

Dans la pratique, on ne peut pas mesurer avec exactitude la proportion d'un caractère dans la population (intention de vote pour un candidat lors d'une élection - nombre de personne ayant regardé une émission de télévision samedi soir...). Pour réaliser ce type de mesure, les statisticiens mesurent le caractère en question sur une petite partie de la population (suffisamment grande **un échantillon**) et en mesurer les fréquences. On parle de **fluctuation d'échantillonnage**.

a. Intervalle de fluctuation

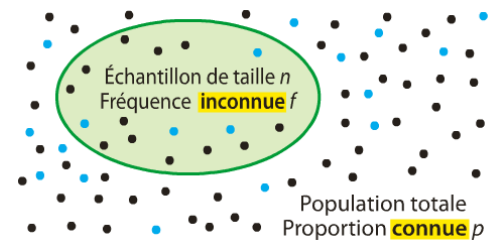
Lors de la simulation d'une expérience aléatoire sur ordinateur, on observe que les fréquences des échantillons créés fluctuent autour de la probabilité.

Définition : (Lorsqu'on connaît la proportion du caractère dans la population p)

On définit l'**intervalle de fluctuation** au seuil de 95% pour des échantillons de taille $n \geq 25$ et pour p compris entre 0,2 et 0,8 par :

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

On admet que pour de tels échantillons la fréquence d'apparition f observé appartient à l'intervalle de fluctuation avec une probabilité d'au moins 0,95.



b. Intervalle de confiance

Définition : (Lorsqu'on **ne connaît pas** la proportion du caractère dans la population p)

On définit l'**intervalle de confiance** au seuil de 95% pour des échantillons de taille $n \geq 25$ et un caractère qui apparaît avec une fréquence f comprise entre 0,2 et 0,8 par :

$$\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right].$$

On peut estimer que la proportion p du caractère dans la population totale est dans l'intervalle de confiance avec une probabilité d'au moins 0,95.

ATTENTION : dans 5% des cas, l'intervalle de confiance ne contient peut-être pas p .

On parlera quelques fois de « preuve statistique ».