**DÉFINITION**

Soit $X \sim \mathcal{B}(n;p)$ et $\alpha \in ]0;1[$.

On appelle ***intervalle de fluctuation au seuil*** $\mathbf{1-\alpha}$ un intervalle $I$ qui vérifie : $p(X\in I)\geqslant 1-\alpha$.

$\alpha$ est le pourcentage d'erreur que l'on accepte

Si l'on souhaite un intervalle de fluctuation bilatéral (l'erreur est partagée « à gauche et à droite » de l'intervalle, à ses bornes), on peut utiliser la définition suivante :

**DÉFINITION**

Soit $X \sim \mathcal{B}(n;p)$ et $\alpha \in ]0;1[$. Un ***intervalle de fluctuation au seuil*** $\mathbf{1-\alpha}$ est $\left[\frac{k_1}{n};\frac{k_2}{n}\right]$ où :

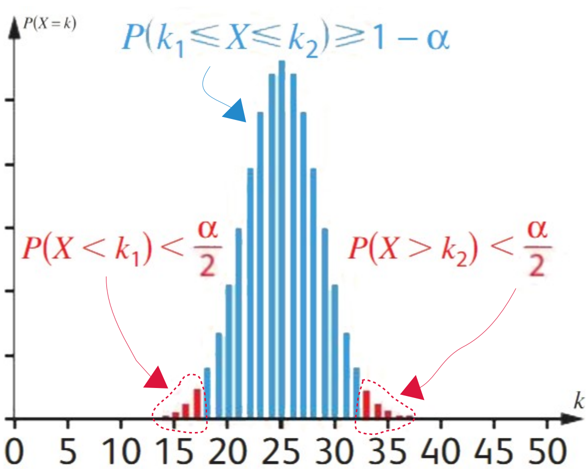
- $k_1$ est le plus grand entier vérifiant $p(X<k_1)\leqslant\frac{\alpha}{2}$
- $k_2$ est le plus petit entier vérifiant $p(X>k_2)\leqslant\frac{\alpha}{2}$.

Une autre définition équivalente, et bien plus pratique, est la suivante :

**DÉFINITION**

Un ***intervalle de fluctuation au seuil*** $\mathbf{1-\alpha}$ est $\left[\frac{k_1}{n};\frac{k_2}{n}\right]$ où :

- $k_1$ est le plus petit entier vérifiant $p(X\leqslant k_1)>\frac{\alpha}{2}$
- $k_2$ est le plus petit entier vérifiant $p(X\leqslant k_2)\geqslant 1-\frac{\alpha}{2}$.




Loi binomiale $\mathcal{B}(50\,;0,5)$

EXEMPLE C1

En première partie de soirée, une série a attiré près de 6,2 millions de téléspectateurs soit 34 % de part d'audience. Paul pense faire un sondage auprès de 100 habitants de son village, pour savoir la proportion de ceux qui ont regardé cette série.
Soit X la variable aléatoire qui correspond au nombre de téléspectateurs qui ont regardé cette série dans un échantillon de 100 personnes ayant regardé la télévision en première partie de soirée.
Le nombre de téléspectateurs en première partie de soirée est suffisamment important pour considérer que la variable X suit la loi binomiale de paramètres $n =$ ……… et $p =$ ……… .
Le plus petit entier $a$ tel que $p(X \leqslant a) > 0{,}025$ est ………
et, le plus petit entier $b$ tel que $p(X \leqslant a) \geqslant 0{,}975$ est ……… .
Un intervalle de fluctuation à 95 % de la fréquence des téléspectateurs qui ont regardé cette série dans un échantillon de taille 100 est donc :

………………………………

De la même manière, déterminer un intervalle de fluctuation à 90 % :

## → ESTIMER UNE PROPORTION : RÈGLE DE DÉCISION ←

On considère un caractère dont la proportion dans la population est supposée être égale à $p$.

La prise de décision consiste, à partir d'un échantillon de taille $n$, à valider ou non cette hypothèse faire sur la proportion $p$ :

**1.** On calcule la fréquence observée $f$ du caractère dans cet échantillon.

**2.** On choisit le risque α, puis on détermine un intervalle de fluctuation au seuil $1-\alpha$, noté I.

**3.** On applique la règle suivante :

→ Si $f \notin \mathrm{I}$ alors on rejette l'hypothèse faite sur $p$.
Dans ce cas, **il y a un risque[1] de se tromper de α :**
la probabilité qu'on rejette à tort l'hypothèse faite sur $p$ alors qu'elle est vraie (probabilité conditionnelle) est égale à α.

→ Si $f \in \mathrm{I}$ alors on accepte l'hypothèse faite sur $p$.
Dans ce cas, **le risque d'erreur n'est pas quantifié[2] !**

1 On parle de « risque de première espèce ». Ce risque est défini à l'avance (le plus souvent 1 % ou 5 %).
2 On parle de « risque de seconde espèce ».
À taille d'échantillon égale, si l'on diminue le risque de première espèce, on augmente le risque de seconde espèce.

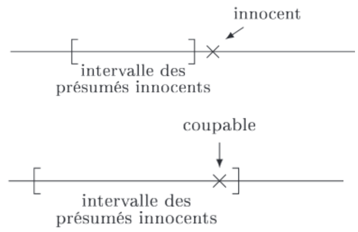
Le tableau ci-dessous représente bien les deux types d'erreurs :

| On fait l'hypothèse $p=p_0$ | | |
|---|---|---|
| Réalité ↓ Décision → | $p=p_0$ | $p\neq p_0$ |
| $p=p_0$ | OK | erreur : rejeter à tort |
| $p\neq p_0$ | erreur : valider à tort | OK |

Le risque de valider à tort n'est pas quantifiable car cela signifie que nous avons validé l'hypothèse $p=p_0$ alors que $p\neq p_0$ : mais alors on ne sait pas combien vaut $p$ ! On ne peut donc pas calculer la probabilité conditionnelle de valider cette hypothèse sachant qu'elle est fausse…

*Pourquoi ne pas abaisser le seuil de rejet d'une hypothèse ?*

On pourrait penser qu'il suffit de réduire le risque d'erreur (de première espèce) de rejeter une hypothèse à tort, de façon à n'avancer que des hypothèses très fiables.
Mais en faisant cela, on augmente le risque de commettre une autre erreur (de seconde espèce) : accepter l'hypothèse alors qu'elle est fausse !
Une analogie simple suffit à faire comprendre la situation : une prise de décision est comme un jugement au tribunal (hypothèse = le prévenu est présumé innocent). Il y a deux risques au jugement : celui de condamner un innocent (rejet à tort de l'hypothèse – première espèce), ou d'innocenter un coupable (acceptation à tort de l'hypothèse – seconde espèce).




Plus on réduit un risque, plus on augmente l'autre : ainsi, la décision que l'on doit prendre est un compromis adapté à la situation.
Voilà pourquoi le seuil de 5 % est souvent utilisé.

## → TESTER LA CONFORMITÉ D'UN ÉCHANTILLON PAR RAPPORT À LA POPULATION ←

On considère un caractère dont la proportion dans la population est connue, égale à $p$.

La prise de décision consiste, à partir d'un échantillon de taille $n$, à valider ou non sa représentativité.

**1.** On calcule la fréquence observée $f$ du caractère dans cet échantillon.

**2.** On choisit le risque α, puis on détermine un intervalle de fluctuation au seuil $1-\alpha$, noté I.

**3.** On applique la règle suivante :

→ Si $f\notin \mathrm{I}$ alors on considère que l'échantillon n'est pas représentatif de la population.

→ Si $f\in \mathrm{I}$ alors on considère que l'échantillon est représentatif de la population.

*Attention : ici, on ne fait pas d'hypothèse sur la probabilité théorique p, puisqu'on la connaît.*

*On ne commet dont aucune erreur.*