# Chapitre 6

# Fluctuation d'échantillonnage

## I. Échantillon et fluctuation

Lorsqu'on étudie un caractère d'une population, la connaissance de la population n'est, en général, pas envisageable.
On doit se contenter de la connaissance d'un **échantillon** de la population.
Pour prendre des décisions fondées sur des théories mathématiques, il est indispensable que l'échantillon soit prélevé au **hasard.**

### 1) Échantillon

**Définition :**

Une **expérience aléatoire** est une expérience que l'on peut reproduire dans les mêmes conditions et dont on connaît à priori tous les résultats (ou **issues**) possibles, sans pouvoir dire, avec certitude, le résultat qui se produira.

**Exemples :** lancer un dé, tirage de boules dans une urne, …

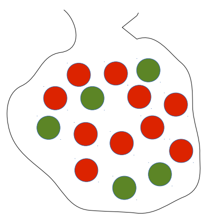
**Définition :**

Lorsque l´on répète $n$ fois, de façon indépendante, une même expérience aléatoire, on obtient une série de $n$ résultats que l´on appelle **échantillon** de taille $n$.

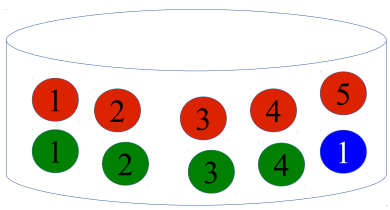
**Remarque :** pour obtenir un échantillon à l'aide d'un tirage, celui-ci doit s'effectuer **avec remise** pour que la proportion du caractère considéré ne change pas.

**Exemples :**

- On a lancé 7 fois un dé, et l'on a obtenu l'échantillon de taille 7 suivant :
  4 ; 4 ; 2 ; 1 ; 2 ; 6 ; 3
- Une urne contient 10 boules rouges et 5 boules vertes.
  On choisit une boule au hasard, on note sa couleur, puis on **la remet** dans l'urne.
  On mélange et on recommence six fois de suite l'expérience.
  On a ainsi un échantillon de taille 7, avec par exemple, 5 boules rouges et 2 boules vertes.



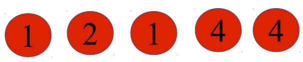
- Une urne contient des jetons verts, rouges et bleus, de forme identique, numérotés.
  La proportion de jetons verts dans l'urne est $p = 0{,}4$.
  On tire au hasard un jeton dans l'urne, on note le résultat, puis on le remet dans l'urne.
  On répète cette expérience aléatoire 5 fois.




*En remettant le jeton dans l'urne, on s'assure que le tirage qui va suivre ne dépend pas du précédent ; les répétitions sont **indépendantes**.*

Si l'on obtient l'échantillon 

, la fréquence de jetons vertes est $f = 0{,}6$.

Si l'on obtient l'échantillon 

, la fréquence de jetons vertes est $f = 0$.

## Taille de l'échantillon

Lorsque la taille de l'échantillon augmente la **fréquence *f*** tend à se **stabiliser** autour d'un nombre.

### 2) Fluctuation d'échantillonnage

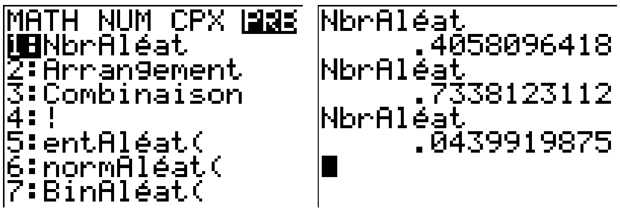
**Définition :**
Pour une population donnée, des échantillons aléatoires produits suivant le même protocole peuvent avoir des compositions différentes : on dit qu'il y a **fluctuation d'échantillonnage**.

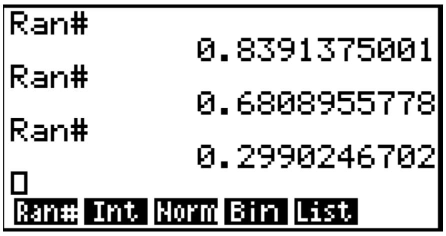
**Remarque :**
La fréquence ***f*** d'un caractère dans l'échantillon n'est pas nécessairement égale à la proportion ***p*** de ce caractère dans la population.

# II. Simulation

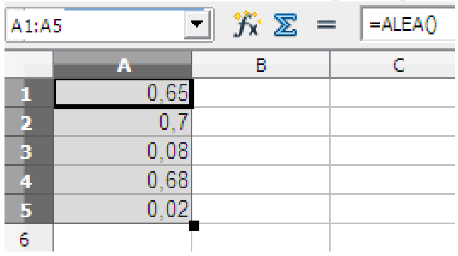
Lorsqu'on veut étudier une population, il n'est pas toujours possible d'y avoir accès.
On est alors amené à produire des échantillons par **simulation**.
Une simulation peut être faite avec un objet usuel (pièce de monnaie, dé, …) ou bien avec des outils de simulation.

<table>
<tr><th>Utilisation de la calculatrice</th><th>Utilisation du tableur</th></tr>
<tr><td>
Générer un nombre aléatoire de [0 ; 1[.

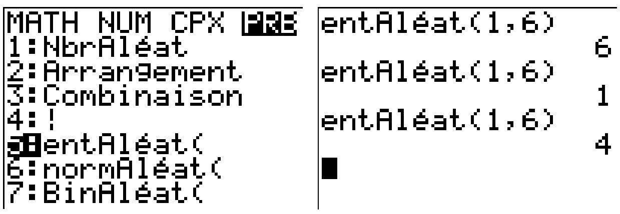


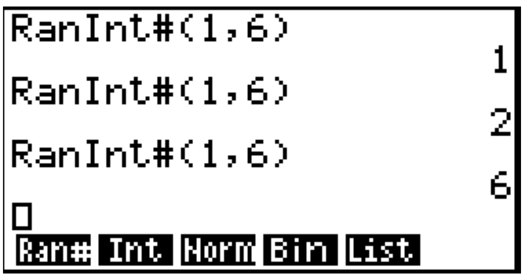



</td><td>
Générer un nombre aléatoire de [0 ; 1[.

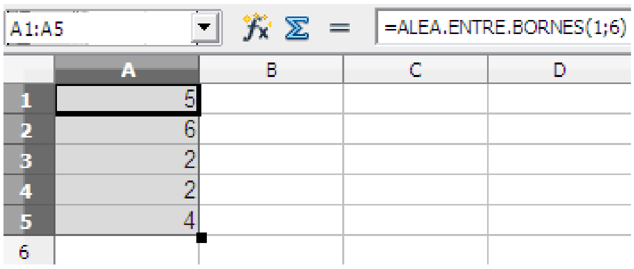


</td></tr>
<tr><td>
Simulation d'un jet de dé.



</td><td>
Simulation d'un jet de dé.

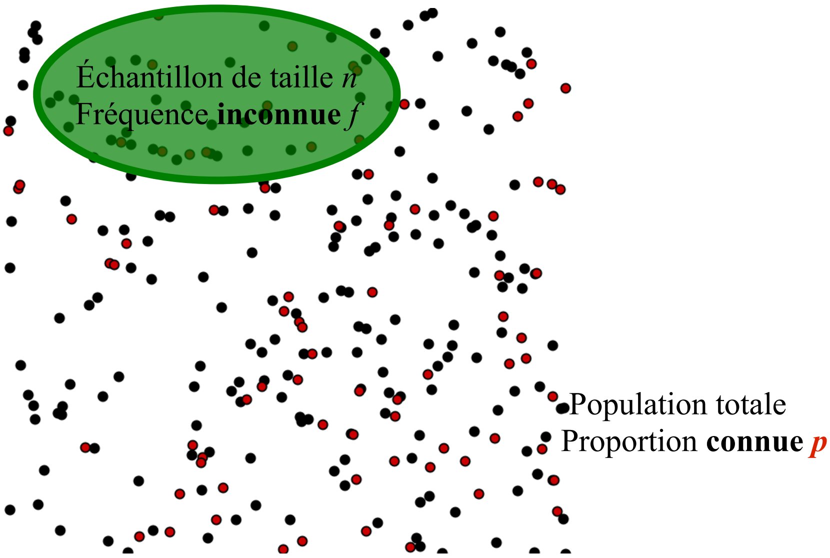


</td></tr>
</table>

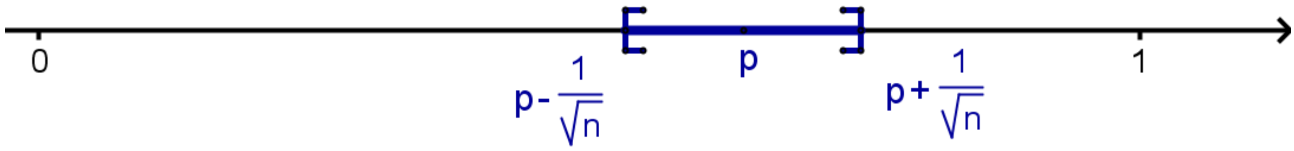
## III. Intervalle de fluctuation

### 1) Étude d'un échantillon

Au sein d'une population, **on connaît la proportion *p*** des individus ayant un caractère donné.
Parmi les échantillons de **taille *n*** extraits de cette population, la fréquence d'apparition ***f*** du caractère varie avec l'échantillon prélevé.

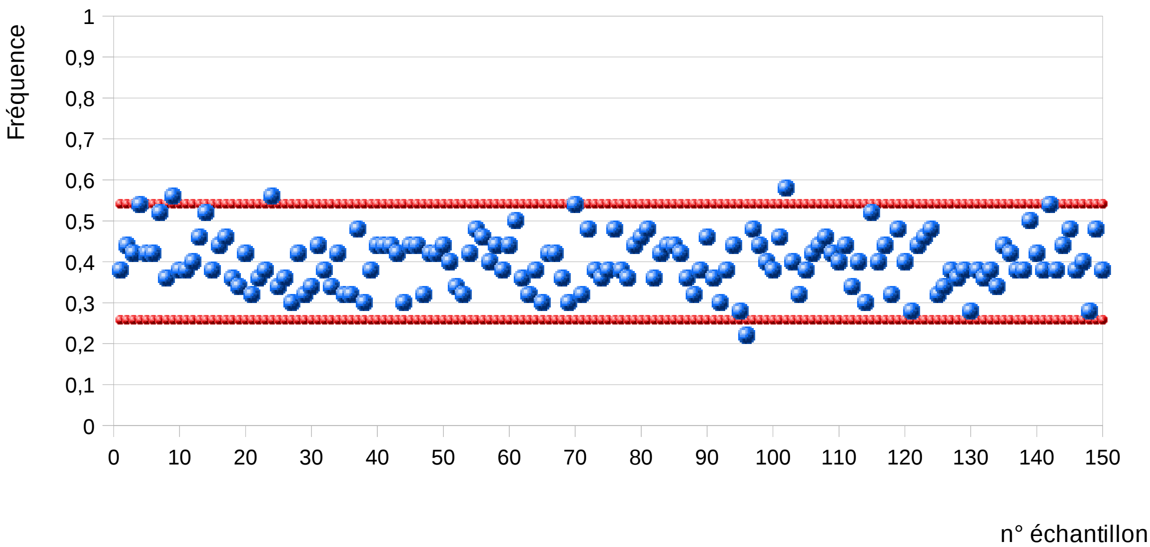



On admet que, pour un échantillon de taille $n \geqslant 25$ et pour ***p*** comprise entre 0,2 et 0,8, la fréquence d'apparition ***f*** observée appartient à l'**intervalle** $\left[p-\frac{1}{\sqrt{n}}; p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95.



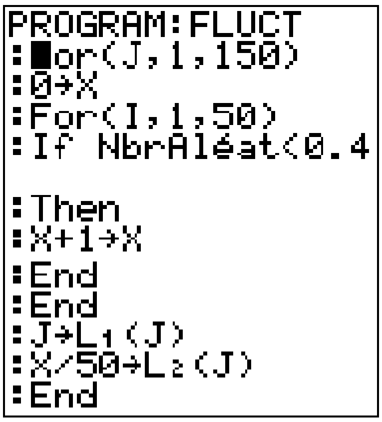

Cet intervalle est appelé **intervalle de fluctuation** au seuil de 95 %.

**Exemple :**
Une urne contient des jetons de même forme.
La proportion de jetons verts est $p = 0,4$.
On a produit 150 échantillons de taille 50 des jetons de l'urne.
Dans chaque échantillon on a mesuré la fréquence de jetons verts.
Les 150 fréquences obtenues sont représentées dans le nuage ci-dessous.

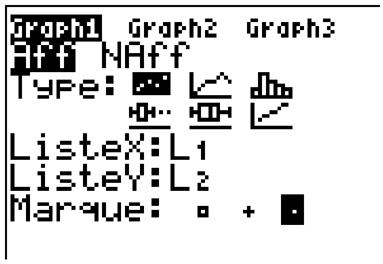


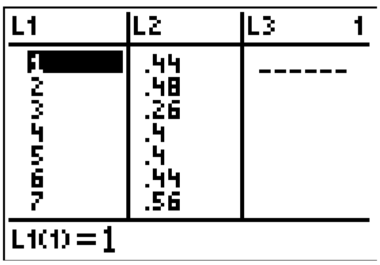

L'intervalle $\left[0,4-\frac{1}{\sqrt{50}};0,4+\frac{1}{\sqrt{50}}\right]$ apparaît avec les lignes rouges.
On peut vérifier que 95 % au moins des fréquences obtenues appartiennent à cet intervalle.

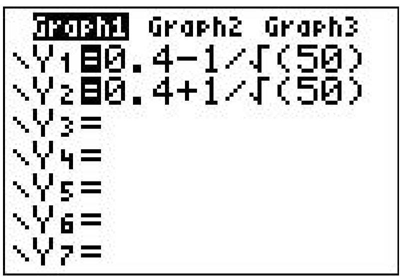
**Calculatrice**

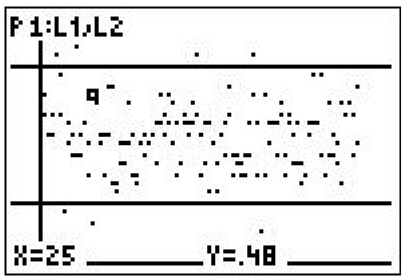


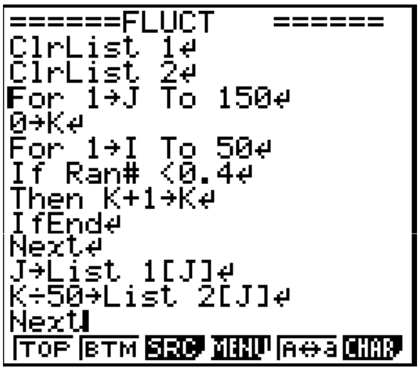


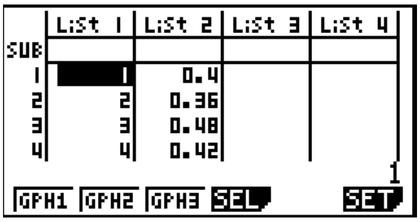


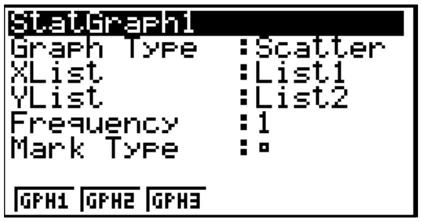



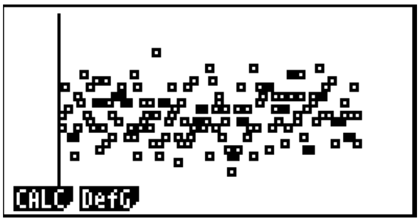



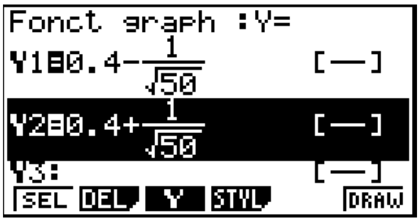



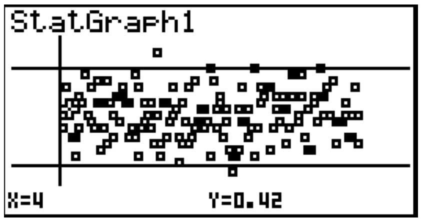








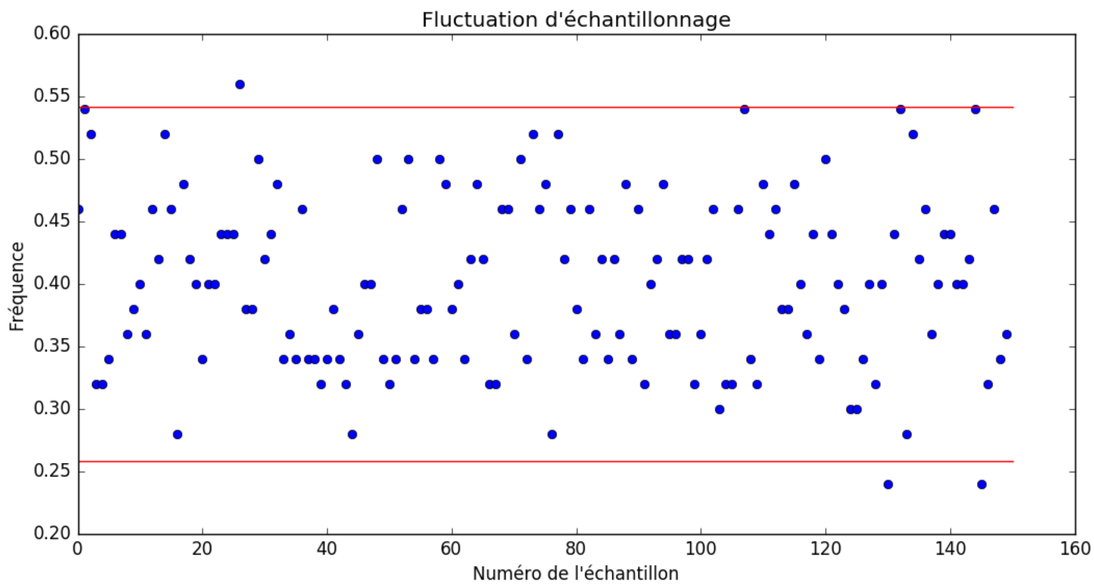





**Python**

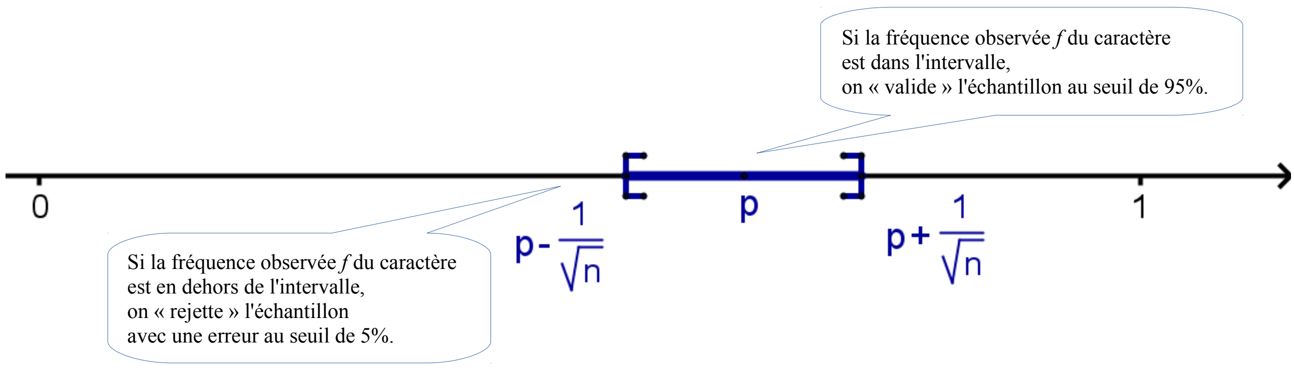
```
import matplotlib.pyplot as plt
from random import random
from math import sqrt

proportion = float(input("Proportion étudiée : "))
taille = int(input("Taille de l'échantillon : "))
nombre = int(input("Nombre d'échantillons : "))
echantillons = []
for j in range(nombre):
    k = 0
    for i in range(taille):
        if random() < proportion:
            k = k + 1
    echantillons += [k / taille]

plt.plot(echantillons, 'o')
plt.plot([0, nombre], [proportion + 1 / sqrt(taille),
                     proportion + 1 / sqrt(taille)], color = 'red')
plt.plot([0, nombre], [proportion - 1 / sqrt(taille),
                     proportion - 1 / sqrt(taille)], color = 'red')
plt.title("Fluctuation d'échantillonnage")
plt.xlabel("Numéro de l'échantillon")
plt.ylabel('Fréquence')
plt.show()
```

```
Proportion étudiée : 0.4
Taille de l'échantillon : 50
Nombre d'échantillons : 150
```




**Remarque :**

**Connaissant la proportion *p*** d'individus dans une population, l'intervalle de fluctuation $\left[p-\frac{1}{\sqrt{n}}\,;\,p+\frac{1}{\sqrt{n}}\right]$ permet d'étudier un échantillon donné par rapport à ce caractère.



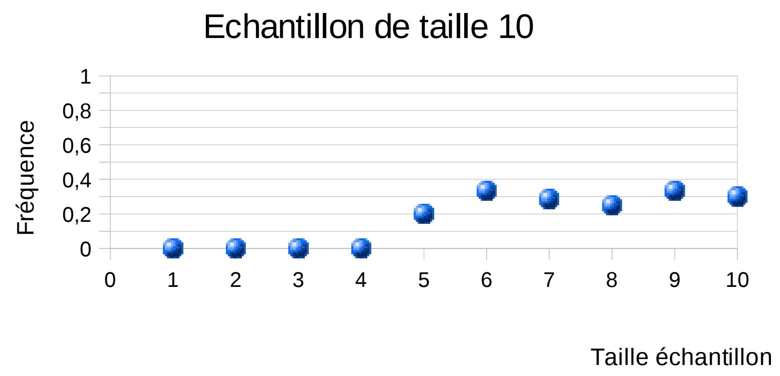
## 2) *Estimation d'une proportion*

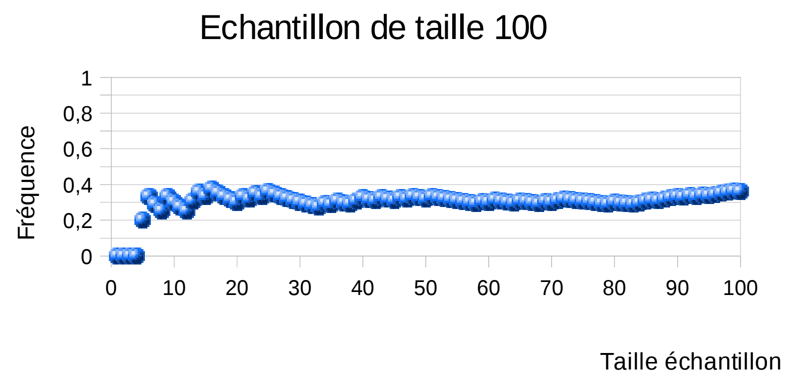
- Si l'on étudie des échantillons de **tailles** de **plus en plus grandes**, on constate alors que l'**amplitude** des fluctuations **diminue** et que les fréquences observées ont tendance à se stabiliser.
- Ces fréquences ont tendance à se stabiliser autour d'une valeur, que l'on peut utiliser comme une **estimation de la proportion *p*** de l'issue étudiée.

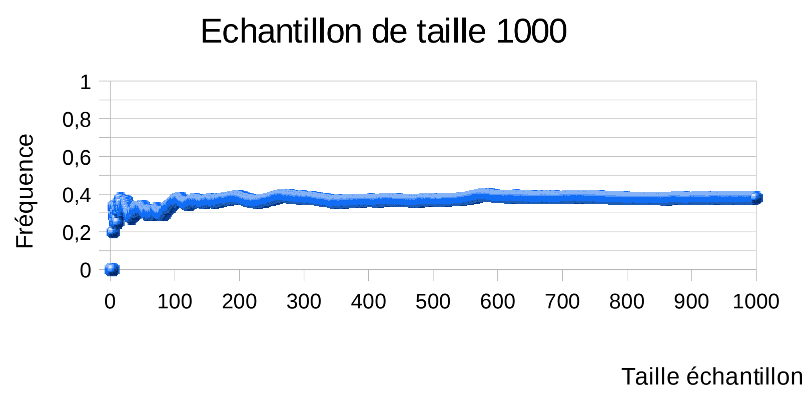
__Exemples :__

- Une urne contient des jetons verts, rouges et bleus, de forme identique, numérotés.
  La proportion de jetons verts dans l'urne est ***p* = 0,4**.

Observons l'évolution des fréquences de boules vertes en fonction de la taille de l'échantillon.

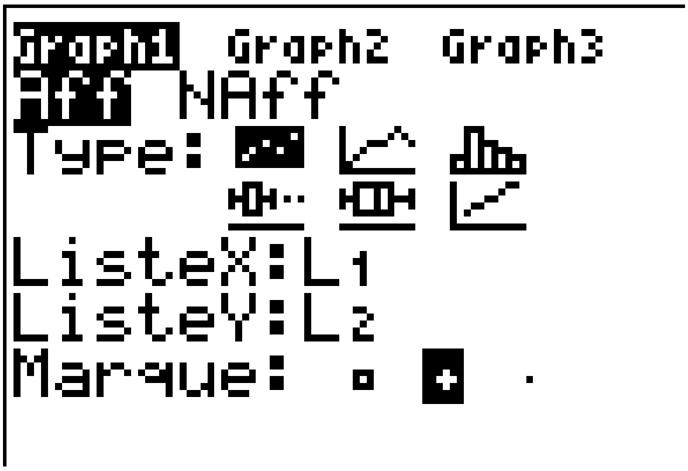


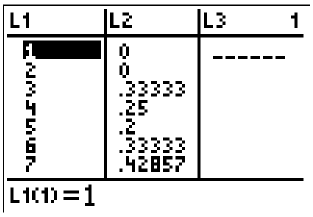



**Calculatrice** :

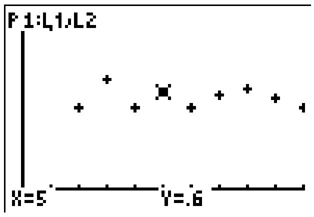
```
PROGRAM:STABIL
:Input "TAILLE:"
,N
:EffListe L1
:EffListe L2
:0→X
:For(I,1,N)
:If NbrAléat<0.4
:Then
:X+1→X
:End
:I→L1(I)
:X/I→L2(I)
:End
```



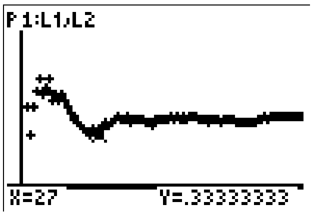

```
prgmSTABIL
TAILLE:10
               Fait
```



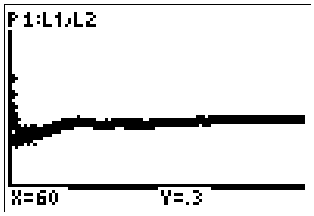

```
FENETRE
 Xmin=.5
 Xmax=10
 Xgrad=1
 Ymin=-.1
 Ymax=1.1
 Ygrad=1
 Xres=1
```



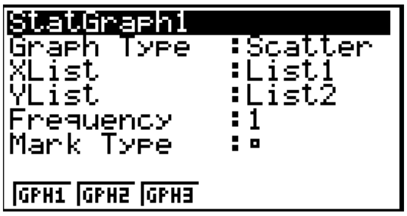

```
prgmSTABIL
TAILLE:100
               Fait
```



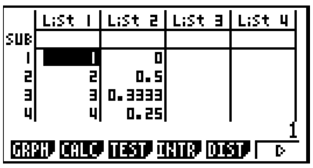

```
prgmSTABIL
TAILLE:999
               Fait
```

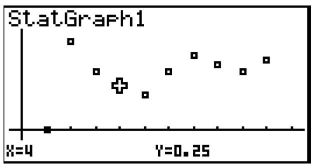



```
======STABIL  ======
ClrList 1↵
ClrList 2↵
"TAILLE:"?→N↵
0→K↵
For 1→I To N↵
If Ran# <0.4↵
Then K+1→K↵
IfEnd↵
I→List 1[I]↵
K÷I→List 2[I]↵
Next
```

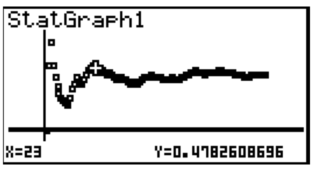



```
TAILLE:?
10
               0.4
```

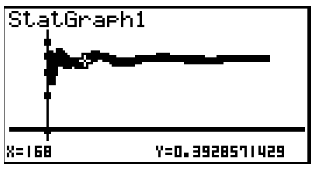



```
TAILLE:?
100
              0.43
```



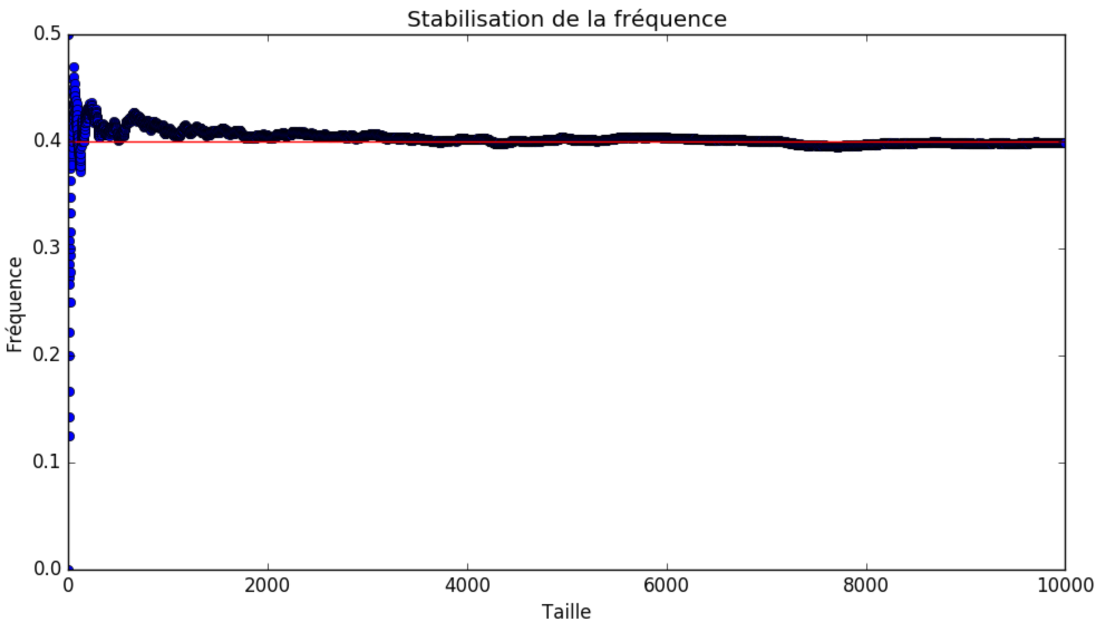

```
TAILLE:?
999
      0.4094094094
```



**Python :**

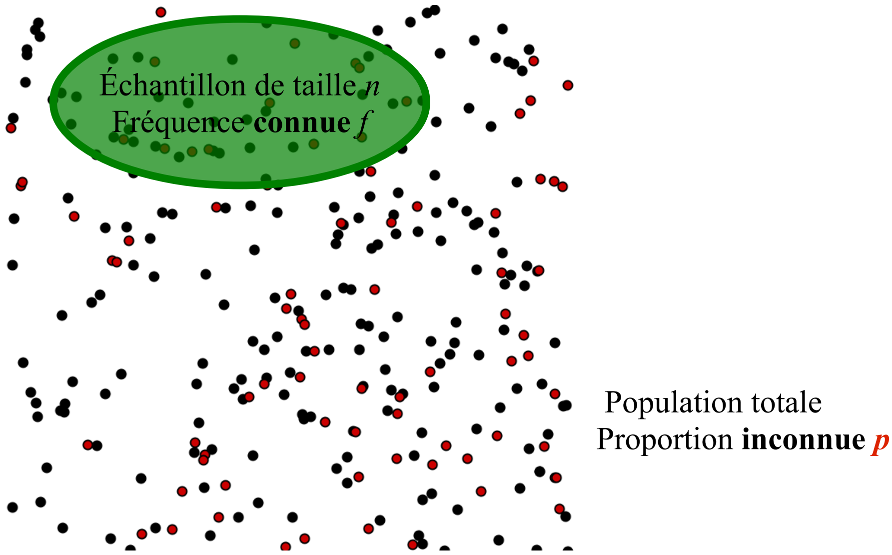
```
import matplotlib.pyplot as plt
from random import random

proportion = float(input("Proportion étudiée : "))
taille = int(input("Taille de l'échantillon : "))
frequences = []
k = 0
for i in range(1, taille + 1):
    if random() < proportion:
        k = k + 1
    frequences += [k / i]

plt.plot(frequences, 'o')
plt.plot([0, taille], [proportion, proportion], color = 'red')
plt.title("Stabilisation de la fréquence")
plt.xlabel("Taille")
plt.ylabel('Fréquence')
plt.show()
```

```
Proportion étudiée : 0.4
Taille de l'échantillon : 10000
```



## Sondage

C'est notamment, à partir de cette propriété (stabilisation de la fréquence $f$ vers la proportion $p$) que les sondages sont réalisés ; on **ne connaît pas la proportion $p$** d'un caractère (on souhaite l'estimer) et on va essayer de l'approcher en étudiant la fréquence du caractère sur un échantillon.




**Exemple** :
Lors d'un sondage effectué auprès de 900 personnes, 51 % d'entre elles déclarent vouloir voter pour le candidat A.
En supposant que les personnes sondées ont répondu sincèrement et qu'elles ne changeront pas d'avis le jour du vote, le candidat A peut-il raisonnablement penser qu'il sera élu au premier tour (c'est à dire avec plus de 50 % des votes).

*On constate que la taille de l'échantillon est* ***supérieur à 25*** *et l'on peut* ***considérer*** *que la proportion p (inconnue) est de l'ordre de grandeur de la fréquence de voix favorable (donc* ***0,2<p<0,8****).*

On sait donc que la fréquence 0,51 appartient à : $\left[p-\frac{1}{\sqrt{900}};p+\frac{1}{\sqrt{900}}\right]$ (intervalle de fluctuation au seuil de confiance de 95 %).
On en déduit donc que :

$$p\in\left[0{,}51-\frac{1}{\sqrt{900}};0{,}51+\frac{1}{\sqrt{900}}\right] \text{ (avec une marge d'erreur de 5 \%)}$$

donc :

$$0{,}48\leqslant p\leqslant 0{,}54 \text{ (arrondis à } 10^{-2}\text{)}$$

Le candidat ne peut donc pas considérer qu'il est assuré d'obtenir plus de 50 % des voix.

## Méthode de Monte Carlo

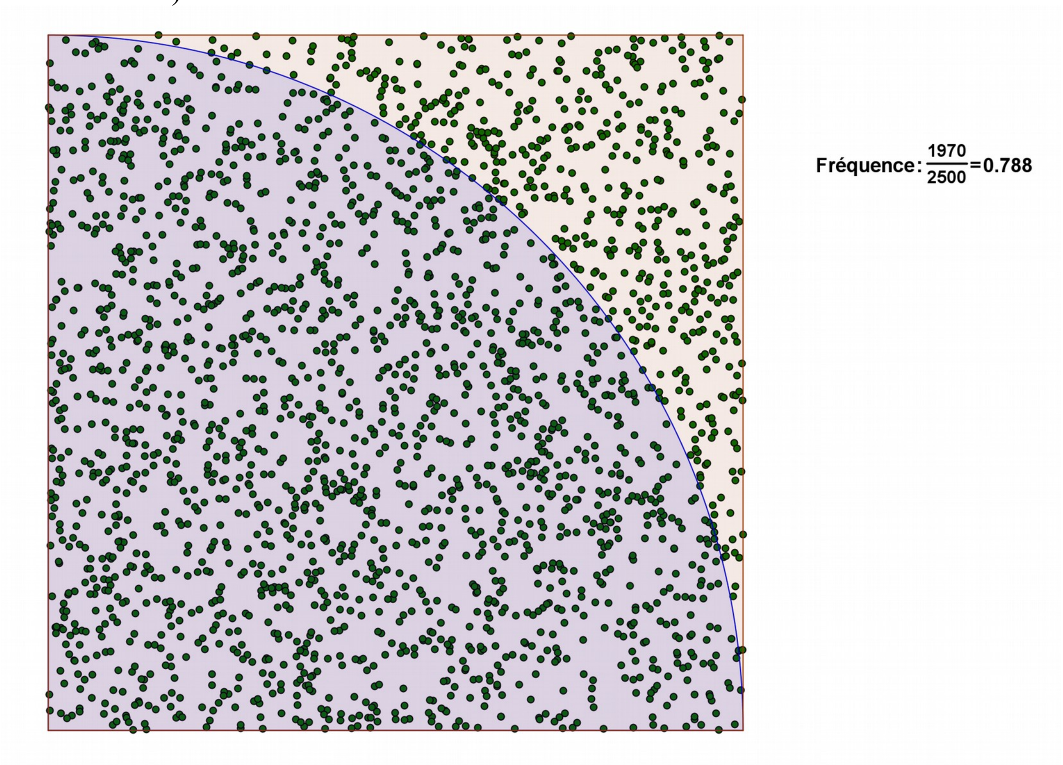
On détermine, par un procédé aléatoire, une fréquence qui nous permet d'approcher une valeur numérique.

**Exemple** :

Approximation de $\frac{\pi}{4}$.

On ne connaît pas la valeur numérique de $\frac{\pi}{4}$, mais on sait que cette valeur correspond à l'aire d'un quart de disque de rayon 1.

On va donc placer des points aléatoirement dans un carré de rayon 1, puis on détermine la fréquence d'appartenance de ces points au quart de disque et on obtient ainsi une approximation (avec une marge d'erreur de 5 %) de la valeur de π.




Sur cette simulation, sur les 2500 points placés dans le carré, 1970 sont dans le quart de disque (donc $f = 0{,}788$)

On sait donc que (avec une marge d'erreur de 5 %),

la fréquence $f$ appartient à $\left[\frac{\pi}{4}-\frac{1}{\sqrt{2500}};\frac{\pi}{4}+\frac{1}{\sqrt{2500}}\right]$ (on rappelle que ici on ne connaît pas $\frac{\pi}{4}$).

Donc $\frac{\pi}{4} \in \left[0{,}788-\frac{1}{\sqrt{2500}};0{,}788+\frac{1}{\sqrt{2500}}\right]$ ainsi $0{,}768<\frac{\pi}{4}<0{,}808$ et donc $3{,}072<\pi<3{,}232$

Voici donc une méthode probabiliste (avec une marge d'erreur de 5 %) permettant de déterminer une approximation de π.

Il suffit alors d'augmenter le nombre de points pour affiner l'approximation.