Intervalle de fluctuation et intervalle de confiance

D.Zancanaro

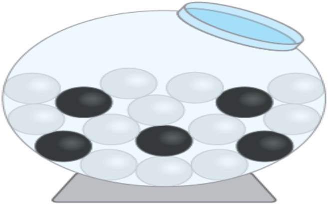
Vers la maîtrise du hasard

12$^{\text{th}}$ May 2020

Introduction : Présentation du problème

On illustre la situation ainsi : on dispose d'une grande urne où se trouvent un très grand nombre de boules noires et blanches.

- **La situation** : On tire au hasard avec remise n boules de l'urne. On obtient donc une fréquence f de boules noires sur cet échantillon.
- **La loi des grands nombres** : On s'attend à ce que la fréquence f observée soit " proche " de p, fréquence théorique et ceux d'autant plus que n est grand. On dit que f fluctue autour de p.
  On est ici dans le cadre d'un échantillonnage ; on vient en effet de réaliser un échantillon de taille n.



Principe de l'échantillonnage : Intervalle de fluctuation au seuil 0.95

Théorème

Soit p la **proportion** *d'un caractère donné dans une grande population.*
Soit f la **fréquence d'apparition** *du caractère dans un échantillon de taille n de cette popuplation.*
Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ alors parmi tous les échantillons de taille n, dans plus de 95 % des cas, la fréquence f se situe dans l'intervalle :

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

On l'appelle **intervalle de fluctuation** *au seuil 95%.*

Principe de l'échantillonnage : Signification de l'intervalle de fluctuation

- **Exemple** : *On considère une pièce de monnaie équilibrée. La fréquence théorique (ou probabilité) de l'issue Pile est* $p = 0,5$*. Imaginons qu'on lance cette pièce* 10000 *fois ; on obtiendrait un échantillon de taille* $n = 10000$*.*
  Comme $n \geq 30$, $np = 5000 \geq 5$ *et* $n(1-p) = 5000 \geq 5$, *on peut calculer l'intervalle de fluctuation au seuil* 95% *du caractère Pile :*

$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right] = \left[0.5 - \frac{1}{\sqrt{10000}}; 0.5 + \frac{1}{\sqrt{10000}}\right] = [0.49; 0.51]$$

  Il y a donc 95% *de chance que la fréquence observée de l'issue Pile sur un échantillon de taille* 10000 *appartienne à cet intervalle.*

Signification de l'intervalle de fluctuation - Simulation en python

On lance une pièce de monnaie bien équilibrée 500 fois on compte le nombre de pile. On reproduit 10 fois un tel échantillon.

```python
# Hello World program in Python
from random import randint


def simulation(n):
    Npile=0
    for i in range(n):
        hasard=randint(0,1)
        if hasard==0:
            Npile=Npile+1
    return(Npile)

print(simulation(500))

L=[]
for i in range(10):
    L.append(simulation(500))
L.sort()
print(L)
```

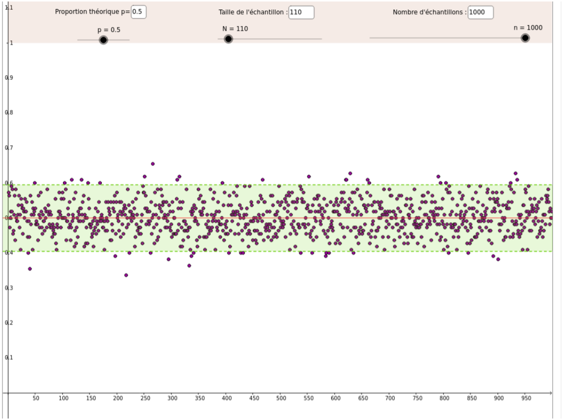
input

Compiled Successfully. memory: 9952 time: 0.34 exit code: 0

```
244
[238, 239, 247, 256, 259, 262, 263, 264, 265, 265]
```

Signification de l'intervalle de fluctuation - Simulation

On lance une pièce de monnaie bien équilibrée 100 fois on compte le nombre de pile. On reproduit 1000 fois un tel échantillon, la zone verte correspond à l'intervalle de fluctuation au seuil 0.95.



Principe de l'échantillonnage : Prise de décision à partir d'un échantillon

Théorème

On considère une population dans laquelle on suppose que la proportion d'un caractère est p. On observe sur un échantillon de taille n une fréquence f du caractère.
On veut tester l'hypothèse : La proportion de ce caractère dans la population est p . Si I est l'intervalle de fluctuation au seuil 95% (en respectant les conditions de validité), la règle de décision est la suivante :

- *Si $f \in I$, on considère que l'hypothèse selon laquelle la proportion est p dans la population n'est pas remise en question (et on l'accepte au seuil de confiance 95%).*
- *Si $f \notin I$, on rejette l'hypothèse selon laquelle cette proportion est p au seuil de confiance 95% (au risque de se tromper dans 5% des cas).*

Prise de décision à partir d'un échantillon : Un exemple

Exemple : L'entreprise A annonce que le pourcentage de moteurs défectueux dans la production est égal à 1%. Afin de vérifier cette affirmation 800 moteurs sont prélevés au hasard. On constate que 15 moteurs sont détectés défectueux. Le résultat de ce test remet-il en question l'annonce de l'entreprise A ? Justifier. On pourra s'aider d'un intervalle de fluctuation.

Prise de décision à partir d'un échantillon : Solution

Solution : Pour une proportion p et un échantillon de taille n, l'intervalle de fluctuation asymptotique au seuil de 95 % est $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ sous les trois conditions: $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

L'échantillon de l'enquête est de taille $n = 800$ et l'entreprise annonce que le pourcentage de moteurs défectueux est égal à 1 % donc $p = 0,01$.
Regardons si les trois conditions sont vérifiées:
$n = 800 \geq 30$, $np = 800 \times 0,01 = 8 \geq 5$ et
$n(1-p) = 800 \times 0,99 = 792 \geq 5$.

L'intervalle est: $I = \left[0,01 - \frac{1}{\sqrt{800}} ; 0,01 + \frac{1}{\sqrt{800}}\right] \approx [0,003 ; 0,017]$.

On constate que 15 moteurs sont détectés défectueux sur 800, ce qui fait une proportion de $\frac{15}{800} = 0,01875$; or $0,01875 \notin I$ donc le résultat de ce test remet en question l'annonce de l'entreprise A.

Principe de l'estimation : Intervalle de confiance

- On ne connait pas p, la proportion de boules noires de l'urne.
- On veut donc estimer la proportion p de boules noires dans l'urne. On tire donc au hasard avec remise n boules de l'urne. On obtient donc une fréquence f de boules noire sur cet échantillon. On s'attend à ce que la probabilité p théorique soit proche de f, fréquence observée. On est ici dans le cadre d'une **estimation**.
- **Loi des grands nombres** : Lorsque n est grand, sauf exception, la fréquence observée est proche de la probabilité.
- $\left[f - \dfrac{1}{\sqrt{n}} ; f + \dfrac{1}{\sqrt{n}}\right]$ est appelé intervalle de confiance au seuil 95% de la proportion p de ce caractère aléatoire de la population.
- Si p est la probabilité d'une issue et f sa fréquence observée dans un échantillon, alors la proportion des cas où l'écart entre p et f est inférieur ou égal à $\dfrac{1}{\sqrt{n}}$ est au moins égale à 95%.

Bilan : les intervalles de fluctuation

Intervalle de fluctuation : Prise de décision

- On teste l'hypothèse : "la probabilité d'un caractère est p".
- On fabrique un échantillon de taille n ; on obtient alors une fréquence f du caractère observée.
- On vérifie les trois conditions les trois conditions: $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ puis on calcule l'intervalle de fluctuation au seuil 0.95 :

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

- On regarde si oui ou non la fréquence f observée appartient à l'intervalle de fluctuation.
  - Si $f \in I$, on considère que l'hypothèse selon laquelle la proportion est p est "valable"
  - Si $f \notin I$, on rejette l'hypothèse au rsique de se tromper dans 5% des cas.

Bilan : les intervalles de confiance

Intervalle de confiance : Estimation

- On ne connaît pas la probabilité théorique p d'apparition d'un caractère.
- On effectue n simulations pour obtenir une fréquence f.
- On estime que p appartient à l'intervalle de confiance :

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

au seuil 95%

Vidéos

- Echantillonnage : https://youtu.be/27dXCURENpE
- Estimation : https://youtu.be/EXEcSJE31QY
- Python : https://www.youtube.com/watch?v=VmOPhT4HFNE&list=PLVUDmbpupCaobTdn2MIqD-DV6AUvEuUll

A lire

- Cours
  http://wicky-math.fr.nf/pdf2/Coronatravail/TSTMG/Cours_ICD.pdf
- Cours (autre version)
  http://wicky-math.fr.nf/pdf2/Coronatravail/2nde/CM_7.pdf
- La page qui résume tout
  http://wicky-math.fr.nf/index.php/coronatravail

- Thank You.