Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2012 – 2013

Statistiques II

Correction du contrôle continu n°1, mars 2013

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

1. **(Sur 16 points)** Soit $\sigma^2 > 0$ et soit $\varepsilon = (\varepsilon_t)_{t\in\mathbf{Z}}$ le processus gaussien centré tel que

$$\mathrm{cov}(\varepsilon_s, \varepsilon_t) = \sigma^2 \frac{1}{1+|t^2-s^2|} \qquad \text{pour} \quad s,t \in \mathbf{Z}.$$

   (a) Déterminer la loi de ε_t pour $t \in \mathbf{Z}$. Les variables aléatoires (ε_t) sont-elles identiquement distribuées? Sont-elles indépendantes?

   (b) Montrer que $(\varepsilon_t)_{t\in\mathbf{Z}}$ n'est pas un processus stationnaire. Déterminer la tendance et la composante saisonnière de (ε_t).

   (c) Soit $(a_i)_{i\in\mathbf{N}}$ une suite bornée de réels. Montrer que

$$\mathrm{var}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i\Big) = \frac{\sigma^2}{n}\big(\sum_{i=1}^n a_i^2 + 2 \sum_{1\le i<j\le n} \frac{a_i a_j}{j^2-i^2+1}\big).$$

   Déterminer la loi de $\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i$.

   (d) Soit (Z_n) une suite de variables aléatoires telles que pour $n \in \mathbf{N}$, $Z_n \sim \mathcal{N}(m_n, \gamma_n^2)$. On suppose que $m_n \underset{n\to\infty}{\longrightarrow} m$ et $\gamma_n^2 \underset{n\to\infty}{\longrightarrow} \gamma^2 > 0$. Montrer que $Z_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(m\,,\, \gamma^2)$.

   (e) Montrer que

$$\sum_{1\le i<j\le n} \frac{1}{j^2-i^2+1} \le \sum_{1\le i<j\le n} \frac{1}{2i}\big(\frac{1}{j-i} - \frac{1}{j+i}\big) \le \sum_{i=1}^{n-1} \frac{1}{2i} \sum_{k=1}^{2i} \frac{1}{k}.$$

   On suppose en plus que $\frac{1}{n}\sum_{i=1}^n a_i^2 \underset{n\to\infty}{\longrightarrow} A^2$ avec $A^2>0$. En déduire que $\displaystyle\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,\, \sigma^2 A^2)$.

Proof. (a) On a pour tout $t \in \mathbf{Z}$, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ **(0.5pts)**. On en déduit que (ε_t) est une suite de v.a.i.d. puisque la loi ne dépend pas de t **(0.5pts)**. En revanche, ce n'est pas une suite de v.a. indépendantes car elles sont corrélées **(0.5pts)**.
(b) Si (ε_t) formait un processus stationnaire, alors $\mathrm{cov}(\varepsilon_0, \varepsilon_1) = \mathrm{cov}(\varepsilon_1, \varepsilon_2)$. Or $\mathrm{cov}(\varepsilon_0, \varepsilon_1) = \sigma^2/2$ et $\mathrm{cov}(\varepsilon_1, \varepsilon_2) = \sigma^2/4$. Donc $\mathrm{cov}(\varepsilon_0, \varepsilon_1) \neq \mathrm{cov}(\varepsilon_1, \varepsilon_2)$ et le processus n'est pas stationnaire **(2pts)**.
On a $\mathbb{E}\varepsilon_t = 0$ donc (ε_t) ne possède ni tendance ni composante saisonnière **(0.5pts)**.
(c) On a $\mathrm{var}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i\Big) = \frac{\sigma^2}{n}\big(\sum_{i=1}^n \sum_{j=1}^n \frac{a_i a_j}{|j^2-i^2|+1}\big)$. Dans cette double somme, on isole le cas où $i=j$, et cela donne $\sum_{i=1}^n a_i^2$. Ensuite, comme il y a symétrie en i et j, on peut considérer que pour l'ensemble des cas où $i \neq j$ la double somme équivaut à 2 fois l'ensemble des cas où $i<j$. On obtient ainsi le second terme **(1.5pts)**.
Comme (ε_t) est un processus gaussien, toute combinaison linéaire des ε_t est de loi gaussienne. De plus $\mathbb{E}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i\Big) = 0$. Ainsi $\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i \sim \mathcal{N}\Big(0\,,\, \frac{\sigma^2}{n}\big(\sum_{i=1}^n a_i^2 + 2\sum_{1\le i<j\le n} \frac{a_i a_j}{j^2-i^2+1}\big)\Big)$ **(1pt)**.
(d) On va utiliser la caractérisation de la convergence en loi par la fonction caractéristique. On sait que la fonction caractéristique $\mathbb{E}(e^{itZ_n})$ de Z_n vaut $\exp(itm_n + t^2\gamma_n^2/2)$. Or pour tout $t \in \mathbf{R}$ fixé, on a $\exp(itm_n + t^2\gamma_n^2/2) \underset{n\to\infty}{\longrightarrow} \exp(itm + t^2\gamma^2/2)$ (la fonction exponentielle étant continue). On en déduit que $Z_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(m\,,\, \gamma^2)$ **(2.5pts)**.
(e) Il est clair que $\frac{1}{j^2-i^2+1} \le \frac{1}{j^2-i^2}$ et $\frac{1}{j^2-i^2} = \frac{1}{2i}\big(\frac{1}{j-i} - \frac{1}{j+i}\big)$. Donc $\sum_{1\le i<j\le n} \frac{1}{j^2-i^2+1} \le \sum_{1\le i<j\le n} \frac{1}{2i}\big(\frac{1}{j-i} - \frac{1}{j+i}\big)$ **(1pt)**.
Ensuite on utilise les suites de termes "téléscopiques" et on a ainsi $\sum_{j=i+1}^n \big(\frac{1}{j-i} - \frac{1}{j+i}\big) = \sum_{k=1}^{2i} \frac{1}{k} - \sum_{k=n-i+1}^{n+i} \frac{1}{k} \le \sum_{k=1}^{2i} \frac{1}{k}$ **(2pts)**.
Pour montrer la convergence en loi, d'après la question précédente, il convient juste de chercher la limite de $\mathrm{var}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^n a_i \varepsilon_i\Big)$

puisque l'espérance de $\frac{1}{\sqrt{n}}\sum_{i=1}^n a_i\varepsilon_i$ est toujours nulle. Or, en utilisant la décomposition de la variance obtenue en (c), le premier terme vaut $\frac{\sigma^2}{n}\sum_{i=1}^n a_i^2$ et converge donc vers $\sigma^2 A^2$ d'après l'hypothèse. Pour le second terme, qui vaut $\frac{2\sigma^2}{n}\sum_{1\leq i<j\leq n}\frac{a_i a_j}{j^2-i^2+1}$, on utilise d'abord le fait que les a_i. Il revient alors d'étudier $\sum_{1\leq i<j\leq n}\frac{1}{j^2-i^2+1}$. On utilise alors les inégalités précédentes. On doit donc chercher la limite de $\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{2i}\sum_{k=1}^{2i}\frac{1}{k}$. Or pour i grand, $\sum_{k=1}^{2i}\frac{1}{k}\sim\log i$, donc $\sum_{i=1}^{n-1}\frac{1}{2i}\sum_{k=1}^{2i}\frac{1}{k}\sim\sum_{i=1}^{n-1}\frac{\log i}{2i}\sim\int_1^{n-1}\frac{\ln x}{x}\,dx\sim(\ln(n))^2$. En conséquence, $\frac{1}{n}\sum_{1\leq i<j\leq n}\frac{1}{j^2-i^2+1}\underset{n\to\infty}{\longrightarrow}0$. Ainsi $\mathrm{var}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n a_i\varepsilon_i\Big)\underset{n\to\infty}{\longrightarrow}\sigma_2 A^2$ et d'après (d) on en déduit que $\frac{1}{\sqrt{n}}\sum_{i=1}^n a_i\varepsilon_i\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0\,,\,\sigma^2A^2)$ **(4pts)**. $\square$

2. **(Sur 12 points)** On suppose maintenant que l'on observe $(X_1,\cdots,X_{12n})$ issu du processus $X=(X_t)_{t\in\mathbf{N}}$ tel que

$$X_t=\theta_0+\theta_1\cos(\frac{\pi}{6}\,t)+\theta_2\cos(\frac{\pi}{3}\,t)+\varepsilon_t\qquad\text{pour tout } t\in\mathbf{Z}.$$

où $\theta_0,\theta_1,\theta_2\in\mathbf{R}$ sont inconnus.

(a) **(Préliminaires)** Montrer que $\sum_{k=1}^{12n}\cos(\frac{\pi}{6}\,k)=0$, $\sum_{k=1}^{12n}\cos^2(\frac{\pi}{6}\,k)=6n$ et $\sum_{k=1}^{12n}\cos(\frac{\pi}{6}\,k)\cos(\frac{\pi}{3}\,k)=0$.

(b) Le processus $(X_t)_{t\in\mathbf{N}}$ est-il stationnaire? Gaussien?

(c) Déterminer la tendance, la saisonalité de X et sa période.

(d) On désire estimer θ_0,θ_1 et θ_2. A l'aide d'une estimation par moindres carrés ordinaires, montrer que

$$\widehat{\theta}_0^{(n)}=\frac{1}{12\,n}\sum_{k=1}^{12n}X_k,\qquad\widehat{\theta}_1^{(n)}=\frac{1}{6\,n}\sum_{k=1}^{12n}\cos(\frac{\pi}{6}k)X_k\qquad\text{et}\qquad\theta_2^{(n)}=\frac{1}{6\,n}\sum_{k=1}^{12n}\cos(\frac{\pi}{3}k)X_k.$$

(e) Déterminer le biais de $(\widehat{\theta}_0^{(n)},\widehat{\theta}_1^{(n)},\widehat{\theta}_2^{(n)})$ et écrire chacun de ces estimateurs en fonction de $(\varepsilon_t)_t$.

(f) En utilisant les résultats du 1.(e), déterminer des théorèmes de la limite centrale vérifiés par $(\widehat{\theta}_0^{(n)},\widehat{\theta}_1^{(n)},\widehat{\theta}_2^{(n)})$.

Proof. (a) On a peut écrire que $\cos(\frac{\pi}{6}\,k)=\mathcal{R}e\Big(e^{i\frac{\pi}{6}\,k}\Big)$, donc $\sum_{k=1}^{12n}\cos(\frac{\pi}{6}\,k)=\mathcal{R}e\big(\sum_{k=1}^{12n}(e^{i\frac{\pi}{6}})^k\big)=\mathcal{R}e\Big(e^{i\frac{\pi}{6}}\frac{1-e^{2i\pi n}}{1-e^{i\frac{\pi}{6}}}\Big)=0$.
$\sum_{k=1}^{12n}\cos^2(\frac{\pi}{6}\,k)=\frac{1}{2}\sum_{k=1}^{12n}(1+\cos(\frac{\pi}{3}\,k))=6n+0$ avec le même raisonnement que ce qui précède **(2.5pts)**.
$\cos(\frac{\pi}{6}\,k)\cos(\frac{\pi}{3}\,k)=\frac{1}{2}\big(\cos(\frac{\pi}{2}\,k)+\cos(\frac{\pi}{6}\,k)\big)$ et comme $\sum_{k=1}^{12n}\cos(\frac{\pi}{2}\,k)+\cos(\frac{\pi}{6}\,k)=0$ on en déduit que $\sum_{k=1}^{12n}\cos(\frac{\pi}{6}\,k)\cos(\frac{\pi}{3}\,k)=0$ **(1.5pts)**.
(b) $\mathbb{E}X_t=\theta_0+\theta_1\cos(\frac{\pi}{6}\,t)+\theta_2\cos(\frac{\pi}{3}\,t)$ donc $\mathbb{E}X_t$ dépend de t: le processus n'est pas stationnaire **(0.5pts)**.
En revanche c'est un processus gaussien car (ε_t) est un processus gaussien **(0.5pts)**.
(c) D'après les préliminaires, la tendance de (X_t) est $a(t)=\theta_0$ et sa saisonnalité est $s(t)=\theta_1\cos(\frac{\pi}{6}\,t)+\theta_2\cos(\frac{\pi}{3}\,t)$ avec un période $T=12$ **(1pt)**.
(d) On définit la matrice Z de taille $(12n,3)$ dont la première colonne est composée de 1, la seconde des $\cos(\frac{\pi}{6}\,k)$ et la troisième des $\cos(\frac{\pi}{3}\,k)$ pour $k=1,\cdots,12n$. En calculant ${}^tZ\,Z$ on obtient grâce aux préliminaires

$${}^tZ\,Z=\left\{\begin{array}{ccc}12n&0&0\\0&6n&0\\0&0&6n\end{array}\right\}.$$

On en déduit $({}^tZ\,Z)^{-1}$ et on montre ainsi que ${}^t\Big(\widehat{\theta}_0^{(n)},\widehat{\theta}_1^{(n)},\theta_2^{(n)}\Big)=({}^tZ\,Z)^{-1}\,{}^tZ\,X$ avec $X={}^t(X_1,\cdots,X_{12n})$ **(2pts)**.
(e) On sait que le biais de ces estimateurs MCO est nul **(0.5pts)** et

$$\widehat{\theta}_0^{(n)}=\theta_0+\frac{1}{12\,n}\sum_{k=1}^{12n}\varepsilon_k,\qquad\widehat{\theta}_1^{(n)}=\theta_1+\frac{1}{6\,n}\sum_{k=1}^{12n}\cos(\frac{\pi}{6}k)\varepsilon_k\qquad\text{et}\qquad\theta_2^{(n)}=\theta_2+\frac{1}{6\,n}\sum_{k=1}^{12n}\cos(\frac{\pi}{3}k)\varepsilon_k\qquad\textbf{(1pt)}.$$

(f) En appliquant le théorème limite du 1.(e) avec $a_i=1$, puis $a_i=\cos(\pi\frac{i}{6})$ et enfin $a_i=\cos(\pi\frac{i}{3})$, avec respectivement $A^2=1$, $1/2$ et $1/2$, on obtient

$$\sqrt{12n}\Big(\widehat{\theta}_0^{(n)}-\theta_0\Big)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2),\qquad\sqrt{12n}\Big(\widehat{\theta}_1^{(n)}-\theta_1\Big)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\frac{1}{2}\sigma^2)\qquad\text{et}\qquad\sqrt{12n}\Big(\widehat{\theta}_2^{(n)}-\theta_2\Big)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\frac{1}{2}\sigma^2)\qquad\textbf{(2pts)}.$$

$\square$