Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2010 − 2011

Statistiques II

Correction du contrôle continu n°1, mars 2011

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

1. **(Sur 13 points + 4 points facultatifs)** Soit $X = (X_t)_{t\in\mathbf{Z}}$ un processus stationnaire centré tel que $\mathrm{var}X_0 = \sigma^2 < \infty$. On suppose que X est M-dépendant, où $M \in \mathbf{N}$, c'est-à-dire que pour tout t tel que $t \geq M+1$, alors $(X_{-k})_{k\in\mathbf{N}}$ est indépendant de $(X_{t+k})_{k\in\mathbf{N}}$.

   (a) Montrer que pour tout $(s,t) \in \mathbf{Z}$ tels que $|s-t| \geq M+1$ alors X_s est indépendant de X_t **(1pt)**.

   (b) On note $r(k) = \mathbb{E}X_0X_k$ pour tout $k \in \mathbf{Z}$. Montrer que $|r(k)| \leq \sigma^2$ pour tout $k \in \mathbf{Z}^2$ **(1pt)**. Que peut-on dire de $r(k)$ si $|k| > M$ **(0.5pts)**?

   (c) Soit $\overline{X}_N := \frac{1}{N}\sum_{i=1}^N X_i$. Montrer que $N\,\mathrm{var}(\overline{X}_N) \to \gamma^2$ quand $N \to \infty$ avec $\gamma^2 = \sigma^2 + 2\sum_{k=1}^M r(k) < \infty$ **(2.5pts)** et en déduire que $\overline{X}_N \xrightarrow[N\to+\infty]{\mathcal{P}} 0$ **(1pt)**.

   (d) Soit $n \geq 2M+1$ tel que $N = r\,n$ où $r \in \mathbf{N}^*$. On considère $Z_i = \frac{1}{n}\sum_{k=1}^{n-M} X_{(i-1)n+k}$ pour $i = 1,\cdots,r$. Montrer que les variables aléatoires Z_i sont des v.a.i.i.d. **(1pt)**. Montrer que pour tout $n \geq 2M$, il existe $\gamma_n^2 < \infty$ que l'on précisera tel que $\sqrt{r\,n}\overline{Z}_r \xrightarrow[N\to+\infty]{\mathcal{P}} \mathcal{N}(0,\gamma_n^2)$ où $\overline{Z}_r = \frac{1}{r}\sum_{i=1}^r Z_i$ **(2.5pts)**.

   (e) Montrer que $\gamma_n^2 \xrightarrow[n\to\infty]{} \gamma^2$ **(0.5pt)**.

   (f) Soit $U_i^{(n)} = \frac{1}{n}\sum_{k=0}^{M-1} X_{in-k}$ pour $i = 1,\cdots,r$. Montrer que les variables aléatoires $U_i^{(n)}$ sont des v.a.i.i.d. **(0.5pts)**. Montrer que $N\mathrm{var}(\overline{U}_r^{(n)}) \xrightarrow[n\to\infty]{} 0$, où $\overline{U}_r^{(n)} = \frac{1}{r}\sum_{i=1}^r U_i^{(n)}$ **(2pts)**.

   (g) **(Question facultative: 4pts)** Déduire de (d), (e) et (f) que
$$\sqrt{N}\,\overline{X}_N \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\gamma^2). \tag{1}$$

   (h) Adapter le théorème (1) au cas où $(X_t)_{t\in\mathbf{Z}}$ est un processus stationnaire M-dépendant tel que $\mathbb{E}X_0 = m$ et $\mathrm{var}X_0 = \sigma^2 < \infty$ **(0.5pts)**.

Proof. (a) Du fait de la stationarité, on a (X_s, X_t) qui a même loi que $(X_0, X_{|t-s|})$ et il suffit alors d'appliquer la définition de la M-dépendance.
(b) On utilise Cauchy-Schwarz et $|r(k)| = |\mathbb{E}X_0X_k| \leq \big(\mathbb{E}X_0^2\mathbb{E}X_k^2)^{1/2} = \sigma^2$. D'après la définition de la M-dépendance et la stationnarité, $r(k) = 0$ pour $|k| \geq M+1$.
(c) On a $N\,\mathrm{var}(\overline{X}_N^2) = \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N r(|i-j|) = \frac{1}{N}\Big(Nr(0) + 2(N-1)r(1) + \cdots + 2(N-M)r(M)\Big)$. Ainsi quand $N \to \infty$, on a bien $N\,\mathrm{var}(\overline{X}_N^2) \to \gamma^2$ car M est un nombre ne dépendant pas de N.
On utilise l'Inégalité de Bienaymé-Tchebytchev, et pour tout $\varepsilon > 0$, $\mathbb{P}(|\overline{X}_N| \geq \varepsilon) \leq \frac{\mathrm{var}(\overline{X}_N^2)}{\varepsilon^2}$ d'où $\mathbb{P}(|\overline{X}_N| \geq \varepsilon) \xrightarrow[N\to\infty]{} 0$ et ainsi $\overline{X}_N \xrightarrow[N\to+\infty]{\mathcal{P}} 0$.
(d) D'après la stationarité de X, il est clair que $(X_1,\cdots,X_{n-M})$ a la même loi que $(X_{(i-1)n+1},\cdots,X_{in-M})$ et donc les Z_i sont identiquement distribuées. De plus, d'après la définition de la M-dépendance et la stationnarité de X, pour $i < j$, $(X_{in-M-k})_{k\in\mathbf{N}}$ est indépendant de $(X_{(j-1)n+k})_{k\in\mathbf{N}}$ donc $(Z_1,\cdots,Z_i)$ est indépendant de $(Z_j,\cdots,Z_r)$: les (Z_i) sont bien des v.a.i.i.d.
Il est clair que comme $n \geq 2M+1$, on a $\mathrm{var}(Z_i) = \frac{1}{n^2}\big(\sigma^2 + 2\sum_{i=1}^M(n-i)r(i)\big)$. Donc $\mathbb{E}Z_i = 0$, $\mathrm{var}(Z_i) < \infty$ et comme les Z_i sont des v.a.i.i.d. on peut appliquer le TCL usuel. On obtient ainsi en posant $\gamma_n^2 = n\,\mathrm{var}(Z_i) = \sigma^2 + 2\sum_{i=1}^M(1-\frac{i}{n})r(i)$ que pour $r \to \infty$, donc quand $N \to \infty$, $\sqrt{rn}\,\overline{Z}_r \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\gamma_n^2)$.
(e) On voit aisément comme $i/n \to 0$ que $\gamma_n^2 \xrightarrow[n\to\infty]{} \gamma^2$.
(f) On montre que les $U_i^{(n)}$ sont des v.a.i.i.d. comme on l'a fait pour les Z_i.

Par ailleurs, comme précédemment $\mathrm{var}(U_i^{(n)}) = \frac{1}{n^2}\Big(\sigma^2 + 2\sum_{i=1}^M (n-i)r(i)\Big)$. Ainsi $\mathrm{var}(\overline{U}_r^{(n)}) = \frac{1}{rn^2}\gamma_n^2$. Comme $N = nr$, on en déduit que $N\mathrm{var}(\overline{U}_r^{(n)}) \underset{n\to\infty}{\longrightarrow} 0$.
(g) On a $\sqrt{N}\,\overline{X}_N = \sqrt{N}\left(\overline{Z}_r + \overline{U}_r^{(n)}\right)$. Prenons par exemple $n = [\sqrt{N}]$ et donc $r = N/n \sim \sqrt{N}$. Comme $\gamma_n^2 \underset{n\to\infty}{\longrightarrow} \gamma^2$, il est clair d'après le Théorème de Slutsky que $\sqrt{N}\,\overline{Z}_r \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\gamma^2)$. De plus d'après (f), $\sqrt{N}\,\overline{U}_r^{(n)} \xrightarrow[n\to\infty]{\mathcal{L}} 0$. Donc d'après le Théorème de Slutsky, $\sqrt{N}\left(\overline{Z}_r + \overline{U}_r^{(n)}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\gamma^2)$, d'où le TLC (1).
(h) Il suffit d'appliquer le TLC (1) à $X_i' = X_i - m$ pour montrer que $\sqrt{N}\left(\overline{X}_N - m\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\gamma^2)$. □

2. **(Sur 14 points)** On considère une suite $(\varepsilon_i)_{i\in\mathbb{N}}$ de variables aléatoires gaussiennes indépendantes identiquement distribuées centrées et de variance commune $\sigma^2 > 0$. On définit le processus $X = (X_n)_{n\in\mathbf{N}^*}$ tel que

$$X_i = a + \varepsilon_{i-1}\,\varepsilon_i \quad \text{pour tout } i \in \mathbf{N}^*,$$

où a est un réel.

(a) Déterminer la tendance et la saisonalité de X **(1pt)**.
(b) Montrer que X est un processus stationnaire **(2pts)**.
(c) Montrer que X est un processus d'ordre 2 **(0.5pts)**. Le processus X est-il gaussien **(0.5pts)**?
(d) Déterminer l'autocovariance de X **(1pt)**. En déduire l'expression de la densité spectrale de X après avoir montré qu'elle existe **(1pt)**. Le processus $(X-a)$ est-il un bruit blanc faible **(0.5pts)**? Un bruit blanc fort **(1.5pts)**?
(e) On suppose que l'on observe une trajectoire $(X_1,\cdots,X_N)$ avec $N\in\mathbf{N}^*$ mais que a, σ^2 et les $(\varepsilon_i)_{0\leq iN}$ sont inconnus. On veut estimer a et σ^2. On définit:

$$\overline{X}_N := \frac{1}{N}\sum_{i=1}^N X_i, \quad \text{et} \quad \widehat{s}_N := \frac{1}{N}\sum_{i=1}^N X_i^2.$$

En utilisant 1.(h), montrer des thèorèmes centraux limite vérifiés par $\overline{X}_N$, par $(\overline{X}_N)^2$ et par $\widehat{s}_N$ **(4pts)**.
(f) Déduire de ceci des estimateurs convergents de a et σ^2 et préciser leurs vitesses de convergence **(2pts)**.

Proof. (a) Il est clair que $\mathbb{E}X_n = a$ donc la tendance de (X_n) est a et (X_n) n'a pas de saisonalité.
(b) Le processus (ε_i) est stationnaire car c'est un bruit blanc. Donc pour $n\in\mathbf{N}^*$, $(t_1,\cdots,t_n)\in\mathbf{N}^n$ et pour tout $c\in\mathbf{N}$, $(\varepsilon_{t_1},\varepsilon_{t_1-1},\varepsilon_{t_2},\varepsilon_{t_2-1},\cdots,\varepsilon_{t_n-1})$ a la même loi que $(\varepsilon_{t_1+c},\varepsilon_{t_1+c-1},\varepsilon_{t_2+c},\varepsilon_{t_2+c-1},\cdots,\varepsilon_{t_n+c-1})$. En considérant la fonction $g:(x_1,\cdots,x_{2n})\in\mathbf{R}^{2n}\mapsto(a+x_2x_1, a+x_4x_3,\cdots,a+x_{2n}x_{2n-1})\in\mathbf{R}^n$ qui est une fonction continue donc mesurable, il est clair que $g(\varepsilon_{t_1},\varepsilon_{t_1-1},\varepsilon_{t_2},\varepsilon_{t_2-1},\cdots,\varepsilon_{t_n-1})$ a la même loi que $g(\varepsilon_{t_1+c},\varepsilon_{t_1+c-1},\varepsilon_{t_2+c},\varepsilon_{t_2+c-1},\cdots,\varepsilon_{t_n+c-1})$, donc $(X_{t_1},\cdots,X_{t_n})$ a la même loi que $(X_{t_1+c},\cdots,X_{t_n+c})$: le processus X est stationnaire.
(c) On a clairement $\mathbb{E}X_i^2 = a^2+\sigma^4<\infty$ donc le processus est bien d'ordre 2.
X n'est pas un processus gaussien car le produit de 2 v.a. gaussiennes indépendantes (ε_i et ε_{i-1}) non dégénérées n'est pas une v.a. gaussienne, donc X_i n'est pas une v.a. gaussienne.
(d) On a pour $k\geq 0$, $r(k) = \mathbb{E}[(X_1-a)(X_{k+1}-a)] = \mathbb{E}[\varepsilon_0\varepsilon_1\varepsilon_k\varepsilon_{k+1}]$. Donc si $k=0$, $r(0)=\sigma^4$, si $k=1$, $r(1)=0$ et pour $|k|\geq 2$, $r(k)=0$.
On a clairement $\sum_k |r(k)|<\infty$ donc la densité spectrale de X existe et est continue. Pour $\lambda\in[-\pi,\pi]$ on a $f_X(\lambda) = r(0) = \sigma^4$.
Comme $\mathbb{E}(X-a)=0$, $r(k)=0$ pour $|k|\geq 1$, $(X-a)$ est bien un bruit blanc faible.
On a $\mathbb{E}[X_1^2X_2^2] = \mathbb{E}[(a+\varepsilon_0\varepsilon_1)^2(a+\varepsilon_1\varepsilon_2)^2] = a^4+2a^2\sigma^4+3\sigma^8$ alors que $\mathbb{E}[X_1^2]\,\mathbb{E}[X_2^2] = (a^2+\sigma^4)^2 = a^4+2a^2\sigma^4+\sigma^8$, donc $\mathbb{E}[X_1^2X_2^2]\neq\mathbb{E}[X_1^2]\,\mathbb{E}[X_2^2]$ et ainsi les X_i ne sont pas des v.a.i.i.d. et $(X-a)$ n'est pas un bruit blanc fort.
(e) Il est clair que X est un processus stationnaire 1-dépendant, d'espérance a et d'autocovariance r. On en déduit d'après 1. (h) que: $\sqrt{N}\left(\overline{X}_N - a\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,r(0)) = \mathcal{N}\big(0,\sigma^4\big)$.
En appliquant la Delta-méthode à ce TCL, on montre également un TCL vérifié par $(\overline{X}_N)^2$ soit: $\sqrt{N}\left((\overline{X}_N)^2 - a^2\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0,4a^2\sigma^4\big)$.
Le processus $(X_i^2)_i$ est également un processus stationnaire 1-dépendant. Son espérance est $a^2+\sigma^4$ et son autocovariance r_2 est telle que $r_2(0) = \mathbb{E}(a+\varepsilon_0\varepsilon_1)^4 - (a^2+\sigma^4)^2 = 4a^2\sigma^4+8\sigma^8$ et $r_2(1) = 2\sigma^8$, $r_2(k)=0$ pour $|k|\geq 0$. On peut donc appliquer le TLC de 1. (h) et on obtient que $\sqrt{N}\left(\widehat{s}_N - (a^2+\sigma^4)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, r_2(0)+2r_2(1)) = \mathcal{N}\big(0,4a^2\sigma^4+12\sigma^8\big)$.
(f) Il est clair que l'on peut choisir $\widehat{a}_N = \overline{X}_N$ comme estimateur convergent de a. Pour estimer σ^2 on choisira naturellement $\widehat{\sigma_N^2} = (\widehat{s}_N - (\overline{X}_N)^2)^{1/2}$. □