ÉCHANTILLONNAGE

OBJECTIFS

- Connaître la définition d'un échantillon de taille n pour une expérience à deux issues.
- Effectuer une première approche de la loi des grands nombres.
- Savoir estimer une probabilité, ou une proportion dans une population, à partir d'une fréquence observée sur un échantillon.

I Échantillonnage

1. Notion d'échantillon

À RETENIR

Définitions

On considère une expérience aléatoire à deux issues, que l'on peut répéter de manière **indépendante** (ie. de sorte que la probabilité de chaque issue ne dépend pas des résultats précédemment obtenus).

Un **échantillon** de taille n est constitué des résultats obtenus par n répétitions de cette expérience aléatoire.

EXEMPLE

On lance un dé équilibré à 6 faces et on regarde si on tombe sur un résultat pair.

On peut répéter cette expérience aléatoire plusieurs fois de manière indépendante. Si on la répète dix fois, on obtient un échantillon de taille 10, par exemple :

$$(I; P; P; I; I; P; P; P; I; P)$$

où P désigne l'issue : « Le nombre obtenu est pair » et I l'issue : « Le nombre obtenu est impair ».

EXERCICE 1

On lance un dé tétraédrique (à quatre faces) équilibré, et on s'intéresse au fait d'obtenir 1. Donner un exemple d'échantillon de taille 5 possible. ....

....

Voir la correction : https://mes-cours-de-maths.fr/cours/seconde/echantillonnage/#correction-1.

2. Simulation

À RETENIR

Remarque

Simuler un échantillon informatiquement permet d'étudier des séries statistiques comportant un très grand nombre de données.

EXEMPLE

Le code ci-contre est composé de deux fonctions.

— `piece()` qui permet de renvoyer aléatoirement 0 (pour Face) ou 1 (pour Pile) afin de simuler un lancer de pièce.

— `echantillon(n)` qui simule n lancers de pièce et renvoie la liste des résultats.

Par exemple, `echantillon(20)` pourra renvoyer la liste

[0,0,1,1,1,0,0,0,1,0,0,1,0,0,1,1,1,0,0,0]

qui présente une fréquence de Pile égale à $\frac{8}{20}$.

```
import random

def piece():
  return random.randint(0, 1)

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

EXERCICE 2

On lance un dé à 6 faces et on s'intéresse à l'événement « Obtenir 6 ».

1. Quelles sont les issues possibles pour cette expérience? Donner la loi de probabilité associée.

2. Écrire une fonction `frequenceDe6(n)` en Python qui construit un échantillon de taille n et qui calcule la fréquence de 6 dans cet échantillon.

3. À quoi sert la fonction `simulation(n, N)` écrite ci-dessous?

....

....

....

....

```
def simulation(n, N):
  resultat = []
  for i in range(N):
    resultat.append(frequenceDe6(n))
  return resultat
```

Voir la correction : https://mes-cours-de-maths.fr/cours/seconde/echantillonnage/#correction-2.

II Échantillons de grande taille

1. Fluctuation

À RETENIR

Définitions

On considère une expérience aléatoire à deux issues et on note p la probabilité d'une issue ω.

Si on réalise plusieurs échantillons de même taille, la fréquence de l'issue ω observée sur chaque expérience varie. C'est ce qu'on appelle la **fluctuation d'échantillonnage**.

Plus la taille des échantillons est grande, plus le phénomène de fluctuation diminue : les fréquences se rapprochent alors de p. C'est la **loi des grands nombres**.

EXEMPLE

On exécute `simulation(1000, 10)` où `simulation(n, N)` est la fonction définie à l'exercice 2, et on obtient le résultat suivant :

```
[0.164,0.186,0.176,0.154,0.178,0.161,0.159,0.176,0.176,0.167]
```

On constate qu'avec une taille d'échantillon suffisamment élevée (1 000 ici), les fréquences se stabilisent autour de la probabilité de $\frac{1}{6} \approx 0,167$.

2. Estimation

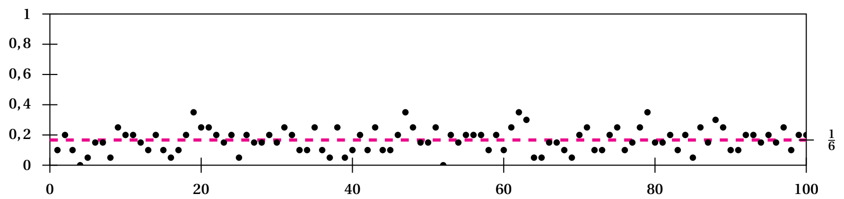
À RETENIR

Définition

On considère une expérience aléatoire à deux issues et on note p la probabilité d'une issue. Soit f la fréquence observée de cette issue dans l'échantillon. Lorsque n est grand, f et p sont proches donc, si l'on ne connaît pas la valeur de p, on peut considérer que f constitue une **estimation**.

EXEMPLE

On exécute `simulation(20, 100)` où `simulation(n, N)` est la fonction définie à l'exercice 2, et on représente le résultat dans le graphique ci-dessous par un nuage de points.



À partir des fréquences observées, on retrouve une approximation de la probabilité d'obtenir 6 : $\frac{1}{6}$.

INFORMATION

Remarque

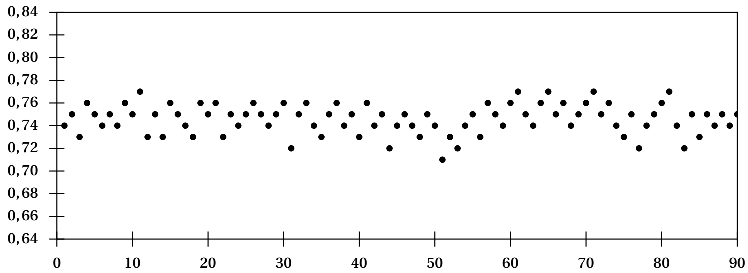
Lorsqu'on approche la probabilité p par la fréquence observée f, l'erreur commise est égale à $|f - p|$. Un résultat mathématique permet d'affirmer que, la plupart du temps,

$$|f - p| \leq \frac{1}{\sqrt{n}}$$

En particulier, plus n est grand, plus f et p sont proches.

EXERCICE 3

Un supermarché souhaite estimer la proportion de ses clients qui paient par carte bancaire (CB). Pour cela, pendant 90 jours, on relève la fréquence de clients payant par CB sur les 1 000 premiers clients, de sorte que l'on a 90 échantillons de taille 1 000. Les résultats sont représentés dans le graphique ci-dessous.



Estimer la proportion des clients payant par CB dans ce supermarché.

Voir la correction : https://mes-cours-de-maths.fr/cours/seconde/echantillonnage/#correction-3.