

# Chapitre :

# Fluctuation



## I. Retour sur la loi normale

---

**Propriété** | Soit  $X$  une variable aléatoire qui suit la loi normale centrée réduite  $\mathcal{N}(0; 1)$ . Pour tout réel  $\alpha \in ]0; 1[$ , il existe un unique réel  $t_\alpha$  tel que  $\mathbb{P}(-t_\alpha \leq X \leq t_\alpha) = 1 - \alpha$ .

**Démonstration (exigible)** : On considère  $f$ , la densité de probabilité de la loi  $\mathcal{N}(0; 1)$ .

Soit  $H : t \mapsto \int_0^t f(x)dx$ . Alors (par symétrie)  $\mathbb{P}(-t \leq X \leq t) = 2\mathbb{P}(0 \leq X \leq t) = 2H(t)$ .

La fonction  $H$  est la primitive de  $f$  sur  $\mathbb{R}$  qui s'annule en 0.

La fonction  $f$  étant positive sur  $[0; +\infty[$ ,  $H$  est donc continue et strictement croissante sur  $[0; +\infty[$ .

Or, par symétrie de la courbe,  $\lim_{t \rightarrow +\infty} H(t) = \mathbb{P}(X \geq 0) = \frac{1}{2}$ .

Par suite, la fonction  $2H$  est continue, strictement croissante sur  $[0; +\infty[$ , et ses valeurs sont comprises entre  $2 \times 0 = 0$  et  $2 \times \frac{1}{2} = 1$ .

Soit  $\alpha \in ]0; 1[$ . Alors  $1 - \alpha$  appartient aussi à  $]0; 1[$ . D'après le corollaire du théorème des valeurs intermédiaires, il existe donc un unique réel  $t_\alpha$  strictement positif tel que  $2H(t_\alpha) = 1 - \alpha$ , autrement dit tel que  $\mathbb{P}(-t_\alpha \leq X \leq t_\alpha) = 1 - \alpha$ .  $\square$

**Propriété** | Avec les notations précédentes,

- Pour  $\alpha = \frac{5}{100} = 0,05$ , alors  $t_\alpha \simeq 1,96$
- Pour  $\alpha = \frac{1}{100} = 0,01$ , alors  $t_\alpha \simeq 2,58$

Autrement dit :

$$\mathbb{P}(-1,96 \leq X \leq 1,96) \simeq 0,95 \quad \text{et} \quad \mathbb{P}(-2,58 \leq X \leq 2,58) \simeq 0,99$$

**Démonstration** : Exercice

Il s'agit de se ramener à la recherche de  $x$  tel que  $\mathbb{P}(X \leq x) = p$  ( $p$  étant donné), dont la calculatrice donne une valeur approchée.  $\square$

**Remarque** Parfois on utilise la notation  $u_\alpha$  à la place de  $t_\alpha$ .

## II. Intervalle de fluctuation asymptotique

---

⊗ **Activité** : 1p356, en groupe, en salle informatique, en utilisant le fichier de tableur associé.

Soit, pour tout  $n \in \mathbb{N}$ ,  $X_n$  une variable aléatoire qui suit la loi binomiale  $\mathcal{B}(n, p)$ . On pose alors

$F_n = \frac{X_n}{n}$ , fréquence de succès.

**Propriété** Avec les notations de la section précédente,

pour tout  $\alpha \in ]0; 1[$ , on définit l'intervalle  $I_n = \left[ p - t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$ .

Alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(F_n \in I_n) = 1 - \alpha$$

**Démonstration (exigible) :** On réécrit :

$$\begin{aligned} F_n \in I_n &\Leftrightarrow p - t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \\ &\Leftrightarrow np - t_\alpha \frac{n\sqrt{p(1-p)}}{\sqrt{n}} \leq X_n \leq np + t_\alpha \frac{n\sqrt{p(1-p)}}{\sqrt{n}} \\ &\Leftrightarrow np - t_\alpha \sqrt{np(1-p)} \leq X_n \leq np + t_\alpha \sqrt{np(1-p)} \\ &\Leftrightarrow -t_\alpha \leq Z_n \leq t_\alpha \text{ en posant } Z_n = \frac{X_n - np}{\sqrt{np(1-p)}} \end{aligned}$$

Ainsi,  $\mathbb{P}(F_n \in I_n) = \mathbb{P}(-t_\alpha \leq Z_n \leq t_\alpha)$ .

Or, d'après le théorème de Moivre-Laplace,  $\lim_{n \rightarrow +\infty} \mathbb{P}(-t_\alpha \leq Z_n \leq t_\alpha) = \int_{-t_\alpha}^{t_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ .

Or  $\int_{-t_\alpha}^{t_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \mathbb{P}(-t_\alpha \leq Y \leq t_\alpha)$  où  $Y$  suit la loi normale centrée réduite  $\mathcal{N}(0; 1)$ .

Ainsi on a bien

$$\lim_{n \rightarrow +\infty} \mathbb{P}(F_n \in I_n) = 1 - \alpha$$

□

**Définition** L'intervalle  $I_n = \left[ p - t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + t_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$  est un **intervalle de fluctuation asymptotique au seuil de confiance  $1 - \alpha$**  de la variable aléatoire fréquence  $F_n$  qui, à tout échantillon de taille  $n$ , associe la fréquence obtenue.

Cet intervalle contient  $F_n$  avec une probabilité d'autant plus proche de  $1 - \alpha$  que  $n$  est grand.

On considère que c'est le cas lorsque  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ .

On considère en général l'intervalle asymptotique au seuil de confiance 95%, donc avec  $\alpha = 0,05$ .

Il s'agit de l'intervalle :

$$\left[ p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

- **Exercice** : 1p364 (vérification des conditions)
- **Exercices** : 2,4p364, 15,16p365 (seuil 95%)
- **Exercices** : 5,6p364, 19p365 (seuil 99%)
- **Exercices** : 20,21p365 (autres seuils)
- **Exercices** : 23,24p366 (algorithmes)

# III. Prise de décision

---

⊗ **Activité** : 2p356 (avec vrai intervalle pour la loi binomiale)

Pour décider si une fréquence  $f$  observée sur un échantillon de taille  $n$  est compatible ou non avec une proportion  $p$  donnée pour la population totale, on teste l'appartenance de  $f$  à un intervalle de fluctuation (asymptotique ou non) au seuil de 95%. Par suite,

- Si  $f$  n'est pas dans l'intervalle de fluctuation, alors on peut rejeter l'hypothèse que l'échantillon est compatible avec le modèle.
- Si  $f$  est dans l'intervalle de fluctuation, alors on accepte l'hypothèse que l'échantillon est compatible avec le modèle.

**Remarque** Quelle que soit la décision prise il y a toujours le risque que ce ne soit pas la bonne décision.

 Pour utiliser l'intervalle de fluctuation asymptotique, il faut que les conditions soient vérifiées!

**Remarque** On peut prendre des décisions avec des seuils différents (voir page 360).

► **Exercices** : 8p364, 27p366

► **Exercices** : 30p366, 31p367 (conditions non vérifiées)

# IV. Estimation

---

L'intervalle de fluctuation s'utilise lorsque l'on connaît la proportion  $p$  dans la population totale. On s'intéresse ici au problème « inverse », c'est à dire qu'à partir d'une fréquence  $f$  observée, on veut pouvoir estimer la proportion  $p$ . C'est le problème posé pour les sondages par exemple.

**Propriété** Soit, pour tout  $n \in \mathbb{N}$ ,  $X_n$  une variable aléatoire qui suit la loi  $\mathcal{B}(n, p)$ , avec  $p \in ]0; 1[$ . Alors il existe un entier  $n_0$  tel que pour tout  $n \geq n_0$ ,

$$\mathbb{P} \left( p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}} \right) > 0,95$$

**Démonstration** : (davantage détaillée page 362). On pose  $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ .

D'après le théorème de Moivre-Laplace, en posant  $a_n = \mathbb{P}(-2 \leq Z_n \leq 2)$ , on a  $\lim_{n \rightarrow +\infty} a_n = \mathbb{P}(-2 \leq Z \leq 2)$ , où  $Z$  suit la loi normale centrée réduite  $\mathcal{N}(0; 1)$ .

Notons  $L = \mathbb{P}(-2 \leq Z \leq 2)$ . On peut obtenir  $L \leq 0,9544$ , soit  $L > 0,95$ .

Soit  $\epsilon > 0$  tel que  $\epsilon < 0,004$ . Puisque la suite  $(a_n)_{n \in \mathbb{N}}$  a pour limite  $L$ , il existe un entier  $n_0$  tel que pour tout  $n \geq n_0$ ,  $a_n \in ]L - \epsilon; L + \epsilon[$ , autrement dit  $L - \epsilon < a_n < L + \epsilon$ .

En particulier,  $a_n > 0,95$ .

$$\text{Or, } a_n = \mathbb{P}(-2 \leq Z_n \leq 2) = \mathbb{P} \left( p - 2 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + 2 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right)$$

La fonction  $\phi : p \mapsto p(1-p)$ , polynomiale de degré 2, admet un maximum en  $\frac{1}{2}$ , qui vaut  $\frac{1}{4}$ .

Ainsi,  $p(1-p) \leq \frac{1}{4}$ , d'où  $\sqrt{p(1-p)} \leq \frac{1}{2}$  puis  $2 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}$ .

Par conséquent,  $\left[ p - 2 \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 2 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right] \subset \left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$ .

Finalement,  $\mathbb{P} \left( p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}} \right) \geq a_n > 0,95$ . □

**Propriété** Avec les notations précédentes, on pose  $F_n = \frac{X_n}{n}$ .

Alors l'intervalle  $\left[ F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$  contient, pour  $n$  assez grand, la proportion  $p$  avec une probabilité au moins égale à 0,95.

**Démonstration** : Il suffit d'appliquer la propriété précédente, en observant que

$$F_n \in \left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] \Leftrightarrow p \in \left[ F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$$

□

**Définition** Soit  $f$  la fréquence observée d'un caractère dans un échantillon de taille  $n$  extrait d'une population dans laquelle la proportion de ce caractère est  $p$ .

Alors l'intervalle  $\left[ f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}} \right]$  est un **intervalle de confiance de la proportion  $p$  au niveau de confiance 95%**.

On considère que les conditions pour utiliser cet intervalle sont :  $n \geq 30$ ,  $nf \geq 5$  et  $n(1-f) \geq 5$ , soit les mêmes que pour l'intervalle de fluctuation en remplaçant  $p$  par  $f$ .

**Remarque** Il s'agit de l'intervalle de confiance qui a pu être vu en seconde.

**Remarque** Il existe d'autres intervalles de confiance, comme :

$$\left[ f - 1,96 \sqrt{\frac{f(1-f)}{n}}; f + 1,96 \sqrt{\frac{f(1-f)}{n}} \right]$$

qu'il n'est pas possible de justifier en terminale S.

► **Exercices** : 10,12p364 et 35,39p367

► **Exercices** : 43,44,45,46p368 (amplitude de l'intervalle, détermination de  $n$ )