STATISTIQUE MATHÉMATIQUE - MS ENSAI
RÉSUMÉ CHAP.3 - 2025

1 Exhaustivité, minimalité, complétude

Définition

Une statistique S est exhaustive pour X relativement à θ si la loi conditionnelle de X sachant $[S = s]$ ne dépend pas de θ.

Théorème — Critère de factorisation de Neyman

S exhaustive ssi, la vraisemblance s'écrit

$$L(x,\theta) = h(x) \times g(S(x),\theta) \quad (1)$$

Théorème

Soit T une statistique exhaustive et ϕ une fonction mesurable. Alors la statistique S vérifiant $T = \phi(S)$ est exhaustive.

Si ϕ est bijective, S et T sont alors équivalentes et T exhaustive $\iff$ S exhaustive.

Définition

S minimale si elle est fonction de toute statistique exhaustive:

$$T \text{ exhaustive minimale } \iff$$
$$\forall S \text{ exhaustive }, \exists\, \phi : \; T = \phi(S),\; \mathbb{P}_\theta - \text{p.s.}, \forall\theta$$

Rapport de vraisemblance (« likelihood ratio ») pour x, y fixés:

$$\theta \longrightarrow LR(x,y,\theta) = \frac{L(x,\theta)}{L(y,\theta)} \quad (2)$$

Théorème — Critère de minimalité

Soit S stat. d'un modèle dominé /

$LR(x,y,\theta)$ ne dépend pas de $\theta \iff T(x) = T(y)$

Alors T exhaustive et minimale.

Définition

S libre / θ si sa loi ne dépend pas de θ.
U ignorable si $\exists$ T exhaustive $\perp\!\!\!\perp$ de U.

Définition

S complète s'il n'existe pas de fonction g non constante, intégrable, à valeurs dans $\mathbb{R}$ qui soit libre :

$$\forall\theta,\; \mathbb{E}_\theta[g(T)] = 0 \Rightarrow g \equiv 0 \;\; \mathbb{P}_\theta - \text{p.s.}$$

Théorème

S complète $\Rightarrow$ S minimale.

Ex: si modèle exponentiel:
- La statistique canonique S est exhaustive.
- Si Λ, espace des paramètres, contient 1 ouvert de $\mathbb{R}^k$ ou 1 repère affine, alors S minimale complète.

2 Information de Fisher

- $\theta = (\theta_1, ..., \theta_p) \in \Theta \subset \mathbb{R}^p$ ouvert.
- $S = (S_1, ..., S_p) \in \mathbb{R}^p$ statistique de θ.
- $g : \Theta \longrightarrow \mathbb{R}$ de classe C^2.
- $\nabla g(\theta)$ gradient en θ de g.
- $H_g(\theta)$ sa matrice hessienne.
- $\mathbb{V}(S)$ matrice de variance/covariance de S.
- $l(X,\theta) = \ln L(X,\theta)$ log-vraisemblance de X.
- $S_c(X,\theta) = \partial l/\partial\theta(X,\theta) = \nabla l(\theta)$ score du modèle.

Définition

Un modèle paramétrique est régulier s'il est:
- dominé, homogène et $\Theta \subset \mathbb{R}^p$ ouvert.
- $L(x,\theta) > 0 \;\forall x \in \mathbb{X}, \forall\theta \in \Theta$.
- $\theta \longrightarrow L(x,\theta)$ de classe C^2 pour p.$\forall x$.
- $\forall B \in \mathbb{X}$, $\theta \longrightarrow \int_B L(x|\theta) d\mu(x)$ C^2 sous $\int$.
- Le score $\nabla l(X,\theta) \in L^2(\mathbb{P}_\theta)$.

Définition

L'information de Fisher $\mathbb{I}(\theta)$ du modèle est la variance du score:

$$\mathbb{I}(\theta) = \mathbb{V}_\theta(S) \quad (3)$$

- $\mathbb{E}_\theta[\nabla l(X,\theta)] = 0$
- En dim.1, $\mathbb{I}(\theta) = \mathbb{E}\left[\nabla l^2\right] = -\mathbb{E}\left[l''(\theta)\right]$.
- En dim.$p > 1$,

Propriété info d'un modèle régulier

- $\mathbb{I}(\theta) = \mathbb{E}_\theta[\nabla l . \nabla l^T] = -\mathbb{E}_\theta \left[\frac{\partial^2 l}{\partial \theta^2} \right] = -\mathbb{E}_\theta [H_l]$

Soit $(\mathbb{X}, \mathfrak{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$ modèle régulier et T stat.
Soit $g(t, \theta)$ densité de T.

Propriété

- Info du modèle image : $I_T(\theta) = \mathbb{V}(S(T, \theta))$
- Score du modèle image : $\frac{d}{d\theta} \ln g(T, \theta)$
- Si $S_1 \perp\!\!\!\perp S_2$: $I_{(S_1,S_2)}(\theta) = I_{S_1}(\theta) + I_{S_2}(\theta)$
- En particulier : $\mathbb{I}_{(X_1,\ldots,X_n)}(\theta) = n \times \mathbb{I}_{X_1}(\theta)$
- Info conditionnelle : $I(\theta) = I_S(\theta) + I_{X|S}(\theta)$
- S exhaustive $\iff I_X(\theta) = I_S(\theta)$
- Reparamétrage : si $\lambda = \phi(\theta)$, $\mathbb{I}_X(\lambda) = J^T \mathbb{I}_X(\theta) J$

J est la matrice jacobienne de ϕ. En dimension 1, on a la formule importante suivante :

$$\mathbb{I}_X(\lambda) = \frac{\mathbb{I}_X(\theta)}{\phi'(\theta)^2} \tag{4}$$

3 Estimateurs optimaux

Différentes fonctions de coût (perte ou loss) :
$c_2(\theta, \theta') = (\theta - \theta')^2$, $c_1(\theta, \theta') = |\theta - \theta'|$,
$c_0(\theta, \theta') = \mathbb{1}_{[|\theta - \theta'| > \epsilon|]}$.

Définition

Le risque moyen d'un estimateur T de θ est l'espérance de la fonction de coût entre T et θ:

$$R(S, \theta) = \mathbb{E}_\theta[c(T, \theta)] \tag{5}$$

Définition

$\mathcal{T}$ classe d'estimateurs de θ, $T \in \mathcal{T}$ admissible dans $\mathcal{T}$ pour θ s'il est uniformément préférable à tout autre estimateur de $\mathcal{T}$:

$$\forall T' \in \mathcal{T}, \forall \theta \in \Theta,\ R(T, \theta) \leq R(T', \theta) \tag{6}$$

On rappelle la formule biais variance :

$$R(T, \theta) = \mathbb{V}_\theta(T) + (\mathbb{E}_\theta[T] - \theta)^2 \tag{7}$$

Définition

T^* est de variance uniformément minimale parmi les estimateurs sans biais de θ (VUMSB) si T^* admissible pour le risque quadratique, dans la classe des estimateurs sans biais :

- $\mathbb{E}[T^*] = \theta$ (8)
- $\forall\ T$ sans biais , $\mathbb{V}_\theta(T^*) \leq \mathbb{V}_\theta(T)$ (9)

Théorème de Rao-Blackwell

T un estimateur sans biais et S stat. exhaustive. Alors

$$T^* = \mathbb{E}_\theta[T|S] \tag{10}$$

estimateur sans biais de θ préférable à T pour le risque moyen quadratique.

Théorème de Lehmann-Scheffé

T estimateur sans biais et S stat. complète. Alors $T^* = \mathbb{E}_\theta[T|S]$ est VUMSB.

Théorème Borne FDCR

$X \sim \mathbb{P}_\theta$ modèle régulier où $\mathbb{I}(\theta)^{-1} \exists\ \forall \theta \in \Theta$.
T un estim. sans biais de $\phi(\theta) \in \mathbb{R}^d$ / :

- $\phi(\theta) = \mathbb{E}_\theta[T]$ différentiable en θ.
- $\mathbb{E}_\theta[T]$ différentiable en θ sous $\int$.

Alors

$$\mathbb{V}_\theta(T) \geq \nabla \phi(\theta) \times \mathbb{I}(\theta)^{-1} \times \nabla \phi(\theta)^T \tag{11}$$

Cas particulier important : $\phi = Id$, c.-à-d. T estim. sans biais de θ, alors $\mathbb{V}_\theta(T) \geq \mathbb{I}(\theta)^{-1}$.

Définition

T estimateur de $\phi(\theta)$ est efficace s'il est sans biais et atteint la borne FDCR.

Le théorème suivant est très important car il permet de généraliser et de remplacer le théorème de la limite centrale, dans des cas où les va ne sont pas indépendantes ou des cas où les estimateurs ne sont pas sous la forme de moyennes empiriques.

Théorème — Efficacité asymptotique de l'e.m.v.

Soit $X_1,\ldots,X_n$ un n-échantillon de $X \sim \mathbb{P}_\theta$, $\theta \in \Theta$, où $f(x,\theta)$ est la densité de $\mathbb{P}_\theta$ par rapport à μ. On suppose que le modèle est régulier. Alors l'e.m.v. $\widehat{\theta}_n$ de θ est $\sqrt{n}$-consistant et AN :

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \overset{\mathbb{P}_\theta}{\rightsquigarrow} \mathcal{N}(0, \mathbb{I}(\theta)^{-1}) \tag{12}$$

En particulier $\widehat{\theta}_n$ est asymptotiquement sans biais et efficace.

Théorème — 2e principe fondamental de la stat.

Mieux vaut un petit biais qu'une grosse variance.

Définition

Le risque moyen *a posteriori* de l'estimateur T associé à la loi *a priori* Π et à l'observation x est

$$R(\Pi, T, x) = \mathbb{E}[c(H, T(x))|X = x] \tag{13}$$

$$= \int_\Theta c(\theta, T(x))L(\theta, x)d\theta \tag{14}$$

Définition

Le risque intégré de l'estimateur T associé à la loi *a priori* Π est

$$R(\Pi, T) = \mathbb{E}[c(H, T)] \tag{15}$$

$$= \int_{\mathbb{X}}\int_\Theta c(\theta, T(x))L(\theta, x)d\theta d\mu(x) \tag{16}$$

Définition

Soit $\mathcal{T}$ classe d'estim. de θ. Risque de Bayes :

$$R(\Pi) = \inf_{T\in\mathcal{T}} R(\Pi, T) \tag{17}$$

L'estim. de Bayes $\widehat{\theta}_\Pi$ associé à la loi *a priori* Π est l'estim. qui minimise le risque intégré :

- $R(\Pi, \widehat{\theta}) = R(\Pi)$ (18)
- $\widehat{\theta} = \mathrm{argmin}_T R(\Pi, T)$ (19)