Chapitre 6

Estimation bayésien

Nous avons vu que lorsqu'on dispose de la lois des observations $p(\boldsymbol{x}|\boldsymbol{\theta})$ et d'une loi *a priori* $p(\boldsymbol{\theta})$, on peut construire la loi *a posteriori* $p(\boldsymbol{\theta}|\boldsymbol{x})$ et utiliser cette loi pour inférer sur $\boldsymbol{\theta}$. Cette loi peut être utilisée non seulement pour définir un estimateur pour $\boldsymbol{\theta}$ ou pour une autre grandeur d'intérêt $h(\boldsymbol{\theta})$, mais aussi tout autres caractéristique (précision, variance, région de confiance, barre d'erreurs, etc.) de cet estimateur. On peut par exemple définir l'estimateur du maximum de vraisemblance *a posteriori* de $\boldsymbol{\theta}$

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ p(\boldsymbol{\theta}|\boldsymbol{x}) \right\},$$

le mode de $p(\boldsymbol{\theta}|\boldsymbol{x})$ qui est aussi le maximum de $l(\boldsymbol{\theta}|\boldsymbol{x})p(\boldsymbol{\theta})$.

Cette extension permet aussi de se conformer au principe de vraisemblance tout en proposant une démarche qui abouti effectivement à un estimateur, qui conserve les propriétes d'optimalité asymptotique de l'estimateur du MV classique (convergence, efficacité). En effet, en générale, lorsque la taille de l'échantillon tend vers l'infini, le terme $l(\boldsymbol{\theta}|\boldsymbol{x})$ devient prépondérant par rapport à $p(\boldsymbol{\theta})$. Ainsi les estimateurs bayésiens sont asymptotiquement équivalent aux estimateurs du MV classiques, mai qui ont de plus l'avantage d'être aussi acceptables et justifiés pour une taille d'échantillon finie.

Il faut dire que d'un point de vue bayésien, la loi *a posteriori* $p(\boldsymbol{\theta}|\boldsymbol{x})$ contient tout information que l'on souhaite sur $\boldsymbol{\theta}$ ou sur n'importe quelle autre grandeur $h(\boldsymbol{\theta})$. Par exemple, si on choisit comme estimateur de $h(\boldsymbol{\theta})$ la moyenne *a posteriori*

$$\delta(\boldsymbol{x}) = \int h(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta},$$

on peut aussi calculer, par exemple, l'erreur quadratique moyenne de cette estimateur

$$\int [\delta(\boldsymbol{x}) - h(\boldsymbol{\theta})]^2 p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta},$$

qui correspond également, dans ce cas, à la variance *a posteriori* de cette estimateur.

On peut aussi utiliser cette loi *a posteriori* pour la prédiction. En effet si une grandeur X suit la loi $p(x|\boldsymbol{\theta})$ et on a observé n échantillons $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ de X et on souhaite inférer sur les échantillons à venir $\boldsymbol{z} = \{z_1, \cdots, z_m\} = \{x_{n+1}, \cdots, x_{n+m}\}$, on peut tout d'abord calculer la loi *a posteriori*

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \frac{p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\displaystyle\int p(\boldsymbol{x}|\theta)p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}}$$

Sensibilité de la moyenne *a posteriori* vis-à-vis du choix d'une loi *a priori*

Vraisemblance $p(x\|\theta)$	Reference prior $\pi(\theta)$	Posterior mean $\mathrm{E}[\theta\|x]$
$\mathbf{N}(x\|\theta, n\lambda)$	$\mathbf{N}(\theta\|\mu, \lambda_0)$	$\mathrm{E}[\theta\|x] = \frac{n\lambda x+\lambda_0\mu}{n\lambda+\lambda_0}$,
$\mathbf{N}(x\|\theta, n\lambda)$	$\mathbf{St}(\theta\|\mu, \lambda_0, \alpha)$	$\mathrm{E}[\theta\|x] = x - \frac{(\alpha+1)(x-\mu)}{n\lambda\left[\frac{\alpha}{\lambda_0}+(x-\mu)^2\right]}$,
$\mathbf{N}(x\|\theta, n\lambda)$	$\mathbf{Exd}(\theta\|\frac{\sqrt{2}}{\nu}, \mu)$ $= \frac{\nu}{\sqrt{2}}\exp\left[-\frac{\sqrt{2}}{\nu}\|\theta-\mu\|\right]$	$\mathrm{E}[\theta\|x] = w(x)(x-b) + [1-w(x)](x-b)$, with $b = \frac{\sqrt{2}}{n\nu\Lambda}$ and $0 \leq w(x) \leq 1, x-b \leq \mathrm{E}[\theta\|x] \leq x+b$

TAB. 6.1 – *Sensibilité de la moyenne* a posteriori *vis-à-vis du choix d'une loi* a priori

et définir la distribution prédictive de $\boldsymbol{z}$ par

$$p(\boldsymbol{z}|\boldsymbol{x}) = \int p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{x})\,\mathrm{d}\boldsymbol{\theta}$$

et en déduire toute inférence sur $\boldsymbol{z}$.