Échantillonnage / Fluctuation

1ère STMG

Simulation

On lance un dé à 6 faces n fois de suite et on observe le nombre de fois que le dé s'arrête sur la face "1".

On considère donc comme **"succès"** le fait d'obtenir un **1**.

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

On va simuler l'expérience à l'aide d'un programme qui renvoie une liste composée d'un échantillon de n lancers de dé :

from random import *

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

On exécute le programme avec $n = 10$:

```
>>> echantillon(10)
[3, 1, 3, 1, 5, 2, 2, 6, 4, 6]
>>>
```

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

```
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)
0.0
```

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

```
>>> echantillon(10)
0.0
```

```
>>> echantillon(100)
0.17
```

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

```
>>> echantillon(10)
0.0

>>> echantillon(100)
0.17

>>> echantillon(1000)
0.171
```

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

```
>>> echantillon(10)
0.0

>>> echantillon(100)
0.17

>>> echantillon(1000)
0.171

>>> echantillon(10000)
0.171
```

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

```
>>> echantillon(10)
0.0

>>> echantillon(100)
0.17

>>> echantillon(1000)
0.171

>>> echantillon(10000)
0.171

>>> echantillon(100000)
0.16761
```

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

```
>>> echantillon(10)
0.0
```

```
>>> echantillon(100)
0.17
```

```
>>> echantillon(1000)
0.171
```

```
>>> echantillon(10000)
0.171
```

```
>>> echantillon(100000)
0.16761
```

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)
```

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):
        f=echantillon(n)
        L.append(f)
    return(L)
```

On exécute le programme pour 10 échantillons de taille 50 :

```
>>> simulation(10,50)
[0.18, 0.32, 0.12, 0.04, 0.12, 0.18, 0.16, 0.14, 0.22, 0.22]
```

On vient de calculer la fréquence d'aparition de 1 sur 50 lancers, ceci 10 fois de suite.

Fluctuation d'échantillonnage

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

Fluctuation d'échantillonnage

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**.

Fluctuation d'échantillonnage

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**.

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

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

```
>>> simulation(10,50)
[0.16, 0.04, 0.1, 0.22, 0.26, 0.14, 0.26, 0.06, 0.06, 0.2]
```

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

```
>>> simulation(10,50)
[0.16, 0.04, 0.1, 0.22, 0.26, 0.14, 0.26, 0.06, 0.06, 0.2]
```

```
>>> simulation(10,500)
[0.164, 0.17, 0.17, 0.16, 0.176, 0.16, 0.182, 0.168, 0.154, 0.162]
```

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

```
>>> simulation(10,50)
[0.16, 0.04, 0.1, 0.22, 0.26, 0.14, 0.26, 0.06, 0.06, 0.2]

>>> simulation(10,500)
[0.164, 0.17, 0.17, 0.16, 0.176, 0.16, 0.182, 0.168, 0.154, 0.162]

>>> simulation(10,5000)
[0.1712, 0.1648, 0.168, 0.1566, 0.16, 0.1682, 0.1602, 0.1574, ...
... 0.1676, 0.166]
```

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

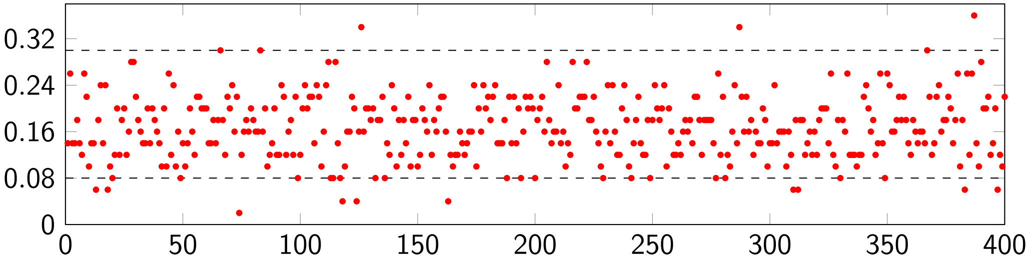
```
>>> simulation(10,50)
[0.16, 0.04, 0.1, 0.22, 0.26, 0.14, 0.26, 0.06, 0.06, 0.2]

>>> simulation(10,500)
[0.164, 0.17, 0.17, 0.16, 0.176, 0.16, 0.182, 0.168, 0.154, 0.162]

>>> simulation(10,5000)
[0.1712, 0.1648, 0.168, 0.1566, 0.16, 0.1682, 0.1602, 0.1574, ...
... 0.1676, 0.166]

>>> simulation(10,50000)
[0.16866, 0.16704, 0.16736, 0.1657, 0.16726, 0.16718, 0.1647, ...
... 0.16526, 0.16508, 0.16738]
```

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

```
>>> simulation(10,50)
[0.16, 0.04, 0.1, 0.22, 0.26, 0.14, 0.26, 0.06, 0.06, 0.2]

>>> simulation(10,500)
[0.164, 0.17, 0.17, 0.16, 0.176, 0.16, 0.182, 0.168, 0.154, 0.162]

>>> simulation(10,5000)
[0.1712, 0.1648, 0.168, 0.1566, 0.16, 0.1682, 0.1602, 0.1574, ...
... 0.1676, 0.166]

>>> simulation(10,50000)
[0.16866, 0.16704, 0.16736, 0.1657, 0.16726, 0.16718, 0.1647, ...
... 0.16526, 0.16508, 0.16738]
```

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 $0,30$.

Dispersion des résultats

p est la **proportion théorique** dans un échantillon de taille n.

s est **l'écart-type** de la série des fréquences obtenues. On pourra prendre $s \approx \dfrac{1}{2\sqrt{n}}$

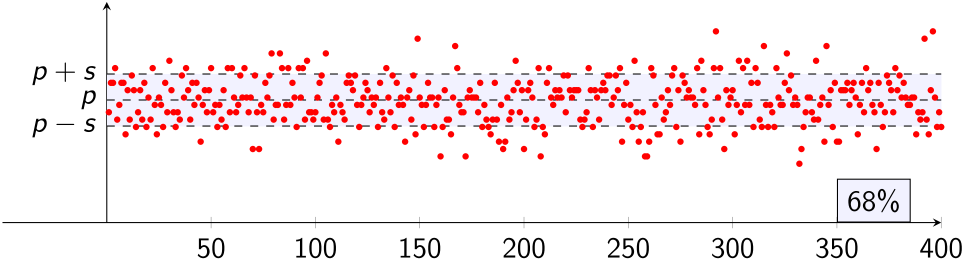
Dispersion des résultats

p est la **proportion théorique** dans un échantillon de taille n.

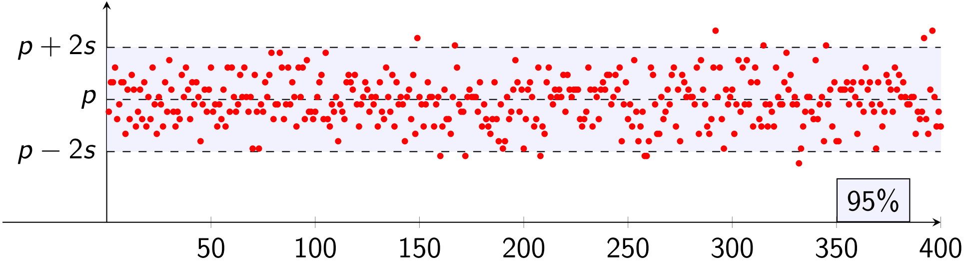
s est **l'écart-type** de la série des fréquences obtenues. On pourra prendre $s \approx \dfrac{1}{2\sqrt{n}}$

En moyenne,

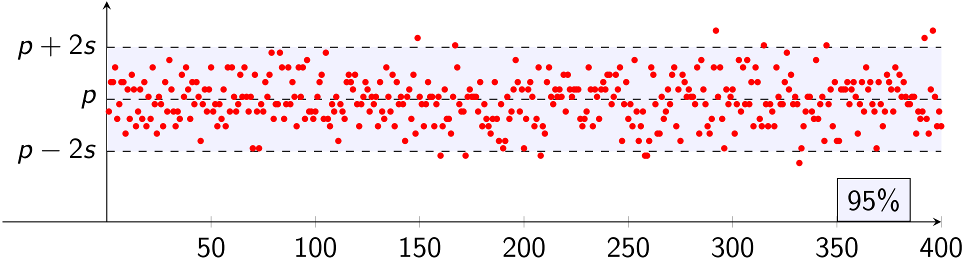
- 68% des fréquences appartiennent à l'intervalle $[p - s\ ; p + s]$.



- 95% des fréquences appartiennent à l'intervalle $[p - 2s\ ;\ p + 2s]$.



- 95% des fréquences appartiennent à l'intervalle $[p - 2s\ ;\ p + 2s]$.



- 99% des fréquences appartiennent à l'intervalle $[p - 3s\ ;\ p + 3s]$.