1. Échantillonnage

Dans ce paragraphe, X désigne une variable aléatoire suivant la loi $\mathscr{B}(n\ ;\ p)$ et a et b deux réels.

DÉFINITION

On dit que $[a\ ;\ b]$ est un **intervalle de fluctuation au seuil de** 95 % **ou 0,95** du nombre de succès si $P(X \in [a\ ;\ b]) \geqslant 0{,}95$.

REMARQUES :

- Cela veut dire que si l'on réalise l'expérience, il y a plus de 95 % de chance que le nombre de succès soit dans l'intervalle $[a\ ;\ b]$.
- On définit de la même manière un intervalle de fluctuation à d'autres seuils mais en pratique on utilisera le plus souvent les seuils 95 % et 99 %.
- Pour un seuil donné, il existe plusieurs et même une infinité d'intervalles de fluctuation.

PROPRIÉTÉ

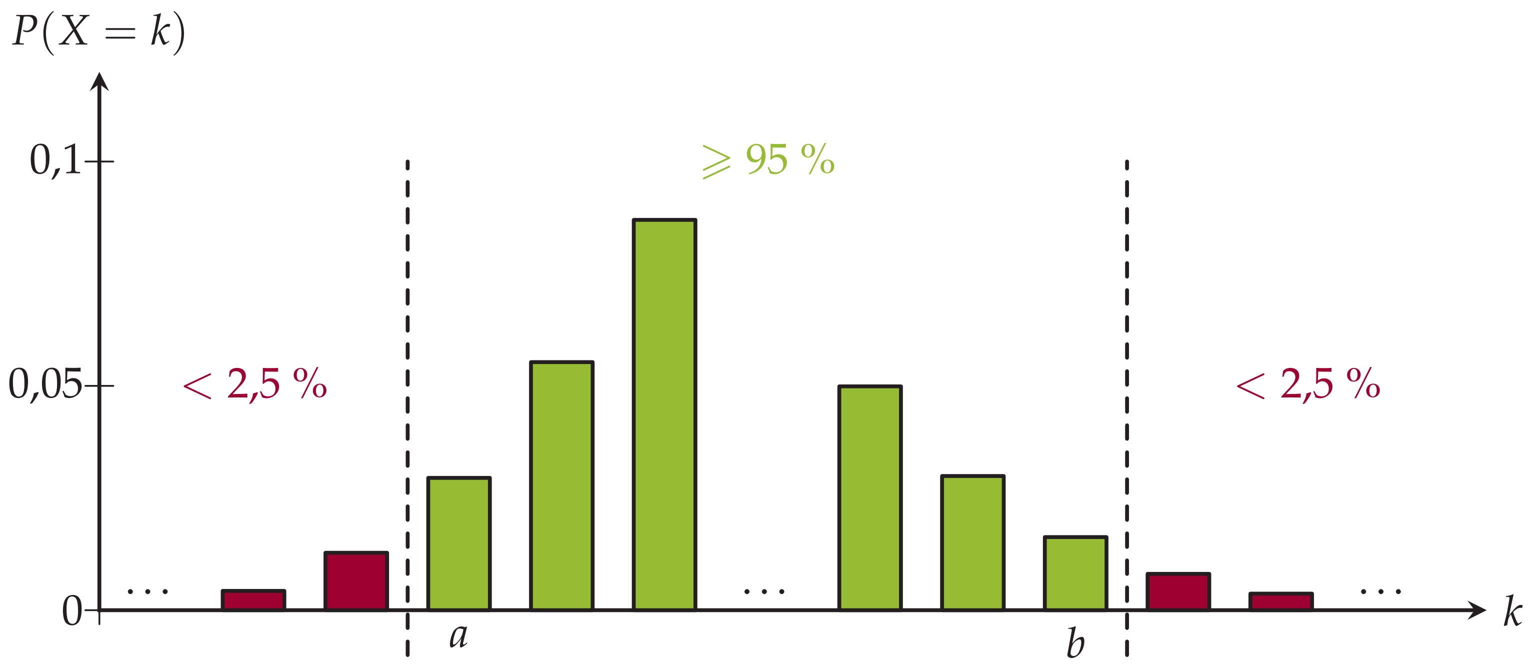
L'intervalle $[a\ ;\ b]$ où :

- a est le plus petit entier tel que $P(X \leqslant a) > 0{,}025$
- b est le plus petit entier tel que $P(X \leqslant b) \geqslant 0{,}975$

est un intervalle de fluctuation au seuil de 95 %.

REMARQUES : Dans la propriété précédente :

- on peut représenter la situation par le graphique suivant où le diagramme en bâtons représente la loi binomiale :



- c'est cet intervalle que l'on donnera généralement quand on demandera de déterminer l'intervalle de fluctuation au seuil de 95 %, même s'il en existe d'autres.

Dans la suite, l'intervalle $[a\ ;\ b]$ désigne l'intervalle de la propriété précédente.

PROPRIÉTÉ

On considère la variable aléatoire $F = \dfrac{X}{n}$ donnant la fréquence de succès.

L'intervalle $\left[\dfrac{a}{n}\,;\,\dfrac{b}{n}\right]$ est un intervalle de fluctuation au seuil de 95 % de cette fréquence.

PREUVE On a $\dfrac{a}{n} \leqslant \dfrac{X}{n} \leqslant \dfrac{b}{n} \Leftrightarrow a \leqslant X \leqslant b$ donc $P\left(\dfrac{a}{n} \leqslant \dfrac{X}{n} \leqslant \dfrac{b}{n}\right) = P(a \leqslant X \leqslant b)$.

Comme $P(a \leqslant X \leqslant b) \geqslant 0{,}95$, on en déduit que $P\left(\dfrac{a}{n} \leqslant F \leqslant \dfrac{b}{n}\right) \geqslant 0{,}95$.

PROPRIÉTÉ

On considère une population dans laquelle la proportion d'un certain caractère est p.
Si la population est suffisamment grande, quand on prélève un échantillon de n individus, on peut considérer que le nombre d'individus ayant le caractère suit une loi binomiale de paramètres n et p.
Ainsi, l'intervalle de fluctuation au seuil de 95 %

- du nombre d'individus ayant le caractère dans l'échantillon est $[a\,;\,b]$;
- de la fréquence du caractère dans l'échantillon est $\left[\frac{a}{n}\,;\,\frac{b}{n}\right]$.

REMARQUES :

- Le fait que la population soit suffisamment grande permet d'assimiler le prélèvement d'un échantillon de taille à n à **des tirages avec remise** donc indépendants.
- Si, après prélèvement de l'échantillon, on observe que la fréquence est bien dans l'intervalle de fluctuation, on dit que cet échantillon est **représentatif** de la population, sinon, on dit qu'il ne l'est pas, au seuil de 95 %.

Exemple Dans la population française, il y a 24,4 % de « moins de 20 ans » (source : *ined*).

1) On prélève au hasard un échantillon de 250 personnes dans la population française.
   Donner un intervalle de fluctuation au seuil de 95 % de la fréquence des « moins de 20 ans » dans cet échantillon.
2) Dans un village de 250 habitants, la proportion de « moins de 20 ans » est 28,5 %.
   Ce village est-il représentatif de la population française sur ce critère ?

Correction

1) Comme la population française est suffisamment grande pour assimiler le prélèvement de cet échantillon à un tirage avec remise, le nombre de « moins de 20 ans » dans cet échantillon suit la loi binomiale de paramètres $n = 250$ et $\dfrac{24,4}{100} = 0{,}244$ d'après la propriété précédente.
   - En tabulant cette loi avec la calculatrice, on obtient qu'un intervalle de fluctuation au seuil de 95 % du nombre de « moins de 20 ans » dans cet échantillon est $[48\ ;\ 75]$.
   - Un intervalle de fluctuation au seuil de 95 % de la fréquence de « moins de 20 ans » dans cet échantillon est donc $\left[\dfrac{48}{250}\ ;\ \dfrac{75}{250}\right]$ soit $[0{,}192\ ;\ 0{,}3]$.
2) Comme $\dfrac{28{,}5}{100} = 0{,}285 \in [0{,}192\ ;\ 0{,}3]$, ce village est représentatif de la population française.

Remarque : Si l'on avait utilisé la formule de l'intervalle de fluctuation vue en seconde, on aurait obtenu $\left[0{,}024\,4 - \dfrac{1}{\sqrt{250}}\ ;\ 0{,}024\,4 + \dfrac{1}{\sqrt{250}}\right]$ soit approximativement $[0{,}18\ ;\ 0{,}31]$.
Ce résultat est proche de $[0{,}192\ ;\ 0{,}3]$ obtenu avec la loi binomiale : ce sera toujours le cas dans les conditions d'application de cette formule ($0{,}2 \leqslant p \leqslant 0{,}8$ et $n \geqslant 25$).

MÉTHODE Rejeter ou non une hypothèse

Dans une population, on peut faire l'hypothèse que la proportion d'un certain caractère est p puis, à l'aide d'un intervalle de fluctuation et de l'effectif ou de la fréquence dans un échantillon prélevé, choisir de rejeter ou non cette hypothèse.

Exercice d'application

Un candidat à l'élection présidentielle affirme que 54 % de la population lui est favorable.
Lors d'un sondage réalisé sur 1 000 personnes, 523 personnes se sont déclarées favorables à ce candidat.
Peut-on rejeter ou non la proportion de 54 % donnée par le candidat, au seuil de 95 % ?

Correction

- Sous l'hypothèse que la proportion de personnes favorables au candidat est bien de 54 %, en tabulant la loi $\mathscr{B}(1000\ ;\ 0{,}54)$ avec la calculatrice, on obtient que l'intervalle de fluctuation au seuil de 95 % dans un échantillon de taille 1000 est $[509\ ;\ 571]$.
- Le nombre de personnes favorables dans l'échantillon est 523 qui appartient bien à $[509\ ;\ 571]$ donc, au seuil de 95 %, on ne peut pas rejeter l'hypothèse que la proportion est de 54 %.

REMARQUES :

- Par abus de langage, on dit parfois que l'on « accepte l'hypothèse » plutôt que l'on « ne rejette pas l'hypothèse ».
- Plutôt que dire que l'on rejette ou non une hypothèse « au seuil de 95 % », on peut aussi dire qu'on la rejette ou non « au risque d'erreur de 5 % ».