resultat est de : 2 blanches 3 bleus



? ou 2 bleus 3 blanches

Blanches	Bleu
1111111	1
11111	11111

2

	Blanc	Bleu	Total
36 fois	||||||||||||||||||||	||||||||||||||||	
	20	16	36

Blanc $= \frac{20}{36} = 0,5 = 55\,\%$

Bleu $= \frac{16}{36} = 44\,\%$

Blanc	$55\,\% \times 5 = 2,75$	36
Bleu	$44\,\% \times 5 = 2,2$	26

Lelia
Edem
Lihem.

- Nous avons vu deux billes blanches et une bille bleue. Nous pensons donc qu'il y a plus de billes blanches que de billes bleues.

- Pour savoir si il y a plus de billes blanches ou de billes bleues, on doit ~~t~~ secouer ~~le m~~ au maximum pour être le plus précis possible.

- Sur ~~ces~~ 100 tests, nous sommes tombés 53 fois ~~sur~~ la blanche et 47 fois sur la bleue. Entre 53 et 47, il n'y a pas beaucoup d'écarts et comme il y a 5 billes, il y a forcément 3 blanches et 2 bleues.

- Nous pensons qu'il y autant de billes bleues que de billes blanches car sur 31 tests, il y a 15 billes blanches et 16 billes bleues. Il y a très peu d'écart qui sépare ces deux nombres.

Quiterie BERANGER
Justine GARNIER
Lima COUDIERE
Groupe 2GT1

Mathématique : probabilité

15/04/18

5 billes

Combien de billes blanche combien de billes bleu ?

Nous allons faire des essaie avec la bouteille pour voirre laquelle couleur revien le plus souvent pour etablir une hypothèse.

Essaie :

1 : blanche
2 : blanche
3 : blanche
4 : blanche
5 : bleu
6 : bleu
7 : blanche
8 : bleu
9 : bleu
10 : blanche

Totale : blanche : $\frac{6}{10}$
bleu : $\frac{4}{10}$

On peut etablir une frequence et une hypothèse sur le faite qu'il y a plus de billes blanche que de bleu, même un nombre 2 billes bleu pour 3 billes blanche.

Mais pour être plus sûre nous allons faire plus d'essaie. (100)

Après 100 essaie nous avons eu comme resultat :

bleu : 40
blanche : 60

Comme notre 1er essaie ~~10 fois~~

40% de chance de tomber sur une bille bleu et 60% de chance sur une bille blanche ce qui fait 2 billes bleu et 3 billes blanche car 40% de 5 est egale a 2. Notre hypothèse et donc exact.

nurielle
Sabrina
Amadou
Alice

5 billes

2/5 ou 1/5	qu'il y ait une boule blanche
3/5 ou 1/5	bleue

1
2
3
4
5
6
7
8
9
10

~~4 bleu~~
~~1 blanches~~

1 bleu
4 blanches

1 Bleu	9 Blan	17 Bleu
2 Bleu	10 Blan	18 Blan
3 Bleu	11 Ble	19 Bleu
4 Bleu	12 Ble	20 Bleu
5 Blanche	13 Blan	
6 Bleu	14 Blan	
7 Blanche	15 Blan	
8 Blan	16 Bla	

Bleu: 10/20

On a tiré 5 fois les boules on a trouvé 4 fois des boules blanches et une fois une boule bleu.

4 blanches

1 bleu

On a 4 chances sur 5 pour avoir la bille blanches

1 chance sur 5 pour avoir la bille blanches

~~Plusieurs issues~~

2 bleus
3 blanches

on a tiré 20 fois, on a obtenu 10 billes bleues et 10 billes blanches mais avec 5 billes dans la brique de jus on ne peut pas obtenir autant de boules blanches que de boules bleues

Conclusion

On ne sait pas si on a tiré moins de 100

Si on tire plus de 100

on sera sûr qu'il y a autant de bille blanche que de bille bleu.

Murielle
Alice
Sabrina
Amadou

1 Blan
2 Blan
3 Blan
4 Blan
5 Blan
6 Blan
7 Blan
8 Bleu
9 Blan
10 Blan

9/10
Blanc

1 Bleu
2 Ble
3 Blan
4 Blan
5 Blan
6 Blan
7 Blan
8 Blan
9 Ble
10 Blan

7/10
Blanc

1 Blan
2 Blan
3 Blan
4 Blan
5 Blan
6 Blan
7 Bleu
8 Ble
9 Blan
10 Blan

8/10
Blanc

1 Ble
2 Blan
3 Blan
4 Blan
5 Blan
6 Bleu
7 Bleu
8 Blanc
9 Blan
10 Blan

7/10

1 Bleu
2 Blan
3 Bleu
4 Blanc
5 Blanc
6 Blanc
7 Bleu
8 Blanc
9 Bleu
10 Blanc

6/10

Blanc

1 Blanc
2 Blan
3 Bleu
4 Blan
5 Blan
6 Blan
7 Blan
8 Blan
9 Blan
10 Bleu

8/10
Blanc

Enzo, Anton, Mehdi et Sakko

Bille blanche : 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1

Bille bleue : 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1

53

$\frac{53}{100} \times 5 = 2,65 \approx 3$

47

$\frac{47}{101} \times 5 = 2,35 \approx 2$

On lance 100 billes ce qui represent l'effectif cumulé croissant, pour les billes bleues et blanches, on trouve 53 billes blanches et 47 billes bleues. On fait la formule $\frac{\text{effectif}}{\text{effectif cumulé croissant}}$ pour obtenir la frequence On l'a mutiplie par 5 car il y a 5 billes. On trouve donc 3 billes blanches e 2 billes bleues

100

100

Annexe 4 : Fiches de lecture

(Dutarte, Octobre 2013) : "Du bon usage d'un intervalle de fluctuation", de Philippe Dutarte, dans APMEP n°505

Article écrit par un IA-IPR de l'académie de Créteil, très intéressant parce qu'il explicite la philosophie qui sous-tend l'introduction de l'échantillonnage au lycée. Il est très fourni en repères historiques (l'un des premiers exemples, celui du médecin écossais John Arbuthnot, est d'ailleurs un problème de sex-ratio), notamment sur les débats fondateurs de la discipline entre Ficher d'un côté, Neyman et Pearson de l'autre, au cours du XXème siècle. Il contient quelques exemples intéressants d'exercices, mais davantage à destination des premières ou des terminales.

Voilà le passage le plus pertinent pour cette étude :

"La mise en œuvre d'un intervalle de fluctuation, pour la prise de décision, devrait répondre, au lycée, à un certain nombre de critères, garantissant le respect de l'esprit de la démarche statistique, susceptible, selon les orientations choisies, d'être poursuivie dans l'enseignement supérieur.

– Le premier critère nous semble devoir être de poser un problème, une situation, ayant quelque écho avec le monde réel. Cette situation doit être telle que la proportion *p* dans la population est supposée connue : c'est l'hypothèse de départ, avec laquelle on construit la règle de décision. Cela exclut les situations où l'on n'a aucun *a priori* sur la valeur de *p* (type « sondage ») […]

– Il faut ensuite que cette situation soit de type « bilatérale » : par rapport à la valeur de *p* « visée » (celle de l'hypothèse) on s'intéresse à un écart trop important à gauche et à droite.

– La règle de décision doit être élaborée, autant que possible, avant la prise d'échantillon. Il est plus honnête de décider d'une règle avant de jouer, qu'après la partie. […]

– L'échantillon doit être prélevé par tirage au hasard avec remise (équiprobabilité garantie par randomisation).

– En cas de rejet de l'hypothèse, il faut savoir que l'on peut interpréter le 5% comme la probabilité de commettre une erreur de décision (probabilité de rejet de l'hypothèse sachant qu'elle est vraie). En cas d'acceptation de l'hypothèse, il faut savoir que l'estimation de la probabilité d'erreur de décision est plus compliquée."

Puis en conclusion : "L'enseignement au lycée de la notion d'intervalle de fluctuation est essentiel, tout d'abord dans le cadre de la prise de décision, ensuite comme préliminaire à l'investigation de la notion d'intervalle de confiance. Il faut faire comprendre « que toute décision s'accompagne d'un risque, mais que ce risque peut être évalué ». Dans cet enseignement, l'apport principal du professeur de mathématiques doit être de dégager la logique qui préside à la démarche statistique, les raisonnements mis en œuvre. On pourrait confier les détails des calculs à un logiciel (ou programmer un algorithme), là n'est pas l'essentiel. Partir de situations-problèmes « concrètes » est indispensable pour mettre en perspective les principes introduits, comprendre les définitions et les précautions d'utilisation (rôle de la zone de rejet, distinction entre la règle de décision et la prise d'échantillon, signification des pourcentages 95%, 5%, …). Il ne s'agit pas d'élaborer des théories mathématiques « à vide », dans un cadre purement probabiliste : à l'arbitraire de cette introduction s'ajouterait l'incapacité d'appliquer à bon escient ces concepts.

(Cerclé, Avril 2013) "Quelques interrogations du professeur de lycée autour des intervalles de fluctuation", de Véronique Cerclé, dans Repères – IREM n°91, 2013

Cet article transcrit les questions légitimes qui apparaissent à l'enseignant à la lecture des programmes et y propose des réponses grâce à une analyse poussée.

Il insiste tout d'abord sur les différentes définitions données dans les différentes classes du lycée :

En Seconde	En Première	En Terminale
IF2 = [$p-\frac{1}{\sqrt{n}}$; $p+\frac{1}{\sqrt{n}}$]	IF1 = [$\frac{a}{n}$; $\frac{b}{n}$] déterminé à l'aide de la loi binomiale (tableur ou algorithme) de sorte que $P(X<a) \leqslant 0{,}025$ et $P(X>b) \leqslant 0{,}025$	IFT = [$p-1{,}96\sqrt{\frac{p(1-p)}{n}}$; $p+1{,}96\sqrt{\frac{p(1-p)}{n}}$]

Il précise ainsi que "L'IF de Seconde est donc le résultat de deux opérations : d'abord l'approximation de la loi binomiale par une loi normale, puis la majoration simplifiant la formulation des bornes."

IF1 *(déterminé par la loi binomiale)*
↓ *approximation*
IFT *(formule)*
↓ *majoration*
IF2 *(formule)*

Il se penche ensuite sur les cas, rares certes, où, du fait des approximations successives, l'intervalle de fluctuation proposé en 2^{nde} est mis en défaut c'est-à-dire contient moins de 95% des échantillons (alors que les valeurs de n et p sont dans les bornes imposées). Par exemple pour n = 30 et p = 0,55, ou n= 99 et p = 0,53. Il précise alors à juste titre que "Pour le lycéen, ce qui est important est qu'un intervalle de fluctuation sert à indiquer la fourchette dans laquelle se trouve la ≪ grande majorité ≫ des fréquences des échantillons observés, le seuil de 95% étant fixe arbitrairement. Néanmoins, pour le professeur, il faudra réfléchir à une formulation satisfaisante pour la Seconde."

Il réfléchit sur la notion d'échantillon, pas si évidente, et insiste sur le fait que lorsqu'on réalise des simulations pour obtenir l'intervalle de façon approchée, on constitue des échantillons d'échantillons et ce n'est pas toujours clair.

Puis conclut :

" Compte tenu des éléments évoqués, on ne peut pas écrire en Seconde que ≪ la fréquence F appartient à IF2 avec une probabilité d'au moins 95% ≫ car c'est faux. De plus, même si dans la pratique elle ne semble pas présenter de difficulté aux élèves de Seconde, il convient d'être conscient de la complexité cachée de cette définition qui renvoie une probabilité à une autre, s'appuie sur le lien entre probabilité et fréquence établi en Troisième tout en utilisant le mot échantillon sans précaution.

Cette complexité me semble opacifier l'intérêt de l'IF, alors que l'idée importante à ce niveau est qu'il s'agit d'un ≪ intervalle dans lequel la fréquence F de l'échantillon qu'on va prélever est censée se trouver avec une forte probabilité (ici de 95 %) ≫.

Il me semble donc qu'une formulation suffisante en Seconde pourrait être : « environ 95% des échantillons ont une fréquence qui appartient à IF », « intervalle dans lequel se trouvent l'essentiel (soit environ 95%) des fréquences constatées sur les échantillons ». Une telle définition est correcte quel que soit le sens donné à chaque terme, elle suffit à faire comprendre la notion d'intervalle de fluctuation à ce niveau, tout en permettant son utilisation pratique. "

En annexe, il propose des exemples intéressants et les analyse.

(Sotura, Janvier 2013) Une activité pour initier à la statistique inférentielle en classe de seconde, de Brigitte Sotura dans APMEP n°502

C'est le document qui a servi de référence dans la mise en œuvre de l'activité qui fait l'objet de la deuxième partie de ce mémoire. Il rappelle le contexte dans lequel s'inscrit l'échantillonnage, décrit l'énoncé de l'activité ainsi que du traitement informatique qui en est fait, rend compte du déroulé de l'expérimentation réalisée par l'auteure en 2012, et en souligne les points saillants. Il nous a donc très largement inspiré et il faudrait presque le citer en entier.

Parmi les passages importants, autour de la notion importante de questionnement : "La classe de seconde est le moment privilégié pour faire comprendre l'enjeu de cet enseignement qui sera décliné sur les trois années de scolarité au lycée pour la plupart des élèves. Le programme de seconde précise que l'objectif est d'amener les élèves à un questionnement lors des activités de prise de décision à partir d'un échantillon ou d'estimation d'une proportion. C'est ce point de vue qui a été choisi pour cette activité de la classe de seconde."

Puis : "Enfin une difficulté et non des moindres est la conclusion : pour certains élèves le fait que le hasard puisse produire un résultat égal ou inférieur à 46/132 les conduit à dire qu'il n'y a pas lieu de considérer la situation comme anormale ; ils ne prennent pas en compte le caractère peu probable de cet événement. Il y aura nécessité de déterminer un seuil à partir duquel on considérera que l'événement est trop peu probable. Parvenir à le faire comprendre (sans nécessairement aller jusqu'à le quantifier) c'est préparer ce qui suivra ensuite dans les classes de première et terminale.

Donner prématurément une règle de décision selon que la fréquence observée est ou n'est pas dans l'intervalle $[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}]$ risque de priver les élèves de tout questionnement : or comme le dit explicitement le programme de seconde l'objectif est d'amener les élèves à un questionnement. C'est ce point de vue qui a prévalu dans cette activité afin de faciliter la compréhension de la démarche qui sera vue en première à partir de la loi binomiale."

(Parnaudeau, Octobre 2013) Statistiques inférentielles : un débat scientifique en classe de seconde, de Jean-Marie Pernaudeau dans APMEP n°505

Un bel exemple de l'intérêt d'un débat – entre élèves, et avec le professeur – pour faire émerger une règle commune et lui donner du sens, ce qui n'est pas le cas si elle est "parachutée" d'en haut. L'énoncé de l'activité proposée est le suivant : "Une personne possède une pièce de monnaie. Elle pense que cette pièce n'est pas bien équilibrée. Proposer une méthode (ou une expérience) pour répondre à cette question d'un point de vue statistique." Les réponses des élèves permettent de mieux anticiper les difficultés qu'ils rencontrent.

Annexe 5 : Documents élèves pour les activités de la séquence échantillonnage

Un problème de pollution ?

Problème :

Dans la réserve indienne d'Aamjiwnaang, située au Canada à proximité d'industries chimiques, il est né, entre 1999 et 2003, 132 enfants dont 50 garçons.

Cette situation doit-elle être considérée comme anormale et interroger les autorités sanitaires ?

Questions - indices :

1) Combien de naissances de garçons serait-il normal d'attendre ? N'y a-t-il qu'un seul résultat normal ?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

2) Le hasard est-il capable d'expliquer la situation de la réserve indienne ? Est-ce que cela vous paraît probable ?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

3) Comment pourrait-on justifier cette réponse en s'appuyant sur des données chiffrées ? Quelles expériences peut-on réaliser ?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

TP : Simulation et échantillonnage

Nom : ………………….   Prénom : ……………………

<u>Attention : votre feuille de calcul est à sauvegarder à la fin de l'heure dans le dossier 2GT1/Rendu/Maths</u>

Pour répondre au problème de la réserve indienne, on se propose de simuler le hasard sur tableur, c'est-à-dire ici de simuler le résultat de 132 naissances.

1) Ouvrez le tableur, et rentrez dans la première cellule la formule =ALEA( ) qui permet de générer un nombre aléatoire entre 0 et 1
2) Utiliser la fonction =SI pour afficher dans la cellule d'à côté (colonne B) « Garçon » si le nombre généré est inférieur à 0,5 et « Fille » sinon
3) Copier-glisser les deux cellules pour générer 132 naissances
4) Dans la cellule B134, utiliser la fonction =NB.SI pour compter le nombre de garçons obtenus avec votre simulation.
5) Relancer le calcul à l'aide de la touche F9. Entre quelles valeurs semble varier le nombre de garçons sur 132 naissances ?

……………………………………………………………………………………..

6) Quelle réponse pouvez-vous donner au problème posé ? Est-ce une situation anormale ?

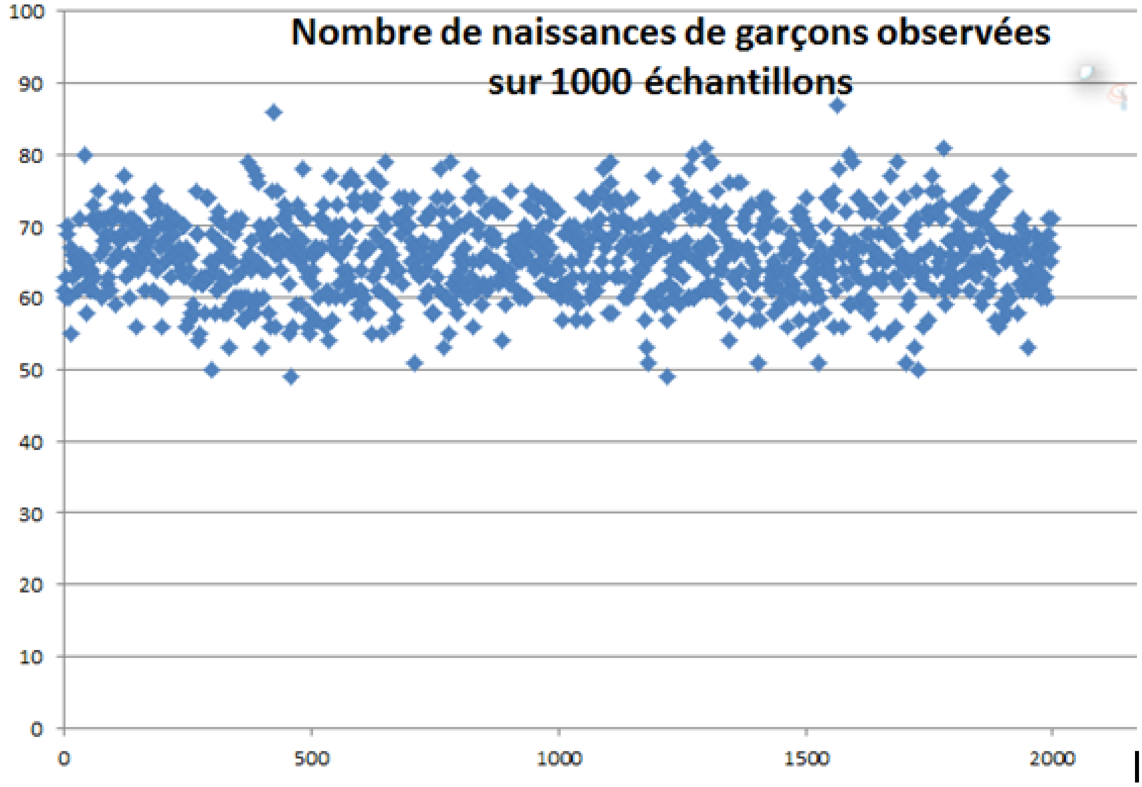
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………

Pour affiner ces observations, on se propose maintenant de simuler 1000 échantillons de 132 naissances

7) Sélectionnez l'ensemble de vos données dans les colonnes A et B et faites un copier-glisser jusqu'à la colonne BXX (c'est loin !)
8) Sélectionnez la ligne 134, celle qui contient le nombre de garçons de chaque échantillon et cliquer dans les onglets en haut sur Insérer puis sur Nuage pour représenter la série.

   Cela doit donner un diagramme comme celui de la page suivante. Faites valider par le professeur.

9) À l'aide du graphique reproduit ci-dessous répondre aux questions suivantes :



a) Est-ce que le hasard peut donner un nombre de garçons égal ou inférieur à 50 ?

....

b) Pour 132 naissances, dans quel intervalle se trouve le plus souvent le nombre de garçons ?

....

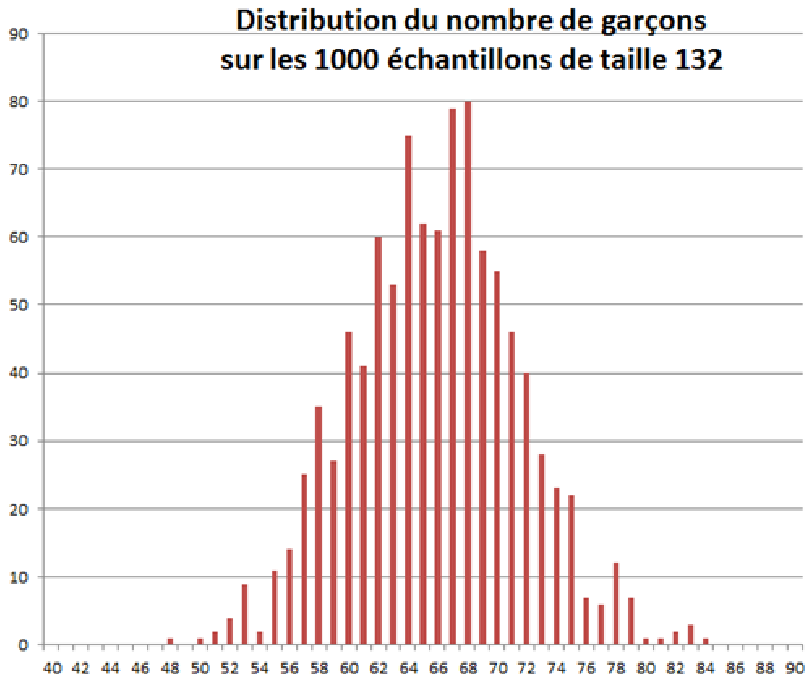
c) Sur les 1000 échantillons combien y en a-t-il pour lesquels ce nombre est en dehors de l'intervalle [50 ; 80] ?

....

d) Diriez-vous que ce qui se passe dans la réserve indienne d'Aamjiwnaang doit être considéré comme le fait du hasard ou diriez-vous que la situation doit être considérée comme anormale et qu'une enquête doit être menée pour déterminer ce qui pourrait en être la cause ?

....
....
....
....

10) On se propose de regarder plus précisément la façon dont se distribue le nombre de garçons sur ces 1000 échantillons. On a réalisé le diagramme en bâtons ci-dessous à partir de la simulation de 1000 échantillons de taille 132.



a) Qu'a-t-on indiqué sur l'axe des abscisses ?

……………………………………………………………………………………

b) Sur l'axe des ordonnées ?

……………………………………………………………………………………

c) Un des bâtons correspond à 70 en abscisse et 55 en ordonnée. Interprétez ces données.

……………………………………………………………………………………

……………………………………………………………………………………

d) Réalisez vous-même ce type de diagramme sur le tableur à partir de la simulation de vos 1000 échantillons. Faites valider par le professeur.

11) Il est plus intéressant de raisonner sur la fréquence de garçons que sur le nombre de garçons.

a) Quelle est la fréquence de garçons nés dans la réserve et quelle est la fréquence théorique attendue ?

……………………………………………………………………………………

……………………………………………………………………………………

b) Dans la cellule B135, calculer la fréquence correspondant au nombre de garçons obtenu par simulation

c) Copiez-glisser pour faire de même avec les 1000 échantillons

d) Représentez le nuage de points de la série des fréquences à la manière de la question 9)

Activité : Influence de la taille de l'échantillon

Reprenons la situation de la réserve indienne en la modifiant un peu.

1) Dans un village qui a eu 10 naissances dans l'année, 4 des nouveau-nés sont des garçons, c'est-à-dire 40% des naissances. Cette situation vous parait-elle anormale et doit-on informer les autorités sanitaires ?

………
………
………

2) Et si à l'échelle de la France, il y avait un million de naissances et 40% de celles-ci sont des garçons ? Cette situation vous parait-elle anormale ?

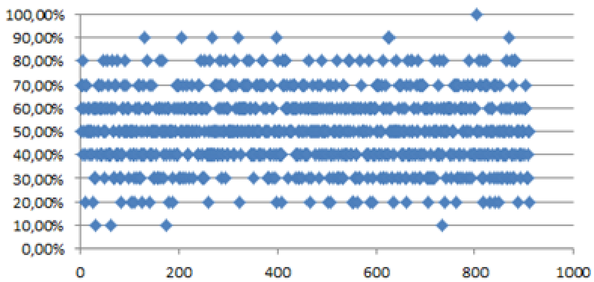
………
………
………
………

3) Expliquer en quoi le nombre de naissances influe sur votre décision

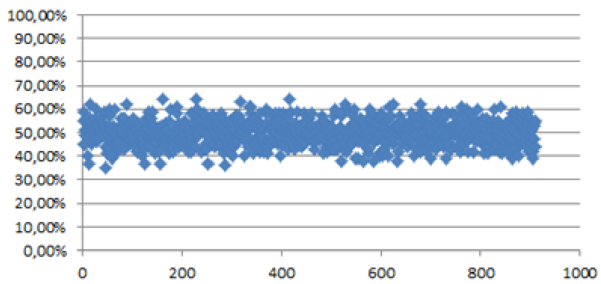
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

Pour étudier plus en détail cette influence, on a simulé (à la manière du problème de la réserve indienne) un grand nombre d'échantillons de 10 naissances, puis de 100, 1000 et 10000 naissances, et on a représenté les fréquences de garçons obtenues dans des nuages de points :

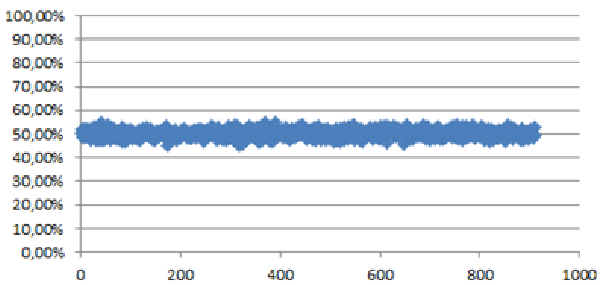
Fréquences de garçons simulées pour des villages de 10 habitants



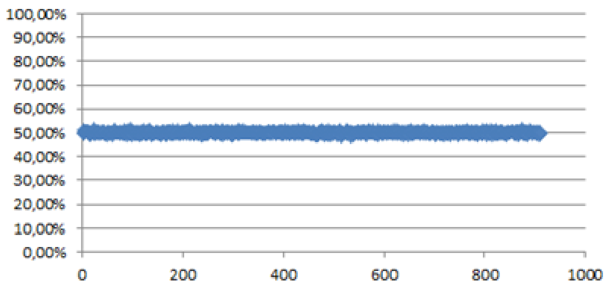
Fréquences de garçons simulées pour des villages de 100 habitants



Fréquences de garçons simulées pour des villages de 1000 habitants



Fréquences de garçons simulées pour des villes de 10000 habitants



4) Dans quel intervalle se trouvent l'essentiel des fréquences quand on simule :

10 naissances ? ……………………………………………

100 naissances ? ……………………………………………

1000 naissances ? ……………………………………………

10000 naissances ? ……………………………………………

5) Pour lesquels de ces villages, la fréquence de 40% de garçons vous parait-elle anormale

……………………………………………………………………………………………………

……………………………………………………………………………………………………

6) S'il naît 40% de garçons pour 10 naissances, peut-on affirmer que c'est uniquement à cause du hasard ?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

7) S'il naît 40% de garçons pour 10 000 naissances, peut-on affirmer que ce n'est pas uniquement à cause du hasard ?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

8) A l'aide de la calculatrice, calculer :

$\frac{1}{\sqrt{10}} \approx$ ………….. puis $0{,}5 - \frac{1}{\sqrt{10}} \approx$ ………. et $0{,}5 + \frac{1}{\sqrt{10}} \approx$ ……….

$\frac{1}{\sqrt{100}} \approx$ ………….. puis $0{,}5 - \frac{1}{\sqrt{100}} \approx$ ………. et $0{,}5 + \frac{1}{\sqrt{100}} \approx$ ……….

$\frac{1}{\sqrt{1000}} \approx$ ………….. puis $0{,}5 - \frac{1}{\sqrt{1000}} \approx$ ………. et $0{,}5 + \frac{1}{\sqrt{1000}} \approx$ ……….

$\frac{1}{\sqrt{10000}} \approx$ ………….. puis $0{,}5 - \frac{1}{\sqrt{10000}} \approx$ ………. et $0{,}5 + \frac{1}{\sqrt{10000}} \approx$ ……….

9) Que constatez-vous par rapport à vos réponses de la question 4) ?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Définition : On appelle intervalle de fluctuation un intervalle dans lequel se trouve l'essentiel (environ 95%) des fréquences constatées pour les échantillons.

Remarque : On parle aussi d'intervalle de fluctuation au seuil de 95%, cela signifie qu'on prend un risque de 5% en considérant que le hasard n'est pas seul responsable d'une fréquence constatée en dehors de cet intervalle.

Théorème : Au sein d'une population, si l'on connaît la proportion p des individus ayant un caractère donné, alors si l'on tire un échantillon de n individus, la fréquence f d'apparition du caractère dans cet échantillon est comprise dans l'intervalle $\left[\boldsymbol{p} - \frac{1}{\sqrt{n}} ; \boldsymbol{p} + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 95% (à condition que $0{,}2 \leq p \leq 0{,}8$ et $n \geq 25$)