M2 - Bayésien non-paramétrique

Examen, le 05/04/2018

Durée 3h. Les documents ne sont pas autorisés.

Questions de cours

1. Définir en une ou deux lignes pour une loi a posteriori $\Pi[\cdot \mid X_1, \ldots, X_n]$, sans redéfinir le cadre et les notations, la notion de convergence à vitesse $\varepsilon_n \to 0$.
2. Interpréter les trois hypothèses du théorème GGV rappelé en annexe (1-2 lignes chaque).
3. Dans le modèle fondamental $X \mid \theta \sim \mathcal{N}(\theta, 1)^{\otimes n}$, avec pour loi a priori $\theta \sim \mathcal{N}(a, 1) = \Pi$, où $a \in \mathbb{R}$ est fixé, déterminer la loi a posteriori $\Pi[\cdot \mid X]$. Le théorème BvM est-il vérifié ici ?

Problème

On se place dans le modèle d'estimation de densité sur $[0, 1]$. Dans le cadre bayésien, pour $X = X^{(n)} = (X_1, \ldots, X_n)$,

$$\begin{aligned} X \mid f \sim & \; P_f^{\otimes n}, \\ f \sim & \; \Pi, \end{aligned}$$

où $dP_f(x) = f(x)dx$ avec f densité sur l'intervalle $[0, 1]$, et Π une loi a priori sur les densités à choisir.

Notations. Pour $K \geq 1$ un entier, on note $I_1 = [0, \frac{1}{K}]$ et $I_j = (\frac{j-1}{K}, \frac{j}{K}]$ pour $j = 2, \ldots, K$ la subdivision régulière de $[0, 1]$ en K blocs. On note $\mathcal{H}_K$ l'ensemble des histogrammes sur cette subdivision, et $\mathcal{H}_K^1$ les éléments de $\mathcal{H}_K$ qui sont des densités. On note $\mathcal{D} = \{f : [0, 1] \to (c, M] \;\text{ mesurable}, \; \int_0^1 f(u)du = 1\}$ pour $0 < c \leq M < \infty$ des constantes fixées. Pour $\beta \in (0, 1]$, on pose

$$H_D(\beta) = \{f \in \mathcal{D}, \; |f(x) - f(y)| \leq D|x - y|^\beta, \; \forall \, x, y \in [0, 1]\}.$$

On considère la loi a priori Π sur les densités engendrée par, pour $x \in [0, 1]$ et $a > 0$,

$$f(x) = \sum_{k=1}^{K} Y_k K \mathbb{1}_{I_k}(x), \quad (Y_1, \ldots, Y_K) \sim \text{Dir(a,a,...,a)},$$

où K est un entier à choisir dans la suite en fonction de n.

La loi a posteriori $\Pi[\cdot \mid X]$ est la loi de f sachant X. On étudie cette loi en probabilité sous P_{f_0}, c'est-à-dire sous l'hypothèse qu'il existe f_0, "vraie" valeur du paramètre, telle que $X \sim P_{f_0}^{\otimes n}$. Dans la suite, on supposera

$$f_0 \in H_D(\beta).$$

1. Pour $f \in L^2[0,1]$, montrer que sa projection orthogonale f_K pour la norme $\|\cdot\|_2$ sur $\mathcal{H}_K$ est
$$f_K = \sum_{k=1}^{K} \bar{f}_k 1\!\!1_{I_k}, \quad \bar{f}_k = K \int_{I_k} f.$$
2. Pour $f \in H_D(\beta)$ montrer que, pour D_1 à déterminer, et $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$,
$$\|f - f_K\|_\infty \leq D_1 K^{-\beta}.$$
3. On pose
$$\mathcal{F}_{n,K} = \left\{ f \in \mathcal{H}_K^1,\ f = \sum_{k=1}^{K} a_k 1\!\!1_{I_k}, \quad \max_{1\leq k\leq K} a_k \leq K \right\}.$$
Vérifier que $\Pi[\mathcal{F}_{n,K}] = 1$.
4. On *admet* que, pour $\|\cdot\|_K$ la norme euclidienne sur $\mathbb{R}^K$, soit $\|u\|_K^2 = \sum_{k=1}^K u_k^2$, et $B_\infty(r) = \{u \in \mathbb{R}^K,\ \max_{1\leq k\leq K} |u_k| \leq r\}$,
$$N(\varepsilon_n, \mathcal{F}_{n,K}, h) \leq N(\varepsilon_n, B_\infty(\sqrt{K}), \|\cdot\|_K).$$
   (a) Montrer que $B_\infty(\sqrt{K}) \subset B_2(K)$, où $B_2(K)$ est définie dans l'annexe.
   (b) En déduire une majoration de $N(\varepsilon_n, \mathcal{F}_{n,K}, h)$ en fonction de K et ε_n.
5. Montrer que pour $f \in \mathcal{H}_K^1$, $f = \sum_{k=1}^K a_k 1\!\!1_{I_k}$, on a $\bar{f}_k = a_k$, puis que
$$\|f - f_0\|_\infty \leq \sum_{k=1}^{K} |\bar{f}_k - \bar{f}_{0,k}| + \|f_0 - f_{0,K}\|_\infty.$$
6. On *admet* l'inclusion suivante, pour $C > 0$ une constante,
$$\{f \in \mathcal{H}_K^1,\ \|f - f_0\|_\infty \leq C\varepsilon_n\} \subset B_{KL}(f_0, \varepsilon_n).$$
   (a) En utilisant les questions précédentes, en déduire que
$$\Pi[B_{KL}(f_0,\varepsilon_n)] \geq P\left(K \sum_{k=1}^{K} \left| Y_k - \frac{\bar{f}_{0,k}}{K} \right| + D_1 K^{-\beta} \leq C\varepsilon_n \right).$$
   (b) En déduire, en utilisant la propriété de la loi de Dirichlet donnée en Annexe, une minoration de $\Pi[B_{KL}(f_0,\varepsilon_n)]$ en fonction de K et ε_n (on pourra supposer K assez petit de sorte que $D_1 K^{-\beta} \leq C\varepsilon_n/2$).
7. Ecrire les conditions sur K, ε_n qui permettent de vérifier les hypothèses du théorème GGV en utilisant ce qui précède.
8. Dans cette question seulement, on suppose β connu et on pose $K = \lfloor n^{\frac{1}{1+2\beta}} \rfloor$. Montrer qu'à un facteur logarithmique près on peut prendre ε_n de l'ordre de $n^{-\beta/(2\beta+1)}$.
9. On pose maintenant $K = \lfloor n^{\frac{1}{1+2\alpha}} \rfloor$ pour $\alpha \in (0,1]$.

(a) Déterminer une vitesse ε_n vérifiant les hypothèses du théorème GGV en fonction de α, β et n.

(b) Quel choix de α donne la vitesse la plus rapide ? Commenter.

(c) A votre avis, pour l'a priori Π considéré, la vitesse de convergence ε_n ci-dessus est-elle améliorable ? Quel type de résultat pourrait-on chercher à établir pour vérifier votre intuition ?

10. Dans cette question, on définit une nouvelle loi a priori Π par le schéma

$$\begin{aligned} f \mid K \sim & \sum_{k=1}^{K} Y_k K 1\!\!1_{I_k}, \qquad (Y_1, \dots, Y_K) \sim \text{Dir(a,a,...,a)} \\ K \sim & \ \pi_K, \end{aligned}$$

où π_K est la loi sur les entiers strictement positifs définie par $\pi_K(k) = ce^{-k \log k}, k \geq 1$, avec c constante de proportionnalité.

(a) On pose, pour K_n à choisir,

$$\mathcal{F}_n = \bigcup_{k=1}^{K_n} \mathcal{F}_{n,k}.$$

Vérifier que $\Pi[f \notin \mathcal{F}_n] = \pi_K[K > K_n]$.

(b) Déterminer une vitesse ε_n vérifiant les hypothèses du théorème GGV pour cette nouvelle loi a priori Π.

(c) Commenter.

Exercice 1

On considère le modèle de suite gaussienne, pour $n \geq 1$,

$$X_k = \theta_k + \frac{\varepsilon_k}{\sqrt{n}}, \qquad k \geq 1,$$

où $\theta = (\theta_1, \theta_2, \dots)$ appartient à $\ell_2(\mathbb{N}^*)$ et $\varepsilon_1, \varepsilon_2, \dots$ sont des variables i.i.d. de loi $\mathcal{N}(0,1)$. Ainsi, le modèle s'écrit $\mathcal{P} = \{P_\theta^{(n)},\ \theta \in \ell_2(\mathbb{N}^*)\}$, où

$$P_\theta^{(n)} = \bigotimes_{k \geq 1} \mathcal{N}(\theta_k, n^{-1}).$$

On pose $X = (X_1, X_2, \dots)$ et on note, pour deux suites de réels $a = (a_k), b = (b_k)$,

$$\langle a, b \rangle = \sum_{k \geq 1} a_k b_k,$$

lorsque la série correspondante converge, et $\|a\|^2 = \langle a, a \rangle = \sum_{k \geq 1} a_k^2$.

Soient θ_1, θ_0 dans $\ell_2(\mathbb{N}^*)$. On note $B(\theta_1, R) = \{\theta \in \ell_2(\mathbb{N}^*),\ \|\theta - \theta_1\| \leq R\}$ et on pose

$$r = \|\theta_0 - \theta_1\|/4.$$

On définit un test ϕ_n par

$$\phi_n = 1\!\!1_{2\langle \theta_1 - \theta_0, X\rangle > \|\theta_1\|^2 - \|\theta_0\|^2}.$$

Enfin, on note E_θ l'espérance sous la loi $P_\theta^{(n)}$ et $\bar{\Phi}(u) = P[\mathcal{N}(0,1) > u]$.

1. Montrer que
$$E_{\theta_0}\phi_n \leq \bar{\Phi}(\sqrt{n}\|\theta_1 - \theta_0\|/2).$$
2. Montrer que pour tout $\theta \in B(\theta_1, r)$, on a
$$\langle \theta_1 - \theta_0, \theta - \theta_0\rangle \geq \frac{3}{4}\|\theta_1 - \theta_0\|^2.$$
3. En déduire que
$$\sup_{\theta \in B(\theta_1, r)} E_\theta(1 - \phi_n) \leq \bar{\Phi}(\sqrt{n}\|\theta_1 - \theta_0\|/4).$$
4. Quelle est l'utilité d'un résultat de ce type ? (on rappelle $\bar{\Phi}(u) \leq e^{-u^2/2}$ si $u \geq 0$)

Exercice 2

Soit $\mathcal{P} = \{P_f^{\otimes n},\ f \in \mathcal{F}\}$ le modèle d'estimation de densité, avec $\mathcal{F}$ un sous-ensemble de densités et d une distance sur $\mathcal{F}$. Soit Π une loi a priori sur $\mathcal{F}$. On note $X = X^{(n)} = (X_1, \ldots, X_n)$. On observe X avec $X \,|\, f \sim P_f^{\otimes n}$. Soit $B(f, r) = \{h \in \mathcal{F},\ d(h, f) \leq r\}$ la boule fermée centrée en f de rayon r.

Pour tout $f \in \mathcal{F}$, on note

$$r_n(f) = \inf\{r > 0,\ \Pi[B(f, r) \,|\, X] \geq 1/2\}.$$

On suppose que $\Pi[\cdot \,|\, X]$ converge sous P_{f_0} vers f_0 pour la distance d à vitesse ε_n.

1. Pour $g \in \mathcal{F}$ fixée, montrer que $r \to \Pi[B(g, r) \,|\, X]$ s'interprète comme la fonction de répartition d'une variable aléatoire (laquelle ? on pourra noter Y une variable aléatoire à valeurs dans $\mathcal{F}$ tirée sous la loi $\Pi[\cdot \,|\, X]$). En déduire que
$$\Pi[B(g, r_n(g)) \,|\, X] \geq 1/2.$$
2. Soit f_n un point de $\mathcal{F}$ tel que
$$r_n(f_n) \leq \inf_{f \in \mathcal{F}} r_n(f) + \varepsilon_n.$$
Montrer que, sur un événement de probabilité qui tend vers 1, l'intersection $B(f_0, \varepsilon_n) \cap B(f_n, r_n(f_n))$ est non vide [*on pourra considérer sa masse a posteriori*].
3. Soit $\hat{f}_n$ le centre de la boule de plus petit rayon contenant encore au moins la moitié de la masse a posteriori. Montrer que, sur un événement de probabilité qui tend vers 1, on a $d(\hat{f}_n, f_0) \leq 3\varepsilon_n$.

Annexe – Notations et rappels

Lois de Dirichlet.

- La loi de Dirichlet Dir(α)=Dir($a_1, \ldots, a_k$) avec $\alpha = (a_1, \ldots, a_k) \in \mathbb{R}_*^{+k}$ est la loi sur le simplexe $\mathcal{S}_k = \{x = (x_1, \ldots, x_k) \in [0,1]^k,\ \sum_{i=1}^k x_i = 1\}$ de densité

$$x \to \frac{1}{B(\alpha)} \prod_{i=1}^{k} x_i^{a_i - 1} 1\!\!1_{\mathcal{S}_k}(x), \qquad B(\alpha) = \frac{\prod_{i=1}^k \Gamma(a_i)}{\Gamma(\sum_{i=1}^K a_i)}.$$

- Soit y un point quelconque du simplexe S_k. Soit $\varepsilon > 0$ tel que $\varepsilon > 1/k$. Alors, si $(Y_1, \ldots, Y_k) \sim$ Dir$(a, \ldots, a)$, et si $a \geq \varepsilon$, on peut montrer que, pour des constantes universelles $c_1, C_1 > 0$,

$$P\left(\sum_{i=1}^{k} |Y_i - y_i| \leq 2\varepsilon\right) \geq C_1 \exp\left(-c_1 k \log \frac{1}{\varepsilon}\right).$$

Entropie.

- L'entropie $N(\varepsilon, \mathcal{F}, d)$ est le nombre minimal de boules de rayon ε pour la distance d nécessaires pour recouvrir $\mathcal{F}$.
- Si $B_2(M) = \{x \in \mathbb{R}^K,\ \|x\|_K \leq M\}$ est la boule de $\mathbb{R}^K$ centrée en 0 de rayon M pour la norme euclidienne $\|x\|_K^2 = \sum_{k=1}^K x_k^2$, on a $N(\varepsilon, B_2(M), \|\cdot\|) \leq (1 + \frac{2M}{\varepsilon})^K$ pour tout $K \geq 1$ et tous $M, \varepsilon > 0$.

Théorème GGV en estimation de densité.

Dans le modèle d'estimation de densité du Problème, le théorème GGV s'énonce comme suit. Soit h la distance de Hellinger sur les densités

$$h(f,g) = \left(\int_0^1 (\sqrt{f} - \sqrt{g})^2(u) du\right)^{1/2}.$$

Soit ε_n une suite telle que $n\varepsilon_n^2 \to \infty$ et $\varepsilon_n \to 0$ quand $n \to \infty$. Soit $\mathcal{F}$ l'ensemble des densités sur $[0,1]$ et soit $f_0 \in \mathcal{F}$. Supposons qu'il existe des ensembles $\mathcal{F}_n \subset \mathcal{F}$, et $C, D > 0$, tels que

$$\begin{aligned} \log N(\varepsilon_n, \mathcal{F}_n, h) &\leq D n \varepsilon_n^2 \\ \Pi[\mathcal{F} \setminus \mathcal{F}_n] &\leq e^{-(C+4)n\varepsilon_n^2} \\ \Pi[B_{KL}(f_0, \varepsilon_n)] &\geq e^{-Cn\varepsilon_n^2}. \end{aligned}$$

Alors pour M assez grand, quand $n \to \infty$,

$$E_{f_0} \Pi[\{f : h(f, f_0) \leq M\varepsilon_n\} \,|\, X] \to 1.$$