cc le 7/11/07

Contrôle continu. Cours de M2 "Statistique non paramétrique."

F. Comte.

3 Octobre 2007, à rendre pour dans deux semaines.

On considère le modèle de régression : $Y_i = m(X_i) + \varepsilon_i$, $i = 1, \ldots, n$. Les $(\varepsilon_i)_{1 \le i \le n}$ sont i.i.d. centrés de variance commune s_ε^2 et les $(X_i)_{1\le i\le n}$ sont i.i.d. de densité commune f. De plue, les $(X_i)_{1\le i\le n}$ et les $(\varepsilon_i)_{1\le i\le n}$ sont indépendantes. On dispose d'un échantillon d'observations $(X_i, Y_i)_{1\le i\le n}$. On se propose d'estimer la fonction $\ell = mf$ au moyen de l'estimateur

$$\hat{\ell}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h}\right)$$

où K est un noyau.

1)a) Donner deux exemples de noyaux.

b) Choisissez une fonction m et une loi f, des ε_i gaussiens, $s_\varepsilon^2 = 0.25$, $n = 1000$. Donner un programme de simulation du modèle. centré

c) Choisissez K, h. Donner un programme de calcul de l'estimateur et une représentation graphique superposant la vraie et l'estimée.

2) Montrer que l'erreur quadratique ponctuelle $\mathbb{E}[(\hat{\ell}_n(x) - \ell(x))^2] = MSE(x)$ se décompose en une somme de deux termes : un terme de variance $\mathrm{Var}(\hat{\ell}_n(x)) = \sigma^2(x)$ plus un terme de biais au carré $[\mathbb{E}(\hat{\ell}_n(x)) - \ell(x)]^2 = b^2(x)$.

3) Calculer $\mathbb{E}(\hat{\ell}_n(x))$ en fonction de K, h, ℓ.

4) Montrer que si f est bornée par M_f, $|m|$ est bornée par M, K est un noyau tel que $\int K(u)du = 1$ et $\int K^2(u)du < +\infty$, alors $\sigma^2(x) \le M_f(M^2 + s_\varepsilon^2)\int K^2(u)du/(nh)$.

5) On note $\lfloor \beta \rfloor$ le plus grand entier strictement inférieur à β. On définit la classe de Hölder $\Sigma(\beta, L)$ sur un intervalle I comme l'ensemble des fonctions $h : I \to \mathbb{R}$ telles que la dérivée d'ordre p, $h^{(p)}$ existe, où $p = \lfloor \beta \rfloor$, et vérifie :

$$|h^{(p)}(x) - h^{(p)}(y)| \le L|x-y|^{\beta - p}, \forall x, y \in I.$$

Montrer que si ℓ appartient à une classe $\Sigma(\beta, L)$, si K est un noyau d'ordre $p = \lfloor \beta \lfloor$ (i.e. $\int K(u)du = 1$ et $\int u^k K(u)du = 0$, pour $u = 1, \ldots p$) et si $\int |u|^\beta ||K(u)|du < +\infty$, alors,

$$\forall x \in \mathbb{R}, \forall h > 0, \forall n \ge 1,\ |b(x)| \le C_2 h^\beta,\ \text{où}\ C_2 = \frac{L}{p!}\int |u|^\beta||K(u)|du.$$

6) En déduire une majoration de $MSE(x)$. Montrer que, si $\beta > 1/2$, cette quantité majorante est minimale pour une valeur h_n^* de h que l'on calculera.