II. Simulation d'échantillons

On lance un dé à 6 faces.

On considère l'épreuve de Bernoulli : « lancer un dé à 6 face » dont le « succès » est l'événement S : « le dé s'arrête sur la face 1 » On a $p(S) = \frac{1}{6}$.

Cette expérience suit donc une loi de Bernoulli de paramètre $p = \frac{1}{6}$

Si on répète n fois de suite cette expérience à 2 issues, on obtient un schéma de Bernoulli de paramètres n et $p = \frac{1}{6}$

On s'intéresse au nombre de fois que le dé s'arrête sur la face « 1 ».

On peut simuler l'expérience à l'aide d'un programme qui renvoie une liste composée d'un échantillon de n lancers de dé. On peut aussi le faire avec un tableur :

```
from random import*

def echantillon(n):
    L=[]                      # crée une liste vide
    for i in range(n):
        x=randint(1,6)        # nombre entier aléatoire compris entre 1 et 6.
        L.append(x)           # ajoute x à la fin de la liste L
    return(L)
```

Que fait l'instruction suivante ?

```
>>> echantillon(10)
[6, 1, 3, 1, 2, 2, 6, 3, 5, 2]
>>>
```

On exécute la fonction échantillon pour n=10, afin de simuler 10 lancers de dé.

On modifie ensuite le programme afin qu'il renvoie en sortie la fréquence de « 1 » obtenus pour un échantillon de taille *n*.

```
from random import*

def echantillon(n):
    c=0                       # on utilise un compteur initialisé à 0 pour dénombrer le nombre de succès obtenus.
    for i in range(n):
        x=randint(1,6)
        if x==1:              # on teste si le lancer s'arrête sur la face 1.
            c=c+1
    return(c/n)
```

A quoi servent les instructions ci-dessous ?

```
>>> echantillon(10)
0.1
>>> echantillon(100)
0.2
>>> echantillon(1000)
0.18
>>> echantillon(100000)
0.16499
>>> echantillon(10000000)
0.1668073
```

On exécute le programme avec des valeurs de n de plus en plus grandes afin de réaliser l'expérience sur des échantillons de plus en plus grands.

Qu'observe-t-on ?

Plus la taille de l'échantillon augmente, plus les fréquences simulées observées se rapprochent de la valeur théorique $p = \frac{1}{6}$

On améliore encore le programme pour simuler N échantillons de taille n et afficher en sortie les fréquences obtenues :

```
from random import*

def echantillon(n):
    c=0
    for i in range(n):
        x=randint(1,6)
        if x==1:
            c=c+1
    return(c/n)

def simulation(N,n):
    L=[]
    for i in range(N):
        L.append(echantillon(n))
    return(L)
```

nbr de fois où on répète l'expérience

taille d'un échantillon.

2e échantillon

1er échantillon

Nᵉ échant.

résultats des n lancers de l'échantillon de taille n.

Que fait l'instruction suivante ?

```
>>> simulation(10,50)
[0.24, 0.18, 0.16, 0.26, 0.14, 0.16, 0.18, 0.24, 0.18, 0.14]
>>>
```

On exécute le programme pour créer 10 échantillons de 50 lancers chacun.

IV. Fluctuation d'échantillonnage

1. définition

La simulation précédente nous montre que si l'on réalise plusieurs échantillons de même taille,…la fréquence observée du succès fluctue.………………………………

C'est ce qu'on appelle la…fluctuation d'échantillonnage.……………………

Plus la taille de l'échantillon est grande, plus les fréquences …se rapprochent…de la probabilité théorique.

Que réalise l'instruction suivante ?

```
>>> simulation(10,1000)
[0.194, 0.187, 0.179, 0.186, 0.172, 0.166, 0.177, 0.163, 0.178, 0.178]
>>>
```

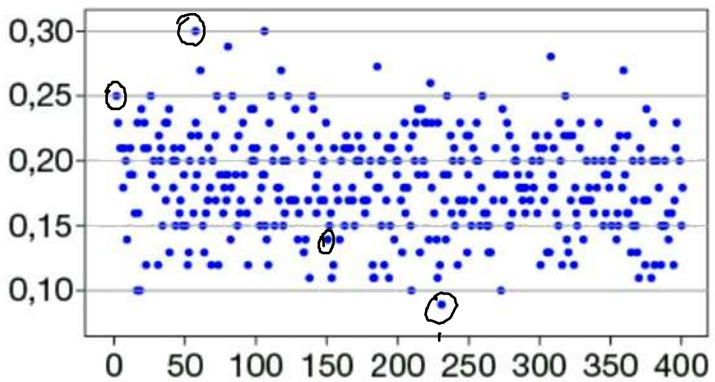
..On exécute la fonction simulation pour N = 10 et n = 1000 afin de simuler 10 échantillons de taille 1000 chacun.

Que remarque – t – on ?

……Le phénomène de fluctuation diminue : les résultats sont moins dispersés autour de p.

Le nuage de points ci-dessous représente la simulation de 400 échantillons de taille 50.

On peut lire que les fréquences …observées simulées..fluctuent entre 0,08 et 0,30.



2. Dispersion des résultats

On note *p* la …proportion théorique…d'un caractère dans un échantillon de taille n .

On note *s* …l'écart-type…de la série des fréquences obtenues.

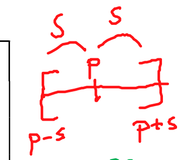
Standard deviation : écart-type

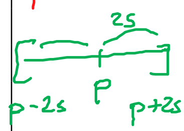
On admet que s $\approx \dfrac{1}{2\sqrt{n}}$

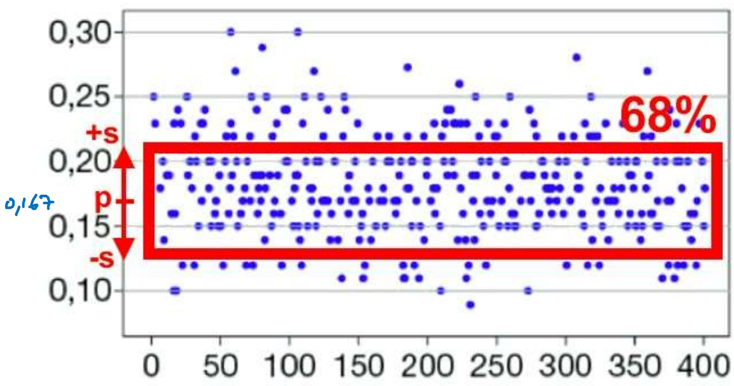
Remarque : l'écart-type est souvent noté …σ (sigma)…

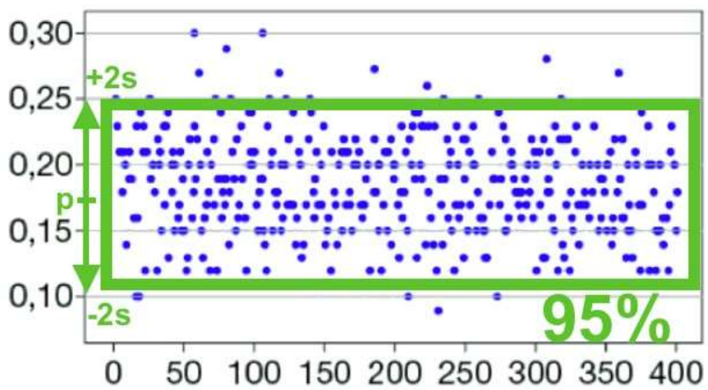
Dans un échantillon de taille n, pour une proportion théorique p, en moyenne,

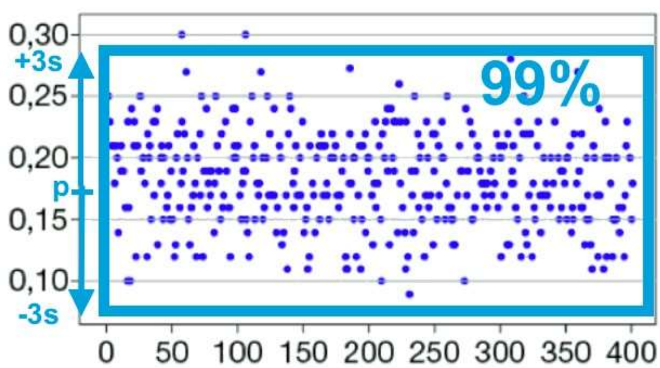
- 68% des fréquences observées appartiennent à l'intervalle …$[p-s\,;\,p+s]$…
- 95% des fréquences observées appartiennent à l'intervalle …$[p-2s\,;\,p+2s]$…
- 99% des fréquences appartiennent à l'intervalle …$[p-3s\,;\,p+3s]$…











$s = \frac{1}{2\sqrt{n}}$ $2s = \cancel{2} \times \frac{1}{\cancel{2}\sqrt{n}} = \frac{1}{\sqrt{n}}$

Rappels de seconde :

On prélève un échantillon dans la population et on note f la fréquence d'apparition du caractère observé dans cet échantillon.

INTERVALLE DE FLUCTUATION AU SEUIL DE 95%

Un intervalle de fluctuation de la fréquence f au seuil de 95% relatif aux échantillons de taille n est un intervalle I tel que, pour au moins 95% de l'ensemble des échantillons possibles, la fréquence observée appartient à I.

PROPRIÉTÉ

Pour une proportion p comprise entre 0,2 et 0,8, et des échantillons de taille $n \geqslant 25$, l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ est un intervalle de fluctuation au seuil de 95% de la fréquence f observée.

$[p - 2s \,;\, p + 2s]$

PRISE DE DÉCISION

Selon la situation étudiée, l'intervalle de fluctuation au seuil de 95% peut permettre :

— de décider si l'échantillon est représentatif de l'ensemble de la population ;

— d'accepter ou pas l'hypothèse que la proportion d'individus présentant le caractère étudié est égale à p.

$\boxed{P_{théo} = 60\%}$ échantillon de taille $n = 1000$.

$f_{obs} \in \left[60\% - \frac{1}{\sqrt{1000}} ; 60\% + \frac{1}{\sqrt{1000}}\right]$

$f_{obs} \in [0,568 ; 0,632]$

si f_{obs} n'appartient pas à cet intervalle, $P_{théo}$ est remise en question.