Expérimentations : simuler des échantillons

Fiche méthode

(Spécialité maths 1re · Probabilités · Expérimentations)

La démarche pour étudier une simulation

1. Je relis la loi simulée : valeurs, probabilités (seuils cumulés), puis μ et σ.

2. Je repère les tailles : n (taille d'un échantillon) et, s'il y a lieu, N (nombre d'échantillons).

3. Je calcule $\dfrac{2\sigma}{\sqrt{n}}$ et j'écris l'intervalle centré sur μ.

4. Je conclus dans le contexte : moyenne dans l'intervalle ou non ; proportion de l'ordre de 0,95, qui fluctue. En approfondissement : « compatible avec le modèle » ou « modèle mis en doute ».

Le réflexe — une conclusion se rédige avec les bons mots : « compatible avec le modèle » ou « le modèle est mis en doute », jamais « le modèle est démontré » ou « le modèle est faux ».

Cas 1 · Simuler une loi

Réflexe — seuils = probabilités cumulées ; la dernière valeur va dans le else ; randint(a, b) inclut b.

Exemple. Question type : « Écrire une fonction qui simule Y ». Exemple : Y vaut -1, 2 ou 5 avec les probabilités 0,25 ; 0,6 ; 0,15.

Seuils cumulés : 0,25 puis $0{,}25 + 0{,}6 = 0{,}85$.

```
def Y():
    u = random()
    if u < 0.25:
        return -1
    elif u < 0.85:
        return 2
    else:
        return 5
```

$Y = 2$ quand $u \in [0{,}25\,;0{,}85[$, de longueur 0,6 ; $Y = 5$ quand $u \in [0{,}85\,;1[$, de longueur 0,15.

Attention — écrire elif u $<$ 0.6 donnerait une longueur 0,35 au lieu de 0,6.

Cas 2 · Moyenne ou fréquence d'un échantillon

Réflexe — accumulateur à 0 avant la boucle, une valeur ajoutée par tour, division par n après la boucle.

Exemple. Question type : « Écrire une fonction qui renvoie la fréquence de pile sur n lancers ». Exemple : pièce truquée, pile avec la probabilité 0,6.

```
def freq(n):
    c = 0
    for k in range(n):
        if random() < 0.6:
            c = c + 1
    return c / n
```

Le compteur `c` ajoute 1 à chaque pile : `freq(n)` est la moyenne d'un échantillon de taille n de la variable de Bernoulli de paramètre 0,6, d'espérance $\mu = 0{,}6$.

Attention — return indenté dans la boucle = la fonction s'arrête au premier tour.

Cas 3 · Situer une moyenne observée

Réflexe — μ, σ, n → $2\sigma/\sqrt{n}$ → intervalle centré sur μ → phrase de conclusion dans le contexte.

Exemple. Question type : « L'observation remet-elle en cause l'annonce ? ». Exemple : une enseigne annonce 90 % de clients satisfaits ; sur 900 clients, 783 le sont.

Bernoulli : $\mu = 0{,}9$ et $\sigma = \sqrt{0{,}9 \times 0{,}1} = \sqrt{0{,}09} = 0{,}3$.
$\dfrac{2 \times 0{,}3}{\sqrt{900}} = \dfrac{0{,}6}{30} = 0{,}02$: intervalle $[0{,}88\,;0{,}92]$.
$f = \dfrac{783}{900} = 0{,}87 < 0{,}88$: la fréquence sort de l'intervalle ; en approfondissement, l'annonce de l'enseigne est mise en doute.

Attention — intervalle centré sur μ, pas sur la fréquence observée ; σ et non σ².

Cas 4 · Choisir la taille de l'échantillon (approfondissement)

Réflexe — $2\sigma/\sqrt{n} \leq a$, puis $\sqrt{n} \geq 2\sigma/a$, puis $n \geq (2\sigma/a)^2$, arrondi à l'entier supérieur.

Exemple. Question type : « Quelle taille minimale pour une demi-largeur au plus égale à a ? ». Exemple : $\sigma = 7$ et $a = 1$.

$\dfrac{14}{\sqrt{n}} \leqslant 1 \iff \sqrt{n} \geqslant 14 \iff n \geqslant 196$ (les deux membres sont positifs).
Taille minimale : 196. Pour diviser cette demi-largeur par 2, il faudrait 4 fois plus d'individus.

Attention — arrondir n à l'entier supérieur, jamais au plus proche.

Cas 5 · Lire le résultat d'une simulation

Réflexe — le test abs(m - mu) <= e s'écrit $|m - \mu| \leq 2\sigma/\sqrt{n}$; c / N est une proportion d'échantillons.

Exemple. Question type : « Interpréter la valeur renvoyée ». Exemple : `proportion(500, 64)` renvoie 0,954.

Sur 500 échantillons de taille 64, $0{,}954 \times 500 = 477$ vérifient $|m - \mu| \leqslant \dfrac{2\sigma}{8}$.
C'est proche de 0,95, comme l'annonce le cours ; une autre exécution donnerait une valeur un peu différente.

Attention — N = nombre d'échantillons, n = taille de chacun ; 95 % porte sur les échantillons.

Les pièges

- Seuils non cumulés dans une simulation : la loi simulée n'est plus la bonne.
- randint(1, 6) peut renvoyer 6 : les deux bornes sont comprises.
- Confondre μ (fixe, le modèle) et m (la moyenne observée, qui fluctue).
- Prendre la variance au lieu de l'écart type dans $2\sigma/\sqrt{n}$.
- Doubler n ne divise pas la demi-largeur par 2 : il faut multiplier n par 4.
- Une moyenne dans l'intervalle ne démontre pas le modèle ; une moyenne hors de l'intervalle le met en doute, avec un risque d'erreur d'environ 5 %.

Ce document est la propriété intellectuelle de Ben, professeur particulier (benprofparticulier.fr), protégée par le Code de la propriété intellectuelle (art. L. 111-1 et L. 122-4). Il est mis à disposition pour un usage personnel de révision. Toute reproduction, diffusion, mise en ligne ou revente, totale ou partielle, sans autorisation écrite est interdite et constitue une contrefaçon (art. L. 335-2).