Statistique non-asymptotique
Partie 1 (statistique non-paramétrique)

Examen

13 novembre 2017, 8h-10h.

Les appareils électroniques et les documents ne sont pas autorisés. Le sujet comporte **deux** pages.

Dans tout l'examen, on considère $(X_i)_{i=1,\ldots,n}$ un échantillon de variables aléatoires réelles indépendantes identiquement distribuées de densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$, inconnue.

Question de cours (2.5 points).

On suppose que $f \in L^2(\mathbb{R})$. Soit un sous espace vectoriel $S_D \subset L^2(\mathbb{R})$ engendré par une base orthonormée $\{\varphi_1, \ldots, \varphi_D\}$, (avec $\varphi_j : \mathbb{R} \to \mathbb{R}$). Pour toute fonction $t \in \mathbb{L}^2(\mathbb{R})$, on définit la fonction de contraste suivante :

$$\gamma_n(t) = \|t\|^2_{L^2(\mathbb{R})} - \frac{2}{n}\sum_{i=1}^n t(X_i).$$

Démontrer que $\hat{f}_D = \arg\min_{t \in S_D} \gamma_n(t)$ est défini de façon unique et satisfait $\hat{f}_D = \sum_{j=1}^D \hat{a}_j \varphi_j$, avec $\hat{a}_j = n^{-1}\sum_{i=1}^n \varphi_j(X_i)$, $j = 1, \ldots, D$.

Exercice. Estimation des dérivées d'une densité (7.5 points + 2 points hors barème).

Soient $\alpha, L \in \mathbb{R}_+^*$. On suppose que f est bornée sur $\mathbb{R}$ et appartient à la classe de Hölder $\mathcal{H}_{\mathbb{R}}(\alpha, L)$. On rappelle que cela signifie, en notant $\ell = \lfloor \alpha \rfloor$, que f est $\ell-$fois dérivable sur $\mathbb{R}$ et, quels que soient $x, x' \in \mathbb{R}$,

$$\left| f^{(\ell)}(x) - f^{(\ell)}(x') \right| \leq L|x - x'|^{\alpha - \ell}.$$

On cherche à estimer la dérivée $f^{(s)}$ d'ordre s de f pour tout entier $s < \ell$. Pour cela, on pose, pour tout réel $h > 0$, et tout entier $s < \ell$

$$\widehat{f}_{h,s}(x) = \frac{1}{nh^{s+1}}\sum_{i=1}^n Q\left(\frac{X_i - x}{h}\right),\ x \in \mathbb{R},$$

où Q est une fonction telle que,

(i) $Q \in L^2(\mathbb{R})$;

(ii) pour tout $j \in \{0, 1, \ldots, \ell\}$, la fonction $u \mapsto u^j Q(u)$ est intégrable ;

(iii) pour tout $j \in \{0, 1, \ldots, \ell\}\backslash\{s\}$, $\int_{\mathbb{R}} u^j Q(u)du = 0$;

(iv) $\int_{\mathbb{R}} u^s Q(u)du = s!$.

On considère le risque quadratique ponctuel en $x_0 \in \mathbb{R}$ fixé de $\widehat{f}_{h,s}$ pour l'estimation de $f^{(s)}$

$$MSE(x_0, \widehat{f}_{h,s}) = \mathbb{E}\left[\left(\widehat{f}_{h,s}(x_0) - f^{(s)}(x_0)\right)^2\right].$$

0. Est-ce que la fonction Q est un noyau ?

1. Montrer que $MSE(x_0, \widehat{f}_{h,s}) = B^2(\widehat{f}_{h,s}(x_0)) + \mathrm{Var}(\widehat{f}_{h,s}(x_0))$, avec

$$\begin{cases} B^2(\widehat{f}_{h,s}(x_0)) = \left(\mathbb{E}\left[\widehat{f}_{h,s}(x_0)\right] - f^{(s)}(x_0)\right)^2, \\ \mathrm{Var}(\widehat{f}_{h,s}(x_0)) = \mathbb{E}\left[\left(\widehat{f}_{h,s}(x_0) - \mathbb{E}\left[\widehat{f}_{h,s}(x_0)\right]\right)^2\right]. \end{cases}$$

2. Montrer qu'il existe une constante $C_V > 0$ telle que

$$\mathrm{Var}(\widehat{f}_{h,s}(x_0)) \leq \frac{C_V}{nh^{2s+1}}.$$

3. On souhaite maintenant majorer $B^2(\widehat{f}_{h,s}(x_0))$.

(a) Montrer que $\mathbb{E}\left[\widehat{f}_{h,s}(x_0)\right] = \frac{1}{h^s}\int_{\mathbb{R}} Q\,(u)\,f(x_0+uh)du$.

(b) Montrer que

$$\mathbb{E}\left[\widehat{f}_{h,s}(x_0)\right] = f^{(s)}(x_0) + \frac{1}{h^s}\int_{\mathbb{R}} Q(u)\frac{(uh)^\ell}{\ell!}f^{(\ell)}(x_0+\tau uh)du,$$

pour un certain réel $\tau \in]0,1[$. On pourra effectuer un développement de Taylor-Lagrange et utiliser les hypothèses sur la fonction Q.

(c) Conclure qu'il existe une constante $C_B > 0$, telle que

$$B^2(\widehat{f}_{h,s}(x_0)) \leq C_B h^{2(\alpha-s)}.$$

4. Déduire des questions précédentes qu'il existe une valeur $h*$ de h bien choisie, dépendant de α pour laquelle

$$MSE(x_0, \widehat{f}_{h*,s}) \leq Cn^{-\frac{2(\alpha-s)}{2\alpha+1}}.$$

5. Comparer la vitesse de convergence de $\widehat{f}_{h,s}$ pour l'estimation de $f^{(s)}$ avec celle de l'estimateur à noyau classique de la densité f.

6. **(Bonus - hors barème)** Soit $(Q_m)_{m\in\mathbb{N}}$ la famille des polynômes de Legendre. On rappelle que cette famille de polynômes forme une base hilbertienne de $L^2([-1,1])$, est échelonnée en degré (Q_m est de degré m pour tout $m \in \mathbb{N}$). Montrer que la fonction

$$Q\,:\,u\in\mathbb{R}\longmapsto\sum_{m=0}^{\ell} Q_m^{(s)}(0)Q_m(u)\mathbf{1}_{u\in[-1,1]}$$

vérifie les conditions (i) à (iv).