Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2017 – 2018

Econométrie II

Contrôle continu n°1, avril 2018

Examen de 1h30. Tout document ou calculatrice est interdit.

Soit une fonction $f : [0,1] \to \mathbf{R}$ connue, de classe $\mathcal{C}^1$ sur $[0,1]$ et telle que $f(0) \neq f(1)$. On dispose d'une variable $Y = {}^t(Y_1, \cdots, Y_n)$ observée pour n individus, émanant du modèle:

$$Y_i = \theta_0 + \theta_1 \, f\big(\frac{i}{n}\big) + \varepsilon_i \quad \text{pour } i = 1, \dots, n,$$

où $\theta = {}^t(\theta_0, \theta_1)$ est un vecteur de paramètres inconnus et $\varepsilon = {}^t(\varepsilon_1, \cdots, \varepsilon_n)$ est un vecteur gaussien composé de n v.a.i.i.d. centrées de variance σ_ε^2 inconnue. On note $\| \cdot \|$ la norme euclidienne classique de $\mathbf{R}^n$.

1. Montrer que $\Big| \frac{1}{n} \sum_{i=1}^n f\big(\frac{i}{n}\big) - \int_0^1 f(t)dt \Big| \leq \frac{1}{2n} \sup_{x \in [0,1]} |f'(x)|$ pour tout $n \in \mathbf{N}$ **(2.5 pts)**. En déduire $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n f\big(\frac{i}{n}\big)$ **(0.5 pts)**.
2. On effectue la régression linéaire par moindres carrés de Y. Donner une expression matricielle de $\widehat{\theta} = {}^t(\widehat{\theta}_0, \widehat{\theta}_1)$ **(1 pt)** en justifiant son existence pour n suffisamment grand **(2 pts)**, puis l'expression exacte de $\widehat{\theta}_0$ et de $\widehat{\theta}_1$ **(1 pt)**. Montrer le théorème de la limite centrale
$$\sqrt{n}\big(\widehat{\theta} - \theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0\,, \frac{\sigma_\varepsilon^2}{\int_0^1 f^2(x)dx - (\int_0^1 f(x)dx)^2} \begin{pmatrix} \int_0^1 f^2(x)dx & -\int_0^1 f(x)dx \\ -\int_0^1 f(x)dx & 1 \end{pmatrix}\right) \qquad \textbf{(2 pts)}.$$
3. Rappeler l'expression exacte (non matricielle) d'un estimateur $\widehat{\sigma}_\varepsilon^2$ non biaisé de σ_ε^2 **(1 pt)** dont on précisera la loi **(0.5 pts)**.
4. Donner l'expression exacte (non matricielle) de la statistique $\widehat{T}$ de Student permettant de tester $H_0 : \ \theta_1 = 0$ contre $H_1 : \ \theta_1 \neq 0$ **(1 pt)**. Lorsque $n \to \infty$ déterminer en fonction de $\widehat{\theta}_1$, $\widehat{\sigma}_\varepsilon^2$, n et f l'expression de la p-value de ce test **(1 pt)**.
5. On note ici $F = {}^t\big(f(1/n), \dots, f(1)\big)$, $\mathbf{1} = {}^t(1, \dots, 1)$, $[\mathbf{1}]$ le sous-espace vectoriel de $\mathbf{R}^n$ engendré par la droite vectorielle $\mathbf{1}$ et P_A la matrice de la projection orthogonale sur un sous-espace vectoriel A de $\mathbf{R}^n$. Enfin, pour $U = {}^t(U_1, \dots, U_n) \in \mathbf{R}^n$, on note $\overline{U}_n = \frac{1}{n}(U_1 + \cdots + U_n)$ et $\overline{U} = \overline{U}_n \, \mathbf{1}$.
   (a) Montrer que $\frac{1}{n}\big\|Y - \overline{Y}\big\|^2 = \frac{\theta_1^2}{n} \|F - \overline{F}\|^2 + \frac{2\theta_1}{n} {}^tF \, P_{[1]^\perp} \varepsilon + \frac{1}{n} \|P_{[1]^\perp}\varepsilon\|^2$ **(1.5 pts)**.
   (b) Montrer que $\frac{1}{n}\|F - \overline{F}\|^2 \xrightarrow[n\to\infty]{} \int_0^1 f^2(t)dt - \big(\int_0^1 f(t)dt\big)^2$ **(1.5 pts)**.
   (c) Montrer que $\frac{1}{n}\|P_{[1]^\perp}\varepsilon\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_\varepsilon^2$ **(1 pt)**.
   (d) Déterminer les espérance **(1 pt)** et variance **(2 pts)** de $\frac{2}{n} {}^tF \, P_{[1]^\perp}\varepsilon$. En déduire que $\frac{2}{n} {}^tF \, P_{[1]^\perp}\varepsilon \xrightarrow[n\to+\infty]{\mathcal{P}} 0$ **(1 pt)**.
   (e) Déduire de tout ce qui précède que le coefficient $\widehat{R}^2$ vérifie $\widehat{R}^2 \xrightarrow[n\to+\infty]{\mathcal{P}} 1 - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \theta_1^2\big(\int_0^1 f^2(t)dt - \big(\int_0^1 f(t)dt\big)^2\big)}$ **(1.5 pts)**. Quelle est sa limite lorsque $\sigma_\varepsilon^2 \to \infty$ **(0.5 pts)**?
6. Déduire de ce qui précède que si $\theta_1 \neq 0$, alors $\mathbb{P}\big(\widehat{BIC}(m_0) \leq \widehat{BIC}(m_1)\big) \xrightarrow[n\to\infty]{} 0$, où m_0 est le modèle sans F et m_1 le modèle avec F **(3 pts)**. Qu'en déduire quant à la probabilité de choisir un faux modèle avec le critère BIC **(0.5 pts)**?