Chapitre XIV
Échantillonnage (1s)

Table des matières

Présentation :

L'échantillonnage est un domaine assez récent des mathématiques, à la croisée des statistiques et des probabilités. Ce domaine intervient dans les situations où l'on veut étudier certaines propriétés d'une population de trop grands effectifs pour être observé de façon exhaustive.
On observe alors un échantillon de cette population, échantillon à partir duquel on souhaite obtenir des informations sur la population totale avec un certain degré de précision.
Un exemple omniprésent dans les médias est la réalisation de sondage d'opinion.

L'intérêt est de minimiser le nombre d'individus à étudier au sein d'une population afin d'en tirer les informations les plus fiables possibles.
Cela pourra être, par exemple, dans une entreprise qui fabrique des pièces automobiles, savoir le nombre de pièces à tester afin de savoir si notre production journalière, hebdomadaire ou mensuelle est de bonne qualité, sans avoir à tester l'ensemble des pièces.

I. Échantillonnage

Définitions :

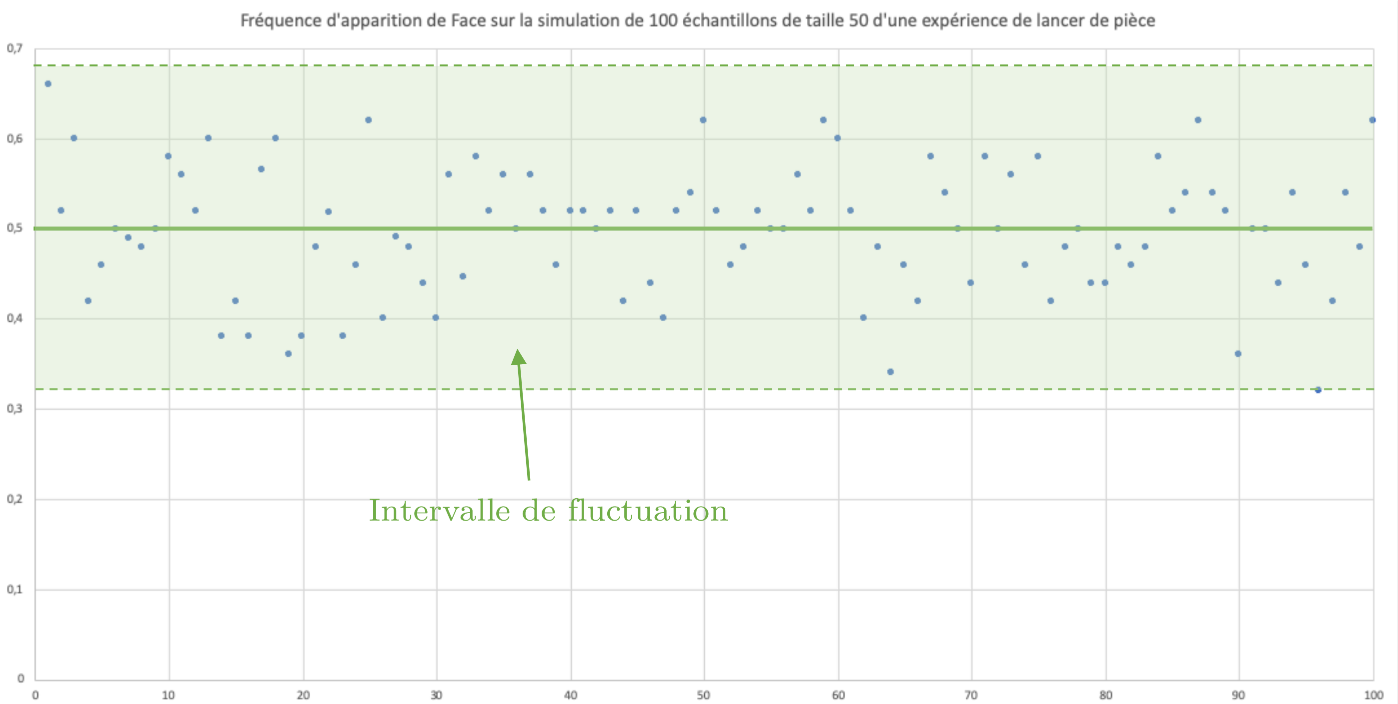
- Réaliser un échantillon de taille n d'une expérience aléatoire, c'est répéter cette même expérience n fois à l'identique, dans les mêmes conditions, et noter la liste des n résultats obtenus.
- La fréquence de réalisation de l'événement à et le nombre de fois où l'événement a et réaliser diviser par la taille de l'échantillon.

Exemple :
On tire 5 fois une pièce de monnaie. À chaque lancer, les 2 issues possibles sont Pile (P) et Face (F).
Le résultat « PFPPF » constitue un échantillon de taille 5 pour cette expérience aléatoire à 2 issues.
La fréquence de réalisation de P est : $\frac{3}{5} = 0{,}6$.

II. Fluctuation d'échantillonnage – vers la loi des grands nombres

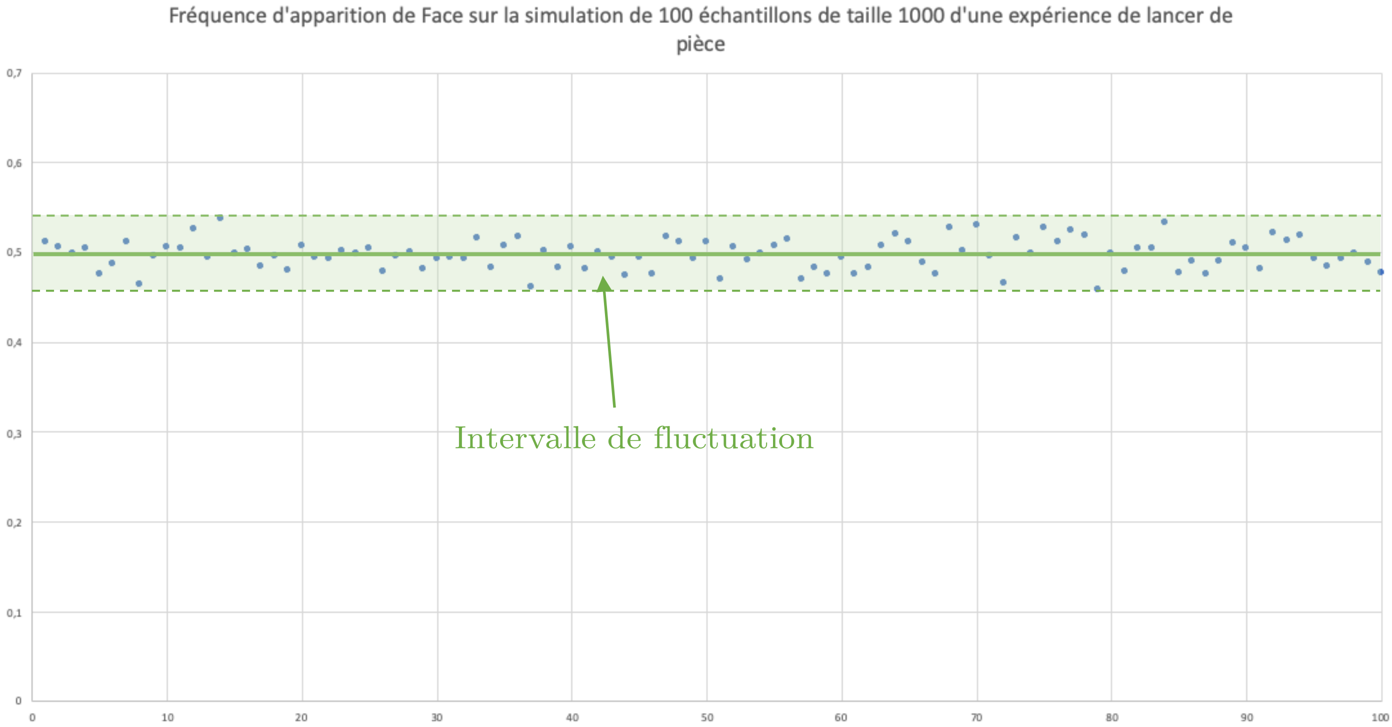
J'ai recueilli les fréquences d'apparition de Face dans des lancers de pièces.

Voici les résultats pour 100 simulations de 50 lancers (taille des échantillons : $n=50$) :



On observe qu'il y a une fluctuation importante et que nos résultats se trouvent dans l'intervalle de fluctuation suivant : $[0{,}5 - 0{,}18; 0{,}5 + 0{,}18]$.

Voici les résultats pour 100 simulations de 1000 lancers (taille des échantillons : *n=1000*) :



On observe cette fois-ci, qu'avec des échantillons plus grands, la fluctuation est moindre. Notre intervalle de fluctuation est ici : $[0{,}5 - 0{,}041; 0{,}5 + 0{,}0{,}41]$.

On peut observer :

- Nos valeurs fluctuent dans les 2 cas autour de la probabilité *p=0,5*.
- Lorsque la taille n de l'échantillon augmentent, il semble que la fluctuation diminue : les fréquences deviennent proches de p, notre intervalle de fluctuation est plus petit.

Lois des grands nombres :
Lorsque la taille n de l'échantillon devient grande, sauf exception, les fréquences f observées sont moins dispersées et demeurent proches de la probabilités p.

La plupart du temps, l'écart entre p et f est inférieur à $\frac{1}{\sqrt{n}}$. Cela signifie que notre intervalle de fluctuation à 0,95 est : $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ (le 0,95 signifie que nos fréquences ont 95% de chance de tomber dans notre intervalle).

Exercice corrigé :

On admet que, lorsque la taille de l'échantillon est supérieur à 2500, la fréquence f de réalisation d'un événement est une valeur approchée de la probabilité p de cette événement à environ $\frac{2}{100}$ près.
Autrement dit, $f - 0{,}02 \leq p \leq f + 0{,}02$.
Une urne contient 100 boules de couleur noir ou orange. On tire au hasard une boule dans l'urne, on note sa couleur puis on la remet dans l'urne.
On note f la fréquence d'obtention d'une boule noire.
Sur 2500 tirages, on a obtenu 1775 boules noires.
1.a. Quelle est la taille de l'échantillon étudié.
b. Que vaut f dans cet exemple ?
2. Encadrer p en utilisant le résultat admis. Que représente p ?
3. Que peut-on en déduire quant au nombre de boules noires dans l'urne ?

Correction :
1.a. Il y a eu 2500 tirages, la taille de notre échantillon est $n = 2500$.
b. f *est la fréquence d'obtention d'une boule noire.* $f = \frac{1775}{2500} = 0{,}71$.

2. Avec le résultat admis : $f - 0{,}02 \leq p \leq f + 0{,}02$.
On a : $0{,}71 - 0{,}02 \leq p \leq 0{,}71 + 0{,}02$ *soit un encadrement de p :* $0{,}69 \leq p \leq 0{,}73$.

3. En multipliant notre encadrement par les 2500 tirages on a :

$0{,}69 \times 2500 \leq p \times 2500 \leq 0{,}73 \times 2500$ *soit* $1725 \leq \textit{nombre de boules noires} \leq 1825$.
On peut estimer que le nombre de boules noires est comprises entre 1725 et 1825.

III. Simuler le hasard avec une calculatrice ou un ordinateur

Sans rentrer dans les détails, lorsque nous allons utiliser une calculatrice ou un logiciel pour générer des nombres, ceux-ci seront générés de façon pseudo-aléatoire.

Il y a plusieurs moyens de le faire.

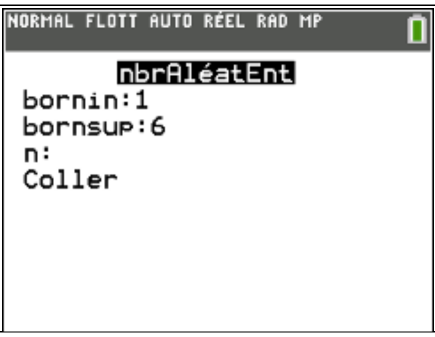
- En utilisant un décalage par rapport à l'horloge de l'ordinateur.
- En utilisant des algorithmes plus ou moins sophistiqués qui les généreront. Par exemple, Excel va utiliser l'algorithme de Mersenne pour générer des nombres pseudo-aléatoires.

Nous ne rentrerons pas plus dans les détails, mais il est important de ne pas perdre de vu que nous travaillons en pseudo-aléatoire et non en aléatoire complet comme ça serait le cas avec une vrai simulation d'épreuve.

Pour générer un nombre entre 1 et 6 (par exemple, simuler un lancer de dé) :

Avec la calculatrice :

TI-83 :
Touche Maths/PROB/choix : 5 :nbreAléatEnt(

		bornin : borne inférieur de notre intervalle de choix. bornsup : borne supérieur de notre intervalle de choix. n : nombre de lancers simulé Coller : appuyé sur Coller une fois les valeurs entrées.

Casio Graph 90 :
Dans la fenêtre Exe-Mat(fenêtre de calcul classique) :
Touche OPTN/F6(⊳)/F3 (PROB)/F4 (RAND)/F2 Int
RanInt# (doit s'afficher).

La syntaxe de RanInt# est : RanInt#(borne inférieur de notre intervalle de choix,borne supérieur de notre intervalle de choix,nombre de lancers simulé)

Remarque : le nombre de lancers simulés n'est pas obligatoire. Si vous ne mettez pas de valeurs, vous obtiendrez un lancer.
Attention à bien utiliser la virgule (,) pour séparer vos valeurs.

Pour simuler un lancer d'un dé à 6 faces on a donc : RanInt#(1,6)

NUMWORKS :
Dans la fenêtre calcul :
Choisissez la toolbox :


Descendre avec les flèches et choisir : Aléatoire et approximation.
Flèche droite (▷)pour rentrer dans ce menu.
Choisir randint(a,b).
Remplir les vides pour indiquer vos bornes de choix.

Pour simuler un lancer d'un dé à 6 faces on a donc : randint(1,6)

Tableur :
ALEA.ENTRE.BORNES(borne inférieur de notre intervalle de choix ;borne supérieur de notre intervalle de choix)

Pour simuler un lancer d'un dé à 6 faces on a donc : ALEA.ENTRE.BORNES(1;6)

Entrainez-vous à simuler des lancers de dés avec votre calculatrice, et avec un tableur si vous en avez un.