Échantillonnage

I. Modélisation

1.1 Définitions :

- Lorsqu'on étudie une partie de la population, on dit qu'on étudie un ***échantillon***.
- Le nombre d'individus composant l'échantillon est la ***taille de l'échantillon***.
  Un échantillon de taille n est obtenu par n répétitions indépendantes de la même expérience.

— Une **expérience aléatoire** est une expérience que l'on peut reproduire dans les mêmes conditions et dont on connaît à priori toutes les résultats possibles, sans pouvoir dire avec certitude le résultat qui se produira.
— **Simuler** une expérience, c'est la remplacer par une autre, plus simple à réaliser, et qui permettra d'obtenir des résultats similaires.

Remarque : Pour cela, on utilise généralement un générateur de nombres au hasard. Il y en a de plusieurs types :

— les tables de nombres au hasard (très peu utilisés de nos jours)
— la commande ALEA.ENTRE.BORNES d'un tableur
— la commande RANDOM des calculatrices

1.2 Utilisation d'un tableur :

Pour simuler un tirage au sort dans un ensemble à p éléments à l'aide d'un tableur, on peut utiliser la méthode suivante :
— On attribue à chaque élément un nombre entre 1 et p.
— On simule le tirage aléatoire en utilisant la commande =ALEA.ENTRE.BORNES(1,p)
— Pour calculer la fréquence d'une issue pour N tirages au sort, on utilise la commande =NB.SI(plage;critère)/N
où la plage est l'ensemble des cellules où ont été effectués les simulations.

1.3 Utilisation de la calculatrice :

Toute calculatrice scientifique possède une fonction RANDOM qui génère une nombre réel au hasard entre 0 et 1 (avec 1 exclu). Combinée avec la fonction Ent qui donne la partie entière d'un nombre, elle permet de simuler de nombreuses expériences.

	Casio	T.I.
Nombre au hasard (`RANDOM`)	Dans le menu `OPTN`, sélectionner le menu `PROB`, puis la commande `RAN#`. Appuyer sur la touche `EXE` pour exécuter la commande	Dans le menu `MATH`, sélectionner le menu `PRB`, puis la commande `NbreAleat` (ou `Rand`). Appuyer sur la touche `EXE` pour exécuter la commande
Partie entière (`Ent`)	Dans le menu `OPTN`, sélectionner le menu `NUM`, puis la commande `Int`.	Dans le menu `MATH`, sélectionner le menu `NUM`, puis la commande `Ent` (ou `Int`).

Exemple : Simuler, à l'aide de la fonction RANDOM, 10 lancers successifs d'un dé à 6 faces numéroté de 1 à 6.
Il s'agit d'obtenir au hasard un nombre compris entre 1 et 6. Or, la touche RANDOM permet d'obtenir un nombre x tel que $0 \leq x < 1$.

— 6×RANDOM permet d'obtenir un nombre y tel que $0 \leq y < 6$
— Ent(6×RANDOM) permet d'obtenir un nombre entier n tel que $0 \leq n \leq 5$
— Ent(6×RANDOM)+1 permet d'obtenir un nombre entier p tel que $1 \leq p \leq 6$

Il suffit donc d'entrer la fonction Ent(6×RANDOM)+1 dans la calculatrice et d'appuyer 10 fois de suite sur la touche EXE pour obtenir la simulation voulue.

Propriété 1

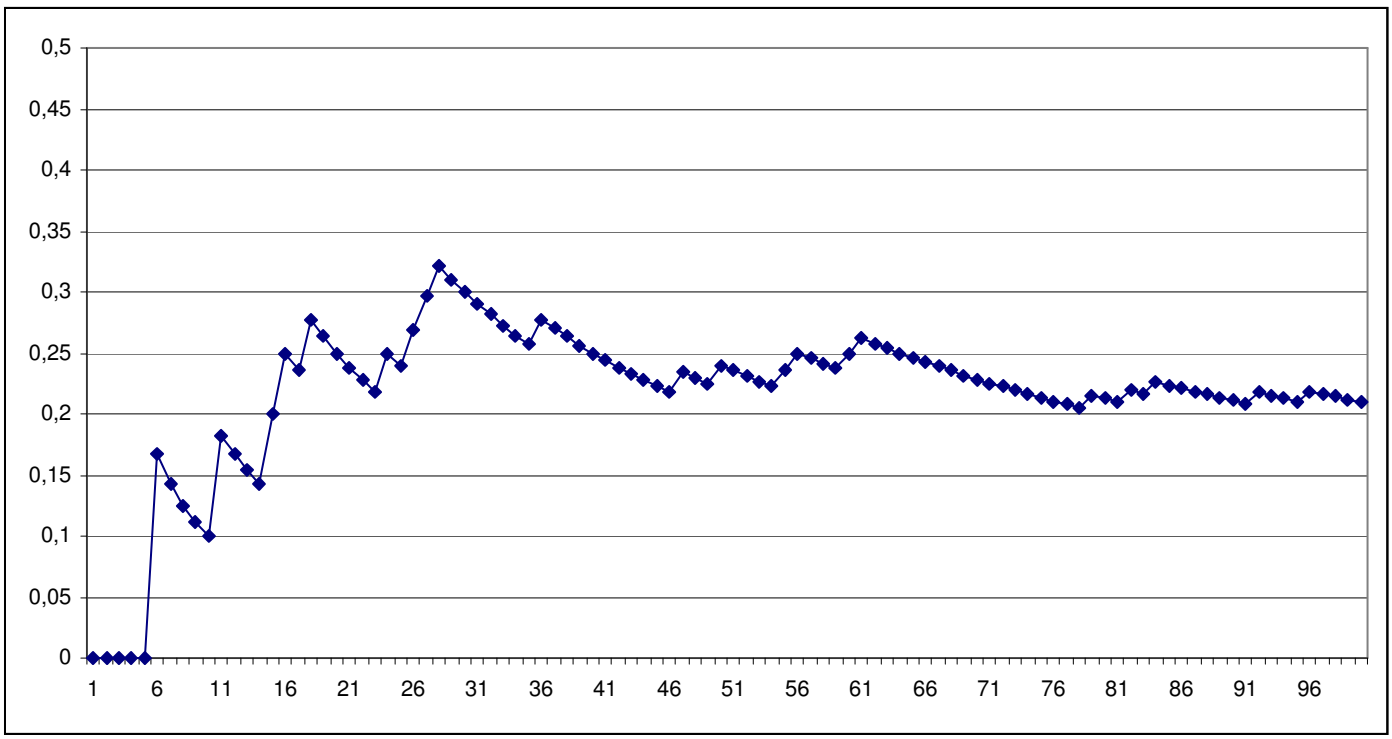
Soit p la proportion de la population vérifiant le critère étudié.
Soit $\widehat{p}$ la proportion de l'échantillon vérifiant ce même critère.
Plus la taille de l'échantillon est grande, plus $\widehat{p}$ se rapproche de p.

Exemple
En lançant deux pièces de monnaie correctement équilibrées, on peut obtenir :

- deux côtés Pile
- deux côtés Face
- un côté Pile et un côté Face

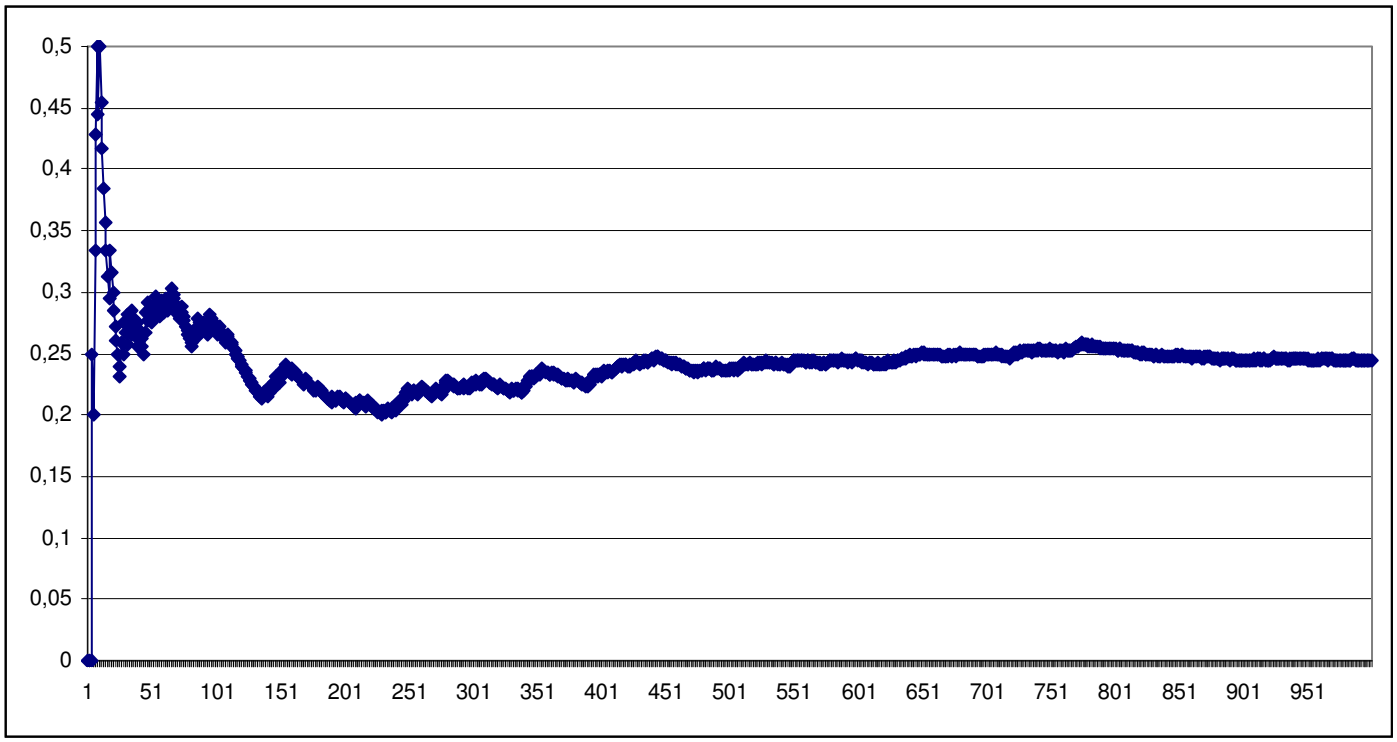
On étudie la fréquence d'obtenir deux côtés Pile.
Pour cela, on simule cette expérience à l'aide d'un tableur.
On étudiera quatre tailles d'échantillon : 100, 200, 500 et 1000.
Remarque
La valeur théorique pour cette expérience est égale à 0,25.

La taille du premier échantillon est égale à 100.



Après 100 lancers, la fréquence d'obtenir deux côtés Pile semble être légèrement supérieure à 0,2.

La taille du quatrième échantillon est égale à 1000.



Après 1000 lancers, la fréquence d'obtenir deux côtés Pile semble être proche de 0,25.

II. Intervalle de fluctuation et intervalle de confiance

Propriété 2 *(théorème de l'intervalle de fluctuation – théorème admis)*

- On considère un échantillon de taille n supérieure ou égale à 25.
- Le nombre p doit appartenir à l'intervalle [0,2 ; 0,8].

Dans ces conditions, le nombre $\widehat{p}$ appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}} \right]$ dans plus de 95 % des cas.

Exemple :

Dans un casino, sur 2 500 lancers de dé, 1 150 ont donné un nombre pair. Y a-t-il lieu de faire une enquête pour utilisation de dés truqués ?

Si les dés ne sont pas truqués, il y a autant de chances d'obtenir un nombre pair qu'un nombre impair.
On assimile la situation à un échantillon de taille $n = 2500$ avec une probabilité $p = 0,5$ d'obtenir un résultat pair.

Comme $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{2500}} = \frac{1}{50} = 0,02$, alors l'intervalle de fluctuation est : $I = [0,48 ; 0,52]$.

La fréquence observée est $f = \frac{1150}{2500} = 0,46$.

Cette fréquence n'appartient pas à l'intervalle I. On rejette donc l'hypothèse que les dés ne sont pas truqués. Une enquête s'impose.

Attention : cette décision comporte un risque d'erreur puisque dans 5 % des cas, 2 500 lancers avec des dés non truqués peuvent donner une fréquence de nombres pairs n'appartenant pas à l'intervalle I.

Propriété 3 *(théorème de l'intervalle de confiance – théorème admis)*

- On considère un échantillon de taille n supérieure ou égale à 25.
- Le nombre $\widehat{p}$ doit appartenir à l'intervalle [0,2 ; 0,8].

Dans ces conditions, le nombre p appartient à l'intervalle $\left[\widehat{p}-\frac{1}{\sqrt{n}}\,;\,\widehat{p}+\frac{1}{\sqrt{n}}\right]$ dans plus de 95 % des cas.

Exemple :

Les 10 000 employés d'une entreprise doivent être consultés sur la nouvelle couleur du sigle de l'entreprise : vert ou rouge. Inquiète du résultat, la direction, qui a fait campagne pour le rouge, a interrogé un échantillon de 100 employés sur leur choix et 54 % ont répondu rouge.
Que peut-on en conclure ?

On assimile la situation à un échantillon de taille $n = 100$, avec une probabilité p inconnue.

La fréquence observée est $\widehat{p}=0,54$. $\frac{1}{\sqrt{n}}=\frac{1}{\sqrt{100}}=\frac{1}{10}=0,1$.

On obtient donc un intervalle de confiance $I=[0,44\,;0,64]$.

Il y a donc de très fortes chances que la proportion réelle p soit comprise dans l'intervalle I.
Comme 0,5 appartient à cet intervalle, on ne peut pas savoir qui des partisans du vert ou du rouge vont l'emporter.