



## Chapitre XI – Échantillonnage et estimation

Bacomathiques — <https://bacomathiqu.es>

### TABLE DES MATIÈRES

<b>I - Définitions</b>	<b>1</b>
<b>II - Théorème de Moivre-Laplace</b>	<b>1</b>
<b>III - Intervalles de fluctuation</b>	<b>2</b>
<b>IV - Intervalles de confiance</b>	<b>4</b>

## I - Définitions

Lorsque l'on travaille sur une population de grande taille, il est rarement possible d'avoir accès à toutes les données de chacun des individus. C'est pourquoi on prélève **un échantillon** de cette population : c'est **l'échantillonnage**.

### À RETENIR 🔦

Un échantillon de taille  $n$  représente  $n$  individus choisis au hasard dans une population.

Il existe deux manières de réaliser un échantillonnage : sans remise (on prélève  $n$  individus différents) et avec remise (il est possible de prélever plusieurs fois le même individu).

## II - Théorème de Moivre-Laplace

Soient  $n \in \mathbb{N}$ ,  $p \in [0; 1]$  et  $X_n$  une suite de variables aléatoires qui suivent la loi binomiale de paramètres  $n$  et  $p$  (voir chapitre précédent). On définit alors la variable aléatoire  $Z_n$  :

### À RETENIR 🔦

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

Soient  $a$  et  $b \in \mathbb{R}$  tels que  $a < b$ , on a :

### À RETENIR 🔦

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

### À LIRE 📖

Cela signifie que si on est dans les bonnes conditions d'approximation ( $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ ), alors on peut avoir une bonne approximation de la variable aléatoire  $X_n$  (i.e. la loi binomiale de paramètre  $n$  et  $p$ ) avec la loi normale de paramètres  $(np; np(1-p))$ .

### III - Intervalles de fluctuation

Soient  $Z$  une variable aléatoire suivant la loi normale centrée réduite,  $\alpha$  un réel et  $u_\alpha$  un réel positif vérifiant  $p(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$ . On se donne également une variable aléatoire  $X_n$  suivant une loi binomiale  $B(n; p)$  et on pose  $I_n$  l'**intervalle de fluctuation asymptotique** au seuil  $1 - \alpha$  :

À RETENIR

$$I_n = \left[ p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

On donne les conditions suivantes qui doivent être satisfaites :

À RETENIR

- $n \geq 30$
- $np \geq 5$
- $n(1-p) \geq 5$

En particulier, pour  $\alpha = 0,05$ , un intervalle de fluctuation au seuil de 95% de la fréquence d'apparition d'un caractère dans un échantillon aléatoire de taille  $n$  est :

À RETENIR

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

Cet intervalle  $J_n$  sera celui qui sera privilégié en classe de Terminale.

## À LIRE

**Exemple :** Dans un lac dans lequel ne sont présents que deux types de poisson (truites et saumons), un groupe de pêcheurs réussit à attraper 50 poissons dans une journée. On estime qu'il y a environ 40 truites et 10 saumons.

Ils prélèvent au hasard 30 poissons de leur prise totale. Quel est l'intervalle de fluctuation asymptotique au seuil de 95% de la fréquence de saumon ?

**Résolution :** On a 50 poissons, la proportion de saumons est  $p = \frac{10}{50} = 0,2$ . La taille de l'échantillon est  $n = 30$ .

On a bien  $n = 30 \geq 30$ ,  $n \times p = 30 \times 0,2 = 6 \geq 5$  et  $n \times (1 - p) = 30 \times (1 - 0,2) = 24 \geq 5$ .

Voici donc l'intervalle de fluctuation asymptotique au seuil de 95% :

$$I = \left[ 0,2 - 1,96 \frac{\sqrt{0,2 \times 0,8}}{\sqrt{30}}; 0,2 + 1,96 \frac{\sqrt{0,2 \times 0,8}}{\sqrt{30}} \right] \approx [0,057; 0,343]$$

Ainsi, cela signifie que la fréquence de saumons a 95% de chances de se situer dans l'intervalle  $I$ .

Ce type d'intervalle peut servir à prendre des décisions. En effet, soit  $I$  un intervalle de fluctuation asymptotique au seuil de 95%. On souhaite avoir une certaine fréquence  $f$  d'un certain caractère. On peut dire qu'il est impossible d'avoir ce caractère si  $f \notin I$  et qu'il possible d'avoir ce caractère si  $f \in I$  avec toujours 5% de chances de se tromper.

## IV - Intervalles de confiance

Soient une expérience de Bernoulli dont on veut estimer la probabilité de succès  $p$  et  $f_n$  la fréquence de succès après  $n$  répétitions indépendantes de l'épreuve. Alors  $p$  a 95% de chances d'appartenir à l'**intervalle de confiance**  $I_C$  suivant :

À RETENIR 🔦

$$I_C = \left[ f_n - \frac{1}{\sqrt{n}}; f_n + \frac{1}{\sqrt{n}} \right]$$

On donne les conditions suivantes qui doivent être satisfaites :

À RETENIR 🔦

- $n \geq 30$
- $nf_n \geq 5$
- $n(1 - f_n) \geq 5$

À LIRE 📖

**Exemple :** On dispose d'un paquet de 52 cartes. On les prend une par une et on les retourne jusqu'à ce qu'il ne reste plus que 22 cartes dans le paquet (on a donc tiré 30 cartes en tout).

On obtient 18 cartes rouges et 12 cartes noires. Dans quel intervalle de confiance au seuil de 95% se situe la probabilité  $p$  de tirer une carte rouge ?

**Résolution :** La taille de l'échantillon est  $n = 30$ . On a 18 cartes rouges, la fréquence observée de cartes rouges est donc  $f_n = \frac{18}{30} = 0,6$ .

On a bien  $n = 30 \geq 30$ ,  $n \times f_n = 30 \times 0,6 = 18 \geq 5$  et  $n \times (1 - f_n) = 30 \times (1 - 0,6) = 12 \geq 5$ .

La probabilité  $p$  de tirer une carte rouge se situe donc dans l'intervalle  $I_C$  avec une marge d'erreur de 5

$$I_C = \left[ 0,6 - \frac{1}{\sqrt{30}}; 0,6 + \frac{1}{\sqrt{30}} \right] \approx [0,417; 0,783]$$

**Remarque :** Dans un jeu de cartes classique, on a autant de chances de tirer une carte rouge que de tirer une carte noire. La "vraie" probabilité est donc de 0,5. Notre estimation est donc bonne car  $0,5 \in I_C$ .