Université Paris I, Panthéon - Sorbonne

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

Statistiques II

Examen final, avril 2013

Examen de 3h00. Tout document ou calculatrice est interdit.

1. **(8 points)** Soit $(\varepsilon_n)_{n\in\mathbf{Z}}$ un bruit blanc gaussien de variance σ_ε^2, où $\sigma_\varepsilon^2>0$ est inconnu et soit $X=(X_n)_{n\in\mathbf{Z}}$ le processus défini par
$$X_n=\alpha X_{n-1}+\varepsilon_n \quad \text{pour} \quad n\in\mathbf{Z},$$
avec $|\alpha|<1$ un réel inconnu.
   (a) Quel processus est X? Est-il centré? stationnaire? gaussien? (justifier)
   (b) Pour $n\in\mathbf{Z}$, déterminer $\mathbb{E}(X_n\,|\,X_{n-1})$ et $\mathbb{E}(X_n^2\,|\,X_{n-1})$. En déduire que $\mathrm{var}(X_n\,|\,X_{n-1})=\sigma_\varepsilon^2$ (on rappelle que $\mathrm{var}(X_n\,|\,X_{n-1})=\mathbb{E}(X_n^2\,|\,X_{n-1})-\big(\mathbb{E}(X_n\,|\,X_{n-1})\big)^2$) et la loi de X_n sachant X_{n-1}.
   (c) On observe une trajectoire $(X_1,\cdots,X_N)$ de X. Montrer que la log-vraisemblance de $(X_2,\cdots,X_N)$ sachant X_1 vaut:
$$LV_{\alpha,\sigma_\varepsilon^2}(X_2,\cdots,X_N)=-\frac{(N-1)}{2}\big(\log(2\pi)+\log(\sigma_\varepsilon^2)\big)-\frac{1}{2\sigma_\varepsilon^2}\sum_{n=2}^N(X_n-\alpha X_{n-1})^2.$$
   En déduire les estimateurs $\widehat{\alpha}_N$ et $\widehat{\sigma}_N^2$ de α et σ_ε^2 maximisant cette log-vraisemblance (on montrera que $\widehat{\alpha}_N=\frac{\sum_{n=2}^N X_nX_{n-1}}{\sum_{n=2}^N X_{n-1}^2}$).
   (d) On rappelle que pour un processus ARMA stationnaire causal, $\sqrt{N}\big(\widehat{r}_N(k)-r(k)\big)_{0\le k\le m}\xrightarrow[N\to\infty]{\mathcal{L}}\mathcal{N}_{m+1}\big(0\,,\,\Sigma\big)$, où $\Sigma=\big(4\pi\int_{-\pi}^{\pi}f^2(\lambda)\cos(i\lambda)\cos(j\lambda)d\lambda\big)_{1\le i,j\le m}$, avec r et f l'autocovariance et la densité spectrale du processus et $\widehat{r}_N(k)=\frac{1}{N}\sum_{n=1}^N X_nX_{n+k}$. En déduire un théorème de la limite centrale vérifiée par $\widehat{\alpha}_N$.

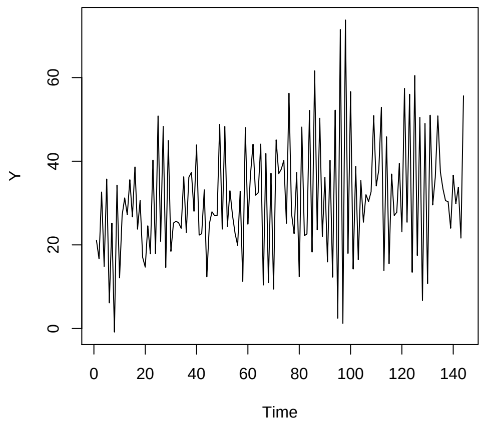
2. **(8 points)** On considère maintenant $(\xi_n)_{n\in\mathbf{Z}}$ un bruit blanc gaussien de variance 1 admettant un moment d'ordre 4 et soit $Y=(Y_n)_{n\in\mathbf{Z}}$ le processus défini par
$$Y_n=\xi_n\sqrt{a_0+a_1Y_{n-1}^2} \quad \text{pour} \quad n\in\mathbf{Z},$$
où a_0 et a_1 sont deux réels strictement positifs inconnus.
   (a) Quel processus est Y? Rappeler la condition portant sur a_1 permettant à Y d'être stationnaire d'ordre 2 et causal.
   (b) On se place désormais sous cette condition. Calculer alors $\mathbb{E}(Y_0)$ et $\mathrm{cov}(Y_0,Y_k)$ pour tout $k\in\mathbf{Z}$.
   (c) Pour $n\in\mathbf{Z}$, déterminer $\mathbb{E}(Y_n\,|\,Y_{n-1})$ et $\mathbb{E}(Y_n^2\,|\,Y_{n-1})$. En déduire la loi de Y_n sachant Y_{n-1}. Que vaut $\mathrm{cov}(Y_n,Y_{n+k}\,|\,Y_{n+k-1})$ pour $k\in\mathbf{N}$?
   (d) Soit le processus $Z=(Z_n)_{n\in\mathbf{Z}}$ tel que $Z_n=Y_n^2-a_0(1-a_1)^{-1}$ pour $n\in\mathbf{Z}$. Montrer que Z vérifie $Z_n=a_1Z_{n-1}+u_n$, où $u_n=(\xi_n^2-1)\big(a_0+a_1(Z_{n-1}+a_0(1-a_1)^{-1})\big)$ pour $n\in\mathbf{Z}$. Montrer que (u_n) est un bruit blanc faible.
   (e) On observe une trajectoire $(Y_1,\cdots,Y_N)$ de Y. En utilisant 1.(c) et (d), proposer un estimateur de a_1 fonction de $(Y_1,\cdots,Y_N)$ et dont on donnera un théorème de la limite centrale (on admettra que le Théorème obtenu en 1.(d) est valide pour (Z_n)), puis en utilisant la valeur de $\mathrm{var}(Y_0)$ proposer un estimateur convergeant de a_0.

3. **(8 points)** Voici des simulations effectuées avec le logiciel R.

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

```
epsilon=3*rnorm(144)
X=0
for (j in c(1:143))
 {X[j+1]=-0.7*X[j]+epsilon[j+1]-2*epsilon[j]}
t=c(1:144)
Y=15+4*log(t)+7*cos(pi*(t-2)/6)+X
ts.plot(Y)
```

Voici le graphe obtenu:



Questions: Quel est le processus simulé par le vecteur X (préciser tous ses paramètres)? Qelles sont les tendances et saisonnalités de Y? Quelle est sa variance (théorique)?

(b) Voici les commandes tapées ensuite:

```
Z1=t
Z2=t^2
Z3=sqrt(t)
Z4=cos(pi*t/6)
Z5=sin(pi*t/6)
reg=lm(Y~Z1+Z2+Z3+Z4+Z5)
summary(reg)
```

Voici les résultats obtenus:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.6447975 11.1173261   0.957 0.339991
Z1          -0.4077850  0.5735870  -0.711 0.478322
Z2           0.0009389  0.0017558   0.535 0.593669
Z3           5.3872548  5.0063068   1.076 0.283763
Z4           3.4346042  1.5587953   2.203 0.029227 *
Z5           6.0576026  1.5746636   3.847 0.000182 ***

Residual standard error: 13.22 on 138 degrees of freedom
Multiple R-squared: 0.1789,     Adjusted R-squared: 0.1492
```

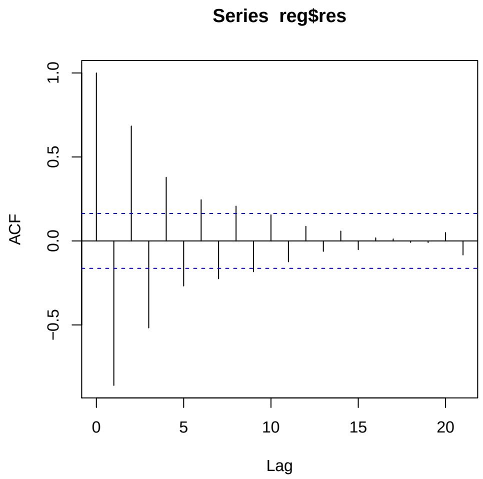
Questions: Qu'a-t-on fait par ces commandes? Que représentent les valeurs 10.6447975, 0.593669 *et* 0.1492*? Que conclure à la lecture de ces résultats?*

(c) On tape enfin les commandes suivantes:

```
reg=lm(Y~Z1+Z3+Z4+Z5)
summary(reg)
reg=lm(Y~Z3+Z4+Z5)
summary(reg)
acf(reg$res)
```

Voici les résultats obtenus ainsi que le graphe tracé:

```
lm(formula = Y ~ Z1 + Z3 + Z4 + Z5)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.2777     6.9492   2.198   0.0296 *
Z1           -0.1099     0.1367  -0.804   0.4225
Z3            2.9408     2.0282   1.450   0.1493
Z4            3.4318     1.5548   2.207   0.0289 *
Z5            5.9595     1.5599   3.820   0.0002 ***

Residual standard error: 13.19 on 139 degrees of freedom
Multiple R-squared: 0.1772,     Adjusted R-squared: 0.1536

lm(formula = Y ~ Z3 + Z4 + Z5)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.1762     3.3436   6.034 1.35e-08 ***
Z3            1.3403     0.3928   3.412 0.000843 ***
Z4            3.4173     1.5527   2.201 0.029384 *
Z5            5.9084     1.5567   3.796 0.000219 ***

Residual standard error: 13.17 on 140 degrees of freedom
Multiple R-squared: 0.1734,     Adjusted R-squared: 0.1557
```



Questions: Expliquez pourquoi on a répété deux fois la commande lm et quel(s) critère(s) guide(nt) cette répétition? Est-on satisfait du modèle obtenu? Pouvait-on s'attendre aux valeurs numériques 3.4173 *et* 5.9084*? Expliquez ce que représente le graphe. La valeur numérique décrite par la seconde barre est approximativement de* -0.85*. Expliquez ce que l'on pouvait espérer comme valeur théorique.*