Claudine Vergne

Formation des PLC2 en Statistique : Analyse de quelques situations

IUFM de Montpellier

Résumé :

Une première phase de l'atelier sera consacrée à la présentation d'une séance de formation effectivement proposée à des PLC2. Le choix des contenus et de la structure de cette séance est motivé par un travail de recherche qui sera brièvement résumé.
Dans une deuxième phase, les participants seront invités à échanger sur la ressource présentée et à faire part de leurs propositions.

~ INTRODUCTION ~

Ce document est la version écrite d'un atelier présenté au colloque CORFEM 2007 selon l'ordre du jour suivant :

PREMIERE PARTIE

Résumé d'un travail de recherche

Compte rendu d'une formation de PLC2 en statistique

DEUXIEME PARTIE

Discussion

Bilan d'un questionnaire relatif à cette formation

Les organisateurs m'ont demandé de rédiger un compte rendu "aussi détaillé que possible", ce que je me suis efforcée de faire. Le document revêt de ce fait une forme particulière, peu habituelle.

~ PREMIERE PARTIE ~

I Résumé d'un travail de recherche

En 2003/2004, dans le cadre d'un DEA (Master), j'ai été amenée à travailler sur l'enseignement de la statistique en classe de seconde, et à rédiger un mémoire ayant pour titre "La notion de variabilité dans le nouveau programme de seconde : étude de conditions de viabilité".
L'objectif était de déterminer si la notion de variabilité, et plus largement, la partie novatrice du programme de statistique de seconde, était viable, à sa sortie en septembre 2000, dans l'Enseignement Mathématique Secondaire Français (noté EMS). Le cadre théorique choisi est celui de la didactique des mathématiques ; l'étude de textes centrés sur l'**écologie** et l'**économie** des systèmes didactiques a permis de proposer une **liste de conditions de viabilité d'un objet d'enseignement** (dans EMS), puis une **grille d'observation et d'analyse** des conditions de viabilité d'un objet d'enseignement[9].

L'analyse, à l'aide de la grille, du programme de 2° et de son document d'accompagnement fait apparaître des obstacles majeurs à la viabilité du nouveau programme de statistique en 2°. L'étude, à l'aide de la même grille, de deux manuels scolaires édités en 2000 montre qu'ils présentent certaines faiblesses et que, de fait, ils apportent aux enseignants une aide limitée et peu fiable. Enfin, des observations de séquences menées dans des classes de seconde en 2003/2004 rendent compte des grandes difficultés qu'éprouvent certains enseignants, même chevronnés, à enseigner à cette partie du programme.

Cela confirme, si besoin était, que l'enseignement du nouveau programme de statistique de 2° (et plus largement d'une statistique plus réaliste, plus moderne) ne peut pas faire l'économie de la formation des enseignants. Quelle formation ? Bien sûr, celle-ci devra inclure des compléments théoriques. Mais certaines pertinences même dans les prescriptions du programme déstabilisent les pratiques des enseignants (par exemple *ne pas faire cours*, *adopter une démarche expérimentale, faire faire un cahier de statistiques*, …) Or, les observations effectuées permettent de penser qu'il est extrêmement difficile pour un professeur de modifier **à la fois** les contenus et les pratiques de son enseignement (plus il est mal à l'aise avec les contenus, plus il risque de se raccrocher à ses "manières de faire" habituelles).
C'est dire que, concernant l'enseignement de la statistique, une formation efficace, qui aide réellement les enseignants ne devrait pas porter seulement sur les contenus théoriques mais devrait aussi s'intéresser aux pratiques d'enseignement.
Notons aussi que du point de vue épistémologique, il y a un lien très fort entre l'enseignement de la statistique et la théorie des situations didactiques[10]. Il paraît pertinent de mettre en place, pour l'enseignement de la statistique, des situations adéquates.

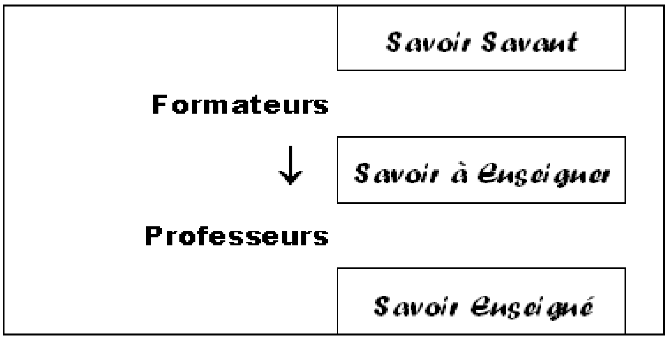
---

[9] Pour plus de précisions, on peut consulter VERGNE C. (2005), *La notion de variabilité dans le nouveau programme de seconde : étude de conditions de viabilité*, Mémoire de DEA, IREM de Montpellier.

[10] On peut lire à ce sujet : BROUSSEAU G., BROUSSEAU N., WARFIELD G. (2002), An experiment on the teaching of statistics and probability, *Journal of mathematical behavior*, 20.

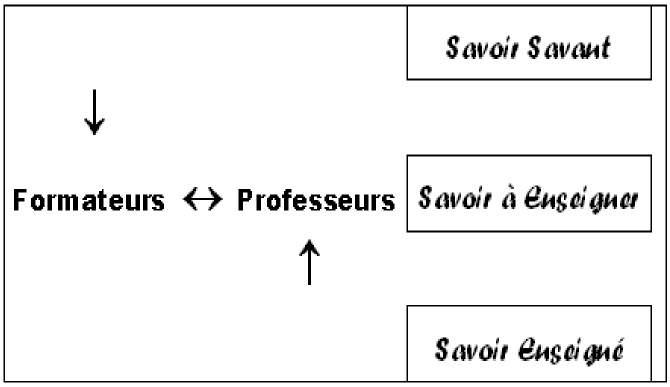
Pour ces raisons, on peut penser que dans ce cas, une formation continue qui se réduirait à des compléments théoriques, le plus souvent sur le mode transmissif, ne peut suffire.

Formation : Schéma 1



D'où la proposition d'un autre type de formation, où formateurs et professeurs travailleraient de façon collaborative à la mise au point de situations d'enseignement : le formateur garantit la validité d'un point de vue cognitif et épistémologique, apporte au besoin des compléments théoriques, tandis que le professeur, qui connaît les réalités du terrain, s'assure des conditions de faisabilité.

Formation : Schéma 2



II Compte rendu d'une formation en statistique des PLC2

1. Introduction

Depuis 2003, je suis amenée à animer, **en 3 heures**, une formation en statistique pour des groupes d'une vingtaine de PLC2.

Il s'agit cette fois de formation initiale et non de formation continue.
Concernant les contenus théoriques, le problème est à peu près le même qu'en formation continue ; la plupart des stagiaires n'ont pas de connaissances approfondies en statistique, rares sont ceux qui se sont spécialisés dans ce domaine durant leur cursus. Toutefois, ils manifestent en général moins d'hostilité vis à vis de la statistique que des enseignants de générations antérieures.
Concernant les pratiques, pour beaucoup de stagiaires, elles sont peu stables, parfois peu adaptées et même indigentes. Cette formation est une opportunité de promouvoir une démarche de type expérimental, démarche préconisée dans les programmes de la 6° à la terminale, et dans tous les domaines des mathématiques, pas seulement en statistique.

Il semble indispensable de parler du nouveau programme de seconde, car il est emblématique (au moins pour l'instant[11]) de la statistique qui doit vivre dans EMS et parce qu'il est clair que sa mise en œuvre pose problème. Toutefois, les stagiaires enseignent à des niveaux divers, certains au collège, d'autres en lycée et beaucoup ne connaissent pas ce nouveau programme. Il faudra donc décontextualiser suffisamment pour intéresser tout le monde, et apporter à chacun des éléments de formation assez consistants qu'il pourra réinvestir éventuellement.

---

[11] En septembre 2008, entrera en vigueur un nouveau programme de 3°, où est envisagée la "*notion de probabilité*".

Si l'on considère qu'une formation efficace en statistique est celle qui favorise sa viabilité dans EMS, la grille qui a servi à repérer des obstacles à son enseignement peut devenir, si on change de point de vue, un **outil de remédiation**. En effet, elle indique les points sensibles à travailler tout particulièrement en formation et suggère qu'il faudra

- vaincre certaines résistances relatives à la raison d'être de statistique dans EMS ;
- faire comprendre les intentions du nouveau programme ;
- convaincre que les organisations mathématiques prescrites permettent de faire faire des mathématiques aux élèves ;
- faire des propositions relatives au travail des techniques et à l'évaluation (exercices, devoir maison, devoir de contrôle) ;
- proposer des pistes pour pallier les difficultés repérées relatives à la gestion du temps, la gestion de la classe, l'utilisation des TIC, le contenu du cahier de statistique, etc. ;
- apporter éventuellement des compléments théoriques ;
- promouvoir une certaine façon d'enseigner qui s'appuie sur des (vraies) questions, qui permette aux élèves de prendre des initiatives, des responsabilités et qui les engage dans une démarche de type expérimental.

En 3 heures, il n'est bien sûr pas possible de travailler avec les participants à la mise au point de situations d'enseignement, mais il n'est pas non plus question pour moi de faire un cours.
Compte tenu de toutes ces contraintes, la séance est conduite sous formes d'échanges à partir de documents distribués (cf. annexes) et selon l'ordre du jour qui suit.

2. Ordre du jour

STATISTIQUES EN COLLEGE/LYCEE

http://www.irem.univ-montp2.fr/avostat

HASARD	vient de l'arabe	AZ-ZAHR	qui signifie	JET de DE
ALEA	vient du latin	ALEA	qui signifie	COUP de DE
CHANCE	vient du latin	CADERE	qui signifie	CHOIR, TOMBER

Bibliographie

I Point sur les programmes

II Spécificités de la classe de seconde

III Etude de quelques situations

Parité des sexes
Joueur A ou B
Une distribution de fréquence

IV Quelques simulations (travail en groupe)

Autres ...

En fait, je commence par la fin, c'est à dire par la bibliographie (cf. ANNEXE I), un peu par provocation mais surtout parce que je ne veux pas l'oublier à la fin de la séance. Je pense qu'à un certain niveau, et en particulier au niveau des stagiaires, toute formation contient une part d'auto formation, j'essaie par là même de les inciter en proposant des pistes. La bibliographie/sitographie est commentée rapidement, les stagiaires sont invités à consulter pendant la pause certains des ouvrages que j'ai apportés à cet effet.

J'ai dit plus haut que je ne voulais pas "faire un cours". Certes je vais exposer et parler beaucoup, mais je voudrais que les stagiaires interviennent et puissent poser des questions quand ils le souhaitent. Certains ne me connaissent pas, ne m'ont jamais vue. Pour installer cette possibilité d'échanges, je commence en racontant l'anecdote suivante, qui est véridique.

> *En septembre, lors du premier GAP (Groupe d'Accompagnement Professionnel), au cours d'une table ronde, une stagiaire PLC2 raconte la prise de contact avec son tuteur ; ils parlent de la progression en seconde ; le tuteur ouvre le livre à la table des matières, prend un crayon, barre le chapitre Statistique en disant " Ça, ce n'est pas la peine de le faire, ça ne sert à rien !"*

Je demande aux stagiaires ce qu'ils en pensent et il y a toujours des réactions ! A partir de ces réactions, il est possible de rebondir et de se demander quelle place occupent les statistiques dans le cursus du collège/lycée. C'est l'objet de la première partie "**Point sur les programmes**".

3. Point sur les programmes

Statistique en collège/lycée

Collège

Sixième	Cinquième	Quatrième	Troisième
Tableaux à 2, plusieurs colonnes, double entrée. Lire et interpréter des graphiques. Diagrammes (bâtons, circulaire, demi-circulaire, cartésien)	Effectifs, fréquences, classes d'égale amplitude. Tableaux de données. Représenter, lire, interpréter. Diagrammes divers, Histogrammes amplitude constante Car. qualitatif (diag. tuyau d'orgue, bandes, secteurs) Car quantitatif (diag. en bâtons) Car. continu (histogramme égale amplitude)	Effectifs cumulés, fréquences cumulées. Moyennes pondérées. Initiation à l'usage des tableurs-grapheurs. Valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles.	Caractéristique de position d'une série statistique. Approche de caractéristiques de dispersion d'une série statistique. Initiation à l'utilisation des tableurs-grapheurs en statistique.
Nouveaux programmes		Anciens programmes	

Seconde

- Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de seconde à l'étendue).
- Définition de la **distribution des fréquences** d'une série prenant un petit nombre de valeurs et de la fréquence d'un événement.
- **Simulation** et **fluctuation d'échantillonnage**.

Cycle terminal

	L	ES	S	STG
1°	-Diagramme en boite, intervalle interquartile - Variance et écart-type Pour l'interprétation des valeurs de ces paramètres, on gardera à l'esprit qu'ils **fluctuent** d'une série à l'autre	- Etude de séries de données - Lissage par moyennes mobiles - Histogrammes à pas non constants - Diagrammes en boites - Effets de structure - Intervalle interquartile, écart-type - Tableau à double entrée /…/	- Variance et écart-type - Diagramme en boite, intervalle interquartile - Influence sur l'écart-type et l'intervalle interquartile d'une transformation affine des données. /…/ - Lien entre loi de probabilité et	- Histogrammes - Diagrammes en boite, secteurs, bâtons - Tendance centrale (moyenne, médiane) - Dispersion (quartiles, déciles, intervalle interquartile, interdécile, écart-type) - Tableaux croisés d'effectifs /…./

	- Tableaux croisés	- Lien entre loi de probabilité et **distribution des fréquences** - Modélisation et **simulation**	**distribution des fréquences** Modélisation et **simulation**	- Probabilité (faire le lien avec les propriétés des **fréquences**) - Expérimentation et **simulation**
T.		- Série statistique à deux variables - Ajustement affine (moindres carrés) **- Simulation** - Conditionnement et indépendance -Modélisation (on retravaillera les expériences de référence vues en 2°) - Statistique et **simulation** (un exemple traitant de l'adéquation à une loi équirépartie) ….	- Conditionnement et indépendance - Statistique et modélisation (expériences indépendantes, répétitions d'expériences indépendantes) - Lois de probabilité - Statistique et **simulation** (un exemple traitant de l'adéquation à une loi équirépartie) …	- Série statistique à deux variables - Ajustement affine - Séries chronologiques … - Probabilité conditionnelle …

Le document ci-dessus est distribué aux stagiaires et commenté.
En collège (à ce jour, nouveau programme en sixième et cinquième, ancien programme en quatrième et troisième), pour l'instant, on trouve seulement des statistiques descriptives. Celles-ci sont reprises en 2°, mais en 2° il faut aussi parler de **distribution des fréquences**, **simulation**, **fluctuation d'échantillonnage**. Si l'on envisage le cycle terminal (seules sont envisagées ici les séries les plus fréquentées), on voit que ces thèmes sont présents. Le lien demandé entre distribution des fréquences et loi de probabilité permettra de faire en 1° une approche fréquentiste des probabilités, la notion de simulation est reprise, tout particulièrement à l'occasion de l'adéquation à une loi équirépartie en Terminale.
La classe de 2° apparaît comme une charnière entre les statistiques descriptives du collège et la prise en compte de l'aléatoire au lycée. C'est le lieu d'une première rencontre, ne pas la ménager priverait de sens l'enseignement de Statistique/Probabilité du cycle terminal. On ne peut donc pas affirmer comme dans l'anecdote citée plus haut que "*Ça ne sert à rien* !" Plus précisément, que prévoit le programme de 2° ? C'est l'objet de la deuxième partie "**Spécificités de la classe de seconde**".

4. Spécificités de la classe de seconde

Le document ci-dessous est distribué aux stagiaires et commenté. Il débute par la belle phrase de Claudine Schwartz "*L'esprit statistique naît lorsque l'on prend conscience de la fluctuation d'échantillonnage*" qui, de mon point de vue, éclaire tout le texte et les choix qu'il contient. La lecture commentée permet de mettre en exergue des passages importants, qui posent problème ou font débat *: consacrer 1/8 du temps à la statistique, voir la notion de fluctuation d'échantillonnage sous l'aspect de la distribution des fréquences, simuler à l'aide du générateur aléatoire d'une calculatrice, d'une liste de chiffres au hasard.*

EXTRAITS DU PROGRAMME DE LA CLASSE DE SECONDE

L'esprit statistique naît lorsque l'on prend conscience de l'existence de fluctuation d'échantillonnage.
(*Claudine Schwartz*)

Introduction
Le programme est composé de trois grands chapitres : statistique, calcul et fonction, géométrie…A titre indicatif, le temps à consacrer aux différents chapitres pourrait être de **1/8 pour les statistiques**, le reste se répartissant équitablement entre les deux autres chapitres…
Un des apports majeurs de l'informatique réside dans la puissance de simulation des ordinateurs ; la **simulation** est ainsi devenue une pratique scientifique majeure : une approche en est proposée dans le chapitre statistique…
En seconde, le travail sera centré sur :

– la réflexion conduisant au choix de résumés numériques d'une série statistique quantitative ;
– la notion de **fluctuation d'échantillonnage** vue ici sous l'aspect élémentaire de la **variabilité de la distribution de fréquences**.
– la **simulation à l'aide du générateur aléatoire d'une calculatrice**. La simulation remplaçant l'expérimentation permet, avec une grande économie de moyens, d'observer des résultats associés à la réalisation d'un très grand nombre d'expériences. On verra ici la diversité des situations simulables à partir d'une **liste de chiffres.**
L'enseignant traitera des données en nombre suffisant pour que cela justifie une étude statistique ; il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts.
La notion de fluctuation d'échantillonnage et de simulation ne doit pas faire l'objet d'un cours.

CONTENUS	CAPACITES ATTENDUES	COMMENTAIRES
Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de seconde à l'étendue.	Utiliser les propriétés de linéarité de la moyenne d'une série statistique. Calculer la moyenne d'une série à partir des moyennes de sous-groupes Calcul de la moyenne à partir de la distribution des fréquences.	L'objectif est de faire réfléchir les élèves sur la nature des données traitées, et de s'appuyer sur des représentations graphiques pour justifier un choix de résumé. On peut commencer à utiliser le symbole Σ. On commentera quelques cas où la moyenne et la médiane diffèrent sensiblement. On remarquera que la médiane d'une série ne peut se déduire de la médiane de sous-séries. Le calcul de la médiane nécessite de trier les données, ce qui pose des problèmes de nature algorithmique.
Définition de la **distribution des fréquences** d'une série prenant un petit nombre de valeurs et de la fréquence d'un événement. **Simulation** et **fluctuation d'échantillonnage.**	Concevoir et mettre en œuvre des simulations simples à partir d'échantillons de chiffres au hasard.	La touche **random** d'une calculatrice pourra être présentée comme une procédure, qui, chaque fois qu'on l'actionne, fournit une liste de n chiffres (composant la partie décimale du chiffre affiché). Si on appelle la procédure un très grand nombre de fois, la suite produite sera sans ordre ni périodicité et les fréquences des dix chiffres seront sensiblement égales. Chaque élève produira des simulations de taille n (n allant de 10 à 100 suivant les cas) à partir de sa calculatrice ; ces simulations pourront être regroupées en une simulation ou plusieurs simulations de taille N, après avoir constaté la variabilité des résultats de chacune d'elles. L'enseignant pourra alors éventuellement donner les résultats de simulations de même taille N préparées à l'avance et obtenues à partir de simulations sur ordinateurs.

Passant sur le secteur I (*Résumé numérique …*) sur lequel pourtant il y aurait fort à dire, on aborde les contenus du secteur II en s'attardant sur les commentaires. Ils sont composés de deux parties. La première concerne la touche random ; je signale qu'il y a un problème et que nous en parlerons plus tard. La deuxième partie propose une méthode pour le regroupement d'échantillons ; ici aussi, il y a un problème, je ne précise pas lequel et je renvoie à plus tard.
Arrivés là, certains stagiaires sont tout à fait déstabilisés car ils ne savent pas de quoi on parle ! Aussi, quand je propose de présenter des séquences effectivement enseignées à des élèves de 2°, de rendre compte des réactions des élèves, ils acceptent avec empressement et parfois… soulagement !

Les trois situations qui suivent sont conçues pour couvrir l'ensemble des thèmes du secteur II du programme de 2°, avec une volonté affirmée de mathématiser cette partie du programme. Elles proposent des pistes pour la motivation des types de tâche demandées aux élèves, la gestion de la classe, le recueil de résultats. La première montre quel type de réponse peut apporter la statistique à une question posée. La deuxième illustre le fait qu'observer le hasard peut être source de connaissance, elle envisage sur un exemple la notion de sondages électoraux, ce qui paraît un objectif à atteindre en 2° si l'on veut participer à la formation du citoyen. La troisième est proposée comme devoir commencé en classe et terminé à la maison, elle aborde la notion de distribution de fréquences et peut être prolongée, avec les stagiaires, pas avec les élèves de 2°, pour travailler le lien entre distribution de fréquences et loi de probabilité.

On s'attaque à la première situation, étant entendu que les stagiaires peuvent poser les questions qu'ils jugent nécessaires et qu'il vaut mieux regarder les choses en détail plutôt que de tout survoler. A partir de là, on rentre dans un fonctionnement un peu schizophrénique, où, sous couvert de parler de l'enseignement proposé aux élèves, en fait, je veux faire de la formation pour les stagiaires, formation portant sur des contenus théoriques et aussi sur des pratiques d'enseignement. C'est l'artifice que j'utilise pour pallier le manque de temps. Dans la deuxième partie, la question de son efficacité sera posée.

5. Première situation "Parité des sexes"

a. Préambule

Dans l'assemblée que nous formons, la parité des sexes est-elle respectée ?

Où que l'on pose cette question, dans une classe, dans un groupe de PLC2, dans un atelier … du colloque CORFEM, la réaction des participants est toujours la même : ils se comptent, déterminent le nombre de femmes, le nombre d'hommes, les comparent à la moitié de l'effectif et le plus souvent … rejettent l'idée que la parité est respectée.
Ce comportement rend compte d'une conception arithmétique de la parité des sexes. Cette conception a le mérite d'être simple à appréhender mais quand l'effectif est impair, elle n'est pas opérationnelle. Dans ce cas, elle est corrigée en disant qu'il y a parité des sexes si le nombre de femmes, ou le nombre d'hommes est égal à la partie entière de la moitié de l'effectif.

Mais voyons, quand se détermine le sexe d'un individu ?

La plupart d'entre nous pense que c'est au moment de la conception, celui de la rencontre entre un spermatozoïde et un ovule. Est-ce qu'à ce moment-là le spermatozoïde et l'ovule ont une perception arithmétique du monde qui les entoure ? On peut penser que non ! Et alors, qu'est ce qui détermine le sexe de l'individu conçu ?
En l'absence d'autres informations ou de connaissances plus poussées, on est amené à considérer que ce phénomène est aléatoire ; en première approche, on peut considérer qu'à la surface de la terre, il naît autant d'hommes que de femmes (ce n'est pas tout à vrai). Du coup, on pourrait proposer une autre définition de la parité des sexes dans une assemblée, qui serait la suivante : il y a parité des sexes si, pour chaque individu, il y a autant de chance que ce soit un homme ou une femme.

Avec cette définition, pouvez-vous avoir une idée de la composition d'une assemblée de n personnes ?

Non, parce que cela dépend du hasard.
Pouvons-nous observer de telles assemblées préalablement constituées ?
Non.
Comment faire alors ?
Eh bien, il ne reste plus qu'à **simuler** de telles assemblées puis à observer leur composition.
Comment faire pour simuler un phénomène aléatoire qui n'a que deux résultats possibles, chacun ayant la même "chance" d'apparaître ?
"*Y a qu'à*" tirer à Pile ou Face et coder, par exemple : Face-Fille, Pile-Garçon !

C'est ainsi que commence dans ma classe de seconde l'étude des phénomènes aléatoires et de la première situation qui a pour titre "**Parité des sexes**".
L'expérience dont je rends compte dans la suite concerne une classe de 34 élèves ; P désigne "le professeur".

b. Pile ou Face

Après ce préambule, les élèves sont chargés de noter dans leur agenda un exercice à faire pour le lendemain, dont l'énoncé est le suivant :

> Simuler en tirant à Pile ou Face la répartition des sexes dans une assemblée de 34 personnes où la parité des sexes, au sens où nous l'avons définie, serait respectée (en codant Face-Fille et Pile-Garçon).

Ceci étant noté, on passe à un autre travail, qui concerne un autre chapitre. Une élève lance :

> Mais, il se peut qu'il y ait 99% de garçons !

Qu'en pensez-vous ?
Le lendemain, P fait circuler une feuille que les élèves renseignent pendant que le travail sur un autre chapitre se poursuit.
Le surlendemain, la feuille suivante est distribuée à chaque élève.

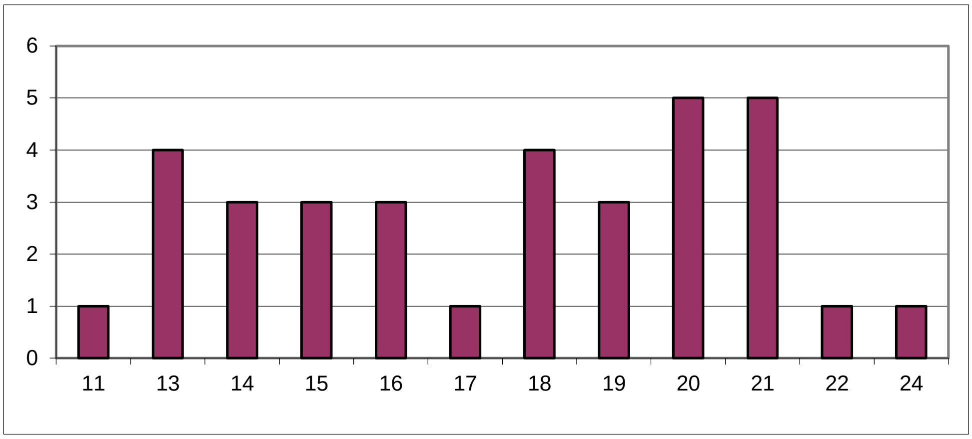
	34 lancers nombre de piles
Arnaud	18
Caroline	19
Brice	22
Laura	20
Elodie	16
Lydie	14
Mathilde	20
M-Charlotte	20
Lisa	19
Lucie	21
Emmanuelle	13
Chloé	14
Anne	15
Lirone	21
Cynthia	20
Elodie	16
Manon	17
Bertrand	15
David	13
Lise	21
Anaïs	18
Nicolas	19
Eve	18
Sylvain	24
Loïs	14
Elodie	13
Marine	20
Céline	18
Céline	11
Franck	13
Fanely	16
Olivier	21
Laure	15
Angélique	21

On commente les résultats. Les élèves sont étonnés des différences de résultats (de la **variabilité**). Pour mieux appréhender la variabilité, P propose d'étudier la série obtenue.

> Chaque élève a effectué un **échantillon** de taille 34et a compté le nombre de Pile. Ces nombres de pile forment une **série** de 34 nombres. Pour cette série, quels sont le mode ? la médiane ? la moyenne ? l'étendue ?

L'expérience m'a montré qu'il vaut mieux avoir retravaillé auparavant avec la classe, au moins rapidement, les notions de statistique descriptive du collège car certains élèves ne les ont plus du tout en tête. On obtient le tableau et le graphique suivants :

Dépouillement de la série															
Valeurs prises															
Effectifs															



Le graphique est commenté, les mots mode, moyenne, médiane, étendue prennent sens. C'est le moment d'institutionnaliser :

*Quand une expérience dépend du hasard, les échantillons n'ont pas tous la même composition, il y a **fluctuation d'échantillonnage**.*

Les élèves ne s'attendaient pas à autant de variabilité, de désordre. P peut alors déclarer que l'objet de la statistique c'est d'étudier cette variabilité et de dégager, à l'intérieur même de celle-ci, des propriétés, des régularités, dans la mesure du possible.

c. Mise en évidence d'un phénomène

P explique que pour observer un phénomène intéressant dans cette variabilité, il faudrait observer des échantillons de taille plus grande. Il n'est plus temps de lancer encore des pièces, d'où la question :

Pouvez-vous, à partir des résultats déjà obtenus par la classe, produire des échantillons de taille double ?

A chaque fois, il se trouve un élève pour proposer de multiplier par 2, c'est à dire dupliquer. Or si on duplique un échantillon, on ne peut plus prétendre qu'il a été produit au hasard. Que faire ?

*On peut **cumuler** des échantillons obtenus de manière **indépendante** pour obtenir un échantillon de taille plus grande.* *Si on duplique un échantillon, il n'est plus produit au hasard.*

"De manière indépendante" signifie ici que chaque élève a produit son échantillon indépendamment des autres élèves. On aborde ici un problème soulevé à la lecture des commentaires du programme de seconde (cf. II 4.) dans lesquels la question de l'indépendance des échantillons cumulés n'est pas

soulevée. P se trouve devant un choix difficile : ne pas évoquer l'indépendance des échantillons peut laisser croire aux élèves que le cumul peut se faire sans précaution, ce qui est faux ; parler d'indépendance sans la définir (c'est impossible ici) peut laisser croire que cette notion est intuitive et laissée à l'appréciation de chacun, ce qui bien sûr ne correspond pas du tout à la notion d'indépendance stochastique. Dans tous les cas, on risque de créer un obstacle didactique ; il faut faire un choix, qui n'a rien d'anodin.
Les élèves se groupent par 2 et forment des échantillons de taille 68. Puis chaque élève répond à la question suivante :

> Y a-t-il proportionnellement plus de Pile dans mon échantillon que dans celui de mon groupe ?

Ceci est l'occasion de passer aux fréquences pour la suite de la séquence.

> Pour comparer des échantillons de taille différentes, il est utile de passer aux fréquences.

P recueille les fréquences des échantillons de taille 68.
Le lendemain, la feuille suivante est distribuée à chaque élève. Dans la colonne A on trouve les fréquences produites par les élèves.

Fréquence de Pile dans 17 échantillons					
	A de taille 68	B de taille 340 68×5	C de taille 1020 340×3	D de taille 1360 340+1020	E de taille 2000
1	**0,49**				0,49
2	**0,47**		0,50	0,49	0,50
3	**0,63**		0,48	0,49	0,50
4	**0,56**	0,51	0,48	0,51	0,52
5	**0,51**	0,49	0,51	0,48	0,49
6	**0,47**	0,66	0,49	0,50	0,47
7	**0,65**	0,49	0,55	0,50	0,50
8	**0,50**	0,48	0,46	0,47	0,49
9	**0,47**	0,46	0,50	0,51	0,50
10	**0,60**	0,51	0,49	0,54	0,49
11	**0,46**	0,49	0,50	0,49	0,52
12	**0,37**	0,51	0,50	0,50	0,49
13	**0,41**	0,47	0,49	0,48	0,50
14	**0,54**	0,52	0,52	0,48	0,50
15	**0,60**	0,54	0,51	0,48	0,50
16	**0,46**	0,56	0,50	0,50	0,49
17	**0,56**	0,50	0,49	0,51	0,50

> Colonne B : Dans les cases déjà remplies, les " fréquence de Pile " ont été obtenues par le professeur par des moyens qui seront explicités plus tard.
> Peut-on compléter les cases vides à l'aide des résultats de la classe ?

Il s'agit là encore de cumuler des échantillons obtenus de manière indépendante. Les élèves proposent de cumuler les 5 premiers, puis les cinq autres, puis les cinq suivants. Il y a ici un travail **d'algèbre** à mener sans précipitation pour que les élèves comprennent comment établir le calcul de la fréquence de Pile dans l'échantillon de taille 340 à partir des fréquences de Pile dans les échantillons de taille 68. Le fait que les échantillons de départ aient tous la même taille joue bien sûr un rôle important.

Colonne C : Peut-on compléter la case vide à l'aide des résultats de la classe ?

Les élèves proposent de cumuler les trois échantillons précédents B1, B2, B3 ; on calcule la nouvelle fréquence.

Colonne D : Peut-on compléter la case vide à l'aide des résultats de la classe ?

Beaucoup sont tentés de cumuler les échantillons C1 et B1. Mais….il y a un problème, ces deux échantillons n'ont pas été obtenus de manière indépendante, B1 ayant servi à construire C1 ! La réponse à la question est : non. Pour obtenir tout de même un échantillon de taille 1360, les élèves proposent de prendre C1 et B4 ; on peut aussi prendre C1 et B5, etc. Les élèves sont très étonnés de voir que chacun peut produire un échantillon différent, qu'il n'y a une réponse unique, la même pour tout le monde. En fait, le phénomène que l'on va mettre en évidence ne dépend pas des données individuelles. Ceci est inhabituel pour les élèves (et aussi pour la plupart des professeurs de mathématiques) !
Ce tableau va permettre de produire un **graphique** qu'il s'agit de construire pas à pas, pour que les observations prennent sens.

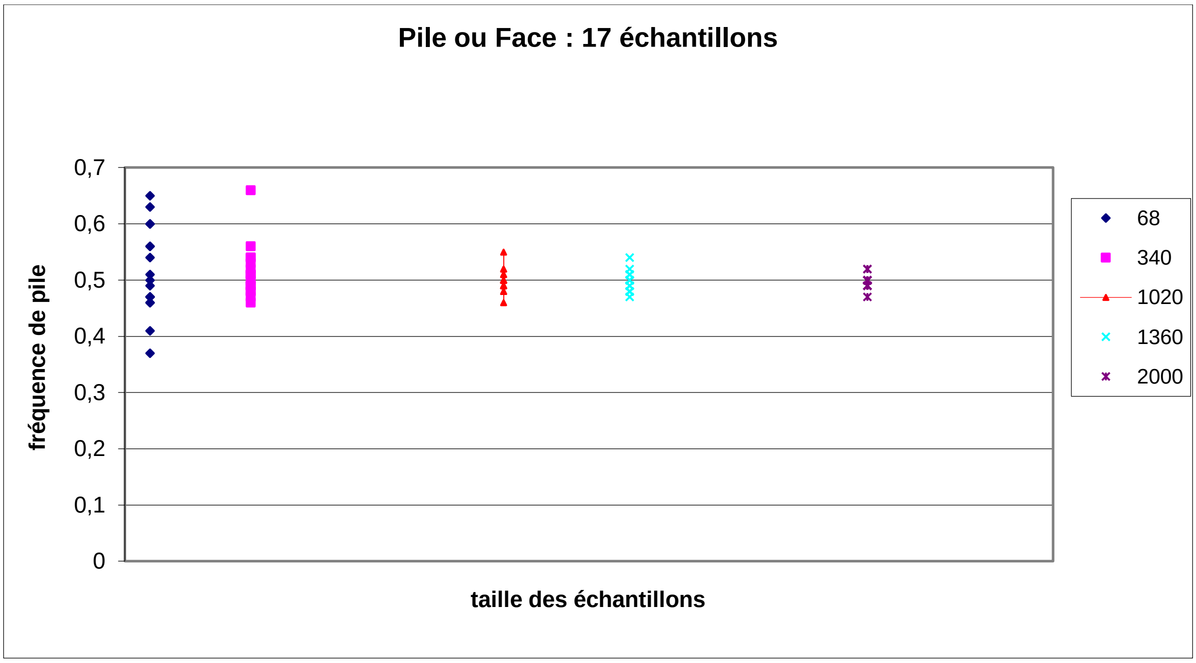
Dans un graphique, placer les points ayant pour abscisses la taille 68 et pour ordonnée les fréquences obtenues dans la colonne A. Qu'est-ce qui rend compte de la fluctuation d'échantillonnage sur le graphique ? Comment la mesurer ?

Certains élèves répliquent que la consigne est impossible à réaliser ! Pourquoi ?
Ces élèves ont étudié auparavant les fonctions et ont retenu (à la satisfaction de P) "*qu'à la verticale d'une valeur en abscisse, on ne peut trouver qu'un point*". Ceci est vrai lorsqu'on étudie un phénomène fonctionnel, et l'occasion est donnée de dire qu'ici, le phénomène n'est pas de ce type. On a ici un exemple de lien avec une autre partie du programme de seconde, ces liens contribuent à donner du sens à l'enseignement de notre discipline.
C'est justement le fait d'avoir plusieurs points sur la même verticale qui rend compte de la fluctuation d'échantillonnage et les élèves proposent, pour la mesurer, de calculer la différence entre la plus grande et la plus petite des fréquences, soit l'étendue de la série des 17 fréquences.

Dans le graphique précédent , placer les points ayant pour abscisses la taille 340 et pour ordonnée les fréquences obtenues dans la colonne B. Qu'est-ce qui rend compte de la fluctuation d'échantillonnage sur le graphique ? Comment la mesurer ?

Si l'on a accepté de passer du temps sur la question précédente, les élèves ont compris et vont cette fois beaucoup plus vite. On peut alors leur demander de terminer le graphique à la maison.

Terminer le graphique. Pouvez-vous faire des remarques ?



Les élèves n'ont aucun mal à interpréter le graphique, et il y a accord sur le constat suivant :

> Il semble que l'ampleur des fluctuations diminue quand la taille des échantillons augmente.

Ce graphique n'est pas dans les manuels scolaires de seconde. Je l'ai imaginé car les graphiques que j'y trouvais ne me satisfaisaient pas : ils illustraient plutôt la stabilisation de la fréquence quand la taille de l'échantillon augmente, ce qui n'est pas tout à fait la même chose et correspond plus exactement au programme des classes de première. Il me paraît important de signaler ceci aux stagiaires et de réfléchir un instant au statut des manuels scolaires, auxquels ils ont tendance à faire grandement confiance. L'occasion se présente ici de dire qu'il est nécessaire d'avoir toujours un point de vue critique sur les contenus des manuels, et qu'il ne faut hésiter à faire preuve de créativité pour construire "son" cours.

La suite de la séquence introduit des thèmes qui sont considérés comme facultatifs dans le programme. Ils auraient pu être passés sous silence, mais ils permettent d'une part de mathématiser de façon importante l'ensemble de la séquence, d'autre part de montrer quel type de réponse peut apporter une démarche statistique à un problème posé. C'est pourquoi ils ont été retenus.

Revenant au graphique et aux échantillons de taille 68, P affirme : "on démontre que la plupart des fréquences obtenues appartiennent à l'intervalle $[\frac{1}{2}-\frac{1}{\sqrt{68}} ; \frac{1}{2}+\frac{1}{\sqrt{68}}]$" , plus précisément

> 95% environ des fréquences obtenues appartiennent à l'intervalle $[\frac{1}{2}-\frac{1}{\sqrt{68}} ; \frac{1}{2}+\frac{1}{\sqrt{68}}]$.

Les élèves n'apprécient guère de voir apparaître une racine carrée, et s'empressent de chercher une valeur approchée à la calculatrice. Mais traîner 10 chiffres après la virgule n'est pas non plus très réjouissant et P propose de travailler avec deux chiffres après la virgule, **à la condition toutefois** que l'intervalle écrit avec des valeurs approchées ne soit pas "plus petit" que l'intervalle [

$\frac{1}{2}-\frac{1}{\sqrt{68}}$; $\frac{1}{2}+\frac{1}{\sqrt{68}}$]. Cette condition (justifiée par le fait que restreindre l'intervalle risquerait de faire oublier des fréquences) oblige à se poser des problèmes d'inclusion d'intervalles, à réfléchir à des questions d'ordre, à utiliser soit des troncatures, soit des valeurs approchées par excès. Cela n'a rien d'évident pour les élèves, ce travail est coûteux en temps mais il présente l'intérêt d'utiliser, en tant qu'outils, des objets du domaine **numérique**. Il faut beaucoup de temps pour arriver à un consensus :

> 95% environ des fréquences obtenues appartiennent à l'intervalle [0,37 ; 0,63].

L'intervalle [0,37 ; 0,63] peut être tracé verticalement sur le graphique.

On recommence avec les échantillons de taille 340 :

> 95% environ des fréquences obtenues appartiennent à l'intervalle
> $[\frac{1}{2}-\frac{1}{\sqrt{340}} ; \frac{1}{2}+\frac{1}{\sqrt{340}}]$.
> 95% environ des fréquences obtenues appartiennent à l'intervalle [0,44 ; 0,56].

Etc.
Lorsque les élèves ont bien compris, on peut énoncer la **propriété** :

> Pour un phénomène aléatoire dont on connaît la fréquence théorique p, on forme des échantillons de taille n. Dans environ au moins 95 % des cas, l'intervalle $[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}]$ contient la fréquence observée dans l'échantillon (ceci est d'autant mieux vérifié que la taille ne des échantillons et le nombre de tirage sont plus élevés).

d. Retour à la question de départ

On peut enfin revenir à la question de départ, celle de la parité des sexes dans la classe et, après discussion, terminer sur cette conclusion :

> Si l'on considère qu'il y a parité des sexes quand chaque individu a autant de chances d'être un garçon ou une fille, au "niveau de confiance 0,95", on ne rejettera pas l'idée qu'il y a parité dans la classe si la fréquence de filles observée appartient à l'intervalle $[\frac{1}{2}-\frac{1}{\sqrt{34}} ; \frac{1}{2}+\frac{1}{\sqrt{34}}]$.

La réponse revêt une forme très différente de celle donnée au début.

e. Remarques

Il arrive, qu'en formation, lorsqu'on étudie cette situation, un stagiaire demande un complément d'information sur mode, ou médiane, etc. Si c'est le cas, il est utile de ménager une digression et d'étudier pendant quelques instants le document intitulé "**SERIE STATISTIQUE A UNE VARIABLE REELLE**" (cf. ANNEXE II) qui permet de faire une synthèse. Ce document pourrait très bien figurer dans le "cahier de statistique" de 2° pour être repris en 1° et complété par "écart type"et "variance". Cela confèrerait de la cohérence à l'enseignement dispensé aux élèves en lycée et ferait gagner du temps aux professeurs de 1°. Mais bien sûr, cela n'est possible que si les professeurs d'un même établissement travaillent en équipe et s'entendent sur des définitions et des notations communes….

Le document présentant la situation "**Parité des sexes**" effectivement distribué aux stagiaires se trouve en ANNEXE III.

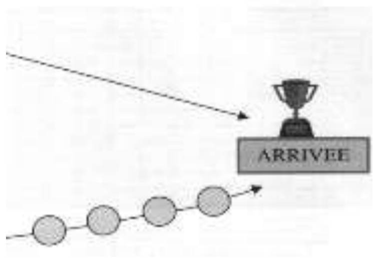
6. Deuxième situation : Vaut-il mieux être le joueur A ou le joueur B ?

a. Un jeu

La question suivante est soumise aux élèves.

Voici un jeu , pour lequel on utilise un dé à 6 faces, équilibré. On le lance.
• Si le dé tombe sur 6, le joueur B gagne directement la partie.
• Si le dé tombe sur 1 ou 2 ou 3 ou 4 ou 5, le joueur A avance d'une case. On relance. Pour que A gagne la partie, il faut qu'il ait avancé 5 fois.
Vaut-il mieux être le joueur A ou le joueur B ?

B



A

Il est important de comprendre que l'objectif n'est pas de jouer à ce jeu mais de **prendre une décision.**

Les élèves ont du mal à s'approprier l'énoncé et il est nécessaire de faire quelques parties avec eux pour clarifier les règles du jeu . On peut utiliser une grille préparée à l'avance pour gagner du temps.

Lancer de dé	1	4	6	5	4	4	2	1	3	3	6	2

Lancer de dé	1	1	4	3	5	4	6	3	2	5	3

Lancer de dé	6	1	2	4	1	5	6	2	4	4	2

Comment savoir si le jeu est plus favorable à A ou à B ou bien s'il est équitable ?
Après réflexion et discussion, les élèves proposent de jouer un nombre assez grand de parties pour voir qui de A ou de B gagne plus souvent. Seulement il y a un problème, aucun n'a de dé ! P propose alors :

b. Un outil pour simuler le hasard : la touche random de la calculatrice

Après avoir repéré la touche random sur leur calculatrice, les élèves sont invités à l'actionner plusieurs fois de suite.

Suivant les modèles, il peuvent voir apparaître

.1216853324
.4930062012
.534806747
.6448564206

ou bien

0,153
0,25
0,287
0,533

Si on laisse les élèves décrire ce qu'ils voient apparaître, ils disent qu'ils voient des nombres décimaux, avec presque toujours le même nombre de décimales. On aborde ici un autre problème, laissé sous silence dans les commentaires (cf. II 4.). En fait, la calculatrice est programmée pour éditer un nombre fixe de décimales, mais elle est aussi programmée pour ne pas écrire les zéros "inutiles" dans une écriture décimale. Et dans les nombres qui semblent avoir moins de décimales, il faut toujours restituer les zéros manquants. Les parties décimales des nombres affichés fournissent une liste de chiffres, dont P doit préciser le statut.

On admet que
- On ne peut pas prévoir les chiffres qui vont apparaître (**imprédictibilité**)
- La sortie d'un chiffre ne dépend pas des chiffres qui sont déjà sortis (**indépendance ... et répétition des chiffres possible**)
- Sur un grand nombre de tirages, on observe une relative **équirépartition** des chiffres (à condition d'avoir restitué les zéros manquants!)

Avec le nouvel outil que constitue une liste de chiffres au hasard, comment simuler le lancer d'un dé ? Certains élèves devinent ce qu'il faut faire, mais rien dans leur culture antérieure ne leur permet d'être sûrs de leur procédure et P doit valider.

Après avoir restitué les zéros manquants, on conserve les chiffres 1, 2, 3, 4 , 5 , 6 et on barre les autres. On admet que les chiffres restants simulent des lancers aléatoires d'un dé équilibré.

Certains élèves font remarquer, avec raison, qu'il ne servait à rien de restituer les zéros manquants puisque après, on les barre. C'est une faiblesse de cette situation qui propose en première utilisation de la liste de chiffres au hasard un cas où les zéros ne sont pas utiles.

c. Exercice

Chaque élève note dans son agenda l'exercice à faire pour la fois suivante.

Simuler 10 parties du jeu et déterminer, dans l'échantillon obtenu, la fréquence de gain de A et la fréquence de gain de B.

Notons que le nombre de chiffres nécessaires pour simuler 10 parties ne peut pas être connu à l'avance, cela dépend des parties, et cet aspect présente une difficulté pour certains élèves.
Le lendemain, P relève les résultats de la classe, le surlendemain, la feuille récapitulative est distribuée à chaque élève.

Echantillons de taille 10	fréq. gain A	fréq. gain B
Chloé	0,3	0,7
Franck	0,3	0,7

Anne	0,6	0,4
Fanely	0,4	0,6
Lirone	0,4	0,6
Cynthia	0,4	0,6
Olivier	0,5	0,5
Laure	0,4	0,6
Elodie	0,6	0,4
Manon	0,1	0,9
Angélique	0,4	0,6
Bertrand	0,3	0,7
Celine	0,2	0,8
Marine	0	1
Lucie	0,5	0,5
M- Charlotte	0,2	0,8
Loïs	0,3	0,7
Mathilde	0,3	0,7
Sylvain	0,4	0,6
David	0,3	0,7
Caroline	0,5	0,5
Lise	0,4	0,6
Brice	0,4	0,6
Anaïs	0,2	0,8
Laura	0,2	0,8
Nicolas	0,2	0,8
Eve	0,4	0,6
Celine	0,3	0,7
Lydie	0,5	0,5
Emmanuelle	0,2	0,8
Elodie	0,4	0,6
Elodie G	0,2	0,8
Lisa	0,5	0,5
Arnaud	0,4	0,6
Echantillon de taille …	….	…..

A partir de ces résultats, on peut cumuler les échantillons pour obtenir un échantillon de taille 340, calculer la fréquence de gain de A grâce à une formule étudiée dans la première situation, remarquer qu'on obtient facilement la fréquence de gain de B par complémentaire à 1.
Pour autant, peut-on prendre une décision ? P peut alors énoncer un nouveau théorème qui diffère de celui de la première situation par le fait qu'ici on ne connaît pas la fréquence du phénomène étudié.

d. Propriété fondamentale et définition

On considère un phénomène aléatoire dont on ignore la fréquence p et pour lequel, dans un échantillon de taille n, ($n > 30$), on a observé une fréquence f ($0,3 < f < 0,7$).
Quelle que soit la valeur de p, dans au moins 95% des cas environ, l'intervalle $[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}]$ contient p inconnue. Cet intervalle est appelé intervalle de confiance*, au niveau de confiance 0,95.
* ou fourchette de sondage

Cette propriété est appliquée dans les questions suivantes :

> A l'aide de l'échantillon obtenu par la classe, déterminer la fourchette au niveau de confiance 0,95 pour fréquence de A. Sur une droite graduée, construire l'intervalle.
> Déterminer la fourchette au niveau de confiance 0,95 pour fréquence de B. Sur la même droite graduée, construire l'intervalle.
> Pouvez-vous décider s'il vaut mieux être le joueur A ou le joueur B ?

Les deux fourchettes ne se recouvrent pas, les élèves concluent qu'il vaut mieux choisir le rôle de B. Il est important alors de demander :
▪ à partir de l'étude précédente, est-on sûr que le rôle de B est plus enviable ? (réponse "non", l'étude théorique faite en première ou terminale permettra de démontrer qu'il en est ainsi).
▪ vous êtes le joueur B, vous savez que son rôle est plus enviable, vous engagez une partie, êtes vous sûr de gagner ? (réponse "non" bien sûr, on est dans un domaine où aucune certitude n'est possible).

e. Prolongement

L'étude menée permet de terminer par cet exercice dont l'intérêt relativement à la formation du citoyen est incontestable.

> Lors d'une élection présidentielle, le candidat A et le candidat B sont en concurrence.
> 1) Une semaine avant l'élection, un journaliste annonce à la télévision : "**un sondage donne 51% d'intentions de vote pour A et 49% pour B**" . Ton frère te dit : "Je pense que A va être élu !" Le sondage a été réalisé sur un échantillon de taille 1000.
> a) Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat A et trace en rouge cette fourchette sur un axe. Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat B et trace en vert cette fourchette sur le même axe.
> b) Que réponds - tu à ton frère?
>
> 2) Le soir de l'élection à 20 h, un journaliste annonce à la télévision : "**un sondage donne 51% de voix pour A et 49% pour B**" . Ton frère te dit : "Je pense que A va être élu !" Le sondage a été réalisé sur un échantillon de taille 100 000.
> a) Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat A et trace en rouge cette fourchette sur un axe. Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat B et trace en vert cette fourchette sur le même axe.
> c) Que réponds - tu à ton frère?

Remarque : Le document présentant la situation "**Vaut-il mieux être le joueur A ou le joueur B ?** " effectivement distribué aux stagiaires se trouve en ANNEXE IV.

7. Troisième situation : Etude d'une distribution de fréquences

On s'intéresse au lancer d'un dé équilibré.

a. Distribution de fréquences dans deux échantillons de taille 10

Constituer un échantillon de taille 10 d'un lancer de dé et placer les résultats dans le tableau ci-dessous dans la ligne "Xavier".

Distribution des fréquences pour 10 lancers						
	1	**2**	**3**	**4**	**5**	**6**
Echantillon de Xavier						
Echantillon d'Yves	**0**	**0,2**	**0,1**	**0,2**	**0,2**	**0,3**

Ici, chaque élève va travailler à partir de ses propres résultats.

Construire un diagramme en bâtons comparant l'échantillon X et l'échantillon Y. Constatez-vous une fluctuation d'échantillonnage ? Qu'est ce qui en rend compte sur le graphique.

Il est important de prendre le temps d'interpréter le graphique et de traduire en mots ce qu'il suggère.

Comment mesurer l'ampleur des fluctuations ?

Généralement, les élèves proposent de calculer les différences de fréquences pour chaque occurrence et d'en faire la somme. Il est facile de voir avec eux que les termes de cette somme peuvent être tantôt positifs, tantôt négatifs et que la somme peut être un nombre nul ou négatif, dont la signification n'est pas opérationnelle. On est vite amené à utiliser la **valeur absolue** des différences, qui garantit que chaque terme sera positif et on peut proposer comme mesure de l'ampleur des fluctuations la somme des valeurs absolues des différences de fréquences (suivant la réceptivité des élèves, on peut aussi proposer la moyenne des valeurs absolues des différences de fréquences ou bien même le maximum des valeurs absolues des différences de fréquences). A nouveau, un objet mathématique du **numérique** apparaît comme un outil efficace pour étudier une question de statistique.

Calculer la moyenne des points obtenus pour l'échantillon X et pour l'échantillon Y. Placer ces deux valeurs en abscisse.

Cette question est là pour faire apparaître que la fluctuation d'échantillonnage porte non seulement sur la distribution de fréquences, mais aussi sur les paramètres associés.
Une fois que les élèves ont bien compris cette étude, on peut donner la suite du travail en devoir maison.

b. Distribution de fréquences dans deux échantillons de tailles plus grandes

b1. Distribution des fréquences pour 100 lancers (Résultats arrondis à 10^{-2} près obtenus par simulation) :

	1	2	3	4	5	6
échantillon X	**0,11**	**0,13**	**0,16**	**0,15**	**0,21**	**0,24**
échantillon Y	**0,17**	**0,13**	**0,15**	**0,15**	**0,25**	**0,15**

Construire un diagramme en bâton comparant l'échantillon X et l'échantillon Y. Constatez-vous une fluctuation d'échantillonnage ? Calculer une mesure de l'ampleur des fluctuations.
Calculer la moyenne des points obtenus pour l'échantillon X et pour l'échantillon Y. Placer ces deux valeurs en abscisse.

b2. Distribution des fréquences pour 1000 lancers (Résultats arrondis à 10^{-3} près obtenus par simulation) :

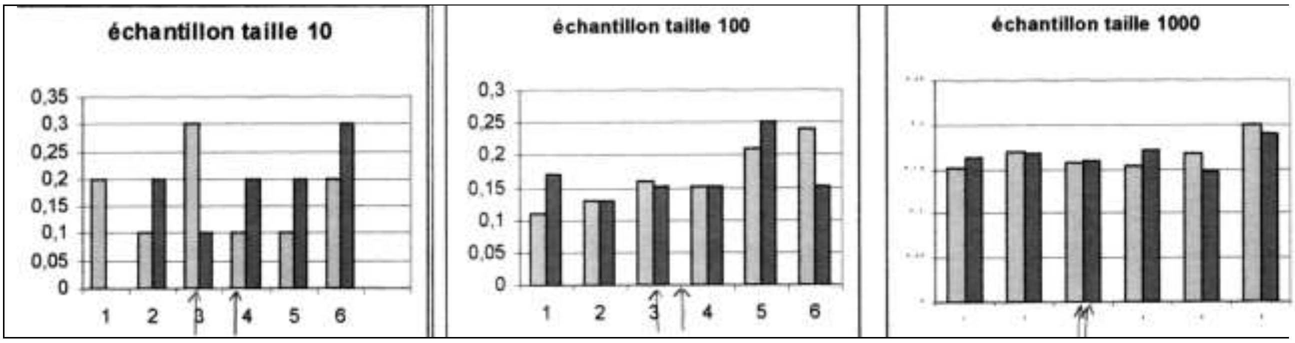
	1	2	3	4	5	6
échantillon X	**0,152**	**0,170**	**0,158**	**0,153**	**0,168**	**0,199**
échantillon Y	**0,164**	**0,167**	**0,160**	**0,172**	**0,148**	**0,189**

Construire un diagramme en bâton comparant l'échantillon X et l'échantillon Y. Constatez-vous une fluctuation d'échantillonnage ? Calculer une mesure de l'ampleur des fluctuations.
Calculer la moyenne des points obtenus pour l'échantillon X et pour l'échantillon Y. Placer ces deux valeurs en abscisse.

b3. Quelles remarques peut-on faire en comparant les trois études précédentes?

b4. Blaise te propose un jeu : tu mises 3 euros, tu lances un dé, si le chiffre apparu sur le dé est x, tu remportes x euros. Es-tu prêt à jouer 1 fois, 10 fois, 100 fois, 1000 fois ? Pourquoi ?

Voici les résultats obtenus par un élève :



Ils sont aisément interprétables. Les élèves répondent à la question b3, avec des formulations plus ou moins heureuses. Par contre, la question b4, qui aborde sans le dire la notion d'espérance mathématique, est difficile pour eux.

c. Prolongement

Avec les stagiaires (non avec les élèves de 2°), on peut profiter de cette étude pour aborder le lien préconisé dans les programmes de 1ere entre distribution de fréquences et loi de probabilité et regarder l'extrait du document d'accompagnement du programme de 1ere qui suit.

Distribution de fréquences et loi de probabilité

Extrait du document d'accompagnement du programme de 1°S, ES

Distribution de fréquences sur E ={ $x_1, x_2, \dots x_r$}					**Loi de probabilité sur E = { $x_1, x_2, \dots x_r$}**				
$\mathbf{x_i}$	x_1	x_2	…	x_r	$\mathbf{x_i}$	x_1	x_2	…	x_r
$\mathbf{f_i}$	f_1	f_2	…	f_r	$\mathbf{p_i}$	p_1	p_2	…	p_r

$(f_1,\dots,f_i)$ $f_i \geq 0; \sum f_i = 1$ A$\subset$ E, fréquence de A : $f(A) = \sum_{x_i \in A} f_i$ Événement complémentaire : $f(\bar{A}) = 1 - f(A)$ Événements A et B disjoints : $f(A \cup B) = f(A \text{ ou } B) = f(A) + f(B)$ Cas numérique : Moyenne empirique : $\bar{x} = \sum f_i x_i$ Variance empirique : $s^2 = \sum f_i (x_i - \bar{x})^2$ Écart type empirique : $s = \sqrt{\sum f_i (x_i - \bar{x})^2}$	$(p_1,\dots,p_r)$ $p_i \geq 0; \sum p_i = 1$ A$\subset$ E, probabilité de A : $P(A) = \sum_{x_i \in A} p_i$ Événement complémentaire : $P(\bar{A}) = 1 - P(A)$ Événements A et B disjoints : $P(A \cup B) = P(A \text{ ou } B) = P(A) + P(B)$ Cas numérique : Espérance d'une loi P : $\mu = \sum p_i x_i$ Variance d'une loi P : $\sigma^2 = \sum p_i (x_i - \mu)^2$ Écart type d'une loi P : $\sigma = \sqrt{\sum p_i (x_i - \mu)^2}$

Remarque : Le document présentant la situation "**Une distribution de fréquences**" effectivement distribué aux stagiaires se trouve en ANNEXE V.

8. Quelques simulations

Si le temps le permet, mais c'est rarement le cas, la séance de formation de trois heures peut s'achever sur un travail de groupe au cours duquel les stagiaires produisent des simulations demandées, certains avec l'aide d'une liste de chiffres au hasard, d'autres avec une calculatrice. Il est intéressant, lors de la mise en commun, de comparer les méthodes et de constater que les connaissances mises en jeu dans l'un et l'autre cas sont de nature très différentes. Voici des exercices possibles :

1. Simuler 20 lancers de Pile ou Face d'une pièce équilibrée.

2. Dans une urne, il y a 3 boules rouges, et 7 boules vertes. On tire au hasard une boule dans l'urne, on note sa couleur, on la remet dans l'urne. Simuler 20 tirages.

3. Dans une urne, il y a 3 boules rouges, et 5 boules vertes. On tire au hasard une boule dans l'urne, on note sa couleur, on la remet dans l'urne. Simuler 20 tirages.

4. Dans une urne, il y a 3 boules rouges, et 2 boules vertes et 4 boules noires. On tire au hasard une boule dans l'urne, on note sa couleur, on la remet dans l'urne. Simuler 20 tirages.

5. On lance deux pièces équilibrées et on regarde les marques. Simuler 20 lancers de ces deux pièces.

6. Simuler un tirage du loto.

7. Simuler le choix au hasard de 3 élèves parmi 30.

8. Dans un pays imaginaire, une politique nataliste est mise en place : les naissances dans chaque famille s'arrêtent soit à la naissance du premier garçon, soit à la naissance du quatrième enfant. Par exemple, G ou bien FFG, ou bien FFFF sont des compositions de famille possibles. Quelle conséquence peut-on prévoir sur la répartition entre les sexes ?

~ DEUXIEME PARTIE ~

L'exposé a été suivi d'une discussion à la fin de laquelle a été présenté le bilan qui suit.

Fin mai, par mail, un questionnaire a été soumis aux stagiaires PLC2 ayant participé à la formation en novembre 2006 ou en mars 2007 ; un quart des stagiaires environ a répondu.

Questionnaire

1. Selon vous, en quoi devrait consister une formation en statistique à l'IUFM pour les PLC2 ?

2. Le …., vous avez assisté à un GFD sur l'enseignement de la statistique en collège/lycée. Quel est votre point de vue sur cette séance ? (forme, contenus, documents, autres, etc.)

3. Autre

Cette démarche n'a aucune valeur statistique, elle présente des biais importants. En particulier, les personnes qui répondent sont, le plus souvent, satisfaites de la formation. Concernant ceux qui n'ont pas répondu, on ignore s'ils étaient mécontents de la formation et pourquoi ou si, simplement, ils n'avaient pas envie de répondre.
Toutefois, les réponses exprimées présentent un certain intérêt, en voici une synthèse (en *italique* des extraits des réponses exprimées).

1. Selon vous, en quoi devrait consister une formation en statistique à l'IUFM pour les PLC2 ?

a. Compléments théoriques

- apporter des connaissances en stats
- rappeler les notions à connaître

b. Informations sur l'utilisation des statistiques dans la société

- montrer en quoi la statistique est utile
- comprendre le travail des statisticiens
- recevoir des statisticiens d'IPSOS ou autre pour qu'ils expliquent leur façon de travailler

pour

- que notre enseignement ait du sens, soit plus vivant
- permettre aux élèves de développer un esprit critique vis-à-vis des sondages, diagrammes dont ils peuvent être abreuvés au quotidien

c. Etude des programmes en statistique

- saisir l'esprit de ce que l'on nous demande d'enseigner
- travailler sur les liens à établir entre chaque classe pour qu'il n'y ait ni rupture ni répétition dans l'enseignement des statistiques entre collège et lycée
- mettre la statistique en relation avec les autres thèmes mathématiques

d. Promotion de la statistique

faire prendre conscience aux stagiaires qu'il s'agit une véritable branche des mathématiques, non plus une « sous-matière » que l'on traite rapidement en fin d'année.

e. Elaboration de séquences

- travailler sur des exemples d'activités que l'on pourrait donner aux élèves et les analyser
- montrer comment aborder ce thème avec les élèves
- réfléchir, sur des cas d'espèce, à la manière de l'enseigner
- sortir avec un large éventail de propositions spécifiques aux statistiques

f. Et aussi …

J'ai fait un mémoire sur le test de Durbin Watson. Pour cela, on a simulé des nombres à l'aide d'ordinateurs via des lois et des outils de probabilités et statistiques. J'ai toujours beaucoup de difficulté à admettre qu'un ordinateur peut générer du "vrai" hasard. J'aimerais une formation qui nous permette de savoir comment ça fonctionne.

2. Le …., vous avez assisté à un GFD sur l'enseignement de la statistique en collège/lycée. Quel est votre point de vue sur cette séance ? (forme, contenus, documents, autres, etc.)

a. Point de vue favorable

très bien, excellent, intéressant, concret, bien mené, utile, cohérent, instructif, pertinent,

Particulièrement sur :

- la bibliographie *: indispensable pour qui souhaite aller plus loin*
- le rappel sur les programmes, *car on ne les a pas toujours en tête*
- le tableau récapitulatif des compétences attendues en fonction des classes
- l'exemple sur la notion de parité, *bien que je pense que la parité homme/femme en politique n'est pas comparable à la parité pile/face d'une pièce*
- l'exemple sur les sondages
- le fonctionnement de la statistique avec la taille de l'échantillon et le niveau de confiance
- des traces écrites : *c'est très intéressant pour la suite de notre enseignement des statistiques surtout en classe de seconde*
- les exemples d'activité *pour intéresser les élèves, comment l'aborder et où les élèves allaient avoir des difficultés*
- la donnée d'une séance de cours complète

  une solide base de travail et de recherche pour fabriquer par la suite notre propre séance
  une séance clefs en main à mettre en place dans notre classe

b. Mais

▪ On aurait pu :

- réduire le temps de lecture des programmes
- nous demander de faire des propositions avant la séance de manière à mettre en connexion les lacunes propres de notre groupe et les réponses que l'ensemble pouvait apporter
- nous donner un cours de première et TS (ou ES) accompagné d'exercices types pour avoir un aperçu de ce vers quoi doit tendre l'enseignement des stat

▪ Trois heures, c'est trop court. Avec un temps de formation plus long, on pourrait :

- approfondir davantage *des points clés de l'enseignement des statistiques*
  leur utilisation dans la formation de l'esprit critique des élèves
- ménager une participation plus importante des stagiaires, *pour leur faire exprimer leur vision des statistiques, qui est souvent erronée, et leur permettrait de mieux comprendre vos propos*
- construire nous-même des préparations de séquences, les expérimenter, en rendre compte (sous forme de fil rouge).

c. Conclusion

▪ De l'enthousiasme

- J'ai appris ce que l'on pouvait faire en seconde.
- Je suis désormais convaincu de l'importance des statistiques dans l'enseignement !
- Je pense que plus de statistiques pourrait rendre les mathématiques moins "aride " (pour les élèves). En effet l'expérimentation est bien plus présente.

▪ A modérer !

Sur le coup, j'ai eu l'impression que l'on pouvait rendre cet enseignement des statistiques très simple. Je me suis dit que j'allais tenter de le faire en prenant vos polycops comme base de travail pour ma classe : l'activité de départ était très motivante, mais je me suis aperçu que personne n'était au même niveau concernant le vocabulaire de base (moyenne, médiane, etc.) alors j'ai choisi de clarifier ces termes pour tous [...].

J'ai également rencontré des problèmes pour justifier le tableau de fréquence de Pile dans 17 échantillons lors du passage à des échantillons de plus grande taille (II. 2).
Du coup, mon cours a pris une forme très différente, mais je ne suis pas sûre que ce soit satisfaisant par rapport aux programmes (je vous le joins, pour que vous puissiez me dire ce que vous en pensez...).

~ Bibliographie ~

ASSUDE T. (2004), *L'étude du curriculum de mathématiques entre changements et résistances. Liens entre Ecologie et économie didactique*. Document pour l'habilitation à diriger des recherches.

BROUSSEAU G., *Stratégies de l'analyse statistique*. LADIST Université Bordeaux I.

BROUSSEAU G., BROUSSEAU N., WARFIELD G. (2002), An experiment on the teaching of statistics and probability, *Journal of mathematical behavior*, 20.

CHEVALLARD Y. (2002) Organiser l'étude, Cours1&3. Structures et fonctions. In Dorier JL et alii (ed) *Actes de la XI° Ecole d'été de didactique des mathématiques)*. La Pensée Sauvage.

CREM (Commission de Réflexion sur l'Enseignement des Mathématiques présidée par JP. Kahane), (2001), *Rapport d'étape "Statistique et Probabilités*"/ (2003), *Rapport d'étape "Statistique ".*

ESCOUFIER Y. (2000), La formation à la statistique. In Académie des Sciences, *La statistique*, rst n°8. Tec & Doc.

MEN (1999), *Programme de seconde*, B.O. Hors-Série n° 6 du 16 août 1999/ (2000), A*ccompagnement des programmes, Mathématiques, classe de seconde*. CNED.

ROBERT C. (1999), A propos de l'introduction de l'enseignement de la statistique dans les lycées. *Bulletin APMEP n°425.*

VERGNE C. (2005), *La notion de variabilité dans le nouveau programme de seconde : étude de conditions de viabilité*, Mémoire de DEA, IREM de Montpellier.

~ ANNEXE I ~

BIBLIOGRAPHIE

LIVRES

ROBERT C. (2003), *Contes et décomptes de la statistique*, Vuibert.
une approche plaisante de la statistique

SCHWARTZ D.(1997, 1°édition 1994), *Le jeu de la science et du hasard*, Flammarion.
présente quelques idées incontournables sur les statistiques, facile à lire

EKELAND I. (1991), *Au hasard*, Points Sciences.
des passages passionnants (chiffres au hasard, déterminisme, suites contingentes…)

LAHANIER-REUTER D.(1999), *Conceptions du hasard et enseignement des probabilités et statistiques*, PUF.
où l'on définit 5 conceptions du hasard

BESSON et al.(1992), *La cité des chiffres*, Autrement.
un autre regard sur les statistiques, très intéressant

CII Statistique/Probabilités, *Autour de la modélisation en probabilités*, PUFC, Collection Didactiques.

DROESBEKE et TASSI (1990), *Histoire de la statistique*, Que sais-je.
pour se faire une idée de l'évolution des statistiques

VESSEREAU A. (20° édition 1999), *La statistique*, Que sais-je.
pour une première approche des contenus d'un point de vue théorique

BROCHURES

DUTARTE /PIEDNOIR (2001), *Enseigner la statistique au lycée : des enjeux aux méthodes*, Irem Paris-Nord.
un excellent aperçu de "ce qu'il faut savoir" pour enseigner en lycée

GIRARD/GROS/PLANCHETTE/REGNIER/THOMAS (1998), *Enseigner la statistique du CM à la seconde. Pourquoi, comment ?*, IREM de Lyon.
des points de vue passionnants ; le point sur les histogrammes

REVUES

Tangente n° spécial 47 (sur les probabilités) et **Tangente n° spécial 77** (sur les statistiques).
certains articles sont intéressants

Dossier pour la science, *Le hasard*, et Dossier pour la science, *Les mathématiques sociales*. Belin.
certains articles sont intéressants

Sciences et Avenir (oct-nov 2001), *Dieu joue-t-il aux dés ?*
différents points de vue sur les probabilités

SITES

Centre de ressources, lieu de partage et de mutualisation pour l'enseignement de la statistique :
www.statistix.fr

Des simulations des thèmes statistiques de seconde :
http://www.irem.univ-mrs.fr
http://perso.wanadoo.fr/jpq

Des simulations dynamiques de quelques concepts de base en statistique :
http://www.kuleuven.ac.be/ucs/java/index.htm

GEPS :
http://www.ac-poitiers.fr/gtdmaths/
http://www.ac-strasbourg.fr/pedago/gtd/

Un logiciel pour la formation (articles, cours, simulation dynamiques) :
http://www.inrialpes.fr/sel/

~ ANNEXE II ~

STATISTIQUE A UNE VARIABLE REELLE

<table>
<tr><td></td><td colspan="7">Variable discrète</td><td colspan="6">Variable continue ou considérée comme telle</td></tr>
<tr><td></td><td>Série brute</td><td colspan="6">Série dépouillée (non chronologique)</td><td colspan="6">Regroupement par classe</td></tr>
<tr><td rowspan="4">Données</td><td rowspan="4">Valeurs : v_1 , v_2 , …….,v_n
L'effectif est n.</td><td>Valeurs*</td><td>x_1</td><td>x_2</td><td>…</td><td>x_r</td><td></td><td>Classes*</td><td>$[a_0, a_1[$</td><td>$[a_1, a_2[$</td><td>….</td><td>$[a_{r-1}, a_r[$</td><td></td></tr>
<tr><td>Effectifs partiels</td><td>n_1</td><td>n_2</td><td>…</td><td>n_r</td><td>n</td><td>Effectifs partiels</td><td>n_1</td><td>n_2</td><td>…</td><td>n_r</td><td>n</td></tr>
<tr><td>Fréquences</td><td>$f_1 = \frac{n_1}{n}$</td><td>$f_2 = \frac{n_2}{n}$</td><td>…</td><td>$f_r = \frac{n_r}{n}$</td><td>1</td><td>Fréquences</td><td>$f_1 = \frac{n_1}{n}$</td><td>$f_2 = \frac{n_2}{n}$</td><td>…</td><td>$f_r = \frac{n_r}{n}$</td><td>1</td></tr>
<tr><td colspan="6">*dans l' ordre croissant</td><td colspan="6">* sauf mention contraire dans certains cas pour les classes extrêmes</td></tr>
<tr><td>Représe n--tation graphique</td><td></td><td colspan="6">Diagramme en bâtons ou en barres :

A chaque valeur de x_i portée en abscisse, on fait correspondre un segment vertical de longueur proportionnelle à l'effectif n_i ou à la fréquence f_i.
L'épaisseur des segments est quelconque.</td><td colspan="6">Histogramme :
Ensemble de rectangles contigus, chacun ayant pour base la largeur de la classe et une aire proportionnelle à la fréquence de la classe (ou à l'effectif).
Chaque rectangle a une hauteur proportionnelle à la densité de fréquence $\frac{f_i}{a_i - a_{i-1}}$ (ou d'effectif).
L'aire totale de l'histogramme représente l'unité (ou l'effectif total).</td></tr>
<tr><td>Mode</td><td colspan="7">Toute valeur de la série qui a le plus grand effectif.</td><td colspan="6">Classe modale :
classe dont la densité de fréquence est la plus élevée.</td></tr>
<tr><td>Moyenne</td><td>Erreur ! = Erreur !</td><td colspan="6">Erreur ! = Erreur !
Erreur ! = $f_1x_1 + … + f_rx_r$</td><td colspan="6">Pour une valeur approchée, par convention : on remplace chaque classe par son centre* ($\frac{a_{i-1} + a_i}{2}$) que l'on nomme x_i puis on applique les formules ci-contre.
*sauf mention contraire dans certains cas pour les classes extrêmes.</td></tr>
<tr><td>Médiane</td><td colspan="7">La série est ordonnée par ordre croissant
► Pour les calculatrices :
▪ s'il y a un nombre impair de valeurs, la médiane est la valeur du milieu
▪ s'il y a un nombre pair de valeurs, la médiane est la ″demi-somme des valeurs du milieu″
► Il y a au moins 50% des valeurs qui sont inférieures ou égales à Me, il y a au moins 50% des valeurs qui sont supérieures ou égales à Me.</td><td colspan="6">Classe médiane :
première des classes pour laquelle la fréquence cumulée croissante atteint ou dépasse 0,5.</td></tr>
</table>

Etendue	$e = v_{max} - v_{min}$	$e = x_r - x_1$	
Ecart type	Noté s (standard déviation) ou σ ou σ_x (calculatrices) s = **Erreur !**	Noté s (standard déviation) ou σ ou σ_x s = **Erreur !** s = **Erreur !**	Pour une valeur approchée, par convention : on remplace chaque classe par son centre* que l'on nomme x_i puis on applique les formules ci-contre. *sauf mention contraire dans certains cas pour les classes extrêmes.
Variance	Le carré de l'écart type (s^2 ou V) $s^2 = \frac{v_1^2 + \ldots + v_n^2}{n}$ – **Erreur !**^{2}	Le carré de l'écart type (s^2 ou V) $s^2 = \frac{n_1x_1^2 + \ldots + n_rx_r^2}{n}$ – **Erreur !**^{2} $s^2 = f_1x_1^2 + \ldots + f_rx_r^2$ – **Erreur !**^{2}	Le carré de l'écart type (s^2 ou V)

~ ANNEXE III ~

SIMULATION

ET

FLUCTUATION D' ECHANTILLONNAGE

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

I PILE ou FACE

1) Exercice (à faire pour demain)

Simuler en tirant à Pile ou Face la répartition des sexes dans une assemblée de 34 personnes (Face : fille ; Pile : garçon).

Remarque de Manon : mais , il se peut qu'il y ait 99% de garçons !

2) Recueil des résultats

Le lendemain, P. fait circuler une feuille que les élèves renseignent.

3) 34 échantillons de taille 34

(La feuille ci-contre est distribuée à chaque élève).

Chaque élève a effectué un **échantillon** de taille 34 et compté le nombre de Pile. On obtient une série de 34 nombres.

Pour cette série, quels sont le mode ?

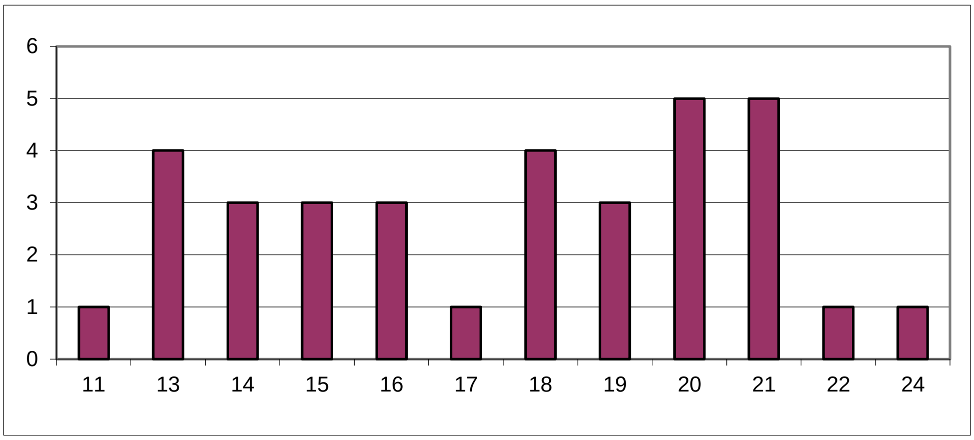
la médiane ?

la moyenne ?

l'étendue ?

	34 lancers nombre de piles
Arnaud	18
Caroline	19
Brice	22
Laura	20
Elodie	16
Lydie	14
Mathilde	20
M-Charlotte	20
Lisa	19
Lucie	21
Emmanuelle	13
Chloé	14
Anne	15
Lirone	21
Cynthia	20
Elodie	16
Manon	17
Bertrand	15
David	13
Lise	21
Anaïs	18
Nicolas	19
Eve	18
Sylvain	24
Loïs	14
Elodie	13
Marine	20
Céline	18
Céline	11
Franck	13
Fanely	16
Olivier	21
Laure	15
Angélique	21

Dépouillement de la série														
Valeurs prises														
Effectifs														



Remarques ? **Quand une expérience dépend du hasard, les échantillons n'ont pas tous la même composition, il y a "fluctuation d'échantillonnage".**

II MISE EN EVIDENCE D'UN PHENOMENE

1) Echantillons de taille 68

a) P. Pouvez-vous, à partir des résultats déjà obtenus, produire des échantillons de taille 68 ?

-

-

On peut cumuler des échantillons obtenus de manière indépendante pour obtenir un échantillon de taille plus grande.
Si on duplique un échantillon, il n'est plus produit au hasard.

b) Les élèves se groupent par 2 et forment des échantillons de taille 68. Puis chaque élève répond à la question suivante :
Y a-t-il proportionnellement plus de pile dans mon échantillon que dans celui de mon groupe ?

Pour comparer des échantillons de taille différentes, il est utile de passer aux fréquences.

2) Echantillons de plus grande taille (La feuille suivante est distribuée à chaque élève)

Fréquence de Pile dans 17 échantillons					
	A de taille 68	B de taille 340 68×5	C de taille 1020 340×3	D de taille 1360 340+1020	E de taille 2000
1	**0,49**				0,49
2	**0,47**		0,50	0,49	0,50
3	**0,63**		0,48	0,49	0,50
4	**0,56**	0,51	0,48	0,51	0,52
5	**0,51**	0,49	0,51	0,48	0,49
6	**0,47**	0,66	0,49	0,50	0,47
7	**0,65**	0,49	0,55	0,50	0,50
8	**0,50**	0,48	0,46	0,47	0,49
9	**0,47**	0,46	0,50	0,51	0,50
10	**0,60**	0,51	0,49	0,54	0,49
11	**0,46**	0,49	0,50	0,49	0,52
12	**0,37**	0,51	0,50	0,50	0,49
13	**0,41**	0,47	0,49	0,48	0,50
14	**0,54**	0,52	0,52	0,48	0,50
15	**0,60**	0,54	0,51	0,48	0,50
16	**0,46**	0,56	0,50	0,50	0,49
17	**0,56**	0,50	0,49	0,51	0,50

a) Colonne A : résultas obtenus par les élèves.

b) Colonne B:
Dans les cases déjà remplies, les « fréquence de Pile » ont été obtenues par le professeur par des moyens qui seront explicités plus tard.
Peut-on compléter les cases vides à l'aide des résultats de la classe?

c) Colonne C : Peut-on compléter la case vide à l'aide des résultats de la classe?

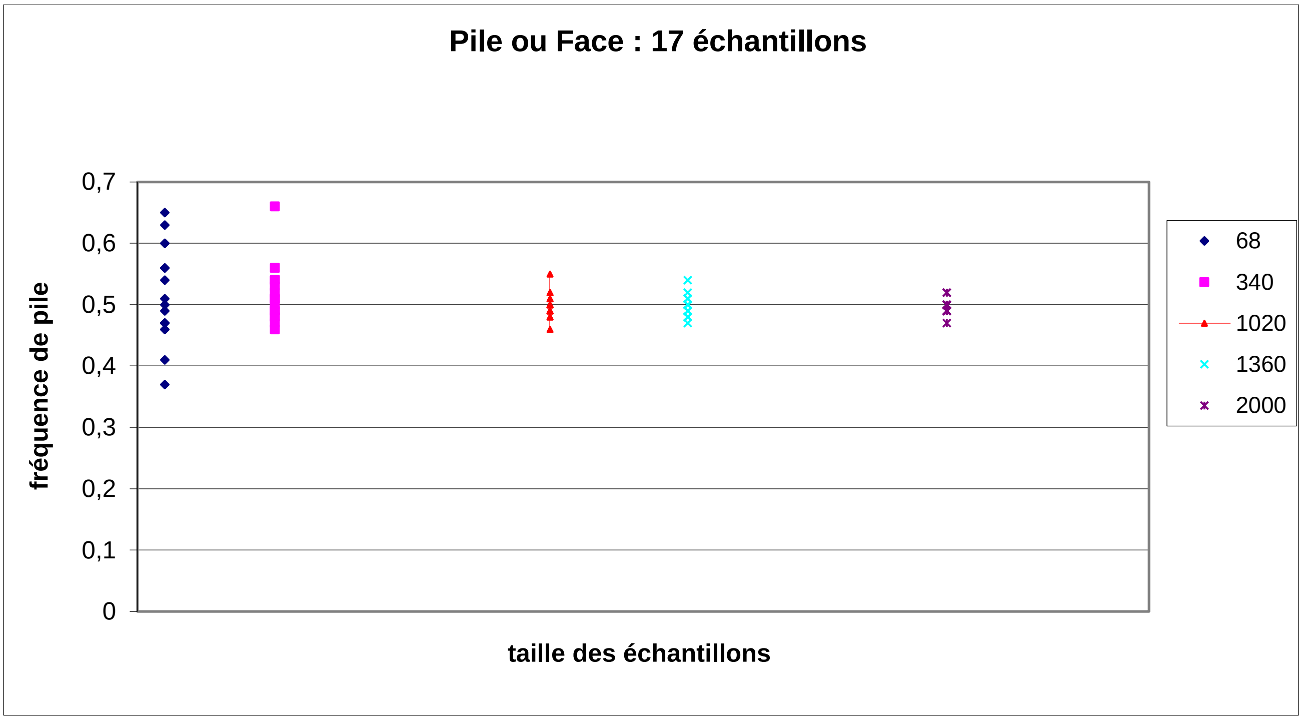
d) Colonne D : Peut-on compléter la case vide à l'aide des résultats de la classe?

3) Graphique

a) Dans un graphique, placer les points ayant pour abscisses la taille 68 et pour ordonnée les fréquences obtenues.
Qu'est-ce qui rend compte de la fluctuation d'échantillonnage sur le graphique ? Comment la mesurer ?

b) Sur le graphique précédent, placez les points ayant pour abscisse la taille 340 et ordonnée les fréquences obtenues.
Qu'est-ce qui rend compte de la fluctuation d'échantillonnage sur le graphique ? Comment la mesurer ?

c) Terminer le graphique. Pouvez-vous faire des remarques ?



Il semble que l'ampleur des fluctuations diminue quand la taille des échantillons augmente.

4) Plus précisément,

a) Taille 340 : la plupart des fréquences obtenues appartiennent à l'intervalle $[\frac{1}{2}-\frac{1}{\sqrt{340}} ; \frac{1}{2}+\frac{1}{\sqrt{340}}]$

Taille 340 : ……………………………………………………………………[0,4457… ; 0,5542 …]
Taille 340: 95% environ des fréquences obtenues appartiennent à l'intervalle [0,44 ; 0,56]

b) Taille 1020 : la plupart des fréquences obtenues appartiennent à l'intervalle $[\frac{1}{2}-\frac{1}{\sqrt{1020}} ; \frac{1}{2}+\frac{1}{\sqrt{1020}}]$

Taille 1020 : ……………………………………………………………………[0,468… ; 0,5313 …]
Taille 1020: 95% environ des fréquences obtenues appartiennent à l'intervalle [0,46 ; 0,54]

c) Taille 1360 : la plupart des fréquences obtenues appartiennent à l'intervalle $[\frac{1}{2}-\frac{1}{\sqrt{1360}} ; \frac{1}{2}+\frac{1}{\sqrt{1360}}]$

Taille 1360 : ……………………………………………………………………[0,4428… ; 0,5271 …]
Taille 1360: 95% environ des fréquences obtenues appartiennent à l'intervalle [0,44 ; 0,53]

d) etc….

5) Propriété fondamentale

Pour un phénomène aléatoire dont on connaît la fréquence théorique p, on forme des échantillons de taille n. Dans environ au moins 95 % des cas, l'intervalle $[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}]$ contient la fréquence observée dans l'échantillon (ceci est d'autant mieux vérifié que la taille n des échantillons et le nombre de tirages sont plus élevés).

6) Application à la question de la parité des sexes

Si l'on considère qu'il y a parité des sexes quand chaque individu a autant de chances d'être un garçon ou une fille, au "niveau de confiance 0,95", on ne rejettera pas l'idée qu'il y a parité dans la classe si la fréquence de filles observée appartient à la fourchette $[\frac{1}{2}-\frac{1}{\sqrt{34}} ; \frac{1}{2}+\frac{1}{\sqrt{34}}]$.

~ ANNEXE IV ~

SIMULATION
ET
FLUCTUATION D' ECHANTILLONNAGE

Vaut-il mieux être le joueur A ou le joueur B ?

1) Un jeu

On utilise un dé à 6 faces, équilibré. On le lance.
- Si le dé tombe sur 1 ou 2 ou 3 ou 4 ou 5, le joueur A avance d'une case. Pour gagner une partie, il faut qu'il ait avancé de 5 cases.
- Si le dé tombe sur 6, le joueur B gagne directement la partie.

Préfères-tu être le joueur A ou le joueur B ?

Exemple :

1	4	6	5	4	4	2	1	3	3	6	2
1	6	4	3	6	4	3	3	2	5		

2) Un outil pour simuler le hasard : la touche random de la calculatrice.

Taper plusieurs fois sur la touche random

.1216853324	0,153
.4930062012	0,25
.534806747	0,287
.6448564206	0,533

On admet que :

- On ne peut pas prévoir les chiffres qui vont apparaître : <u>imprédictibilité</u>
- La sortie d'un chiffre ne dépend pas des chiffres qui sont déjà sortis : <u>indépendance</u> (… et répétition des chiffres possible)
- Sur un grand nombre de tirages, on observe une relative <u>équirépartition</u> des chiffres

3) Comment simuler le lancer d'un dé ?

random

. 5 3 4 8 0 6 7 4 7
. 6 4 4 8 5 6 4 2 0 6

Chiffres affichés										
Dé										
Chiffres affichés										
Dé										

4) Exercice maison : Chaque élève doit simuler 10 parties du jeu et déterminer, dans son échantillon, la fréquence de gain de A et la fréquence de gain de B.

5) Exploitation des résultats

échantillon de taille 10	fréq. gain A	fréq. gain B
Chloé	0,3	0,7
Franck	0,3	0,7
Anne	0,6	0,4
Fanely	0,4	0,6
Lirone	0,4	0,6
Cynthia	0,4	0,6
Olivier	0,5	0,5
Laure	0,4	0,6
Elodie	0,6	0,4
Manon	0,1	0,9
Angélique	0,4	0,6
Bertrand	0,3	0,7
Celine	0,2	0,8
Marine	0	1
Lucie	0,5	0,5
M-Charlotte	0,2	0,8
Loïs	0,3	0,7
Mathilde	0,3	0,7
Sylvain	0,4	0,6
David	0,3	0,7
Caroline	0,5	0,5
Lise	0,4	0,6
Brice	0,4	0,6
Anaïs	0,2	0,8
Laura	0,2	0,8
Nicolas	0,2	0,8
Eve	0,4	0,6
Celine	0,3	0,7
Lydie	0,5	0,5
Emmanuelle	0,2	0,8
Elodie	0,4	0,6
Elodie G	0,2	0,8
Lisa	0,5	0,5
Arnaud	0,4	0,6
échantillon de taille …		

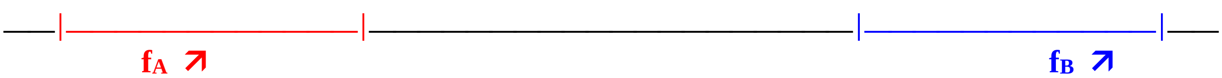
a) Comment P a-t-il calculé la fréquence de gain de A à partir des valeurs de la classe?
b) Pourquoi n'a-t-il pas calculé la fréquence de gain de B?

Propriété fondamentale et définition

On considère un phénomène aléatoire dont on ignore la fréquence p et pour lequel, dans un échantillon de taille n (n > 30), on a observé une fréquence f (0,3 < f < 0,7). Quelle que soit la valeur de p, dans au moins 95% des cas environ, l'intervalle $[\, f - \frac{1}{\sqrt{n}} \,;\, f + \frac{1}{\sqrt{n}} \,]$ contient p inconnue.
Cet intervalle est appelé intervalle de confiance*, au niveau de confiance 0,95.
*** ou fourchette de sondage**

c) A l'aide de l'échantillon obtenu par la classe, déterminer la fourchette au niveau de confiance 0,95 pour fréquence de A. Sur une droite graduée, construire l'intervalle.
d) déterminer la fourchette au niveau de confiance 0,95 pour fréquence de B. Sur la même droite graduée, construire l'intervalle.

e) Peux-tu décider si tu préfères être le joueur A ou le joueur B ?



6) Prolongement

Lors d’une élection présidentielle, le candidat A et le candidat B sont en concurrence.
1) Une semaine avant l’élection, un journaliste annonce à la télévision : **"Un sondage donne 51% d’intentions de vote pour A et 49% pour B"**. Ton frère te dit : " Je pense que A va être élu !" Le sondage a été réalisé sur un échantillon de taille 1000.

a) Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat A et trace en rouge cette fourchette sur un axe. Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat B et trace en vert cette fourchette sur le même axe.
b) Que réponds - tu à ton frère ?
2) Le soir de l'élection à 20 h, un journaliste annonce à la télévision : **"Un sondage donne 51% de voix pour A et 49% pour B"**. Ton frère te dit : " Je pense que A va être élu !" Le sondage a été réalisé sur un échantillon de taille 100000
a) Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat A et trace en rouge cette fourchette sur un axe. Calcule une fourchette de sondage au niveau de confiance 0,95 pour le candidat B et trace en vert cette fourchette sur le même axe.
c) Que réponds - tu à ton frère ?

~ ANNEXE V ~

SIMULATION
ET
FLUCTUATION D' ECHANTILLONNAGE

Etude d'une distribution de fréquences

1) Distribution des fréquences dans deux échantillons de taille 10

Constituer un échantillon de taille 10 d'un lancer de dé et placer les résultats dans le tableau ci-dessous.

Distribution des fréquences pour 10 lancers						
	1	**2**	**3**	**4**	**5**	**6**
échantillon Xavier						
échantillon Yves	**0**	**0,2**	**0,1**	**0,2**	**0,2**	**0,3**

Construire un diagramme en bâton comparant l'échantillon X et l'échantillon Y.
Constatez-vous une fluctuation d'échantillonnage ?
Comment mesurer l'ampleur des fluctuations ?
Calculer la moyenne des points obtenus pour l'échantillon X et pour l'échantillon Y.
Placer ces deux valeurs en abscisse.

2) Distribution des fréquences dans deux échantillons de taille 100

Résultats arrondis à 10^{-2} obtenus par simulation :

Distribution des fréquences pour 100 lancers						
	1	**2**	**3**	**4**	**5**	**6**
échantillon X	**0,11**	**0,13**	**0,16**	**0,15**	**0,21**	**0,24**
échantillon Y	**0,17**	**0,13**	**0,15**	**0,15**	**0,25**	**0,15**

Construire un diagramme en bâton comparant l'échantillon X et l'échantillon Y.
Constatez-vous une fluctuation d'échantillonnage ?
Comment mesurer l'ampleur des fluctuations ?
Calculer la moyenne des points obtenus pour l'échantillon X et pour l'échantillon Y.
Placer ces deux valeurs en abscisse.

3) Distribution des fréquences dans deux échantillons de taille 1000

Résultats arrondis à 10^{-3} obtenus par simulation :

Distribution des fréquences pour 1000 lancers						
	1	**2**	**3**	**4**	**5**	**6**
échantillon X	**0,152**	**0,170**	**0,158**	**0,153**	**0,168**	**0,199**
échantillon Y	**0,164**	**0,167**	**0,160**	**0,172**	**0,148**	**0,189**

Construire un diagramme en bâton comparant l'échantillon X et l'échantillon Y.
Constatez-vous une fluctuation d'échantillonnage ?
Comment mesurer l'ampleur des fluctuations ?
Calculer la moyenne des points obtenus pour l'échantillon X et pour l'échantillon Y.
Placer ces deux valeurs en abscisse.

4) Quelles remarques peut-on faire en comparant les trois études précédentes?

5) Blaise te propose un jeu : tu mises 3 euros, tu lances un dé, si le chiffre apparu sur le dé est x, tu remportes x euros. Es-tu prêt à jouer 1 fois, 10 fois, 100 fois, 1000 fois ? Pourquoi ?

Distribution de fréquences et loi de probabilité

Extrait du document d'accompagnement du programme de 1°S, ES

<table>
<tr><th colspan="5">Distribution de fréquences
sur E ={ $x_1, x_2,\dots x_r$}</th><th colspan="5">Loi de probabilité
sur E = { $x_1, x_2, \dots x_r$}</th></tr>
<tr><td>x_i</td><td>x_1</td><td>x_2</td><td>…</td><td>x_r</td><td>x_i</td><td>x_1</td><td>x_2</td><td>…</td><td>x_r</td></tr>
<tr><td>f_i</td><td>f_1</td><td>f_2</td><td>…</td><td>f_r</td><td>p_i</td><td>p_1</td><td>p_2</td><td>…</td><td>p_r</td></tr>
<tr><td colspan="5">$(f_1,\dots,f_i)$
$f_i \geq 0; \sum f_i = 1$
A$\subset$ E, fréquence de A : $f(A) = \sum_{x_i \in A} f_i$
Événement complémentaire : $f(\bar{A}) = 1 - f(A)$
Événements A et B disjoints :
$f(A \cup B) = f(A \text{ ou } B) = f(A) + f(B)$
Cas numérique :
Moyenne empirique : $\bar{x} = \sum f_i x_i$
Variance empirique : $s^2 = \sum f_i (x_i - \bar{x})^2$
Écart type empirique : $s = \sqrt{\sum f_i (x_i - \bar{x})^2}$</td><td colspan="5">$(p_1,\dots,p_r)$
$p_i \geq 0; \sum p_i = 1$
A$\subset$ E, probabilité de A : $P(A) = \sum_{x_i \in A} p_i$
Événement complémentaire : $P(\bar{A}) = 1 - P(A)$
Événements A et B disjoints :
$P(A \cup B) = P(A \text{ ou } B) = P(A) + P(B)$
Cas numérique :
Espérance d'une loi P : $\mu = \sum p_i x_i$
Variance d'une loi P : $\sigma^2 = \sum p_i (x_i - \mu)^2$
Écart type d'une loi P : $\sigma = \sqrt{\sum p_i (x_i - \mu)^2}$</td></tr>
</table>