Fluctuation d'échantillonnage

Dominique Grihon

La notion de fluctuation d'échantillonnage est désormais au programme sur les trois années du lycée. Cet article permet à tous de mieux cerner les enjeux et les articulations d'un niveau à l'autre.

Voici ce que dit le programme de Seconde :

Échantillonnage Notion d'échantillon. Intervalle de fluctuation d'une fréquence au seuil de 95%*. Réalisation d'une simulation.	• Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. • Exploiter et faire une analyse critique d'un résultat d'échantillonnage.	Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience. À l'occasion de la mise en place d'une simulation, on peut : • utiliser les fonctions logiques d'un tableur ou d'une calculatrice, • mettre en place des instructions conditionnelles dans un algorithme. L'objectif est d'amener les élèves à un questionnement lors des activités suivantes : • l'estimation d'une proportion inconnue à partir d'un échantillon ; • la prise de décision à partir d'un échantillon.

* L'intervalle de fluctuation au seuil de 95 %, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation. Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n > 25$ et des proportions p du caractère comprises entre 0, 2 et 0, 8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}}\ ;p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais elle n'est pas exigible.

Dominique Grihon enseigne au lycée Sud Médoc La Boétie à Le Taillan Médoc (33).

Que peut représenter le terme de « fluctuation » pour les élèves ? C'est la question à laquelle j'essaie de répondre en menant les activités que je vais décrire ici.

Activité 1

On fait jouer effectivement les élèves à pile ou face. Ils lancent leur pièce 20 fois et on recense le nombre de piles pour chaque élève puis la fréquence de piles. Les élèves constatent des résultats très variés d'un élève à l'autre. On introduit la « fluctuation » de la fréquence de piles.

Activité 2

Je propose maintenant de simuler le lancer un certain nombre de fois d'une pièce de monnaie. Pour cela on utilise le nombre aléatoire de la calculatrice. Celui-ci a 10 chiffres après la virgule : chaque chiffre pair représente « pile » et chaque impair « face ». À chaque affichage d'un nombre aléatoire correspond donc la simulation de 10 lancers d'une pièce de monnaie.

Je commence par leur faire refaire les calculs de fréquences avec 20 lancers simulés pour constater que la fluctuation des

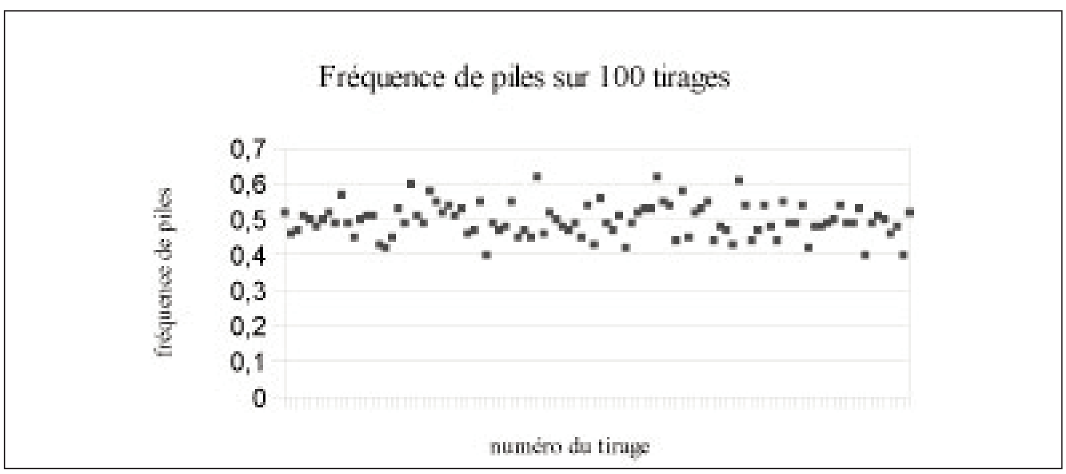
fréquences est toujours là. Puis avant de passer à 100 lancers, je leur demande s'ils ont une idée des résultats qu'ils vont obtenir. Pour beaucoup, les résultats seront aussi variés qu'avec 20 lancers. Pour certains, les résultats vont être plus souvent proches de 0,5.
Les élèves réalisent alors l'expérience de 100 lancers. Ils constatent que les fréquences de piles obtenues sont concentrées entre 0,4 et 0,6, ce qui n'était pas le cas avec 20 lancers.

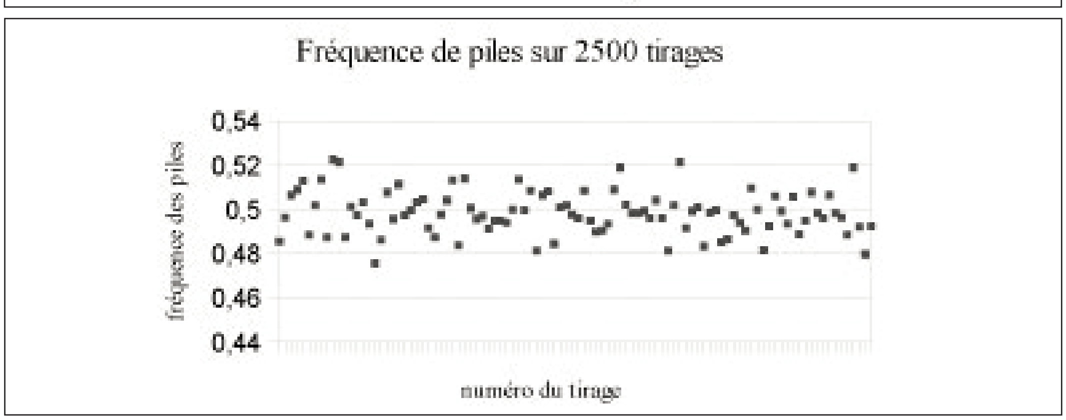
Activité 3

Je présente une simulation faite avec le tableur : 100 fois 100 lancers d'une pièce. L'appui sur la touche F9 permet de refaire autant de simulations qu'on veut. On constate alors qu'on a toujours environ 95 résultats entre 0,4 et 0,6.
Je définis alors l'intervalle de fluctuation à 95 % en expliquant que finalement nous avons remarqué que la fluctuation est « encadrée ». Vient alors la deuxième feuille de tableur avec cette fois-ci 100 fois 2500 lancers. On constate bien un rétrécissement de l'intervalle de fluctuation comme prévu par le calcul : environ 95 % des résultats sont maintenant dans l'intervalle [0,48 ; 0,52].

On trouvera en bas de cette page les graphiques pour 100 lancers puis pour 2 500 lancers.

Viennent ensuite les exercices qui permettent d'acquérir plusieurs compétences :
- Savoir repérer dans un texte les éléments essentiels : le caractère étudié dans la population, la proportion p de ce caractère dans la population, la taille n de l'échantillon ou encore la fréquence f du caractère dans l'échantillon.
- Savoir calculer l'intervalle de fluctuation associé.





- Savoir interpréter le résultat sur l'appartenance ou non de f à l'intervalle de fluctuation. Plus précisément, si f n'appartient pas à l'intervalle de fluctuation, on peut décider que l'échantillon n'est pas conforme à la population pour ce caractère avec un risque de se tromper limité à 5 %. Ce risque est connu, il correspond au fait que dans 5 % des cas, on peut avoir une fréquence en dehors de l'intervalle alors que l'échantillon est bien issu aléatoirement de la population donnée. Par contre, si f appartient à l'intervalle de fluctuation, on ne connaît pas le risque que l'on prendrait à dire que l'échantillon est conforme (voir l'annexe).
- Savoir déterminer une taille d'échantillon permettant d'avoir une amplitude donnée de l'intervalle de fluctuation.

La définition de l'intervalle de fluctuation de seconde ne convient que pour le seuil de 95 %.

En Première, les élèves apprendront à déterminer un intervalle de fluctuation exact à partir de la loi binomiale, car la composition d'un échantillon suit une loi binomiale.

En Terminale, les élèves apprendront à déterminer un intervalle de fluctuation « asymptotique » correspondant au fait que, lorsque n augmente, la loi binomiale centrée réduite « tend » vers la loi normale. Ils pourront alors comprendre d'où vient la formule donnée en Seconde qui est une approximation de l'intervalle de Terminale.

Concernant l'intervalle de confiance, il me semble un peu prématuré de définir la notion en Seconde, sauf à travers un exemple sur les sondages qui permet à tous les élèves de Seconde qui ne feront plus de maths ensuite d'avoir entendu parler de « fourchettes de sondage » et de savoir qu'un résultat brut de sondage politique n'a aucun sens sans la donnée de la fourchette dans laquelle il se situe.

Annexe

Exemple de situation où la décision n'est pas possible :

Un horticulteur commande des sacs de bulbes de crocus de deux couleurs : jaune et violet.

Le fabriquant garantit une composition de 50 % de chaque variété. L'horticulteur plante 625 bulbes et constate que 290 fleurs sont jaunes et que tous les bulbes ont fleuri.

L'intervalle de fluctuation correspondant à cette situation est [0,46 ; 0,54]. La fréquence observé est $f = 0{,}464$; l'horticulteur n'a pas de raison de s'inquiéter.

Cependant, si la proportion réelle de bulbes jaune avait été de 46 %, la fréquence observée serait encore dans l'intervalle de fluctuation, ici : [0,42 ; 0,50]. Cela explique qu'on ne connaît pas le risque qu'on prendrait si on concluait que la proportion de 50 % de chaque couleur était respectée.

NDLR : en 2002, ont été publiées « *11 fiches de statistiques* » dans le document d'accompagnement des programmes pour le niveau Seconde. On les trouve à l'aide d'un moteur de recherche. Ces fiches sont riches d'expériences simples et particulièrement bien décrites pour traiter le thème « Statistiques ».