Estimation bayésienne

Franck Corset

Master 2 - SSD

Définition

Cas multi-dimensionnel

- On suppose que $\theta \in \mathbb{R}^d$ avec $d > 1$ et que $\mathbf{x} = (x_1, \ldots, x_n)$ est l'échantillon observé.
- On va supposer que les lois a priori sont indépendantes.
- On appelle estimation bayésienne de θ la moyenne a posteriori, notée $\mathbb{E}[\theta|\mathbf{x}]$.
- Cette moyenne a posteriori est

$$\mathbb{E}[\theta|(x_1, \ldots, x_n)] = \int_\Theta \theta\pi(\theta|\mathbf{x})\, d\theta = \frac{\int_\Theta \theta f(\mathbf{x}|\theta)\pi(\theta)\, d\theta}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)\, d\theta}$$

- On note $\hat{\theta}_B$, l'estimateur bayésien.

Exemple 1 : Cas Gaussien

- $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma^2)$ inconnue.
- Loi a priori de μ est $\pi(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$
- Loi a priori sur σ^2 est une inverse gamma $\pi(\sigma^2) = IG(a, b)$ de densité
$$f(x) = \frac{b^a}{\Gamma(a)}(1/x)^{a+1}\exp(-b/x)$$
- Ainsi, la loi a priori $\pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2)$.

Exemple 1 : Cas Gaussien

- Loi a posteriori conditionnelle de μ sachant σ^2 :

$$\mathcal{N}(\frac{n\sigma_0^2\bar{x}_n + \sigma^2\mu_0}{n\sigma_0^2 + \sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2})$$

- La loi a posteriori conditionnelle de σ^2 sachant μ :

$$\pi(\sigma^2|\mathbf{x}) = IG(a + n/2, b + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2)$$

Algorithme de Gibbs

- Pour $\theta = (\theta_1, \dots, \theta_p)$, on veut simuler $\pi(\theta)$ à partir de $\pi_i(\theta_i|\theta_{-i})$ pour tout i.
- On initialise avec $\theta^{(0)}$ et à l'étape k, on a :

$$
\begin{aligned}
(\theta_1^{(k)}|\theta^{(k-1)}) &\sim \pi_1(\theta_1^{(k)}|\theta_{-1}^{(k-1)}) \\
(\theta_2^{(k)}|\theta^{(k-1)}, \theta_1^{(k)}) &\sim \pi_2(\theta_2^{(k)}|\theta_1^{(k)}, \theta_3^{(k-1)}, \dots, \theta_p^{(k-1)}) \\
\vdots &\sim \vdots \\
(\theta_p^{(k)}|\theta_{-p}^{(k)}) &\sim \pi_p(\theta_p^{(k)}|\theta_{-p}^{(k)})
\end{aligned}
$$

- On génère ainsi une chaine de Markov dont la loi stationnaire est π.

Algorithme de Gibbs (exemple pour le cas gaussien)

- On initialise en choisissant un couple $(\mu^{(0)}, (\sigma^2)^{(0)})$
- A l'étape k, on simule $\mu^{(k)}$ selon la loi a posteriori conditionnelle $\pi(\mu|(\sigma^2)^{(k-1)}, \mathbf{x})$:

$$\mathcal{N}\left(\frac{n\sigma_0^2\bar{x}_n + (\sigma^2)^{(k-1)}\mu_0}{n\sigma_0^2 + (\sigma^2)^{(k-1)}}, \frac{(\sigma^2)^{(k-1)}\sigma_0^2}{n\sigma_0^2 + (\sigma^2)^{(k-1)}}\right)$$

- puis on simule $(\sigma^2)^{(k)}$ selon la loi conditionnelle $\pi(\sigma^2|\mu^{(k)}, \mathbf{x})$:

$$IG(a + n/2, b + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu^{(k)})^2)$$

- On s'arrête après un grand nombre d'itérations K.
- L'échantillon produit suit alors la loi a posteriori voulue.

Algorithme de Metropolis Hasting

- On se donne une valeur initiale $\theta^{(0)}$ puis à l'étape k : On simule $\theta^{(k)}$ en tirant θ' à l'aide d'une distribution instrumentale, notée $q(.|\theta^{(k-1)})$ et on choisit :

$$\begin{aligned} \theta^{(k)} &= \theta' \text{ avec une probabilité } \alpha \\ &= \theta^{(k-1)} \text{ avec une probabilité 1- } \alpha \end{aligned}$$

où $\alpha = min(\frac{\pi(\theta')}{\pi(\theta^{(k-1)})} \frac{q(\theta^{(k-1)}|\theta')}{q(\theta'|\theta^{(k-1)})}, 1)$.

- Pour la loi instrumentale, on choisira souvent une loi symétrique, par exemple la loi normale centrée en $\theta^{(k-1)}$ à l'étape k, ce qui revient à prendre $\alpha = min(\frac{\pi(\theta')}{\pi(\theta^{(k-1)})}, 1)$, c'est-à-dire vu comme le rapport des lois a posteriori entre le candidat (θ') et l'ancienne valeur $(\theta^{(k-1)})$.