Échantillonnage

Par nature, une expérience aléatoire a un résultat imprévisible avec certitude puisque le hasard intervient. Pour autant, même le hasard obéit à des lois ! Dans ce chapitre, nous allons observer certaines de ces lois. Mieux encore, nous allons les utiliser pour résoudre des problèmes d'une très grande diversité !

1 Loi des grands nombres

Définition 1

Un est la liste des résultats obtenus en la même expérience

Propriété 1

On considère une expérience aléatoire et un E de cette expérience aléatoire.
Lorsque n est sauf exception, la observée de E sur un de taille n est de la de E.

Observation de la loi des grands nombres avec Python.
En lançant une pièce de monnaie la probabilité d'avoir pile est de Donc, vérifions que la observée lors de la simulation d'une série de n lancers avec n est aussi sauf exception, de

```python
from random import*

# Nombre de lancers
nombre_lancers = 1000  # Vous pouvez modifier ce nombre


def simuler_lancers_piece(nombre_lancers):
    compte_pile = 0

    for i_ in range(0, nombre_lancers):
        lancer = randint(0,1)  # 0 pour face, 1 pour pile
        if lancer == 1:
            compte_pile += 1

    frequence_pile = compte_pile / nombre_lancers
    return frequence_pile

# Simulation
frequence_pile = simuler_lancers_piece(nombre_lancers)
print(f"Fréquence d'apparition du côté pile:\n{frequence_pile:.4f}")
```

Remarque 1

La loi des grands nombres peut nous servir de deux façons. D'abord si on connaît déjà la probabilité p avec laquelle un évènement se produit, alors on pourra déterminer si une fréquence observée est compatible avec cette valeur de p. Cela est très utile pour détecter qu'il s'est produit quelque chose d'anormal. Ensuite, quand on ne connaît pas la probabilité p avec laquelle un évènement se produit, alors on pourra l'estimer à partir sa fréquence observée dans un échantillon.

2 Intervalle de fluctuation

Grâce à la loi des grands nombres, on sait que lorsque n est grand, sauf exception, la fréquence f observée de E sur un échantillon de taille n est de la de E. Nous allons dans la propriété suivante préciser ce que signifie que p et f sont proches.

Définition 2

On répète n fois une même expérience où E peut être réalisé avec une On appelle
de cet échantillon l'intervalle

Exercice 1

Déterminer l'intervalle de fluctuation de la fréquence de piles pour un échantillon de 100 lancers de pièces de monnaie équilibrées.

Correction

Ici, $p =$ et $n =$ Donc,

```python
import math
from random import*

def simuler_lancers_piece(nombre_lancers):
    compte_pile = 0

    for i_ in range(0,nombre_lancers):
        lancer = randint(0,1)  # 0 pour face, 1 pour pile
        if lancer == 1:
            compte_pile += 1

    frequence_pile = compte_pile / nombre_lancers
    return frequence_pile

def simuler_n_experiences(nombre_simulations, nombre_lancers):
    intervalle_bas = 0.5 - 1 / math.sqrt(nombre_lancers)
    intervalle_haut = 0.5 + 1 / math.sqrt(nombre_lancers)

    compte_dans_intervalle = 0

    for j_ in range(0,nombre_simulations):
        frequence_pile = simuler_lancers_piece(nombre_lancers)
        if intervalle_bas <= frequence_pile <= intervalle_haut:
            compte_dans_intervalle += 1

    frequence_dans_intervalle = compte_dans_intervalle / nombre_simulations
    return frequence_dans_intervalle

# Paramètres
nombre_simulations = 1000  # Vous pouvez modifier ce nombre
nombre_lancers = 100  # Vous pouvez modifier ce nombre

# Simulation
frequence_dans_intervalle = simuler_n_experiences(nombre_simulations, nombre_lancers)
print("Fréquence des simulations où la fréquence observée est dans l'intervalle:")
print(f"{frequence_dans_intervalle:.4f}")
```

Propriété 2

Si la taille de l'échantillon est assez la observée appartient à avec une probabilité supérieure à

Remarque 2

Cette propriété permet de déceler une statistique quand la fréquence à l'intervalle de fluctuation. Cela permet par exemple de détecter une fraude dans la comptabilité d'une entreprise ou encore en médecine de détecter une pathologie si une mesure est anormalement de la moyenne.

3 Estimation d'une proportion inconnue

Définition 3

Si f est la observée d'un caractère d'un de taille n, l'intervalle de est définie par

Exercice 2

Déterminer l'intervalle de confiance associé à un sondage réalisé sur 1000 personnes dans lequel 52% des sondés affirment vouloir voter pour Monsieur X.

Correction

Ici, $f =$ et $n =$ Donc, $\frac{1}{\sqrt{n}} \approx$ et l'intervalle de vaut

Propriété 3

Si et alors appartient à l'intervalle de de l'échantillon avec un

On peut donc estimer que avec un niveau de confiance de

Remarque 3

L'expression de 95% signifie qu'avant de choisir un échantillon, il y a environ de chances que l'intervalle de contienne la valeur que je cherche à Une fois l'échantillon il n'y a plus de hasard donc on ne peut plus parler de

Remarque 4

C'est sur cette dernière propriété que repose les

Exercice 3

Dans l'exercice précédent, citer deux raisons mathématiques pour lesquelles Monsieur X ne peut pas être certain de gagner malgré le sondage favorable.

Correction

La borne de l'intervalle de confiance est Ce sondage est donc compatible avec une
De plus, le niveau de de 95% est certes mais il n'est pas de On peut réduire le risque en réalisant