Statistique non-asymptotique - Examen

17 décembre 2015, 13h-16h.

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

Partie 1. Statistique paramétrique (S. Pergamenchtchikov)

1. Régression simple (6 points).

1. Donner la définition du modèle de régression simple.
2. On considère le modèle de régression simple avec un bruit de loi gaussienne. Construire un test de niveau $0 < \alpha < 1$ pour vérifier si $a_1 = 1$ ou non.

2. Régression multiple (4 points).

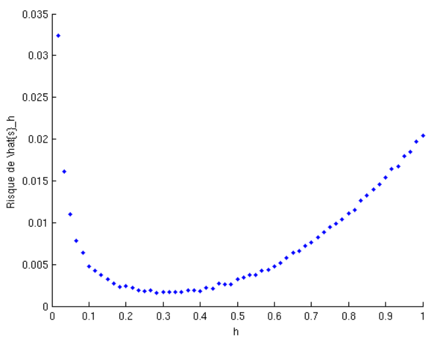
1. Donner la définition du modèle de régression multiple.
2. On considère un modèle de régression multiple d'ordre 3. Construire l'estimateur des moindres carrés pour les paramètres a_1, a_3 et $a_2 + 2a_3$.

Partie 2. Statistique non-parametrique (G. Chagny)

Dans la suite, $X_1, \ldots, X_n$ désigne un échantillon de variables aléatoires réelles indépendantes indentiquement distribuées de densité inconnue f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et de fonction de répartition F (inconnue également).

3. Questions de cours (3 points).

1. Rappeler la définition de l'estimateur à noyau $\hat{f}_h$ de la densité f, associé à un noyau $K : \mathbb{R} \to \mathbb{R}$ et à une fenêtre $h > 0$.
2. Soit $x_0 \in \mathbb{R}$ et $MSE_{x_0}(\hat{f}_h) = \mathbb{E}[(\hat{f}_h(x_0) - f(x_0))^2]$ le risque quadratique au point x_0. Donner (sans démonstration), une majoration de $MSE_{x_0}(\hat{f}_h)$, ainsi que les hypothèses pour obtenir ce résultat.
3. La figure ci-dessous représente le risque $MSE_{x_0}(\hat{f}_h)$ en fonction de la fenêtre h de l'estimateur (courbe obtenue à partir de simulations). Pouvez vous expliquer la forme de la courbe ? Comment choisir la fenêtre h de l'estimateur ?



4. Estimation par projection pour la fonction de répartition (7 points).

Voici les notations utilisées dans la suite.

Norme. Pour $A = \mathbb{R}$ ou $A = [0;1]$, on note $\|.\|_{L^2(A)}$ la norme usuelle de $L^2(A)$, c'est-à-dire $\|g\|_{L^2(A)} = (\int_A g^2(x)dx)^{1/2}$, pour $g \in L^2(A)$.

Sous-espace de projection. Soit $D > 0$ un entier, et pour tout $j \in \{1, \dots, D\}$,

$$\varphi_j(x) = \sqrt{D}\mathbf{1}_{[\frac{j-1}{D}, \frac{j}{D}[}(x),\ x \in [0;1].$$

Soit $S_D = \mathrm{Span}\{\varphi_1, \dots, \varphi_D\}$, et $\Pi_{S_D} F$ la projection orthogonale de F sur S_D. On rappelle que $\Pi_{S_D} F = \sum_{j=1}^D \theta_j \varphi_j$, avec $\theta_j = \int_0^1 F(x)\varphi_j(x)dx$.

Estimateur. On considère l'estimateur suivant pour F :

$$\hat{F}_D = \sum_{j=1}^D \hat{\theta}_j \varphi_j, \text{ with } \hat{\theta}_j = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} \varphi_j(x)\mathbf{1}_{X_i \le x} dx,\ j \ge 1.$$

1. Le but de cette question est d'étudier le risque quadratique intégré de $\hat{F}_D$, défini par

$$MISE(\hat{F}_D) = \mathbb{E}\left[\|\hat{F}_D - F\|^2_{L^2([0;1])}\right].$$

   (a) Calculer $\mathbb{E}[\hat{\theta}_j]$, pour $j \in \{1, \dots, D\}$. En déduire que $\mathbb{E}[\hat{F}_D(x)] = \Pi_{S_D} F(x)$ pour $x \in [0,1]$.

   (b) Justifier que

$$MISE(\hat{F}_D) = \|F - \Pi_{S_D} F\|^2_{L^2([0;1])} + \mathbb{E}\left[\|\hat{F}_D - \Pi_{S_D} F\|^2_{L^2([0;1])}\right].$$

   (c) Prouver que

$$\forall j \in \{1, \dots, D\},\ \left(\hat{\theta}_j - \theta_j\right)^2 \le \int_{\frac{j-1}{D}}^{\frac{j}{D}} \left(\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X_i \le x} - F(x)\right)^2 dx.$$

   (d) En déduire que

$$MISE(\hat{F}_D) \le \|F - \Pi_{S_D} F\|^2_{L^2([0;1])} + \frac{1}{n}.$$

2. Que peut-on conclure quant au meilleur choix possible de la dimension D ? Connaissez vous un autre estimateur de la fonction de répartition F qui permette de justifier ce phénomène ?

3. Pour $g \in \mathbb{L}^2(\mathbb{R})$, on définit la fonction de contraste suivante :

$$\gamma_n(g) = \|g\|^2_{L^2(\mathbb{R})} - \frac{2}{n}\sum_{i=1}^n \int_{\mathbb{R}} g(x)\mathbf{1}_{X_i \le x} dx.$$

   (a) Prouver que $\mathbb{E}[\gamma_n(g)] = \|g - F\|^2_{L^2(\mathbb{R})} - \|F\|^2_{L^2(\mathbb{R})}$, pour tout $g \in \mathbb{L}^2(\mathbb{R})$. En déduire que γ_n convient bien pour estimer la fonction de répartition F.

   (b) Calculer $\tilde{F}_D = \arg\min_{g \in S_D} \gamma_n(g)$, et conclure que $\tilde{F}_D = \hat{F}_D$.