Théorème central limite

1) <u>approche intuitive</u>

Supposons que nous souhaitions estimer la taille moyenne des individus d'une population de grand effectif (1 million d'individus pour fixer les idées) dans laquelle la taille moyenne est $\mu = 175\ cm$ et l'écart-type $\sigma = 4\ cm$.

On procède alors à la mesure des tailles des individus sur un échantillon de taille $n = 1000$ par exemple et on considère l'échantillon comme issu d' un tirage aléatoire avec remise.

Si on note $x_1, x_2, \dots, x_n$ les valeurs des tailles des individus de l'échantillon, elles sont la réalisation de variables aléatoires $X_1, X_2, \dots, X_n$ mutuellement indépendantes de même loi que la variable X associée au tirage aléatoire d'un individu de la population et qui en mesure la taille, d'espérance μ et d'écart-type σ.

On peut alors calculer la moyenne de l'échantillon qui est :

$$\boxed{\boldsymbol{\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}}}$$

Elle est la réalisation d'une variable aléatoire appelée moyenne empirique :

$$\boxed{\boldsymbol{\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}}}$$

d'espérance :

$$\boxed{\boldsymbol{E(\bar{X}_n) = \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} = \frac{n \times E(X)}{n} = E(X) = \mu}}$$

de variance :

$$\boxed{\boldsymbol{V(\bar{X}_n) = \frac{V(X_1) + V(X_2) + \cdots + V(X_n)}{n^2} = \frac{n \times V(X)}{n^2} = \frac{V(X)}{n} = \frac{\sigma^2}{n}}}$$

d'écart-type :

$$\boxed{\boldsymbol{\sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}}}$$

Intuitivement, on peut imaginer que si on procède à un très grand nombre d'échantillonnages, $\bar{X}_n$ va se distribuer de façon symétrique autour de la moyenne μ avec une distribution en forme de loi normale. Le fait de procéder à une moyenne dans un échantillon crée un effet de lissage. Deux personnes avec des valeurs extrêmes contraires, 155 cm et 195 cm par exemple contribuent en s'additionnant pour le calcul de la moyenne comme deux personnes qui aurait la taille moyenne de 175 cm.

Ainsi, intuitivement, on forme l'idée que pour une taille d'échantillon n suffisamment grande (en pratique on considère $n = 30$, $\bar{X}_n$ se distribue selon une loi normale d'espérance μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$.

On peut alors faire intervenir la loi normale centrée réduite en introduisant la variable :

$$\boldsymbol{Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\ \frac{\overline{X}_n - \mu}{\sigma}}$$

Cette variable devient alors pour n suffisamment grand distribuée selon une loi normale centrée réduite. C'est ce qu'on appelle le théorème central limite dont la formulation mathématique précise suit.

2) **Théorème central limite**

Soit $(X_n)_{n\in\mathbb{N}^*}$ une famille de variables aléatoires réelles mutuellement indépendantes et de même loi d'espérance μ et d'écart-type σ. On définit la variable

$$\boldsymbol{\overline{X}_n = \sum_{i=1}^{n} X_i}$$

et la variable :

$$\boldsymbol{Z_n = \sqrt{n}\ \frac{\overline{X}_n - \mu}{\sigma}}$$

Alors Z_n tend en loi vers une loi normale centrée réduite (*)

(*) Rappel : cela signifie qu'en désignant par U une variable distribuée selon une loi normale centrée réduite :

$$\forall\, \boldsymbol{x} \in \mathbb{R} :\ \lim_{n\to+\infty} \boldsymbol{P(Z_n < x) = P(U < x)}$$

En particulier :

$$\forall\, (\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^2 :\ \lim_{n\to+\infty} \boldsymbol{P(a \le Z_n \le b) = P(a \le U \le b)}$$

$$\forall\, (\boldsymbol{c}, \boldsymbol{r}) \in \mathbb{R}^2 :\ \lim_{n\to+\infty} \boldsymbol{P(|Z_n - c| > r) = P(|U - c| > r)}$$

<u>Conséquence</u> : La variable $\overline{X}_n$ tend en probabilité vers μ, ce qui signifie :

$$\forall\, \boldsymbol{\varepsilon > 0} :\ \lim_{n\to+\infty} \boldsymbol{P(|\overline{X}_n - \mu| > \varepsilon) = 0}$$

<u>Preuve de la conséquence</u> :

Soit $\varepsilon > 0$ et $\alpha > 0$

$$P(|\bar{X}_n - \mu| > \varepsilon) = P\left(\left|\sqrt{n}\ \frac{\bar{X}_n - \mu}{\sigma}\right| > \sqrt{n}\ \frac{\varepsilon}{\sigma}\right)$$

$$= P\left(|Z_n| > \sqrt{n}\ \frac{\varepsilon}{\sigma}\right)$$

Or il existe un entier naturel n_0 tel que :

$$P\left(|U| > \sqrt{n_0}\ \frac{\varepsilon}{\sigma}\right) < \frac{\alpha}{2}$$

On a alors pour $n > n_0$, d'une part :

$$P\left(|Z_n| > \sqrt{n}\ \frac{\varepsilon}{\sigma}\right) \leq P\left(|Z_n| > \sqrt{n_0}\ \frac{\varepsilon}{\sigma}\right)$$

d'autre part :

$$\lim_{n\to+\infty} P\left(|Z_n| > \sqrt{n_0}\ \frac{\varepsilon}{\sigma}\right) = P\left(|U| > \sqrt{n_0}\ \frac{\varepsilon}{\sigma}\right) < \frac{\alpha}{2}$$

Donc il existe un entier naturel $n_1 > n_0$ tel que pour $n > n_1$:

$$P(|\bar{X}_n - \mu| > \varepsilon) = P\left(|Z_n| > \sqrt{n}\ \frac{\varepsilon}{\sigma}\right) \leq P\left(|Z_n| > \sqrt{n_0}\ \frac{\varepsilon}{\sigma}\right) < \alpha$$

Ainsi :

$$\lim_{n\to+\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$