Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. $2018-2019$

Econométrie II

Contrôle continu n°1, mars 2019

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

Soit $(Y_i)_{1\le i\le n}$ une famille de variables aléatoires telle que:

$$Y_i = \mu_i^* + \varepsilon_i \qquad \text{pour tout } i \in \{1,\dots,n\}, \tag{1}$$

où $(\varepsilon_i)_{i\in\mathbf{N}}$ est une suite de variables aléatoires identiquement distribuées indépendantes centrées et gaussiennes telles que $\mathrm{var}(\varepsilon_1)=\sigma^2<\infty$ et $\mu^*=(\mu_i^*)_{1\le i\le n}\in\mathbf{R}^n$, vecteur fixé, appartenant à F un sous-espace vectoriel de $\mathbf{R}^n$. Dans toute la suite on notera $\|\cdot\|$ la norme sur $\mathbf{R}^n$ telle que pour tout $Z\in\mathbf{R}^n$, $\|Z\|^2={}^tZ\,Z$ (${}^tA$ désignant la transposée de la matrice A). On définit les estimateurs respectifs de μ^* et σ^2 par

$$\widehat{\mu}=\mathrm{Argmin}_{\mu\in F}\|Y-\mu\| \qquad \text{et} \qquad \widehat{\sigma}_n^2=\frac{1}{n}\,\|Y-\widehat{\mu}\|^2.$$

1. On note $Y=(Y_i)_{1\le i\le n}$. Ecrire le modèle (1) sous une forme vectorielle. Calculer $\mathbb{E}(Y)$ et $\mathrm{cov}(Y)$ (matrice de variance-covariance de Y).

2. Soit $X_0\in\mathbf{R}^n$ tel que ${}^tX_0=(1,1,\dots,1)$. On suppose dans cette question que $F=\mathrm{Vect}(X_0)=[X_0]$, sous-espace engendré par X_0. Montrer alors que $\widehat{\mu}=\overline{Y}_n\,X_0$, où $\overline{Y}_n=\frac{1}{n}\sum_{i=1}^n Y_i$. Calculer le risque quadratique de $\widehat{\mu}$, c'est-à-dire $\mathbb{E}(\|\widehat{\mu}-\mu^*\|^2)$. Calculer $\mathbb{E}(\widehat{\sigma}_n^2)$ et montrer que $\widehat{\sigma}_n^2 \xrightarrow[n\to+\infty]{p.s.} \sigma^2$. Montrer que $(\widehat{\mu},\widehat{\sigma}_n^2)$ est l'estimateur par maximum de vraisemblance de (μ^*,σ^2). Est-ce un estimateur efficace?

3. Soit $X_1\in\mathbf{R}^n$, avec (X_0,X_1) non colinéaires. On suppose maintenant que $F=\mathrm{Vect}(X_0,X_1)=[X_0,X_1]$. Rappeler quelle est la loi de $\|\widehat{\mu}-\mu^*\|^2$, en déduire le risque quadratique de $\widehat{\mu}$. Montrer que $\widehat{\sigma}_n^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma^2$.

4. On suppose que la variable X_1 n'est en réalité pas explicative, c'est-à-dire que $F=[X_0]$ mais que $\widehat{\mu}$ a la même expression que dans la question précédente (donc construit avec X_0 et X_1) car on a cru que X_1 était explicative. Déterminer alors la loi de $\|Y-\widehat{\mu}\|^2$ puis le risque quadratique de $\widehat{\mu}$. Comparer ce risque avec celui obtenu dans le cas où l'on sait à l'avance que seule X_0 est explicative.

5. On veut tester H_0 : $F=\mathrm{Vect}(X_0)$ contre H_1 : $F=\mathrm{Vect}(X_0,X_1)$. Pour cela on considère la statistique
$$\widehat{S}=\frac{\|Y-P_{[X_0]}Y\|^2-\|Y-P_{[X_0,X_1]}Y\|^2}{\frac{1}{n-2}\,\|Y-P_{[X_0,X_1]}Y\|^2},$$
où P_E désigne la matrice de projection orthogonale sur le sev E. Montrer que $\widehat{S}$ est une variable positive, puis que $\widehat{S}$ suit une loi de Fisher $F(1,n-2)$ sous H_0.

6. On se place maintenant sous l'hypothèse H_1 avec précisément $\mu^*=a_0^*X_0+a_1^*X_1$, où $a_1^*\neq 0$. Montrer alors que $\mathbb{E}(\|Y-P_{[X_0]}Y\|^2-\|Y-P_{[X_0,X_1]}Y\|^2)=n(a_1^*)^2\sigma_{1,n}^2+\sigma^2$ avec $\sigma_{1,n}^2=\frac{1}{n}\,{}^tX_1X_1-(\frac{1}{n}{}^tX_0X_1)^2$. On suppose que $\sigma_{1,n}^2 \xrightarrow[n\to\infty]{} \sigma_1^2$ avec $\sigma_1^2>0$. En déduire alors que sous H_1, $\widehat{S} \xrightarrow[n\to+\infty]{\mathcal{P}} +\infty$. Montrer que ce sera notamment le cas si X_1 est la réalisation de n vaiid de variance σ_1^2, indépendantes de ε. Et que se passe-t-il si $X_1={}^t(1,2,3,\dots,n)$?

Proof. 1. On a bien-sûr $Y = \mu^* + \varepsilon$, $\mathbb{E}Y = \mu^*$ et $\mathrm{cov}(Y) = \sigma^2 I_n$, où I_n est la matrice identité **(1.5 pts)**.

2. On a $\widehat{\mu} = X_0({}^tX_0\,X_0)^{-1\,t}X_0Y = \overline{Y}_n\,X_0 = \mu^* + \overline{\varepsilon}_n X_0$ (car ${}^tX_0\,X_0 = n$ et $\mu^* \in \mathrm{Vect}(X_0)$) **(1 pt)**.
De plus, $\mathbb{E}\big(\|\widehat{\mu} - \mu^*\|^2\big) = \mathbb{E}\big(\|P_{[X_0]}\varepsilon\|^2\big) = \sigma^2$ car d'après le Théorème de Cochran, $\|P_{[X_0]}\varepsilon\|^2$ suit un $\sigma^2\chi^2$ avec pour nombre de degrés de liberté $\dim([X_0]) = 1$ **(1 pt)**.
On a $\mathbb{E}(\widehat{\sigma}_n^2) = \frac{1}{n}\,\mathbb{E}\big(\|P_{[X_0]^\perp}\varepsilon\|^2\big) = \frac{n-1}{n}\,\sigma^2$ car d'après le Théorème de Cochran, $\|P_{[X_0]^\perp}\varepsilon\|^2$ suit un $\sigma^2\chi^2$ avec pour nombre de degrés de liberté $n - \dim([X_0]) = n-1$ **(1 pt)**.
On a encore $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n(\varepsilon_i - \overline{\varepsilon}_n)^2 = \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - (\overline{\varepsilon}_n)^2$. Mais les ε_i^2 sont des vaiid d'espérance finie ($= \sigma^2$), on peut donc appliquer la loi forte des grands nombres et $\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \underset{n\to+\infty}{\xrightarrow{p.s.}} \sigma^2$. Ensuite $\overline{\varepsilon}_n \underset{n\to+\infty}{\xrightarrow{p.s.}} 0$ (loi forte des grands nombres appliquée à (ε_i)). Enfin la somme de 2 suites de v.a. convergeant ps coverge ps vers la somme des limites, donc $\widehat{\sigma}_n^2 \underset{n\to+\infty}{\xrightarrow{p.s.}} \sigma^2$ **(2 pts)**.
La log-vraisemblance de $(Y_1,\cdots,Y_n)$ est $L_{\mu,\sigma^2}(Y_1,\cdots,Y_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|Y-\mu\|^2$ que l'on maximise pour trouver $(\widehat{\mu},\widehat{\sigma}^2)$ **(2 pts)**.
Le modèle statistique proposé est un modèle régulier de la famille exponentielle donc l'estimateur du MV est efficace **(2 pts)**.

3. Par le même raisonnement que précédemment, $\mathbb{E}\big(\|\widehat{\mu} - \mu^*\|^2\big) = \mathbb{E}\big(\|P_{[X_0,X_1]}\varepsilon\|^2\big) = 2\sigma^2$ **(1 pt)**.
De plus, $\|P_{F^\perp}Y\|^2$ a pour loi $\sigma^2\chi^2(n-2)$. Mais comme un $\chi^2(n-2)$ peut s'écrire comme une somme de $n-2$ carrés de vaiid gaussiennes centrées réduites, d'après la loi des grands nombres, $\frac{1}{n-2}\chi^2(n-2) \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} 1$, donc $\widehat{\sigma}_n^2 \underset{n\to\infty}{\xrightarrow{\mathcal{L}}} \sigma^2$, ou encore $\widehat{\sigma}_n^2 \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} \sigma^2$ **(1.5 pts)**.

4. $\|Y - \widehat{\mu}\|^2 = \|P_{[X_0,X_1]^\perp}\varepsilon\|^2$ a pour loi un $\sigma^2\chi^2(n-2)$.
Le risque quadratique est encore $2\sigma^2$, donc 2 fois plus grand que le risque obtenu dans le cas où l'on sait à l'avance que seule X_0 intervient **(1 pt)**.

5. On a $\|Y - P_{[X_0]}Y\|^2 - \|Y - P_{[X_0,X_1]}Y\|^2 = \|P_{[X_0]^\perp}Y\|^2 - \|P_{[X_0,X_1]^\perp}Y\|^2$. Comme $[X_0,X_1]^\perp \subset [X_0]^\perp$, on peut toujours écrire que $\|Y - P_{[X_0]}Y\|^2 - \|Y - P_{[X_0,X_1]}Y\|^2 = \|P_{[X_0,X_1]\cap[X_0]^\perp}Y\|^2$ d'après Pythagore, donc la variable $\widehat{S}$ est positive **(1.5 pts)**.
On a sous H_0, $\|Y - P_{[X_0]}Y\|^2 - \|Y - P_{[X_0,X_1]}Y\|^2 = \|P_{[X_0,X_1]\cap[X_0]^\perp}\varepsilon\|^2$ et $\|Y - P_{[X_0,X_1]}Y\|^2 = \|P_{[X_0,X_1]^\perp}\varepsilon\|^2$. Or d'après Cochran, $\|P_{[X_0,X_1]\cap[X_0]^\perp}\varepsilon\|^2 \overset{\mathcal{L}}{\sim} \sigma^2\chi^2(1)$ et $\|P_{[X_0,X_1]^\perp}\varepsilon\|^2 \overset{\mathcal{L}}{\sim} \sigma^2\chi^2(n-2)$, les deux χ^2 étant indépendants car les sous-espaces de projection sont orthogonaux. On en déduit bien que $\widehat{S} \overset{\mathcal{L}}{\sim} F(1,n-2)$ **(2.5 pts)**.

6. On a sous l'hypothèse H_1, $\|Y - P_{[X_0]}Y\|^2 = \|P_{[X_0]^\perp}Y\|^2 = \|P_{[X_0]^\perp}(a_0^*X_0 + a_1^*X_1 + \varepsilon)\|^2 = \|a_1^*P_{[X_0]^\perp} * X_1 + P_{[X_0]^\perp}\varepsilon)\|^2$.
En conséquence, $\mathbb{E}\big(\|Y - P_{[X_0]}Y\|^2\big) = \mathbb{E}\big(\|P_{[X_0]^\perp}Y\|^2\big) = \|a_1^*P_{[X_0]^\perp} * X_1\|^2 + \mathbb{E}\big(\|P_{[X_0]^\perp}\varepsilon)\|^2\big) = (a_1^*)^2\big({}^tX_1X_1 - \frac{1}{n}({}^tX_0X_1)^2\big) + (n-1)\sigma^2$. Comme $\mathbb{E}\big(\|Y - P_{[X_0,X_1]}Y\|^2\big) = (n-2)\sigma^2$, on en déduit le résultat **(3 pts)**.
En reprenant le calcul précédent, on a donc $\|Y - P_{[X_0]}Y\|^2 = (a_1^*)^2\big({}^tX_1X_1 - \frac{1}{n}({}^tX_0X_1)^2\big) + \|P_{[X_0]^\perp}\varepsilon)\|^2 + 2a_1^{*t}X_1P_{[X_0]^\perp}\varepsilon$.
Mais $\mathbb{E}\big({}^tX_1P_{[X_0]^\perp}\varepsilon\big) = 0$ et $\mathrm{var}\big({}^tX_1P_{[X_0]^\perp}\varepsilon\big) = \sigma^2 n\sigma_{1,n}^2$, donc ${}^tX_1P_{[X_0]^\perp}\varepsilon/(n\sigma_{1,n}^2) \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} 0$. On en déduit donc d'après Slutski que $\|Y - P_{[X_0]}Y\|^2 - \|Y - P_{[X_0,X_1]}Y\|^2 \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}}$. Or on sait que le dénominateur de $\widehat{S}$ converge vers σ^2, ce qui montre d'après Slutski que $\widehat{S} \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} \infty$ **(3 pts)**.
D'après la question précédente 2., $\sigma_{1,n}^2 \underset{n\to+\infty}{\xrightarrow{p.s.}} \sigma_1^2$ **(1 pt)**.
Le résultat, $\widehat{S} \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} \infty$, est le même dans le cas où $X_1 = {}^t(1,2,3,\ldots,n)$ **(1 pt)**.

□