Apprentissage statistique
Chapitre 4 : Étude du risque quadratique d'estimateurs

Lucie Le Briquer

29 janvier 2019

Table des matières

1 Définitions et décomposition biais-variance

Définition 1

Soit $(\mathbb{Z}, \mathcal{Z}, \mathcal{P})$ un modèle statistique. $\mathcal{F} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ et T un estimateur de $\theta \mapsto g(\theta) \in \mathbb{R}^q$ tel que $\mathbb{E}_\theta[\|T\|^2] < +\infty \ \forall\theta$.

1. On définit le biais de T par $\forall\theta$:

$$\text{biais}_\theta(T) = \mathbb{E}_\theta(Y) - g(\theta)$$

2. On définit le MSE de T par $\forall\theta$:

$$\text{MSE}_\theta(T) = \mathbb{E}_\theta[\|T - g(\theta)\|^2]$$

Exemple. $(X_1, \ldots, X_n)$ $n-$échantillon de $\{\mathbb{P}_\theta : \theta \in \Theta\}$ et $\forall\theta,\ \mathbb{E}_\theta[X_1^2] < +\infty$. Un estimateur de $\mathbb{E}_\theta[X_1] = g(\theta)$ est :

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

On peut prendre comme etimateur de la variance $\theta \mapsto \text{Var}_\theta(X_1)$:

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

Cet estimateur est biaisé puisque :

$$\mathbb{E}_\theta[S] - \text{Var}_\theta(X_1) = -\frac{1}{n}\text{Var}_\theta(X_1)$$

Remarque. $uv^T = u \otimes v$

Proposition 1

Pour tout $\theta \in \Theta$:

$$\mathbb{E}_\theta[(T - g(\theta))^{\otimes 2}] = (\mathbb{E}_\theta[T] - g(\theta))^{\otimes 2} + \text{Cov}_\theta(T)$$

où $\text{Cov}_\theta(T) = \mathbb{E}_\theta[(T - \mathbb{E}_\theta[T])^{\otimes 2}]$

Preuve.

$$\begin{aligned}
\mathbb{E}_\theta[(T - g(\theta))^{\otimes 2}] &= \mathbb{E}_\theta\big[(T - \mathbb{E}_\theta[T] + \mathbb{E}_\theta[T] - g(\theta))^{\otimes 2}\big] \\
&= \mathbb{E}_\theta\big[(T - \mathbb{E}_\theta[T])^{\otimes 2}\big] + 2\mathbb{E}_\theta\big[(T - \mathbb{E}_\theta[T]) \otimes (\mathbb{E}_\theta[T] - g(\theta))\big] \\
&\quad + \mathbb{E}\big[(g(\theta) - \mathbb{E}_\theta[T])^{\otimes 2}\big]
\end{aligned}$$

On obtient alors le résultat car $(u, v) \mapsto u \otimes v$ est bilinéaire et donc :

$$\mathbb{E}_\theta\big[(T - \mathbb{E}_\theta[T]) \otimes (\mathbb{E}[T] - g(\theta))\big] = \underbrace{\big\{\mathbb{E}_\theta[T - \mathbb{E}_\theta[T]]\big\}}_{=0} \otimes \big\{\mathbb{E}[T] - g(\theta)\big\} = 0$$

■

Corollaire 1

$\forall \theta \in \Theta$,

$$\mathrm{MSE}_\theta(T) = \|\mathbb{E}_\theta[T] - g(\theta)\|^2 + \mathbb{E}\left[\|T - \mathbb{E}_\theta[T]\|^2\right]$$

Preuve.

$$\forall u \in \mathbb{R}^q, \quad \mathrm{Tr}(uu^T) = \|u\|^2$$

Comme Tr est linéaire on peut la rentrer dans l'espérance. Donc,

$$\begin{aligned}
\mathrm{Tr}\left(\mathbb{E}_\theta\left[(T - g(\theta))^{\otimes 2}\right]\right) &= \mathbb{E}_\theta\left[\mathrm{Tr}\left((T - g(\theta)^{\otimes 2}\right)\right] \\
&= \mathbb{E}_\theta\left[\|T - g(\theta)\|^2\right] \\
&= \mathrm{MSE}_\theta(T)
\end{aligned}$$

Les autres termes se traitent de la même manière. ■

Exemple. $X_1, \ldots, X_n$ un $n-$échantillon réel de $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. $\overline{X}_n$ est un estimateur sans biais de la moyenne. Cependant $\forall (\alpha_i)_{i=1}^n \in \Delta_n$ où :

$$\Delta_n = \left\{(\alpha_i) : \sum_{i=1}^n \alpha_i = 1, 0 \leqslant \alpha_i \leqslant 1\right\}$$

$X_{n,\alpha} = \sum_{i=1}^n \alpha_i X_i$ est un estimateur sans biais.
Regardons le plus performant.

$$\mathrm{MSE}_\theta(X_{n,\alpha}) = \mathrm{Var}_\theta(X_{n,\alpha}) \underset{\text{idp}}{=} \sum_{i=1}^n \alpha_i^2 \mathrm{Var}_\theta(X_i) \underset{\text{iid}}{=} \mathrm{Var}_\theta(X_1) \sum_{i=1}^n \alpha_i^2$$

Or,

$$1 = \sum_{i=1}^n \alpha_i \underset{\text{C.S.}}{\leqslant} \left(\sum_{i=1}^n \alpha_i^2\right)^{1/2} \sqrt{n} \qquad \sum_{i=1}^n \alpha_i^2 \geqslant \frac{1}{N}$$

et $\sum_{i=1}^n \alpha_i^2 = \frac{1}{n}$ ssi $(\alpha_i) = \lambda(1, \ldots, 1)$, on trouve $\alpha = \frac{1}{n}$ puisque $\sum_{i=1}^n \alpha_i = 1$.