Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2011 – 2012

Statistiques II

Correction du contrôle continu n°1, mars 2012

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

1. **(Sur 15 points + 7 points facultatifs)** Soit $X = (X_t)_{t\in\mathbf{Z}}$ le processus défini par:

$$X_t = \sum_{i=0}^{\infty} \frac{1}{i+1}\,\varepsilon_{t-i} \qquad \text{pour} \qquad t \in \mathbf{Z},$$

où $(\varepsilon_t)_{t\in\mathbf{Z}}$ est un bruit blanc gaussien de variance σ^2.

(a) Soit $N \in \mathbf{N}$ et soit $(X_t^{(N)})_{t\in Z}$ le processus tel que pour $t \in \mathbf{Z}$, $X_t^{(N)} = \sum_{i=0}^{N} \frac{1}{i+1}\,\varepsilon_{t-i}$. Montrer que $(X_t^{(N)})_{t\in Z}$ est un processus stationnaire.

(b) Montrer que pour tout $t \in \mathbf{Z}$, la suite $(X_t^{(N)})_N$ converge absolument dans $\mathbb{L}^2$. En déduire que (X_t) existe dans $\mathbb{L}^2$ et est un processus stationnaire gaussien.

(c) Déterminer $\mathbb{E}X_t$ et $\mathrm{var}(X_t)$ (on rappelle que $\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$). En déduire la loi de X_t pour $t \in \mathbf{Z}$.

(d) Calculer (sous forme de série) l'autocovariance r_X de (X_t). Le processus (X_t) est-il un bruit blanc?

(e) On utilisant la décomposition $\frac{1}{i(i+k)} = \frac{1}{k}\left(\frac{1}{i} - \frac{1}{k+i}\right)$, montrer que $r_X(k) \simeq \sigma^2\,\frac{\ln k}{k}$ pour $k \to \infty$.

(f) En déduire que la densité spectrale f_X de (X_t) existe sur $[-\pi, 0[\cup]0, \pi[$, mais $\lim_{x\to 0} f_X(x) = \infty$.

(g) **(Question facultative = 7 points)** Soit $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Montrer que pour tout $n \in \mathbf{N}^*$,

$$\mathrm{var}(\overline{X}_n) = \frac{1}{n}\,\mathrm{var}(X_0) + \frac{2}{n^2}\sum_{k=1}^{n-1}(n-k)r_X(k), \text{ puis que } \mathrm{var}(\overline{X}_n) \simeq \sigma^2\frac{\log^2 n}{n} \text{ pour } n \to \infty.$$

En déduire que $\overline{X}_n \xrightarrow[n\to+\infty]{\mathcal{P}} 0$ puis que $\dfrac{\sqrt{n}}{\log n}\,\overline{X}_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

Proof. (a) Pour $n \in \mathbf{N}^*$, $(t_1, \cdots, t_n) \in \mathbf{Z}^n$, (ε_t) est un processus stationnaire donc $(\varepsilon_{t_1}, \varepsilon_{t_1-1}, \cdots, \varepsilon_{t_1-N}, \varepsilon_{t_2}, \varepsilon_{t_2-1}, \cdots, \varepsilon_{t_n}, \varepsilon_{t_n-1}, \cdots, \varepsilon_{t_n-N})$ a la même loi que $(\varepsilon_{t_1+c}, \varepsilon_{t_1-1+c}, \cdots, \varepsilon_{t_n+c}, \varepsilon_{t_n-1+c}, \cdots, \varepsilon_{t_n-N+c})$ pour tout $c \in \mathbf{Z}$. On applique alors la fonction $g : (x_1, \cdots, x_{n(N+1)}) \mapsto \left(\sum_{i=0}^{N}\frac{1}{i+1}x_{i+1}, \cdots, \sum_{i=0}^{N}\frac{1}{i+1}x_{(n-1)(N+1)+i+1}\right)$ qui est continue donc mesurable et ainsi $(X_{t_1}^{(N)}, \cdots, X_{t_n}^{(N)})$ a la même loi que $(X_{t_1+c}^{(N)}, \cdots, X_{t_n+c}^{(N)})$ **(2 pts)**.

(b) On a facilement $\mathbb{E}(X_t^{(N)})^2 = \sigma^2\sum_{i=1}^{N-1}\frac{1}{i^2}$ donc $\mathbb{E}(X_t^{(N)})^2 < \infty$ pour tout $N \in \mathbf{N}^*$. D'où la convergence dans $\mathbb{L}^2$ **(1 pt)**.
La convergence dans $\mathbb{L}^2$ est aussi valable pour tout vecteur $(X_{t_1}^{(N)}, \cdots, X_{t_n}^{(N)})$ (on passe par une combinaison linéaire quelconque) donc (X_t) existe dans $\mathbb{L}^2$ et la propriété de stationnarité est conservée par passage à la limite **(1.5 pts)**.
Enfin, comme chaque $(X_{t_1}^{(N)}, \cdots, X_{t_n}^{(N)})$ est un vecteur gaussien car défini comme une application linéaire (g) d'un vecteur de v.a. indépendantes gaussiennes $(\varepsilon_{t_1}, \varepsilon_{t_1-1}, \cdots, \varepsilon_{t_1-N}, \varepsilon_{t_2}, \varepsilon_{t_2-1}, \cdots, \varepsilon_{t_n}, \varepsilon_{t_n-1}, \cdots, \varepsilon_{t_n-N})$, $(X_t^{(N)})$ est un processus gaussien pour tout $N \in \mathbf{N}^*$. Par passage à la limite de tout vecteur $(X_{t_1}^{(N)}, \cdots, X_{t_n}^{(N)})$, comme cette limite existe, elle est gaussienne **(1.5 pts)**.

(c) On a facilement $\mathbb{E}X_t = 0$ **(0.5 pts)** et $\mathrm{var}(X_t) = \frac{\pi^2}{6}\,\sigma^2$ **(0.5 pts)**. Donc X_t a pour loi $\mathcal{N}(0, \frac{\pi^2}{6}\,\sigma^2)$ **(0.5 pts)**.

(d) On a facilement $r_X(k) = \sigma^2\sum_{i=1}^{\infty}\frac{1}{i(i+k)}$ **(1 pt)**. (X_t) n'est donc pas un bruit blanc car par exemple $r_X(1) = \sigma^2 \neq 0$ **(0.5 pts)**.

(e) On a donc $r_X(k) = \lim_{N\to\infty}\frac{1}{k}\sum_{i=1}^{N}\frac{1}{k}\left(\frac{1}{i} - \frac{1}{k+i}\right) = \lim_{N\to\infty}\frac{1}{k}\left(\sum_{i=1}^{k}\frac{1}{i} - \sum_{i=1}^{k}\frac{1}{N+i}\right)$. Or $\lim_{N\to\infty}\sum_{i=1}^{k}\frac{1}{N+i} = 0$ et $\sum_{i=1}^{k}\frac{1}{i} \simeq \log k$ pour $k \to \infty$. D'où le résultat **(2.5 pts)**.

(f) Pour $x \neq 0$, $f_X(x) = \frac{1}{2\pi}\sum_{k\in\mathbf{Z}} r_X(k)e^{-ikx}$ qui existe (d'après Abel-Dirichlet) et pour $x \to 0$, $f_X(x) \to \frac{1}{2\pi}\sum_{k\in\mathbf{Z}} r_X(k)$ qui diverge d'après le comportement asymptotique de r_X (série de Bertrand) **(1.5 pts)**.

(g) On a $\mathrm{var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} r_X(j-i) = \frac{1}{n}\sum_{i=1}^{n} r_X(0) + \frac{2}{n^2}\sum_{1\le i<j\le n} r_X(j-i)$. Dans cette denière somme, on a $n-1$ fois le terme $r_X(1)$ qui apparaît, $(n-2)$ fois le terme $r_X(2)$, ..., d'où le résultat **(1.5 pts)**.
On a $\sum_{k=1}^{n} r_X(k)$ qui a le même comportement asymptotique que $\int_1^n \frac{\log x}{x}dx$. Par une IPP, on montre que $\int_1^n \frac{\log x}{x}dx = \frac{1}{2}\log^2 n$.

Par ailleurs, $\sum_{k=1}^n k r_X(k)$ a le même comportement asymptotique que $\int_1^n \log x dx \simeq n \log n$. En rassemblant le tout on a bien $\mathrm{var}(\overline{X}_n) \simeq \sigma^2 \dfrac{\log^2 n}{n}$ **(3 pts)**.
On utilise alors l'Inégalité de Bienaymé-Tchebytchev pour montrer la convergence en probabilité **(1 pt)** et le Lemme de Slutski pour montrer le théorème de la limite centrale **(1.5 pt)**.

□

2. **(Sur 9 points)** On suppose maintenant que l'on observe $(Y_1, \cdots, Y_{3n})$ issu du processus $Y = (Y_t)_{t\in\mathbf{N}}$ tel que

$$Y_t = a\, g(t) + X_t \qquad \text{pour tout } t \in \mathbf{N}.$$

où $a \in \mathbf{R}^*$ est inconnu, et $g(3i) = 1$, $g(3i+1) = 3$ et $g(3i+2) = -1$ pour tout $i \in \mathbf{N}$.

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

(b) Déterminer la tendance, la saisonalité de Y et sa période.

(c) On désire estimer a et σ^2. A l'aide d'une estimation par moindres carrés ordinaires, montrer que

$$\widehat{a}_n = \frac{1}{11\,n} \sum_{i=0}^{n-1} \big(3Y_{3i+1} - Y_{3i+2} + Y_{3i+3}\big).$$

En déduire un estimateur non biaisé de σ^2.

(d) Montrer que $\widehat{a}_n$ est un estimateur sans biais. Montrer que si deux suites de variables alétoires $(U_n)_n$ et $(V_n)_n$ sont telles que $U_n \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} 0$ et $V_n \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} 0$, alors $U_n + V_n \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} 0$. En déduire que $\widehat{a}_n \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} a$ (on utilisera les mêmes raisonnements qu'en 1.(g) sans faire de calculs).

(e) En déduire un estimateur convergeant de la tendance et de la saisonnalité de Y.

Proof. (a) Si $a \neq 0$, $\mathbb{E}Y_1 = 3a$ et $\mathbb{E}Y_2 = -a$, donc l'espérance dépend du temps: le processus (Y_t) n'est donc pas stationnaire **(1 pt)**. Il est en revanche gaussien car (X_t) est gaussien **(0.5 pts)**.
(b) On peut écrire que $a(t) = a$ (tendance) et $s(t) = 2a\big(\mathbb{I}_{t=1\,[3]} - \mathbb{I}_{t=2\,[3]}\big)$ (saisonnalité) de période 3 **(1.5 pts)**.
(c) On pose $Z = {}^t\big(3, -1, 1, 3, -1, 1, 3, \cdots, 1\big)$ et alors $Y = aZ + \varepsilon$ (écriture vectorielle du modèle). Alors $\widehat{a}_n = ({}^tZ\,Z)^{-1}\,{}^tZ\,Y$ d'où l'expression proposée de $\widehat{a}_n$ **(1.5 pts)**.
Classiquement on en déduit l'estimateur $\widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{3n} \big(Y_i - \widehat{a}_n\, g(i)\big)^2$ **(0.5 pts)**.
(d) Il est clair que $\widehat{a}_n = a + \frac{1}{11\,n} \sum_{i=0}^{n-1} \big(3X_{3i+1} - X_{3i+2} + X_{3i+3}\big)$ donc $\widehat{a}_n$ est un estimateur non biaisé de a **(1 pt)**.
On a $\mathbb{E}(U_n + V_n)^2 \leq 2(\mathbb{E}U_n^2 + \mathbb{E}V_n^2) \underset{n\to\infty}{\longrightarrow} 0$, donc $U_n + V_n \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} 0$ **(1 pt)**.
Comme en 1.(g), on montre $\mathrm{var}\big(\frac{1}{n}\sum_{i=0}^{n-1} X_{3i+1}\big) \underset{n\to\infty}{\longrightarrow} 0$, $\mathrm{var}\big(\frac{1}{n}\sum_{i=0}^{n-1} X_{3i+2}\big) \underset{n\to\infty}{\longrightarrow} 0$ et $\mathrm{var}\big(\frac{1}{n}\sum_{i=0}^{n-1} X_{3i+3}\big) \underset{n\to\infty}{\longrightarrow} 0$, donc $\frac{1}{11\,n} \sum_{i=0}^{n-1} \big(3X_{3i+1} - X_{3i+2} + X_{3i+3}\big) \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} 0$, donc $\widehat{a}_n \underset{n\to+\infty}{\overset{\mathbb{L}^2}{\longrightarrow}} a$ **(1.5 pts)**.
(e) Comme la tendance est a, alors $\widehat{a}_n$ est un estimateur convergeant de la tendance, comme la saisonnalité est $2a\big(\mathbb{I}_{t=1\,[3]} - \mathbb{I}_{t=2\,[3]}\big)$, alors $2\widehat{a}_n\,\big(\mathbb{I}_{t=1\,[3]} - \mathbb{I}_{t=2\,[3]}\big)$ est un estimateur convergeant de la saisonnalité **(0.5 pts)**.

□