Chapitre Final + 1

L'examen : 5 mai 14h-16h
Zoom ID : 858-2287-6565
Mot de passe : ens

I. Information de Fisher

1) Lien avec EMV

Déf L'information de Fisher d'une variable aléatoire X_1 de loi p_θ, $\theta \in \Theta$ est définie par

$$I(\theta) = \mathbb{E}_\theta[\ell'_{X_1}(\theta)^2]$$

log-vraisemblance d'un 1-échantillon X_1

Dans ce chapitre, on choisit de prouver les résultats avec les VA à densité.

Prop $\mathbb{E}_\theta[\ell''_{X_1}(\theta)] = \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X_1)\right] = -I(\theta)$

Dém : On note f_θ la densité de X_1.

$$\ell'_{X_1}(\theta) = \frac{\partial}{\partial\theta} \ln f_\theta(X_1) = \frac{\partial_\theta f_\theta(X_1)}{f_\theta(X_1)}$$

$$\ell''_{X_1}(\theta) = \frac{\partial_\theta^2 f_\theta(X_1)}{f_\theta(X_1)} - \underbrace{\frac{(\partial_\theta f_\theta(X_1))^2}{f_\theta(X_1)^2}}_{\ell'_{X_1}(\theta)^2}$$

On a $\mathbb{E}_\theta[\ell''_{X_1}(\theta)] = \mathbb{E}_\theta\left[\frac{\partial_\theta^2 f_\theta(X_1)}{f_\theta(X_1)}\right] - I(\theta)$

$$= \underbrace{\int \frac{\partial_\theta^2 f_\theta(x)}{\cancel{f_\theta(x)}} \cancel{f_\theta(x)}\, dx}_{=0\ ?} - I(\theta)$$

En effet, $\int f_\theta(x)\,dx = 1$, en dérivant par rapport à θ deux fois :

$$\int \partial_\theta^2 f_\theta(x)\,dx = 0$$

$\boxed{\text{Th}}$ (Normalité asymptotique de EMV)

$X = (X_1, \dots, X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta$. $\hat{\theta}_n$ EMV de θ,
$I(\theta)$ l'information de Fisher de X_1. Sous certaines hypothèses de régularité,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\infty]{\mathcal{L}_\theta} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

Dém : L'EMV $\hat{\theta}_n$ satisfait $\ell_X'(\hat{\theta}_n) = 0$

Par Taylor,

$$0 = \ell_X'(\hat{\theta}_n) \simeq \ell_X'(\theta) + (\hat{\theta}_n - \theta)\,\ell_X''(\theta)$$

D'où $$\sqrt{n}(\hat{\theta}_n - \theta) \simeq - \frac{\sqrt{n}\,\ell_X'(\theta)}{\ell_X''(\theta)}$$

$$\ell_X'(\theta) = \sum_{i=1}^n \ell_{X_i}'(\theta) \quad \text{avec} \quad \ell_{X_i}'(\theta) = \frac{\partial_\theta f_\theta(X_i)}{f_\theta(X_i)}$$

$\star$ $\mathbb{E}_\theta[\ell_{X_i}'(\theta)] = \int \partial_\theta f_\theta(x)\,dx = 0$

$\star$ $\mathbb{E}_\theta[\ell_{X_i}'(\theta)^2] = I(\theta)$

Ainsi $$\frac{\sqrt{n}\,\ell_X'(\theta)}{n} = \sqrt{n}\left(\frac{\sum_{i=1}^n \ell_{X_i}'(\theta)}{n} - 0\right) \xrightarrow[n\infty]{\mathcal{L}_\theta}{}^{\text{TCL}} \mathcal{N}(0, I(\theta))$$

Or $-\dfrac{l_x''(\theta)}{n} = -\dfrac{\sum_{i=1}^{n} l_{x_i}''(\theta)}{n} \xrightarrow[n\infty]{p.s.} -\mathbb{E}_\theta[l_{x_1}''(\theta)] = I(\theta)$

LGN

Par le lemme de Slutsky,

$$\sqrt{n}(\hat{\theta}_n - \theta) \simeq \frac{\sqrt{n}\,\frac{l_x'(\theta)}{n}}{-\frac{l_x''(\theta)}{n}} \xrightarrow[n\infty]{\mathcal{L}_\theta} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

ex : Soit $(X_1, \dots X_n)$ un n-échantillon de la $\mathcal{B}(\theta)$, $0<\theta<1$.

$l_{x_1}(\theta) = X_1 \ln\theta + (1-X_1)\ln(1-\theta)$ $\rightarrow \ln f_\theta(x_1)$ avec $f_\theta(x) = \begin{cases} \theta & x=1 \\ 1-\theta & x=0 \end{cases}$

$l_{x_1}''(\theta) = -\dfrac{X_1}{\theta^2} - \dfrac{1-X_1}{(1-\theta)^2}$

$I(\theta) = -\mathbb{E}_\theta[l_{x_1}''(\theta)] = \dfrac{1}{\theta} + \dfrac{1}{1-\theta} = \dfrac{1}{\theta(1-\theta)}$

EMV : $\hat{\theta}_n = \overline{X}_n$

Par le th précédent, $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow[n\infty]{\mathcal{L}_\theta} \mathcal{N}(0, \underbrace{\theta(1-\theta)}_{\frac{1}{I(\theta)}})$

ex2 : $(X_1, \dots X_n)$ un n-échantillon de loi $\gamma_{1,\alpha}$ $\leftarrow$ densité $x \longmapsto \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \mathbb{1}_{x>0}$

$$\ell_{X_1}(\alpha) = -\ln \Gamma(\alpha) + (\alpha-1)\ln X_1 - X_1$$

$$\ell''_{X_1}(\alpha) = -\psi'(\alpha) \qquad \text{où } \psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

D'où $I(\alpha) = -\mathbb{E}_\alpha[\ell''_{X_1}(\alpha)] = \psi'(\alpha)$ $\quad (>0)$

Ensuite on cherche EMV :

$$\ell_X(\alpha) = n\left(-\ln \Gamma(\alpha) + (\alpha-1)\overline{\ln X_n} - \overline{X}_n\right)$$

$$\frac{1}{n}\ell'_X(\alpha) = -\psi(\alpha) + \overline{\ln X_n} \quad = 0$$

Donc l'EMV $\hat{\alpha}_n = \psi^{-1}(\overline{\ln X_n})$.

Par le th précédent, $\sqrt{n}\left(\psi^{-1}(\overline{\ln X_n}) - \alpha\right) \xrightarrow[n\infty]{\mathcal{L}_\alpha} \mathcal{N}\Big(0, \underbrace{\frac{1}{\psi'(\alpha)}}_{\frac{1}{I(\alpha)}}\Big)$

2) Borne de Cramér-Rao

Prop (Inégalité de Cauchy) Soient X et Y deux VA réelles, alors $$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\,\mathbb{E}[Y^2]$$ avec égalité ssi $X = \underbrace{c}_{\text{cste}} Y$ p.s.

Th (Borne de Cramér-Rao) $(X_1, \dots X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta$. $\hat{\theta}_n$ un estimateur sans biais de θ. Alors

$$\mathrm{Var}_\theta(\hat{\theta}_n) \geq \frac{1}{n I(\theta)}$$

Dém : $\hat{\theta}_n$ est une fonction de $X_1, \dots X_n$, on écrit

$$\hat{\theta}_n = h(X_1, \dots X_n)$$

$\hat{\theta}_n$ est sans biais

$$\Rightarrow \int_{\mathbb{R}^n} h(x_1, \dots x_n) \overbrace{f_\theta(x_1) \cdots f_\theta(x_n)}^{\exp(\ell_{(x_1, \dots x_n)}(\theta))} dx_1 \cdots dx_n = \theta$$

En dérivant $/\theta$:

$$\underbrace{\int_{\mathbb{R}^n} h(x_1,\dots,x_n)\, \ell'_{(x_1,\dots,x_n)}(\theta)\, f_\theta(x_1)\cdots f_\theta(x_n)\, dx_1\cdots dx_n}_{\mathbb{E}_\theta[\hat\theta_n \ell'_X(\theta)]} = 1$$

De la même façon, on a

$$\mathbb{E}_\theta[\ell'_X(\theta)] = n\, \mathbb{E}_\theta[\ell'_{X_1}(\theta)] = n \int_{\mathbb{R}} \partial_\theta f_\theta(x)\, dx = 0$$

Par conséquent, $\mathbb{E}_\theta[(\hat\theta_n - \theta)\ell'_X(\theta)] = 1$.

On applique l'inégalité de Cauchy :

$$1 = \mathbb{E}_\theta[(\hat\theta_n-\theta)\ell'_X(\theta)]^2 \leq \mathrm{Var}_\theta(\hat\theta_n)\, \mathbb{E}_\theta[\ell'_X(\theta)^2]$$

et $\mathbb{E}_\theta[\ell'_X(\theta)^2] = n\, \mathbb{E}_\theta[\ell'_{X_1}(\theta)^2] = n I(\theta)$

d'où $\mathrm{Var}_\theta(\hat\theta_n) \geq \dfrac{1}{n I(\theta)}$

Rq : L'égalité est atteinte ssi $\hat{\theta}_n = \theta + c(\theta)\, \ell_X'(\theta)$

$c(\theta)$ une fonction de θ.

Rq2 : L'EMV $\hat{\theta}_n$. On a

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow[n\infty]{\mathcal{L}_\theta} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

$\forall f \in \mathcal{C}_b$, $\mathbb{E}[f(X_n)] \xrightarrow[n\infty]{} \mathbb{E}[f(X)]$

Une conséquence est que

$$n\, \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \xrightarrow[n\infty]{} \frac{1}{I(\theta)}$$

Si $\hat{\theta}_n$ est sans biais, ceci implique que

$$\mathrm{Var}_\theta(\hat{\theta}_n) \underset{n\infty}{\sim} \frac{1}{n\, I(\theta)}$$

La borne de Cramér-Rao est atteinte asymptotiquement ! Autrement dit, un EMV sans biais est asymptotiquement optimal (au sens de moindre variance) parmi tous les estimateurs sans biais.

3) Tests du rapport de vraisemblance asymptotique.

Rappel : H_0 : "$\theta \in \Theta_0$" contre H_1 : "$\theta \in \Theta_1$"

Rapport de vraisemblance :

$$T(x) = \frac{\sup_{\theta \in \Theta_1} V_x(\theta)}{\sup_{\theta \in \Theta_0} V_x(\theta)}$$

Le test s'écrit $\mathbb{1}_{T(x) > t_\alpha}$

$\boxed{\text{Th}}$ $(X_1, \dots, X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta \subseteq \mathbb{R}^d$. On considère

H_0 : "$\theta = \theta_0$" contre H_1 : "$\theta \neq \theta_0$". Sous certaines hypothèses de régularité

(pour la consistance et normalité asymptotique de l'EMV),

$$2 \ln T(X) \xrightarrow[n \to \infty]{\mathcal{L}_{\theta_0}} \chi^2(d)$$

Idée : L'EMV $\hat{\theta}_n$. On illustre $d=1$.

$$2\ln T(X) = 2\ell_X(\hat{\theta}_n) - 2\ell_X(\theta_0)$$

Par Taylor,

$$\ell_X(\theta_0) \simeq \ell_X(\hat{\theta}_n) + \underbrace{\ell_X'(\hat{\theta}_n)}_{0}(\theta_0 - \hat{\theta}_n) + \frac{1}{2}\underbrace{\ell_X''(\hat{\theta}_n)}_{\simeq \ell_X''(\theta_0)}(\theta_0 - \hat{\theta}_n)^2$$

$$\simeq \ell_X''(\theta_0) = \sum_{i=1}^n \ell_{X_i}''(\theta_0) \simeq -n\,I(\theta_0)$$

Donc $2\ln T(X) \simeq n\,I(\theta_0)(\theta_0 - \hat{\theta}_n)^2$

Par la normalité asymptotique de $\hat{\theta}_n$:

$$\sqrt{n}\sqrt{I(\theta_0)}\,(\hat{\theta}_n - \theta_0) \xrightarrow[n\infty]{\mathcal{L}_{\theta_0}} \mathcal{N}(0,1)$$

Ainsi $2\ln T(X) \xrightarrow[n\infty]{\mathcal{L}_{\theta_0}} \chi^2(1)$

Cor : Un test du rapport de vraisemblance de

H_0 : "$\theta = \theta_0$" contre H_1 : "$\theta \neq \theta_0$" de niveau α est

$$\mathbb{1}_{2\ln T(X) > c_{1-\alpha}^{(d)}}$$

ex : $X_1, \dots X_n$ i.i.d. $\sim N(\theta, 1)$

H_0 : "$\theta = 0$" contre H_1 : "$\theta \neq 0$"

L'EMV est $\hat{\theta}_n = \bar{X}_n$.

$$
\begin{aligned}
2 \ln T(x) &= 2 \ell_x(\hat{\theta}_n) - 2 \ell_x(0) \\
&= - \sum_{i=1}^n (x_i - \bar{x}_n)^2 + \sum_{i=1}^n x_i^2 \\
&= (\underbrace{\sqrt{n} \bar{X}_n}_{\sqrt{n}\bar{X}_n \sim N(0,1)})^2 \sim \chi^2(1)
\end{aligned}
$$

II. Estimation Bayésienne.

1) Loi a priori et a posteriori

Dans l'approche Bayésienne, on dispose d'une information a priori sur le paramètre inconnu θ. Cette information prend la forme d'une loi sur l'espace des paramètres

Θ, notée p, qui s'appelle la loi <u>a priori</u>. Le paramètre θ devient une variable aléatoire de densité $p(\theta)$.

$\boxed{\text{Déf}}$ La loi jointe des observations $X = (X_1, \dots X_n)$ conditionnelle à θ est notée

$$p(x|\theta) = \begin{cases} f(x_1, \dots x_n|\theta) & \text{continu} \\ \mathbb{P}(X_1 = x_1, \dots X_n = x_n|\theta) & \text{discret} \end{cases}$$

$\boxed{\text{Déf}}$ Un modèle Bayésien est la donnée d'une loi conditionnelle et d'une loi à priori :

$$X|\theta \sim p(x|\theta), \quad \theta \sim p$$

$\boxed{\text{Déf}}$ Dans un modèle Bayésien, on appelle la loi a posteriori la loi de θ conditionnellement aux observations x (réalisation de X), notée

$$p(\theta|x) = \frac{p(x|\theta)\, p(\theta)}{p(x)}$$ où $p(x)$ est la loi marginale de X

définie par $p(x) = \begin{cases} \int p(x|\theta)p(\theta)d\theta & \text{continu} \\ \sum_i p(x|\theta_i)p(\theta_i) & \text{discret} \end{cases}$

Les estimateurs Bayésiens du paramètre θ sont construits à partir de la loi a posteriori $p(\theta|x)$ par minimisation d'une fonction de coût appropriée.

On considère le coût quadratique :

$$C(\hat{\theta}) = E[(\hat{\theta}-\theta)^2 | x]$$

$$= E[((\hat{\theta} - E[\theta|x]) + (E[\theta|x] - \theta))^2 | x]$$

$$= E[(\hat{\theta} - E[\theta|x])^2 | x] + Var(\theta|x)$$

$$\geq Var(\theta|x)$$

L'égalité est atteinte pour $\hat{\theta} = E[\theta|x]$

L'estimateur par la méthode Bayésienne est donc $E[\theta|X]$

▷ Etapes pour construire un estimateur Bayésien :

1. Choisir une loi a priori $p(\theta)$
2. Trouver la loi a posteriori $p(\theta | x)$
3. Poser $\hat{\theta} = \mathbb{E}[\theta | x]$

2) Loi a priori conjuguée.

Comment choisir la loi a priori ?

* Basé sur des expériences du passé ou sur une intuition du statisticien.
* Basé sur la faisabilité des calculs.
* Basé sur la volonté de n'apporter aucune information nouvelle pouvant biaiser l'estimation.

Déf : Une famille $\mathcal{F}$ de lois sur Θ est dite conjuguée pour la loi $p(x|\theta)$ si pour tout $p(\cdot) \in \mathcal{F}$, la loi a posteriori $p(\cdot | x)$ appartient également à $\mathcal{F}$.

ex : $p(\cdot|\theta) \sim \mathcal{B}(\theta)$

$$\underset{(x_1, \dots, x_n)}{\underset{\parallel}{p(x|\theta)}} = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

On considère $\theta \sim \mathcal{B}(\alpha, \beta)$, $\alpha, \beta > 0$.

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \mathbb{1}_{\theta \in (0,1)}$$

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \quad \leftarrow \int_0^1 p(x|\theta)\, p(\theta)\, d\theta$$

$$\propto \theta^{\sum_{i=1}^{n} x_i + \alpha - 1} (1-\theta)^{n - \sum_{i=1}^{n} x_i + \beta - 1} \mathbb{1}_{\theta \in (0,1)}$$

Donc $$p(\theta|x) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum x_i)\Gamma(\beta+n-\sum x_i)} \theta^{\alpha+\sum x_i - 1}(1-\theta)^{\beta+n-\sum x_i - 1} \mathbb{1}_{\theta \in (0,1)}$$

C'est la densité de la loi $\mathcal{B}(\alpha + \sum x_i, \beta + n - \sum x_i)$

[$X \sim \mathcal{B}(\alpha,\beta)$, alors $\mathbb{E}[X] = \dfrac{\alpha}{\alpha+\beta}$]

L'estimateur Bayésien de θ est

$$\hat{\theta} = \mathbb{E}[\theta \mid x] = \frac{\alpha + \sum X_i}{\alpha+\beta+n}$$

$\to$ Si $n \to \infty$, $\hat{\theta} \simeq \dfrac{\frac{\alpha}{n} + \overline{X}_n}{\frac{\alpha+\beta}{n}+1} \simeq \overline{X}_n$

$\to$ Si $n = 0$, $\hat{\theta} = \dfrac{\alpha}{\alpha+\beta} = \mathbb{E}[\mathcal{B}(\alpha,\beta)]$

Rq : On peut aussi construire des intervalles de confiance de θ à partir de la loi $p(\theta|x)$, ou faire des tests d'hypothèses.

ex2 : $p(\cdot|\theta) \sim \mathcal{N}(\theta, \sigma^2)$ $\to$ connu

$p(\cdot) \sim \mathcal{N}(\mu_0, \sigma_0^2)$ à priori

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma^2}\right)$$

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right)$$

Alors $p(\theta|x) \propto p(x|\theta)\,p(\theta)$

$$\propto \exp\left(-\left(\frac{n}{2\sigma^2} + \frac{1}{2\sigma_0^2}\right)\theta^2 + \left(\frac{\sum x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right)\theta \right)$$

$$\propto \exp\left(-\left(\frac{n}{2\sigma^2} + \frac{1}{2\sigma_0^2}\right)\left(\theta - \frac{\frac{\sum x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{2\sigma^2} + \frac{1}{2\sigma_0^2}}\right)^2 \right)$$

On reconnaît la densité d'une loi normale

$$\mathcal{N}\left(\frac{\frac{\sum x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{2\sigma^2} + \frac{1}{2\sigma_0^2}},\ \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \right)$$

$$\hat{\theta} = \mathbb{E}[\theta | X] = \frac{\frac{\sum_{i=1}^{n} X_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{2\sigma^2} + \frac{1}{2\sigma_0^2}}.$$

3) Loi a priori de Jeffreys.

Si on ne veut pas une loi a priori informatif, on pense à la loi uniforme sur

(4). Mais si $\theta \sim \mathcal{U}([0,1])$, et on reparamétrise le modèle par $\phi = -\frac{1}{\lambda}\ln(1-\theta)$, alors $\phi \sim \mathcal{E}(\lambda)$ (TD 2), qui semble beaucoup plus informatif. On voit ainsi que une bonne notion de loi a priori non-informative est une loi invariante par reparamétrisation.

Déf : La loi a priori de Jeffreys est donnée par

$$p(\theta) \propto \sqrt{I(\theta)}$$

avec $I(\theta) = \mathbb{E}\left[\left(\partial_\theta \ln p(x_1|\theta)\right)^2\right]$

Cette loi possède 2 intérêts principaux :

* $I(\theta)$ est un indicateur de la quantité d'information apportée par le modèle $p(x_1|\theta)$, donc au niveau qualitatif, les valeurs de θ pour lesquelles $I(\theta)$ est plus grande doivent être plus probables.

* La loi de Jeffreys est invariante par reparamétrisation. En effet,

Soit $\phi = h(\theta)$ avec h un $\mathcal{C}^1$-difféomorphisme.

ϕ est de loi $\tilde{p}(\phi) = p(\phi)\,|(h^{-1})'(\phi)|$

De plus on a $\tilde{I}(\phi) = I(\phi)\,|(h^{-1})'(\phi)|^2$

Donc $\tilde{p}(\phi) \propto \sqrt{\tilde{I}(\phi)}$