INFERENCE STATISTIQUE

Inférence statistique : Echantillon $\rightarrow$ Population

Hypothèse de l'échantillonnage : Les observations $x_1, \ldots, x_n$ sont considérées comme des réalisations de $X_1, \ldots, X_n$ v.a.i.i.d.

Principe de l'échantillonnage : Etudier les propriétés et les caractéristiques de la suite (X_n), notamment quand n grand

Définition : Une statistique T est une variable aléatoire fonction de $X_1, \ldots, X_n$,

$$T = f(X_1, \ldots, X_n)$$

Exemples : moyenne, fonction répartition empirique,…

FONCTION DE REPARTITION EMPIRIQUE

Soit $F_n(t)$ la proportion des n v.a. $X_1, \ldots, X_n$ qui sont inférieures à t. Pour tout t, $F_n(t)$ est une v.a., donc la fonction

$$\boxed{t \mapsto F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{x_i \leq t\}}}$$

est une fonction aléatoire appelée fonction de répartition empirique de l'échantillon dont les réalisations sont des fonctions en escalier de sauts égaux à 1/n

Propriété : Pour tout t,

$$F_n(t) \underset{p.s.}{\rightarrow} F(t)$$

où F est la fonction de répartition des X_i

Exemple :

FREQUENCE EMPIRIQUE

On considère un événement A tel que P(A)=p.

Sur un échantillon considérons les v.a. $X_1,\ldots,X_n$ telles que $X_i=1$ si l'événement A est réalisé à la $i^{ème}$ expérience ($i^{ème}$ individu) et 0 sinon. Alors la fréquence empirique de l'événement A est définie par

$$\boxed{f_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{K}{n}}$$

où K représente le nombre de fois où l'événement A s'est réalisé au cours des n expériences.

Propriété : $E(f_n) = p$ et $\mathrm{var}(f_n) = \dfrac{p(1-p)}{n}$

Propriété : $f_n \underset{p.s}{\longrightarrow} p$ et $f_n \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}\left(p, \dfrac{p(1-p)}{n}\right)$

Exemple

MOYENNE

Supposons que $E(X_i)=\mu$ et $var(X_i)=\sigma^2$

La moyenne (empirique) de l'échantillon est la v.a. définie par :

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

<u>Propriété</u> : $E(\overline{X}) = \mu$ et $var(\overline{X}) = \dfrac{\sigma^2}{n}$

<u>Propriété</u> : $\overline{X} \underset{p.s}{\longrightarrow} \mu$ et $\overline{X} \underset{\mathcal{L}}{\longrightarrow} \mathcal{N}(\mu, \dfrac{\sigma^2}{n})$

VARIANCE EMPIRIQUE

Supposons que $E(X_i)=\mu$ et $var(X_i)=\sigma^2$

La variance empirique de l'échantillon est la v.a. définie par :

$$\boxed{S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Propriété : $E(S^2) = \frac{n-1}{n}\sigma^2 \Rightarrow \boxed{S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

où $E(S^{*2})= \sigma^2$

Propriété : $S^2 \xrightarrow[p.s]{} \sigma^2$

Propriété : $cov(\overline{X}, S^2) = \frac{n-1}{n^2}\mu_3 \Rightarrow$ asymptotiquement non corrélés

RESUME

Echantillon	Population
Fonction de répartition empirique $t \mapsto F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{x_i \leq t\}}$	Fonction de répartition $t \mapsto F(t) = P(X \leq t)$
Fréquence $f = \frac{K}{n}$	Probabilité p
Moyenne $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$	Espérance $E(X)$
Variance empirique $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ $S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$	Variance $\mathrm{var}(X)$

CAS D'ECHANTILLON GAUSSIEN

Supposons que les X_i suivent une loi $\mathcal{N}(\mu,\sigma^2)$

Propriété : $\overline{X}$ suit une loi $\mathcal{N}(\mu,\frac{\sigma^2}{n})$

$$\Leftrightarrow \quad \sqrt{n}\,\frac{\overline{X}-\mu}{\sigma} \text{ suit une loi } \mathcal{N}(0,1)$$

Propriété : $\frac{nS^2}{\sigma^2}$ suit une loi χ^2_{n-1} ⇒ ***c.f. cours***

$$\Leftrightarrow \quad \frac{(n-1)S^{*2}}{\sigma^2} \text{ suit une loi } \chi^2_{n-1}$$

Propriété : $\frac{\overline{X}-\mu}{S}\sqrt{n-1}$ suit une loi t_{n-1} ⇒ ***c.f. cours***

Propriété : $\overline{X}$ et S^2 sont indépendantes

Exemple