Cours 2

3 Estimateur de Bayes

Comme nous l'avons déjà fait remarquer, la prise d'une décision, ici le choix d'un estimateur, va engendrer un coût que l'on va quantifier à l'aide de la fonction de perte. En pratique, on cherche une décision qui minimise en moyenne la fonction de coût.

Définition 5 – *On appelle estimateur de Bayes associé à un coût L et à une distribution a priori π, toute décision δ^π qui minimise le risque de Bayes $r(\pi, \delta)$.*

On a :

$$\delta^\pi(x) = \underset{\delta \in \mathcal{D}}{Argmin}\ r(\pi, \delta)$$

Remarquons que d'après la proposition (1), une estimateur peut également être défini comme étant une décision qui minimise la moyenne suivant la prédictive $f(x)$ du coût a posteriori.

3.1 Fonctions de coût usuelles

3.1.1 Coût quadratique

Définition 6 – *La fonction de coût quadratique est la fonction définie par :*

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2.$$

Une variante de cette fonction de coût est une fonction de coût quadratique pondérée de la forme : $L(\theta, \delta(x)) = w(\theta)(\theta - \delta(x))^2$.

Proposition 2 – *Sous l'hypothèse d'un coût quadratique, l'estimateur de Bayes $\delta^\pi(x)$ de θ associé à la loi a priori π est la moyenne a posteriori de θ :*

$$\delta^\pi(x) = E^{\pi(.|x)}(\theta) = \int_{\theta \in \Theta} L(\theta, \delta(x))\pi(\theta \mid x)d\theta.$$

Preuve : Par définition, l'estimateur de Bayes minimise le coût a posteriori i.e. $\rho(\pi, \delta) = E^{\pi(.|x)}[L(\theta, \delta(x))]$.
Sous l'hypothèse d'un coût quadratique, on a :

$$\begin{aligned} \rho(\pi, \delta) &= E^{\pi(.|x)}[(\theta - \delta(x))^2] \\ &= E^{\pi(.|x)}(\theta^2) - 2\delta(x)E^{\pi(.|x)}(\theta) + \delta^2(x) \end{aligned}$$

Il s'agit d'un polynôme du second degré en $\delta(x)$. Il sera minimum en $E^{\pi(.|x)}(\theta)$.

Exemple 1 (suite) : Dans cet exemple nous avions obtenu une loi gamma inverse de paramètres $(n+\alpha, \sum_{i=1}^{n} x_i + \beta)$ comme loi a posteriori. Or l'espérance mathématique d'une loi gamma inverse de paramètres (a, b) est égale à $b/(a-1)$, on a donc qu'un estimateur de Bayes $\delta^\pi(x)$ du paramètre θ, sous l'hypothèse d'un coût quadratique, est égale à :

$$\tilde{\theta} = (\sum_{i=1}^{n} x_i + \beta)/(n + \alpha - 1).$$

Autre exemple : Si $X \sim \mathcal{P}(\theta)$ et si $\pi(\theta)$, la loi a priori est une loi Gamma de paramètres (α, β), la loi a posteriori $\pi(\theta \mid x)$ est une loi Gamma de paramètres $(x+\alpha, \beta+1)$ (voir TD). Sous l'hypothèse d'un coût quadratique, un estimateur de Bayes $\delta^\pi(x)$ de θ sera l'espérance a posteriori de θ. Puisque la loi a posteriori est une loi Gamma, l'espérance est le rapport des paramètres et on a :

$$\delta^\pi(x) = (x+\alpha)/(\beta+1).$$

3.1.2 Coût absolu

Définition 7 – *La fonction de coût absolue est la fonction définie par :*

$$L(\theta, \delta(x)) = \begin{cases} k_2(\theta - \delta(x)) & \text{si } \theta > \delta(x), \\ k_1(\delta(x) - \theta) & \text{sinon.} \end{cases}$$

Proposition 3 – *Un estimateur de Bayes associé à π et au coût absolue, est un fractile d'ordre $k_2/(k_1+k_2)$ de $\pi(\theta/x)$.*

Preuve :

$$\begin{aligned} E^{\pi(./x)}[L(\theta, \delta(x))] &= \int_\Theta L(\xi, \delta(x))\pi(\xi|x)d\xi \\ &= \int_{\delta(x)}^{+\infty} k_2(\xi - \delta(x))\pi(\xi|x)d\xi + \int_{-\infty}^{\delta(x)} k_1(\delta(x) - \xi)\pi(\xi \mid x)d\xi \end{aligned}$$

On remarque que : $\pi(\xi \mid x)d\xi = dF(\xi \mid x) = -d(1 - F(\xi \mid x))$ et on écrit donc :

$$\begin{aligned} E^{\pi(.|x)}[L(\theta, \delta(x))] &= [k_2(\xi - \delta(x))(1 - F(\xi \mid x))]_{\delta(x)}^{+\infty} + \int_{\delta(x)}^{+\infty} k_2 P^{\pi(.|x)}(\theta > \xi)d\xi \\ &+ [k_1(\delta(x) - \xi)F(\xi \mid x)]_{-\infty}^{\delta(x)} + \int_{-\infty}^{\delta(x)} k_1 P^{\pi(.|x)}(\theta < \xi)d\xi \\ &= \int_{\delta(x)}^{+\infty} k_2 P^{\pi(.|x)}(\theta > \xi)d\xi + \int_{-\infty}^{\delta(x)} k_1 P^{\pi(.|x)}(\theta < \xi)d\xi \end{aligned}$$

On dérive par rapport à $\delta(x)$, il vient :

$$\begin{aligned}
\frac{\partial}{\partial\delta(x)}\rho(\pi,\delta(x)) &= -k_2P^{\pi(.|x)}(\theta>\delta(x))+k_1P^{\pi(.|x)}(\theta<\xi)=0\\
&\iff -k_2(1-P^{\pi(.|x)}(\theta>\delta(x)))+k_1P^{\pi(.|x)}(\theta<\xi)=0\\
&\iff (k_1+k_2)P^{\pi(.|x)}(\theta<\delta(x)))-k_2=0
\end{aligned}$$

d'où

$$P^{\pi(.|x)}(\theta<\delta(x))=\frac{k_2}{k_1+k_2}.$$

et le coût est donc maximisé pour $\tilde{\theta}=\delta(x)$ tel que $P^{\pi(.|x)}(\theta<\delta(x))=k_2/(k_1+k_2)$.

On rappelle : si f est intégrable sur $[a,b]$ alors $F(x)=\int_a^x f(t)dt$, $a<x<b$, est différentiable p.p. et $F'=f$ p.p.

3.1.3 Coût 0-1

Définition 8 – *On appelle coût 0-1, l'application L définie par :*

$$L(\theta,\delta(x))=\begin{cases}0 & \text{si la décision est bonne}\\ 1 & \text{sinon.}\end{cases}$$

On retrouve en utilisant cette fonction de coût, les résultats de la théorie des tests d'hypothèses. Un problème de test est un problème de choix (de prise de décision) entre $H_0:\theta\in\Theta_0$ et $H_1:\theta\in\Theta_1$.

On définit donc la décision de la manière suivante :

$\delta(X)=1$: on accepte H_0 ;

$\delta(X)=0$: on rejette H_0. (n.b. : ceci ne dépend pas de θ).

On a un espace d'actions de la forme : $A=\{0,1\}$.

Soit W la région de rejet i.e. le sous-ensemble de $\mathfrak{X}$ qui conduit à rejeter H_0. On peut construire une fonction de coût de la manière suivante : supposons $\theta\in\Theta_0$,

si $X\in W$, on prend la décision de rejeter i.e. $\delta(X)=0$, mais la décision n'est pas bonne, on va pénaliser et $L(\theta,\delta(x))=1$.

si $X\notin W$, on ne rejette pas, on prend la décision $\delta(X)=1$, la décision est bonne $L(\theta,\delta(X))=0$.

Le coût s'écrit donc :

$$L(\theta,\delta(x))=\begin{cases}1-\delta(x) & \text{si } \theta\in\Theta_0\\ \delta(x) & \text{sinon.}\end{cases}$$

Ce qu'on peut écrire : $L(\theta,\delta(x))=\mathbb{1}(x\in W)$ et on calcule la fonction de risque :

$$R(\theta,\delta)=E[L(\theta,\delta(X))]=\int_{\mathfrak{X}}L(\theta,\delta(x))dp_\theta(x)=P_\theta(x\in W),\quad \theta\in\Theta_0.$$

On retrouve le risque de première espèce.

□