

Dans cette fiche,

- X_n désigne une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$ où $n \in \mathbb{N}^*$ et $p \in]0, 1[$.
- $F_n = \frac{X_n}{n}$ est la variable aléatoire fréquence associée à X_n .
- $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ est la variable centrée et réduite associée à la variable X_n .

I. Rappels.

1) **Théorème de MOIVRE-LAPLACE.** Pour tous réels a et b tels que $a < b$

$$\lim_{n \rightarrow +\infty} P(a \leq Z_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

ou aussi

$$\lim_{n \rightarrow +\infty} P\left(p + a \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + b \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

2) **Intervalle associé à une probabilité pour $\mathcal{N}(0, 1)$**

Théorème. Soit X une variable aléatoire régie par la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Pour tout réel $\alpha \in]0, 1[$, il existe un réel strictement positif u_α et un seul tel que $P(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$.

On doit connaître en particulier $u_{0,05} = 1,96$ et $u_{0,01} = 2,58$ à 10^{-2} près par défaut.

3) **Conséquences.** Si on combine les résultats précédents, on obtient le résultat suivant :

Pour tout entier naturel non nul n , on considère une variable aléatoire X_n qui suit une loi binomiale $\mathcal{B}(n, p)$ puis

on considère la variable aléatoire fréquence $F_n = \frac{X_n}{n}$. Alors :

$$\lim_{n \rightarrow +\infty} P\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha,$$

et en particulier,

$$\lim_{n \rightarrow +\infty} P\left(p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \geq 0,95.$$

II. Intervalle de fluctuation. Echantillonnage

Dans ce paragraphe, on étudie un caractère C d'une population et on suppose que ce caractère apparaît dans la population avec une probabilité p . On extrait un échantillon et on veut avoir une idée de la fréquence f d'apparition du caractère dans l'échantillon. La variable aléatoire qui à un échantillon de taille n associe la fréquence f d'apparition du caractère C dans l'échantillon est la variable F_n .

Situation où on utilise un intervalle de fluctuation :

Quand on connaît la probabilité p (ou quand on fait une hypothèse sur la valeur de p)
et que l'on veut estimer la fréquence f .

A - Intervalle de fluctuation

1) **L'intervalle de la classe de terminale.**

L'intervalle $I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$ est appelé **intervalle de fluctuation asymptotique** au seuil $1 - \alpha$:

$$\lim_{n \rightarrow +\infty} P(F_n \in I_n) = \lim_{n \rightarrow +\infty} P\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha.$$

En particulier, l'intervalle $J_n = \left[p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$ est un **intervalle de fluctuation asymptotique au seuil 0,95** :

$$\lim_{n \rightarrow +\infty} P(F_n \in J_n) = \lim_{n \rightarrow +\infty} P \left(p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right) \geq 0,95.$$

Conséquence. Pour n grand, la probabilité de l'événement « $p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}$ » vaut environ 0,95. Dans la pratique, on fait cette approximation quand $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$: sous ces conditions, la fréquence f a environ 95% de chances (mais pas au moins 95% de chances) d'être dans l'intervalle J_n .

Théorème. Pour tout $p \in]0, 1[$, il existe $n_0 \in \mathbb{N}^*$ tel que

$$\text{pour tout } n \geq n_0, P \left(p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right) \geq 0,95.$$

2) L'intervalle de la classe de seconde.

Théorème. Pour tout réel $p \in]0, 1[$ et tout entier naturel non nul n , l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ contient l'intervalle $\left[p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$.

Théorème. Pour tout $p \in]0, 1[$, il existe $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$, $P \left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \right) \geq 0,95$.

Théorème. Pour tout $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, $P \left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \right)$ vaut environ 0,95.

B - Prise de décision

On cherche à savoir si la probabilité p d'apparition du caractère C dans la population est égale à un certain nombre p_0 ou pas à partir d'un échantillon de taille n . On fait donc l'hypothèse que $p = p_0$.

- On vérifie d'abord que $n \geq 30$, $np_0 \geq 5$ et $n(1-p_0) \geq 5$.

- On calcule l'intervalle $I = \left[p_0 - 1,96 \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}}, p_0 + 1,96 \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} \right]$.

- On détermine la fréquence f du caractère C dans l'échantillon.

- Si $f \notin I$, on rejette l'hypothèse $p = p_0$ au risque de se tromper d'au plus 5%. Si $f \in I$, on ne peut pas rejeter l'hypothèse $p = p_0$.

III. Intervalle de confiance.

Situation où on utilise un intervalle de confiance :
 Quand on connaît la fréquence f et que l'on veut estimer la probabilité p .

On rappelle que pour n grand, $P \left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right] \right) \geq 0,95$. Puisque

$$p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \Leftrightarrow F_n - \frac{1}{\sqrt{n}} \leq p \leq F_n + \frac{1}{\sqrt{n}} :$$

Théorème.

Pour tout réel $p \in]0, 1[$, il existe un entier $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$, $P \left(p \in \left[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}} \right] \right) \geq 0,95$.

Si on choisit explicitement un échantillon et que le caractère étudié apparaît dans cet échantillon avec une fréquence f , l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}} \right]$ est une réalisation de l'intervalle aléatoire $\left[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}} \right]$.

Quand n est grand, dans au moins 95% des choix d'échantillon, on aura $p \in \left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}} \right]$.

On dit que l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}} \right]$ est un intervalle de confiance de p au niveau de confiance 0,95 (pour n grand). Dans la pratique, on utilise ce résultat quand $n \geq 30$, $nf \geq 5$ et $n(1-f) \geq 5$.