Chap : Introduction : Echantillonnage

Données : $X_1, \dots, X_n$: iid (indép de m loi) ⇒ échantillon

→ But : estimer la loi de X

- Types de statistiques :

① Stat paramétrique

→ Suppose que seulement un ou des paramètres $\theta \in \Theta \subset \mathbb{R}^d$ qui sont inconnus de la loi de X [illegible] (loi est de type connu).

(Par exemple la loi de X Normale $\mathcal{N}(\theta)$ où θ inconnu)

② Stat non paramétrique

Toute la loi est inconnue (à estimer)

Fonction de vraisemblance

= loi de $(X_1, \dots, X_n)$

$\mathcal{L}(x_1, \dots, x_n)$ [illegible] $(x_1, \dots, x_n) \overset{\text{si dép}}{=} \begin{cases} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n), & \text{si la loi est discrète} \\ f_{(X_1, \dots, X_n)}(x_1, \dots, x_n), & \text{si loi continue} \end{cases}$

↳ likelihood

$$\overset{\text{si indép}}{=} \begin{cases} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\ \prod_{i=1}^{n} f_{X_i}(x_i) \end{cases}$$

Log-vraisemblance : $\log \mathcal{L}(x_1, \dots, x_n)$

ex : $X_i \sim \mathcal{B}(\theta)$, $\theta \in]0,1[$

$$\mathcal{L}(x_1, \dots, x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \begin{cases} p = \theta, & \text{si } x_i = 1 \\ 1 - p = 1 - \theta, & \text{si } x_i = 0 \end{cases} = \theta^{x_i} (1-\theta)^{1-x_i}$$

$$\mathcal{L}(x_1, \dots, x_n) = \theta^{\sum_{i=1}^{n} x_i} \cdot (1-\theta)^{n - \sum_{i=1}^{n} x_i}$$

I. Une statistique - Un estimateur :

Une stat est une fct de $(X_i)_i$ (mais pas des paramètres θ)

par ex : $S(X_1, \dots, X_n) = \begin{cases} X_1 - X_1 \\ e^{X_1} + \cos(X_2 X_3) \end{cases}$

Un estimateur : $\hat{T}$ est une stat $\in$ (1-1) (fon de valeurs de paramètres à

Ex $(X_i) \sim B(\theta)$. $X_1 + X_2 - X_3$ est une stat

• $\frac{X_1 + X_2 + X_3}{3}$ est un estimateur de θ ($\in [0,1]$)

→ c'est une V.A (fct des (X_i))

II . Statistique exhaustive (ou "suffisante")

Par déf : une stat $T(X_1, \ldots, X_n)$ est exhaustive pour θ si la loi de $(X_1, \ldots, X_n)$ sachant $T(X_1, \ldots, X_n)$ ne dépend pas de θ

càd : $\forall t$

$\mathcal{L}(X_1, \ldots, X_n / T = t)$ ne dépend pas de θ

⇒ (T contient toute l'information sur θ)

(si $T = \sum_{i=1}^{n} X_i$, $T = \prod X_i$, T = exhaustive)

⇒ résument l'échantillon par la connaissance de θ

réduisant la dimension

Ex : Pile / Face $X_i \sim B(\theta)$

$T = \sum_{i=1}^{n} X_i$ est une stat . exhaustive pour θ ?

Mq

$$\mathcal{L}(X_1, \ldots, X_n \mid \sum_{i=1}^{n} X_i = t) = P(X_1 = x_1, \ldots, X_n = x_n / \sum_{i=1}^{n} X_i = t)$$

$x_1 \ldots x_n$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^{n} X_i)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n) \, 1\!\!1_{\sum_{i=1}^{n} x_i = t}}{P(\sum_{i=1}^{n} X_i = t)}$$

or $(\sum_{i=1}^{n} X_i)$ est de loi Bino $B(n, \theta)$

⇒ $P(\sum_{i=1}^{n} X_i = t) = C_n^t \, \theta^t (1-\theta)^{n-t}$

$$\Rightarrow \mathcal{L}(X_1, \ldots, X_n \mid \sum_{i=1}^{n} X_i = t) = \frac{\theta^t (1-\theta)^{n-t} \, 1\!\!1_{\sum x = t}}{C_n^t \, \theta^t (1-\theta)^{n-t}}$$

$$= \frac{1\!\!1_{\sum_{i=1}^{n} x_i = t}}{C_n^t}$$

→ $(\sum_{i=1}^{n} X_i)$ est une stat exhaustive pour θ

= le nbre (ou proportion de pile) c'est suffisant pour estimer

Déf : Une stat exhaustive T est minimale si ∀ stat exhaustive S est fct de T. $(S = g(T))$

Souv. Trouver une stat. exh. min

→ difficile, mais possible pour une classe de lois : la famille exponentielle

déf par : des lois de la forme de x :

$$\text{loi de } X_i = e^{\langle a(\theta);\, \phi(x)\rangle + b(\theta) + c(x)}$$

fct. s

des fonctions

Exemple ① Si $X_i \sim \mathcal{E}xp(\lambda)$

$$f_{X_i}(x) = \lambda e^{-\lambda x}\mathbb{1}_{x>0} = e^{-\lambda x + \log(\lambda)}\mathbb{1}_{x>0} = e^{a(\theta)x + b(\theta)}$$

si $\theta = \lambda$

Ex ② $X_i \sim \mathcal{B}(\theta)$

$$\mathbb{P}(X_i = x) = \theta^x(1-\theta)^{1-x} = e^{x\log\theta + (1-x)\log(1-\theta)}$$

$$= e^{\underbrace{x}_{\phi(x)}\underbrace{(\log(\theta) - \log(1-\theta))}_{a(\theta)} + \underbrace{\log(1-\theta)}_{b(\theta)}}$$

Ex : lois binom, normale, ...

Théorème : Si famille exponentielle, alors $\sum_{i=1}^n \phi(X_i)$ est une stat. exhaustive minimale pour θ

(de la modèle exponentielle)

Ex : $X_i \sim \mathcal{B}(\theta)$, on a trouvé $\phi(x) = x$ $\Longrightarrow$ $\sum_{i=1}^n X_i$ est une stat exh min pour θ

famille exponentielle $\longrightarrow$ $\sum_{i=1}^n \phi(X_i)$ est une stat exhaustive minimale pour θ

Plus généralement

Une stat $T(X_1, \dots, X_n)$ est exhaustive **ssi** : ∃ 2 fcts g et h telles que

$$f_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = g(\theta; T(x_1,\dots,x_n)) \cdot h(x_1,\dots,x_n)$$ (⇒ Thm de factorisation)

"explication" ≈ loi de $(X_1, \dots, X_n)$ sachant $T(X_1, \dots, X_n)$

loi de $T((x_1, \dots, x_n))$