Université Paris I, Panthéon - Sorbonne

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

Statistiques II

Correction du contrôle continu n°2, avril 2013

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

1. **(Sur 18 (+4) points)** Soit $\varepsilon = (\varepsilon_t)_{t\in\mathbf{Z}}$ un bruit blanc fort centré de variance $\sigma_\varepsilon^2 > 0$, inconnue. On définit, lorsque cela est possible, le processus $X = (X_t)_{t\in\mathbf{Z}}$ tel que

$$X_t - 2a\,X_{t-1} + a^2 X_{t-2} = \varepsilon_t \qquad \text{pour} \quad t \in \mathbf{Z}.$$

avec $a \in \mathbf{R}$.

(a) Dire suivant les valeurs de a quel type de processus est X. On suppose désormais que $|a| < 1$.

(b) Pour $t \in \mathbf{Z}$, déterminer l'expression de X_t en fonction de $(\varepsilon_{t-n})_{n\in\mathbf{N}}$ et de a.

(c) Montrer que pour tout $x \in]-1, 1[$, $\sum_{k=0}^\infty kx^k = x(1-x)^{-2}$ et $\sum_{k=0}^\infty k^2x^k = x(1+x)(1-x)^{-3}$ (on pourra s'aider du développement en série entière de $(1-x)^{-1}$).

(d) En déduire que $r_X(j) = \sigma_\varepsilon^2 a^{|j|}\Big(\frac{(1+a^2)+(1-a^2)|j|}{(1-a^2)^3}\Big)$ pour tout $j \in \mathbf{Z}$ où r_X est l'autocovariance de X. Quelle est la limite de $r_X(j)$ quand $|j| \to \infty$?

(e) On observe finalement $(Y_1, \cdots, Y_n)$ où pour $t \in \mathbf{Z}$, $Y_t = \alpha + X_t + \xi_t$, avec $\alpha \in \mathbf{R}$ inconnu et $\xi = (\xi_t)_{t\in\mathbf{Z}}$ un bruit blanc fort centré de variance σ_ξ^2, indépendant de ε. Montrer que $Y = (Y_t)_{t\in\mathbf{Z}}$ est stationnaire. Déterminer la densité spectrale de Y.

(f) **(Facultatif: 4 points)** Montrer que cette densité spectrale est celle d'un processus ARMA dont on précisera l'ordre.

(g) On définit $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Montrer que $\overline{Y}_n$ est un estimateur non biaisé de α. Montrer que $\overline{Y}_n$ vérifie un théorème de la limite centrale que l'on précisera.

Proof. (a) Si $a = 0$, X est un bruit blanc fort **(0.5 pts)**.
Si $a \neq 0$, comme $P(X) = 1 - 2aX + a^2X^2 = (1-aX)^2$, P admet une unique racine $1/a$, double. On en déduit que si $|a| < 1$, X est un AR(2) stationnaire causal, si $|a| > 1$, X est un AR(2) stationnaire non causal, et si $|a| = 1$, X n'est pas stationnaire **(1.5 pt)**.
(b) On a $(1-z)^{-2} = ((1-z)^{-1})' = (\sum_{k=0}^\infty z^k)' = \sum_{k=0}^\infty (k+1)z^k$. Donc $P(B)^{-1} = \sum_{k=0}^\infty (k+1)a^kB^k$. On en déduit que $X_t = \sum_{k=0}^\infty (k+1)a^k\varepsilon_{t-k}$ **(2 pts)**.
(c) Comme pour $|x| < 1$, $(1-x)^{-1} = \sum_{k=0}^\infty x^k$, en dérivant, on obtient $(1-x)^{-2} = \sum_{k=1}^\infty kx^{k-1}$, donc $x\sum_{k=1}^\infty kx^{k-1} = \sum_{k=0}^\infty kx^k = x(1-x)^{-2}$ **(1 pt)**.
En dérivant cette expression une fois, on obtient que $\sum_{k=0}^\infty k^2x^{k-1} = (x+1)(1-x)^{-3}$ puis en multipliant encore par x, on obtient bien $\sum_{k=0}^\infty k^2x^k = x(1+x)(1-x)^{-3}$ **(1 pt)**.
(d) On a $r_X(j) = \mathbb{E}(X_0X_j) = \mathbb{E}\Big[\sum_{k=0}^\infty\sum_{k'=0}^\infty (k+1)a^k\varepsilon_{-k}(k'+1)a^{k'}\varepsilon_{j-k'}\Big]$. On peut intervertir espérance et séries (Lebesgue) et comme $\mathbb{E}[\varepsilon_i\varepsilon_j] = \sigma_\varepsilon^2\delta_{ij}$, tous les termes de la double somme s'annule, sauf lorsque $-k = j-k'$, donc lorsque $k' = k+j$. En choisissant $j \in \mathbf{N}$, on aboutit à: $r_X(j) = \sigma_\varepsilon^2\sum_{k=0}^\infty a^{2k+j}(k+1)(k+j+1) = \sigma_\varepsilon^2 a^j\big(\sum_{k=0}^\infty k^2a^{2k} + (j+2)\sum_{k=0}^\infty ka^{2k} + (j+1)\sum_{k=0}^\infty a^{2k}\big)$. En utilisant les formules précédentes, on obtient

$$r_X(j) = \frac{\sigma_\varepsilon^2 a^j}{(1-a^2)^3}\big(a^2(1+a^2) + (j+2)a^2(1-a^2) + (j+1)(1-a^2)^2\big) = \frac{\sigma_\varepsilon^2 a^j}{(1-a^2)^3}\big((1+a^2) + j(1-a^2)\big). \quad \textbf{(3 pts)}.$$

En utilisant la parité de r_X on obtient la formule proposée **(0.5 pts)**.
Il est clair que $r_X(j) \to 0$ quand $|j| \to \infty$ **(0.5 pts)**.
(e) On sait que X est stationnaire et ξ également (comme bruit blanc fort). On doit donc montrer que la somme de 2 processus stationnaires indépendants est un processus stationnaire. On peut faire cela en utilisant la fonction caractéristique: si $n \in \mathbf{N}^*$, $(t_1, \cdots, t_n) \in Z^n$, et $c \in \mathbf{Z}$,

$$\begin{aligned}\mathbb{E}\big[e^{i\sum_{k=1}^n u_k(X_{t_k}+\xi_{t_k})}\big] &= \mathbb{E}\big[e^{i\sum_{k=1}^n u_kX_{t_k}}\big]\mathbb{E}\big[e^{i\sum_{k=1}^n u_k\xi_{t_k}}\big] \\ &= \mathbb{E}\big[e^{i\sum_{k=1}^n u_kX_{t_k+c}}\big]\mathbb{E}\big[e^{i\sum_{k=1}^n u_k\xi_{t_k+c}}\big] = \mathbb{E}\big[e^{i\sum_{k=1}^n u_k(X_{t_k+c}+\xi_{t_k+c})}\big],\end{aligned}$$

en utilisant, l'indépendance, puis la stationnarité de X et de ξ, puis encore l'indépendance **(3 pts)**..
Il est clair que comme les 2 processus sont indépendants, la densité spectrale de la somme est la somme des densités spectrales. Donc $f_Y(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{|(1-ae^{i\lambda})^2|^2} + \frac{\sigma_\xi^2}{2\pi}$ **(2 pts)**.
(f) On peut encore écrire $f_Y(\lambda) = \frac{1}{2\pi}\Big(\frac{\sigma_\varepsilon^2+\sigma_\xi^2|(1-ae^{i\lambda})^2|^2}{|(1-ae^{i\lambda})^2|^2}\Big) = \frac{1}{2\pi}\Big(\frac{\sigma_\varepsilon^2+\sigma_\xi^2|(1-ae^{i\lambda})^2|^2}{|(1-ae^{i\lambda})^2|^2}\Big) = \frac{1}{2\pi}\Big(\frac{\sigma_\varepsilon^2+\sigma_\xi^2(1+4a^2+a^4)-4a(1+a^2)\cos(\lambda)+a^2\cos(2\lambda)}{|(1-ae^{i\lambda})^2|^2}\Big)$.
On voit ainsi apparaître un polynôme trigonométrique d'ordre 2 paire, donc Y est un ARMA$(2,2)$ **(4 pts)**.
(g) On a clairement $\mathbb{E}[\overline{Y}_n] = a$ donc l'estimateur est sans biais **(0.5 pts)**.
On a $\overline{Y}_n = a + \overline{X}_n + \overline{\xi}_n$, où $(\overline{X}_n)$ et $(\overline{\xi}_n)$ sont deux processus indépendants. Comme X est un AR(2), on sait que $(\overline{X}_n)_n$ vérifie: $\sqrt{n}\overline{X}_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, 2\pi\, f_X(0)) = \mathcal{N}\big(0, \frac{\sigma_\varepsilon^2}{(1-a)^4}\big)$. Comme ξ est un bruit blanc, on a $\sqrt{n}\overline{\xi}_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_\xi^2)$. Ainsi, comme les 2 processus sont indépendants, la variance de la somme est la somme des variances et on obtient:

$$\sqrt{n}(\overline{Y}_n - a) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\Big(0, \frac{\sigma_\varepsilon^2}{(1-a)^4} + \sigma_\xi^2\Big). \quad \textbf{(2.5 pts)}.$$

□

2. **(Sur 9 points)** Voici des simulations effectuées avec le logiciel R.

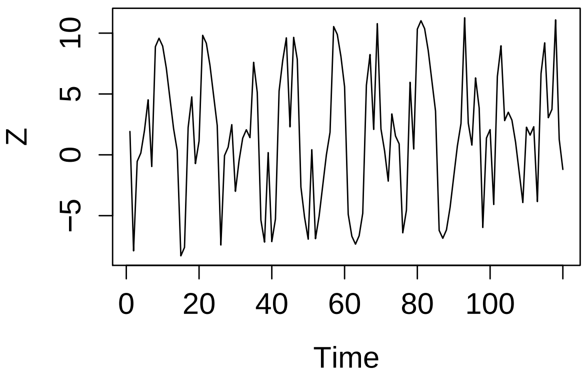
   (a) On tape d'accord les commandes suivantes:

```
X=runif(120)
Y=1
p=0.4
for (i in c(1:119))
{Y[i+1]=Y[i]*(X[i]>p)-Y[i]*(X[i]<p)}
Y
Z=2+0.02*c(1:120)-5*sin(pi*c(1:120)/6)+4*(Y-p)
ts.plot(Z)
```

Voici les premières valeurs de Y obtenues:

```
> Y
  [1]  1 -1  1  1  1  1 -1  1  1  1  1  1  1  1 -1 -1  1  1 -1 -1  1  1
 [23]  1  1  1 -1  1  1  1 -1 -1 -1 -1 -1  1  1 -1 -1  1 -1 -1  1  1  1
 [45] -1  1  1 -1 -1 -1  1 -1 -1 -1 -1 -1  1  1  1  1 -1 -1 -1 -1 -1  1
```

On obtient également le graphe suivant:



Questions: Quel est le processus simulé par le vecteur Y (préciser tous ses paramètres)? Pourquoi considère-t-on $Y - p$? Quelle sont les tendances et saisonnalités de Z?

Proof. Y est tel que Y_n prend ses valeurs dans $\{-1, 1\}$ et quand $Y_{n-1} = 1$, alors $Y_n = 1$ avec une probabilité $(1-p)$ et $Y_n = -1$ avec une probabilité p, quand $Y_{n-1} = -1$ alors $Y_n = 1$ avec une probabilité p et $Y_n = -1$ avec une probabilité $1-p$, ceci pour tout $n > 1$. On en déduit que Y est une chaîne de Markov de matrice de transition $Q = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, où $p = 0.4$, et avec $Y_1 = 1$ **(2 pts)**.
La mesure invariance associée à Y est telle que $(\mu_{-1} 1 - \mu_{-1})Q = (\mu_{-1} 1 - \mu_{-1})$, donc on trouve que la mesure invariante est $(1/2, 1/2)$. Donc asymptotiquement l'espérance de Y_n est 0. Il n'y a donc pas vraiment de raison de considérer $Y - p$. **(1 pt)**
La tendance de Z_t est $2 + 0.02t$, la saisonnalité $-5\sin(\pi t/6)$ de période 12 (et on a bien $\sum_{k=1}^{12} -5\sin(\pi t/6) = 0$) **(1 pt)**. □

   (b) Voici les commandes tapées ensuite:

```
Z1=c(1:120)
Z2=sin(pi*c(1:120)/6)
reg=lm(Z~Z1+Z2)
summary(reg)
```

Voici les résultats obtenus:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.755147   0.732735   2.395   0.0182 *
Z1          -0.001297   0.010518  -0.123   0.9021
Z2          -5.243756   0.515256 -10.177   <2e-16 ***

Residual standard error: 3.98 on 117 degrees of freedom
Multiple R-squared: 0.4706,Adjusted R-squared: 0.4615
F-statistic:    52 on 2 and 117 DF,  p-value: < 2.2e-16
```

Questions: Qu'a-t-on fait par ces commandes? Que représentent les valeurs 1.755147, -0.001297 *et* -5.243756*? Pourquoi s'attendait-on approximativement à ces valeurs?*

Proof. On effectue une régression linéaire de Z par rapport aux variables $Z1_t = t$ et $Z2_t = \sin(\pi t/6)$. Les valeurs obtenues sont ainsi des estimations de 2, -0.02 et -5. On sait que l'estimateur par moindres carrés ordinaires est sans biais et convergeant sous des conditions vérifiées ici **(2 pts)**. □

(c) On tape enfin les commandes suivantes:

```
X=runif(100)
epsilon=2*(X<0.5)-1
Y=0
for (i in c(1:99))
{Y[i+1]=epsilon[i]*sqrt(0.5+0.6*Y[i]^2)}
```

Voici les premières valeurs de Y obtenues:

```
> Y
  [1]  0.0000000  0.7071068  0.8944272 -0.9899495  1.0430724 -1.0736852
  [7]  1.0916410  1.1022740 -1.1086049 -1.1123861  1.1146487 -1.1160040
 [13] -1.1168165  1.1173036  1.1175958  1.1177711  1.1178763 -1.1179394
 [19]  1.1179772  1.1179999  1.1180136 -1.1180217  1.1180266 -1.1180296
 [25]  1.1180313  1.1180324  1.1180330 -1.1180334 -1.1180336 -1.1180338
 [31] -1.1180339 -1.1180339 -1.1180339 -1.1180340  1.1180340  1.1180340
 [37]  1.1180340  1.1180340 -1.1180340  1.1180340  1.1180340 -1.1180340
```

Questions: Expliquez ce qui a été effectué avec ces commandes (en particulier, quelle est la loi de epsilon, quel processus est Y ?). Retrouvez par le calcul la valeur exacte correspondant à 1.1180340.

Proof. Le vecteur *epsilon* est composé de variables aléatoires indépendantes prenant pour valeurs 1 ou -1 avec une probabilité $1/2$ **(0.5 pts)**. C'est donc un bruit blanc fort. Y est donc un processus ARCH(1) tel que $Y_n = epsilon_n \sqrt{0.5 + 0.6 Y_{n-1}^2}$ **(0.5 pts)**.
On a $Y_n^2 = 0.5 + 0.6 Y_{n-1}^2$, donc Y_n^2 est une suite arithmo-géométrique qui converge (raison $= 0.6$ de module < 1) et sa limite ℓ est telle que $\ell = 0.5 + 0.6\ell$ soit $\ell = 5/4$. Donc à la limite Y_n prend pour valeur $\pm\sqrt{5/4} \simeq 1.1180340$ **(2 pts)**. □