TD 3 - Estimation Bayésienne.

Exercice 3.1. Estimateur bayésien. On considère une variable aléatoire X de loi conditionnelle à θ, $X|\theta \sim \mathcal{N}(\theta,\ \sigma^2)$ où le paramètre θ est lui-même distribué selon la loi $\mathcal{N}(\mu, \tau^2)$. On parle de modèle hiérarchique.

1. Donner la loi a posteriori de θ et en déduire l'estimateur bayésien de θ.

2. Étudier et commenter les régimes $\sigma^2 \to +\infty$ et $\tau^2 \to +\infty$, les autres paramètres étant fixés.

3. Reprendre la question 1) avec cette fois l'observation d'un échantillon constitué de variables $X_1, \cdots, X_n$ indépendantes conditionnellement à θ et de même loi que X conditionnellement à θ.

4. Étudier et commenter le régime asymptotique $n \to +\infty$, les autres paramètres étant fixés.

Solution:

1. La loi jointe de X_1 et θ est

$$p_{(X_1,\theta)}(x,\ u) = p(X_1|\theta = u)(x)p_\theta(u)$$

et est donc proportionnelle à

$$\exp(-\frac{1}{2\sigma^2}(x_1 - u)^2)\exp(-\frac{1}{2\tau^2}(u-\mu)^2)\ .$$

La distribution a posteriori s'obtient en effectuant le rapport $p_{(X_1,\theta)}/p_X$ donc s'écrit génériquement $C(x)p_{X|\theta=u}(x)p_\theta(u)$. En incluant dans la constante générique $C(x)$ tout ce qui ne dépend pas de u, on trouve après calculs

$$p_{\theta|X_1=x}(u) = C(x)\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)\left(u^2 - 2u\frac{x/\sigma^2+\mu/\tau^2}{1/\sigma^2+1/\tau^2}\right)\right]$$

On trouve donc une loi gaussienne de moyenne $\dfrac{x/\sigma^2+\mu/\tau^2}{1/\sigma^2+1/\tau^2}$ et de variance $\dfrac{1}{1/\sigma^2+1/\tau^2}$. Ce sont respectivement l'espérance conditionnelle et la variance conditionnelle de θ sachant $X_1 = x_1$:

On peut donc prendre pour estimateur bayésien ponctuel $\hat{\theta}^B$ l'espérance de la distribution a posteriori (pour la fonction de perte quadratique):

$$\hat{\theta}^B = \frac{X_1/\sigma^2+\mu/\tau^2}{1/\sigma^2+1/\tau^2}$$

Cet estimateur ponctuel a donc pour espérance:

$$\mathbb{E}\left(\hat{\theta}^B\right) = \frac{\mu/\sigma^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2} = \mu$$

2. Lorsque $\sigma \to \infty$, $\hat{\theta}^B$ tend vers μ (c'est-à-dire la moyenne a priori) ce qui s'interprète par le fait que l'observation est très dispersée, donc peu informative par rapport au prior. Au contraire, si c'est la variance de la distribution a priori qui tend vers l'infini, c'est l'observation qui devient déterminante et on obtient pour $\hat{\theta}^B$ la valeur observée X_1.

3. La loi jointe de $S = (X_1, \cdots, X_n)$ et θ est

$$p_{(S,\theta)}(s,\ u) = p_{S|\theta=u}(s)p_\theta(u)$$

et est donc proportionnelle à

$$\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - u)^2)\exp(-\frac{1}{2\tau^2}(u-\mu)^2)\ .$$

La distribution a posteriori s'obtient en effectuant le rapport $p_{(S,\theta)}/p_S(s)$ donc s'écrit génériquement $C(s)p_{S|\theta}(s)$. En incluant dans la constante générique $C(s)$ tout ce qui ne dépend pas de u, on trouve après calculs

$$p_{\theta|S=s}(\theta) = C(s)\exp[-\frac{1}{2}(\frac{n}{\sigma^2} + \frac{1}{\tau^2})(u^2 - 2u\frac{n\overline{x}_n/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2})]$$

Solution: (suite) On reconnait le noyau gaussien de moyenne $\frac{n\overline{x}_n/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}$ qui est donc l'espérance a posteriori et donc l'estimateur bayésien est

$$\hat{\theta}_n^B = \frac{n\overline{X}_n/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}.$$

4. Lorsque n tend vers l'infini, l'estimateur bayésien est équivalent à l'estimateur fréquentiste $\overline{X}_n$ et converge en probabilité vers $\mathbb{E}(X_1)$, ce qui est attendu: le poids de l'a priori devient négligeable devant celui des observations.

Exercice 3.2. Estimation bayésienne pour la roue de la fortune. Un premier joueur fait tourner une roue graduée (du type roue de la fortune) aléatoirement et obtient l'angle α, $\alpha \in [0, 2\pi]$. N autres joueurs (N connu) font à leur tour tourner la roue. Parmi ceux-ci, x obtiennent un score (un angle) inférieur à celui du premier joueur. Déterminer la distribution de probabilité a posteriori de α.

Solution:

Modèle pour la vraisemblance: Soit X, la variable aléatoire correspondant au nombre de joueurs obtenant un score inférieur au premier joueur. On a que $X|\alpha \sim B(N, \alpha/2\pi)$. Donc $P(X = x|\alpha) = C_N^x \left(\frac{\alpha}{2\pi}\right)^x \left(1 - \frac{\alpha}{2\pi}\right)^{N-x}$.

Prior: on suppose qu'il n'y a pas de triche a priori et que $\alpha \sim \mathcal{U}_{[0,2\pi]}$.

Posterior: par la formule de Bayes, nous obtenons alors:

$$\begin{aligned} P(a < \alpha < b \text{ et } X = x) &= P(a < \alpha < b \mid X = x)P(X = x) \\ &= P(X = x \mid a < \alpha < b)P(a < \alpha < b) \end{aligned}$$

(On passe par un intervalle $[a, b]$ pour simplifier, car sinon il faut prendre des précautions dans le conditionnement continue vs discrète). Soit:

$$P(a < \alpha < b \mid X = x) = \frac{P(X = x\ ,\ a < \alpha < b)}{P(X = x)}$$

Avec $P(X = x)$ la constante de normalisation:

$$P(X = x) = \int_0^{2\pi} C_N^x \left(\frac{\alpha}{2\pi}\right)^x \left(1 - \frac{\alpha}{2\pi}\right)^{N-x} \frac{d\alpha}{2\pi}$$

et

$$\begin{aligned} P(X = x\ ,\ a < \alpha < b) &= \int_a^b P(X = x|\alpha)p(\alpha)d\alpha \\ &= \int_a^b C_N^x \left(\frac{\alpha}{2\pi}\right)^x \left(1 - \frac{\alpha}{2\pi}\right)^{N-x} \frac{d\alpha}{2\pi} \end{aligned}$$

Soit finalement en posant $U = \alpha/2\pi$:

$$P(\frac{a}{2\pi} < U < \frac{b}{2\pi} \mid X = x) \propto \int_{\frac{a}{2\pi}}^{\frac{b}{2\pi}} u^x (1 - u)^{N-x} du$$

On reconnaît une fonction proportionnelle à la densité d'une loi bêta, donc $U|X = x \sim \mathcal{B}(x + 1, N - x + 1)$.

Exercice 3.3. **Distribution a priori de Jeffreys.** Dans l'estimation bayésienne, quand il n'y a pas de connaissance a priori disponible dont on peut déduire un prior $p(\theta)$ pour le paramètre θ, on peut utiliser des priors dits 'non-informatifs'. Une méthode intéressante pour leur construction est celle de Jeffreys: si $\theta \in \mathbb{R}$, on choisit

$$p(\theta) \propto I(\theta)^{1/2}$$

et si $\theta \in \mathbb{R}^p$, on prend

$$p(\theta) \propto [\det I(\theta)]^{1/2} \ ,$$

où $I(\theta)$ est la matrice d'information de Fisher pour l'observation x, c'est à dire (sous certaines conditions de régularité):

$$I(\theta) = E_\theta\left[\left(\frac{\partial \ \ln\ p(x|\theta)}{\partial\theta}\right)^2\right] = -E_\theta\left[\frac{\partial^2 \ \ln\ p(x|\theta)}{\partial\theta^2}\right]$$

Déterminer les distributions a priori de Jeffreys pour θ pour:

1) la loi binomiale, $X \sim B(n, \theta)$,

2) la loi normale, $X \sim \mathcal{N}(\mu, \sigma^2)$, en prenant $\theta = (\mu, \sigma)$.

Solution:

1)

$$p(x|\theta) = C_n^x \theta^x (1-\theta)^{n-x}$$

Et donc

$$\frac{\partial^2\, ln\, p(x|\theta)}{\partial\theta^2} = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$$

En prenant l'espérance et en se rappelant que $E(x) = n\theta$, nous obtenons:

$$I(\theta) = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}\ .$$

On a donc le prior de Jeffreys:

$$p(\theta) \propto (\theta(1-\theta))^{-1/2}$$

où l'on reconnait une fonction proportionnelle à la densité d'une loi bêta $\mathcal{B}e(1/2, 1/2)$, on prend donc comme prior une $\mathcal{B}e(1/2, 1/2)$.

2) Ici on a en prenant les dérivées partielles secondes et croisées (pour μ et σ):

$$I(\theta) = E_\theta\left[\begin{pmatrix} 1/\sigma^2 & 2(x-\mu)/\sigma^3 \\ 2(x-\mu)/\sigma^3 & 3(\mu-x)^2/\sigma^4 - 1/\sigma^2 \end{pmatrix}\right]\ .$$

Comme $E_\theta(x-\mu) = 0$ et $E_\theta(x-\mu)^2 = \sigma^2$, finalement:

$$I(\theta) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}\ .$$

et on obtient en prenant la racine du déterminant:

$$p(\theta) \propto \frac{1}{\sigma^2}$$

C'est un prior impropre ...