FLUCTUATION D'ÉCHANTILLONNAGE

1) ÉCHANTILLON ET DISTRIBUTION DE FRÉQUENCES

Définition :

Une **expérience aléatoire** est une expérience liée au hasard dont on ne peut pas prévoir l'issue.

Exemples :

- Lancer une pièce de monnaie et s'intéresser à la face visible.
- Lancer un dé et s'intéresser au numéro de la face supérieure.
- Tirer une boule dans une urne contenant des boules de différentes couleurs et s'intéresser à la couleur de la boule tirée.

Définition :

Un **échantillon de taille n** est la série statistique formée des n résultats obtenus lorsqu'on répète n fois une expérience dans les mêmes conditions.

Exemple :

On lance un dé et on note le nombre obtenu sur la face supérieure du dé.
On répète 10 fois cette expérience.
On obtient successivement : 1 , 5 , 2 , 1 , 6 , 3 , 2 , 2 , 6 , 2
Cette série de nombres constitue un échantillon de taille 10.

Remarque :

Pour obtenir un échantillon à l'aide d'un tirage celui-ci doit s'effectuer avec remise pour que la proportion du caractère considéré ne change pas.
Si la taille de l'échantillon est négligeable devant l'effectif total, on peut assimiler un tirage sans remise à un tirage avec remise.

Définition :

La distribution des fréquences associée à un échantillon est la liste des fréquences des issues de l'échantillon.

Exemple :

Dans l'exemple précédent, on peut calculer les fréquences de chacune des issues 1 , 2 , 3 , 4 , 5 et 6.
Ce tableau donne la distribution des fréquences associée à l'échantillon.

Issue	1	2	3	4	5	6
f_i						

2) FLUCTUATION D'ÉCHANTILLONNAGE

Définition :

Si on réalise plusieurs échantillons de même taille pour une même expérience, la distribution des fréquences varie.
C'est ce qu'on appelle **la fluctuation d'échantillonnage.**

Exemple :

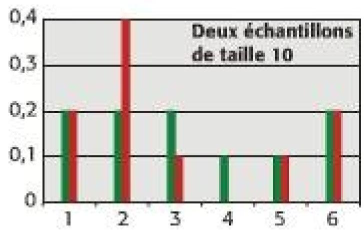
On peut obtenir un nouvel échantillon de taille 10 en relançant 10 fois le dé.
On obtient successivement : 2 , 3 , 2 , 5 , 3 , 4 , 6 , 1 , 6 , 1
On obtient une nouvelle distribution des fréquences.

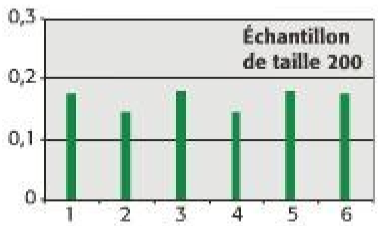
Issue	1	2	3	4	5	6
f_i						

La distribution obtenue est différente.
On dit que les **fréquences fluctuent**.

Remarque :

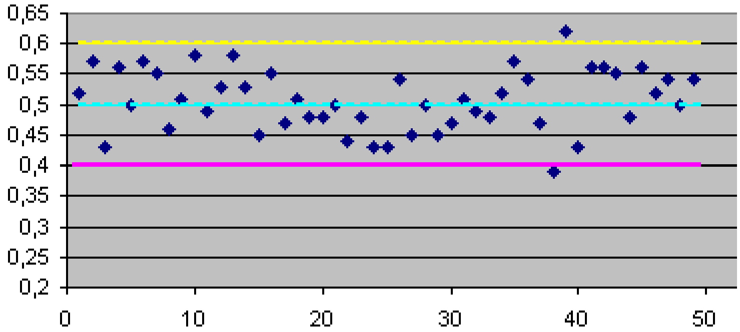
Quand la taille de l'échantillon augmente, l'ampleur des fluctuations des distributions de fréquences calculées sur ces échantillons de taille n diminue et les fréquences ont tendance à se stabiliser.





3) INTERVALLE DE FLUCTUATION

On lance un dé et on regarde si le nombre obtenu sur la face supérieure du dé est pair ou impair.
On simule 50 échantillons de taille 100 de cette expérience et on s'intéresse à la fréquence d'apparition d'un nombre pair.
On se rend compte que les valeurs des fréquences obtenues sont assez proches de la valeur



On constate qu'il y a au moins 95% des échantillons pour lesquels la fréquence est comprise dans l'intervalle

Propriété : (admise)

On considère une population dont la fréquence d'apparition d'un caractère donné est p et un échantillon de taille n , dont la fréquence d'apparition du caractère précédent est f .

Lorsque n est assez grand et p n'est ni proche de 0, ni de 1, il y a environ 95% des échantillons de taille n issus de cette population qui sont tels que la fréquence f du caractère appartient à un intervalle centré en p de la forme :

$$\left]p-\frac{1}{\sqrt{n}}\,;\,p+\frac{1}{\sqrt{n}}\right[$$

Cet intervalle est appelé **intervalle de fluctuation** de la fréquence f au seuil 95% des échantillons.

Remarque :
Plus la taille de l'échantillon est grande, plus les fréquences f du caractère des échantillons sont proches de la proportion p du caractère de la population ,dans 95% des cas.

Propriété :

La largeur de l'intervalle de fluctuation au seuil de 95 % pour un échantillon de taille n est de l'ordre de grandeur de

Applications :

- **Prise de décision à partir d'un échantillon**

A partir d'un échantillon pour lequel la fréquence relative à un caractère donné est f , on regarde si cette fréquence est dans l'intervalle de fluctuation à 95% . Si ce n'est pas le cas, on considère que l'observation n'est pas compatible avec le modèle.

- **Estimer la valeur d'une proportion inconnue**

On a : $p-\frac{1}{\sqrt{n}}\leqslant f\leqslant p+\frac{1}{\sqrt{n}}$

L'équivalence précédente permet d'estimer, au seuil de confiance 95%, la valeur d'une proportion inconnue p dans la population à partir d'une fréquence observée f .

La précision de l'estimation vaut $\frac{1}{\sqrt{n}}$. Elle est d'autant meilleure que n est grand.

$\left]f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right[$ est appelé **intervalle de confiance** au seuil de 95%.