II. Simulation d'échantillons

On lance un dé à 6 faces.

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

Cette expérience suit donc une loi de Bernoulli de paramètre $p = p(\mathrm{S}) = \frac{1}{6}$.

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=[]
    for i in range(n):
        x=randint(1,6)
        L.append(x)
    return(L)
```

Crée une liste vide

nombre aléatoire entre 1 et 6

Ajoute x à la fin de la liste L

On exécute le programme avec n = 10 :

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

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

On exécute le programme avec des valeurs successives de n de plus en plus grandes.

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

Les fréquences simulées semblent de rapprocher de la valeur théorique $\frac{1}{6}$.

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)
```

On exécute le programme pour 10 échantillons de taille 50 :

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

III. Fluctuation d'échantillonnage

1. définition

La simulation précédente nous montre que si l'on réalise plusieurs échantillons de même taille, la fréquence observée de 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]
>>>
```

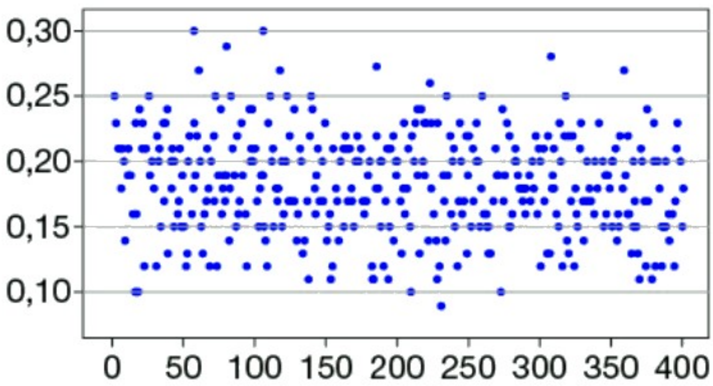
On exécute la fonction simulation avec comme arguments N=10 échantillons de taille n = 1000

Que remarque – t – on ?

On constate alors que le phénomène de fluctuation diminue.

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

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



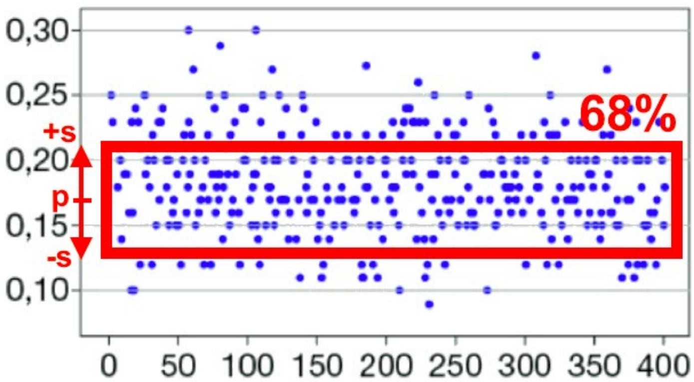
2. Dispersion des résultats

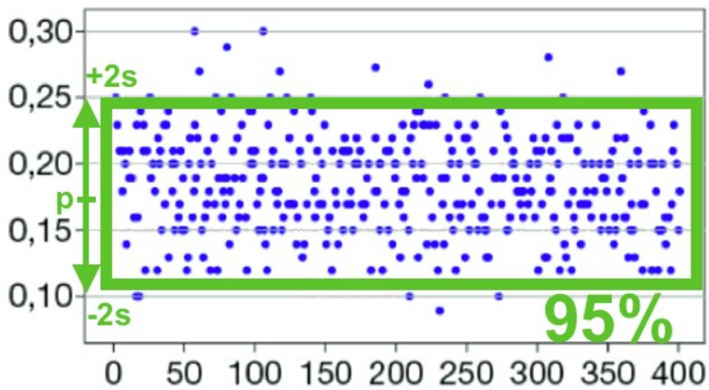
On note p la **proportion théorique** d'un caractère dans un échantillon de taille n .
On note s **l'écart-type** de la série des fréquences obtenues. On admet que $s \approx \frac{1}{2\sqrt{n}}$

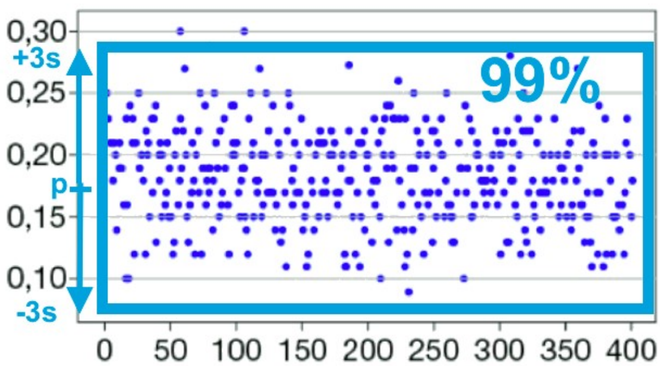
Remarque : l'écart-type est souvent noté σ (sigma)

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].







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.

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.