Chapitre 6 : Estimation ponctuelle

• Estimateurs

Cadre : Modèle statistique

$X_1, \dots, X_n$ v.a.i.i.d dont la loi dépend de θ qu'on cherche à estimer.

on a vu comment donner un encadrement de θ. Maintenant on veut la valeur la plus vraisemblable de θ.

Définition : Y_n estimateur est fortement consistant de θ si

$$Y_n \xrightarrow[n\to\infty]{p.s} \theta \quad \forall \theta \in \Theta$$

faiblement consistant de θ si

$$Y_n \xrightarrow[n\to\infty]{Proba} \theta$$

Définition : L'estimateur Y_n de θ a pour biais : $b(\theta) = E_\theta(Y_n) - \theta$.
L'estimateur est sans biais si $b(\theta) = 0$ i.e. $E_\theta(Y_n) = \theta$ sinon il est biaisé.

↳ $\overline{X_n}$ estimateur fortement consistant et sans biais de $E_\theta(X_1)$.

V_n " " " et biaisé de $Var_\theta(X_1)$.

$V_n^* = \frac{n}{n-1} V_n$ " " " et sans biais de $Var_\theta(X_1)$.

Définition : Si $Y = f(X_1, \dots, X_n)$ est un estimateur de carré intégrable son erreur quadratique moyenne (EQM) est $EQM(Y_n) = E_\theta\left[(Y_n - \theta)^2\right]$.

• Exemple d'estimation par maximum de vraisemblance ↳ voir feuille d'exo. exercice 1+2+3.

• Estimateur du max de vraisemblance pour des v.a. à densité

$X_1, \dots, X_n$ va i.i.d de loi P_θ qui dépend de θ inconnu.

on note f_θ la densité.

si on a observé $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$, quel est le θ le plus vraisemblable ?

si les X_i étaient discrète, on maximiserait la vraisemblance.

$$L(x_1, \dots, x_n ; \theta) = \prod_{i=1}^{n} (X_i = x_i)$$ mais les X_i étant à densité $P_\theta(X_i = x_i) = 0$

↳ Il faut définir la vraisemblance autrement.

$$L(x_1, \dots, x_n ; \theta) = \prod_{i=1}^{n} f_\theta(x_i)$$

- <u>Exemple d'estimation par maximum de vraissemblance.</u>

N manchots sur une plage dont 800 marqués.

500 tirages avec remises : On tire X manchots marqués et on observe $X(w) = 102 \to$ IC à 95 % pour $I(w) = [2631 ; 7692]$.

Maintenant on cherche la valeur N la plus vraissemblable compte tenu des infos disponibles. C'est celle pour laquelle on a le plus de chance d'observer $X(w)$.

$$X \sim Bin\left(500, \frac{800}{N}\right)$$

Pour chaque N on a une proba qu'on note P_N.

$$P_N(X = k) = C_{500}^{k} \left(\frac{800}{N}\right)^{k} \left(1 - \frac{800}{N}\right)^{500-k}$$

La valeur la plus vraissemblable pour N qd on observe $X(w) = k$ est celle pour laquelle $P_N(X = k)$ est maximal.

Notons $f(N) = C_{500}^{k} \dfrac{800^{k}(N-800)^{500-k}}{N^{500}}$ $\qquad f : [801, +\infty[\to \mathbb{R}$

$$\Leftrightarrow \ln f(N) = \cancel{\ln C_{500}^{k}} + k \cancel{\ln 800} + (500-k)\ln(N-800) - 500 \ln N$$

$$\frac{\partial(\ln f(N))}{\partial N} = \frac{500-k}{N-800} - \frac{500}{N} = \frac{-kN + 500 \times 800}{N(N-800)}$$

	801		$\frac{500 \times 800}{k}$		$+\infty$
$\partial \ln f / \partial N$		+	0	−	
$\ln f$		↗		↘	
f		↗		↘	

$N = \dfrac{500 \times 800}{k}$ est la valeur la plus vraissemblable de N observée en $X(w) = k$

on dit que la v.a. $\hat{N} = \dfrac{500 \times 800}{X}$ est l'E.P.M.V. pour N.

$\hat{N}$ maximise la f° de vraissemblance.
$L : (k, N) \longmapsto P_N(X = k)$.

<u>Exercice 1 :</u> EMV pour la proportion de brochet p ds l'étang. $X_1(w) = x_1 \ldots\ X_n(w) = x_n$.

$$L(x_1, \ldots, x_n ; p) = P_p(X_1 = x_1, \ldots, X_n = x_n) \quad X_i \text{ ind.}$$

$$\Leftrightarrow L(x_1, \ldots, x_n ; p) = P_p(X_1 = x_1) \ldots P_p(X_n = x_n)$$

$$= \prod_{i=1}^{n} P_p(X_i = x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$P_p(X_1 = x_1) = \begin{cases} p & \text{si } x_1 = 1 \\ 1-p & \text{si } x_1 = 0 \end{cases}$

$P_p(X_1 = x_1) = p^{x_1}(1-p)^{1-x_1}$

$$= p^{x_1}(1-p)^{1-x_1} \ldots p^{x_n}(1-p)^{1-x_n}$$

$$= p^{x_1 + \cdots + x_n}(1-p)^{(1-x_1) + \cdots + (1-x_n)}$$

$$\Leftrightarrow \ln L(x_1, \ldots, x_n ; p) = (x_1 + \cdots + x_n)\ln p + (n - x_1 - \cdots - x_n)\ln(1-p)$$

$$\Leftrightarrow \frac{\partial \ln L}{\partial p}(x_1, \ldots, x_n ; p) = (\Sigma x_i)\frac{1}{p} + (n - \Sigma x_i)\left(-\frac{1}{1-p}\right) = \frac{\Sigma x_i - np}{p(1-p)}$$

$$\Rightarrow \frac{1}{n}\Sigma x_i \geq p$$

p	0		$\frac{1}{n}\sum x_i$		1
$\frac{\partial \ln L}{\partial p}$		$+$	0	$-$	
$\ln L$		$\nearrow$		$\searrow$	
L		$\nearrow$		$\searrow$	

EMV pour p :
$$\hat{p} = \overline{X_n} = \frac{1}{n}\sum X_i$$

<u>Exercice 2</u> : $X_1, \dots, X_n$ ind de loi géom (p)

EMV $\hat{p}$?

$$L(x_1, \dots, x_n ; p) = P_p(X_1 = x_1, \dots, X_n = x_n)$$
$$= \prod_{i=1}^{n} P_p(X_i = x_i) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}$$
$$= p^n (1-p)^{(\sum x_i) - n}$$

$$\Leftrightarrow \ln L(x_1, \dots, x_n ; p) = n \ln p + ((\textstyle\sum x_i) - n) \ln(1-p)$$

$$\Leftrightarrow \frac{\partial \ln L}{\partial p}(x_1, \dots ; p) = \frac{n}{p} - \frac{\sum x_i - n}{1-p} = \frac{n - p\sum x_i}{p(1-p)}$$

$$\Leftrightarrow \frac{n}{\sum x_i} \geq p$$

L est max en $\frac{n}{\sum x_i} \Rightarrow \hat{p} = \frac{n}{\sum X_i} = \frac{1}{\overline{X_n}}$.

<u>Exercice 3</u> : $X_1, \dots, X_n$ iid de Pois (λ). $\hat{\lambda} = ?$

$$L(x_1, \dots, x_n ; \lambda) = \prod_{i=1}^{n} \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i !}\right) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod (x_i !)}$$

$$\ln L = -n\lambda + (\textstyle\sum x_i) \ln \lambda - \ln(\prod (x_i !))$$

$$\frac{\partial \ln L}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda} \geq 0 \Leftrightarrow \lambda \leq \frac{1}{n}\sum x_i = \overline{x}$$

L est max en $\frac{1}{n}\sum_{i=1}^{n} x_i$ d'où $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X_n}$.

<u>Exercice 4</u> : $X_1, \dots, X_n$ iid de $\mathcal{E}xp(a)$

$$L(x_1, \dots, x_n ; a) = \prod_{i=1}^{n} \left(a e^{-a x_i} \mathbb{1}_{x_i > 0}\right) = a^n e^{-a\sum_{i=1}^{n} x_i} \mathbb{1}_{\min(x_1, \dots, x_n) > 0}$$

Or $P(X_1 > 0) = 1$ si $X_1 \sim \mathcal{E}xp(a)$ donc

$$\ln(L(x_1, \dots ; a)) = n \ln a - a \textstyle\sum x_i$$

$$\frac{\partial \ln L}{\partial a} = \frac{n}{a} - \sum x_i \geq 0 \Rightarrow a \leq \frac{n}{\sum x_i}$$

$$\hookrightarrow \hat{a} = \frac{n}{\sum X_i} = \frac{n}{X_n}$$