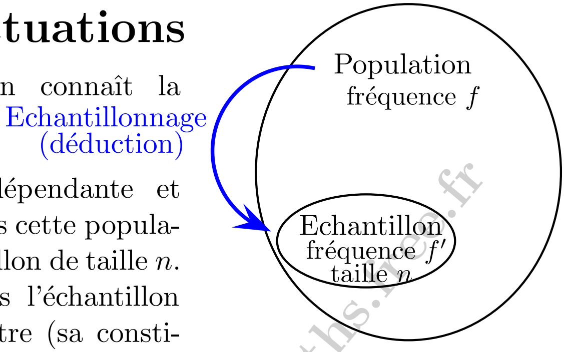
Fluctuations



Dans une population donnée, on connaît la fréquence f d'un caractère.

On répète n fois, de façon indépendante et aléatoire, le choix d'un individu dans cette population de façon à constituer un échantillon de taille n. La fréquence f' du caractère dans l'échantillon peut varier d'un échantillon à l'autre (sa constitution est aléatoire).

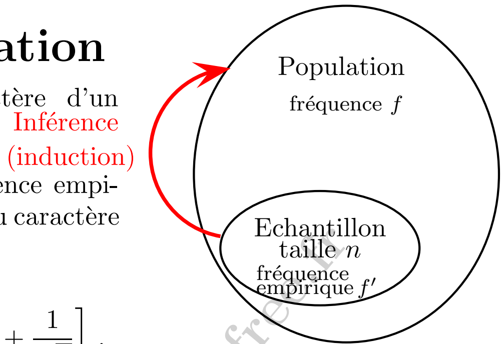
Propriété *L'intervalle de fluctuation au seuil de 95% est environ :*

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

Cet intervalle signifie que dans 95 % des cas (ou avec une probabilité de 0,95) la fréquence observée dans un échantillon prélevé au hasard dans la population sera dans l'intervalle I.

Si pour un échantillon donné la fréquence observée n'appartient pas à I, alors on peut remettre en cause, avec un risque d'erreur de 5 % les hypothèses mises en jeu : l'échantillon est constitué aléatoirement, la fréquence dans la population complète est f,...

Estimation



On connaît la fréquence f' d'un caractère d'un échantillon aléatoire de la population.

A partir de la connaissance de cette fréquence empirique f', on souhaite estimer la fréquence f du caractère dans la population complète.

Propriété *L'intervalle* $I_n=\left[f'-\dfrac{1}{\sqrt{n}}\,;\,f'+\dfrac{1}{\sqrt{n}}\right]$.
contient la fréquence f du caractère dans la population avec une probabilité supérieure ou égale à 0,95.
I_n est l'intervalle au niveau de confiance de 95 %.

Dimensionnement des échantillons (ou sondages) La largeur de l'intervalle de confiance, ou précision de l'estimation est $\dfrac{2}{\sqrt{n}}$.

Ainsi, pour obtenir une estimation (toujours avec un niveau de confiance de 95 %) de la fréquence f avec une précision ε donnée, on doit donc avoir $\dfrac{2}{\sqrt{n}}\leqslant\varepsilon \iff n\geqslant\left(\dfrac{2}{\varepsilon}\right)^2$